INVESTIGATING THE MECHANISM OF HUMAN DEFINITIVE ENDODERM DIFFERENTIATION THROUGH GENOME-SCALE CRISPR SCREENS

by

Qing Li

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

April 2019

_____ _____

Danwei Huangfu, PhD Date

Dissertation Mentor

INVESTIGATING THE MECHANISM OF HUMAN DEFINITIVE ENDODERM DIFFERENTIATION THROUGH GENOME-SCALE CRISPR SCREENS

Qing Li

Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences,

Memorial Sloan Kettering Cancer Center

Human embryonic stem cells (hESCs) utilize TGF-β signaling via SMAD2/3 to promote both pluripotency maintenance and definitive endoderm (DE) differentiation, but how TGF-β signaling is interpreted in two opposing processes remains obscure. Forward genetic screens in model organisms provide a powerful approach for uncovering previously unsuspected regulators of development. Using an iCas9 SOX17-GFP reporter hESC line, we performed comprehensive genome-scale CRISPR screens to identify regulators of DE differentiation.

Unexpectedly we uncovered five JNK/JUN family genes as key barriers of DE differentiation, as genetic inactivation of JNK pathway promotes DE differentiation. The JNK/JUN pathway does not act through directly inhibiting the DE enhancers. Instead it primarily acts on the ESC enhancers. JUN and OCT4 co-occupy many enhancers at the ESC stage, and JUN binding correlates with higher enhancer activities as indicated by increased chromatin accessibility as well as H3K27ac and P300 signals. JNK inhibition during DE differentiation accelerates the destabilization of the pluripotency network, as supported by the

decreased chromatin accessibility at ESC enhancers, which is accompanied by decreased SMAD2/3 occupancy at ESC enhancers thus facilitating the reconfiguration of SMAD2/3 chromatin binding to DE enhancers.

Therefore, the JNK/JUN pathway safeguards pluripotency from precocious DE differentiation. Our work establishes a framework for unbiased interrogation hESC differentiation relevant to human development. Direct pharmacological inhibition of JNK significantly improves the differentiation efficiency of DE and DE-derived pancreatic and lung lineage cells, highlighting the potential of harnessing knowledge gained from unbiased screening efforts to improve hESC/hiPSC directed differentiation for disease modeling and regenerative medicine.

ACKNOWLEDGEMENT

I want to thank my mentor Danwei Huangfu for 6 year's trainings and educations. Danwei has been a patient and thoughtful mentor. I came into her lab without any formal training in human embryonic stem cells, but she believed in my potentials and abilities to succeed in the lab. Danwei is also my scientific role model. I learn how to be a better scientist, how to think, how to write, how to communicate as a scientist from her.

I also want to thank committee members Lorenz Studer, Kathryn Anderson, Lukas Dow for your advice and support throughout my years as a graduate student. Thank you for sharing your insights and experiences with me. You helped me overcome many scientific obstacles and guided my research to the correct path.

I would also like to thank the entire members of Danwei Huangfu's lab. Thank Zhu for being my rotation mentor. Thank Gary Dixon, Nipun Verma, Bess Rosen, Renhe Luo, Chunlong Xu Chew-Li Soh, Dapeng Yang, Abhijit Shukla for helping my with research and manuscripts. I also want to thank my collaborators Qiong Wang, Miriam Gordillo, Miguel Crespo, Qing Xiang, Friderike Dundar, Paul Zumbo, Matthew Witkin, Richard Koche, Doron Betel, Shubing Chen, Joan Massague, Ralph Garippa, Todd Evans and Michael Beer for their supports.

I owe my greatest appreciation to my family. Thank my Mom for her unconditional love. You help me to be where I can be right now. You help me make all the important decisions. Thank my Dad for being with me and taking care of me.

TABLE OF CONTENTS

CHAPTER 3: Investigation of the roles of JNK/JUN pathway in DE differentiation

LISTS OF FIGURES AND TABLES

CHAPTER 1: Introduction

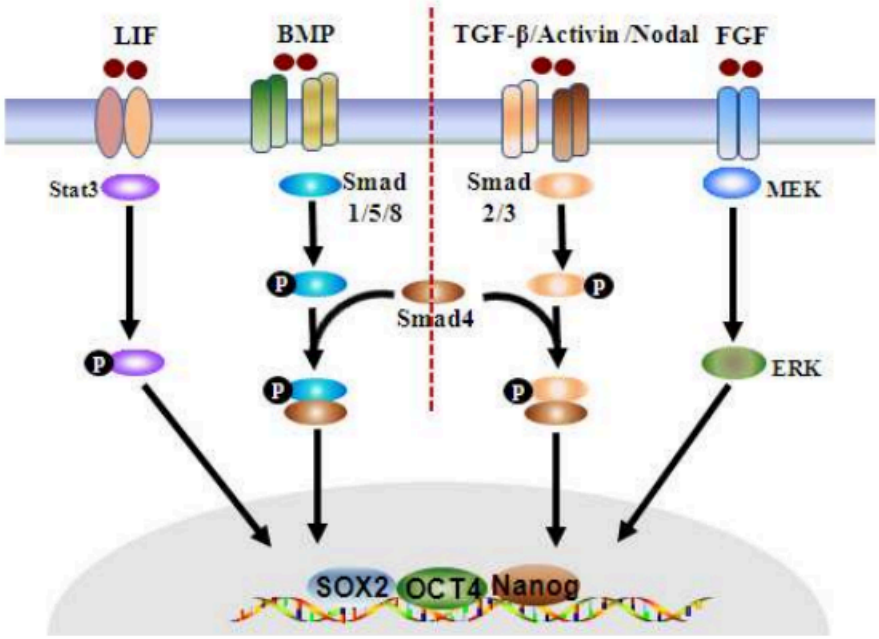
1.1 The unique nature of pluripotent stem cells

1.1.1 Signaling pathways of pluripotency maintenance and self-renewal

Pluripotent stem cells (PSCs) have the ability to unlimitedly self-renew and differentiate to any somatic cell lineage. Pluripotency describes the potential of a cell to give rise to all cell types in an adult organism. During mammalian development, pluripotent cells only exist transiently in early blastocyst embryos before quickly differentiating into the endoderm, mesoderm, and ectoderm lineages. This natural pluripotent state can be maintained in vitro in self-renewing embryonic stem cells (ESCs) derived from the inner cell mass (ICM)[1-3]. Pluripotency can also be experimentally induced by gene overexpression or chemical treatments [4]. As first demonstrated by Yamanaka and colleagues in mouse fibroblasts, somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) that closely resemble ESCs [5]. Pluripotent stem cells, ESCs and iPSCs are capable of unlimited self-renewal while maintaining pluripotency, thus providing an excellent tool for studying the transient early developmental stage that these cells represent. Understanding pluripotency regulation will also enhance the ability to utilize PSCs for studying development and disease, and for further generating disease relevant cells for regenerative medicine.

As pluripotent cells exist only transiently in the embryo, the ability to maintain the pluripotent state in vitro relies on our understanding of the pluripotency network governed by transcription factors and signaling pathways. Mouse ESCs (mESCs) and human ESCs (hESCs) employ different signaling pathways to maintain pluripotency: mESCs can be maintained by LIF activation of the JAK/STAT pathway and BMP signaling[6-9], whereas hESCs are dependent on FGF2 and TGF-β signaling pathways and cannot be derived in mESC growth conditions[3,10-16]. These differences were thought to be due to the intrinsic species differences between mouse and human until the derivation of PSCs from the post-implantation mouse epiblast (mEpiSCs) using FGF2 and Nodal (a member of the TGF-β super-family) which closely resemble the signaling required for hESC maintenance[17,18]. Therefore, the pluripotency state of hESC and mEpiSCs is characterized as primed pluripotency, because the signaling requirements are for epiblast development at the post-implantation stage. While, the pluripotency state of mESC is characterized as naïve pluripotency, because the signaling requirements are for inner cell mass development at the pre-implantation state[19]. Under this primed self-renewal culture condition, hESC has established a unique transcriptional network that is governed by the core transcription factors OCT4, NANOG and SOX2. These lineage transcription factors cooperate signaling transcription factors to sustain gene expression in self-renewal and suppress genes expression related to lineage differentiation (Fig. 1.1).

Figure 1.1 Signaling pathways of pluripotency maintenance and self-renewal

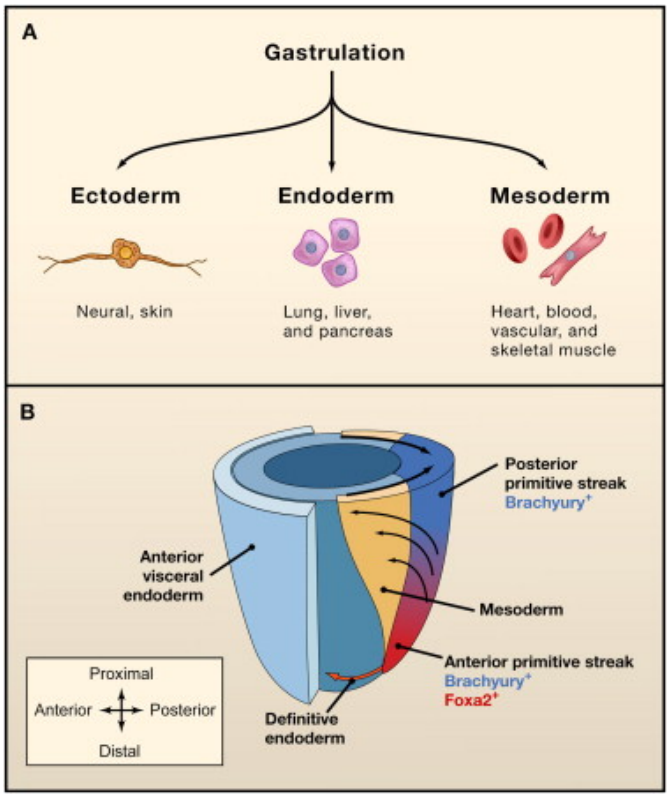


An illustration of signaling pathways in mESC (Lif and Bmp_left) and hESC (Nodal and FGF_right). Lineage transcription factors OCT4, NANOG, SOX2 cooperate with distinct signaling transcription factors to regulate PSC pluripotency. Reprinted from[20] Copyright 2013, with permission from IntechOpen.

1.1.2 Signaling pathways of somatic lineage differentiation

Utilizing model organisms such as the mouse, developmental biologists have uncovered requirements for discrete signaling pathways and precise spatiotemporal coordination during early embryonic development and organogenesis[21]. Using mouse embryo development as an example, somatic lineage specifications occur at the gastrulation stage of embryogenesis. Gastrulation is a complex event organized by different signaling pathways with cell migration events taking place after blastocyst implantation[22]. Wnt and Nodal are the most important pathways in regulating the formation of primitive streak, which is one of the hallmarks of gastrulation[23,24]. Lineage allocation occurs along the primitive streak according to the signaling gradient of Nodal in the anterior-posterior axis. Epiblast cells moving to the most anterior region of primitive streak form the structure of node[25], where Nodal signaling level is the highest. High grade Nodal signaling and Wnt signaling are required for the formation of definitive endoderm[26] (Fig. 1.2). As Nodal signaling gradient decreases and Bmp4 signaling gradient increases from the anterior to posterior primitive streak, different types of mesoderm derivatives like paraxial and lateral plate mesoderm are patterned according to the corresponding gradient [27]. Ectoderm lineage is specified from the cells remaining in the epiblast that are not exposed to strong Nodal and Bmp4 signaling [28,29].

Figure 1.2 Lineage specifications at the gastrulation stage of embryogenesis.



(A) Definitive endoderm, mesoderm and ectoderm are the three somatic lineages specified from gastrulation.

(B) The anatomy and cell movement of a mouse embryo during gastrulation. Red arrow depicted the movement of earliest definitive endoderm formation.

Reprinted from [30] Copyright 2008, with permission from Elsevier.

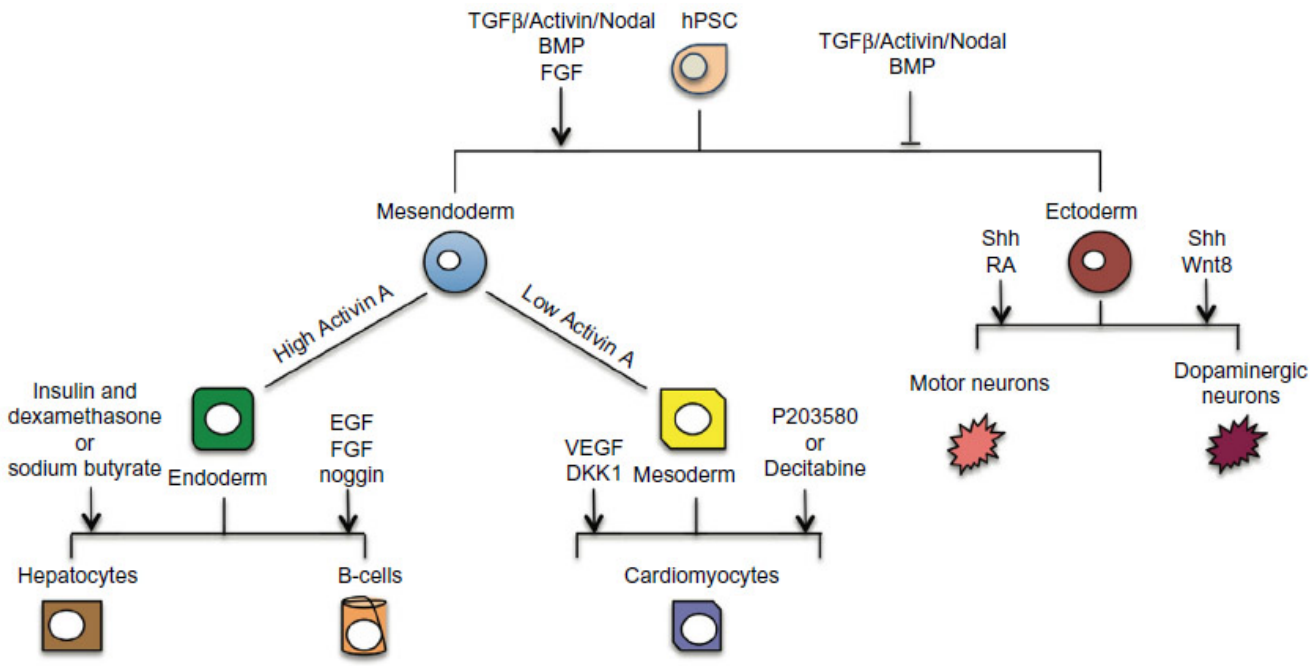
1.1.3 In vitro directed differentiation

These findings form the basis for guided differentiation of human embryonic and induced pluripotent stem cells (hESCs/hiPSCs) into the three embryonic lineages and their derivatives [31,32]. Using recombinant protein and small molecules, many differentiation protocols have been made to induce three germ layer lineages differentiation (Fig. 1.3). For instance, the guided differentiation of hESCs/hiPSCs into definitive endoderm (DE) is induced by Wnt and Activin A [33,34]. Mesoderm lineage derivatives (paraxial and lateral plate mesoderm) can be efficiently induced with the correct dosages of Nodal, BMP4, WNT signaling inputs. Dual inhibition of Nodal and BMP4 signaling by using Smad inhibitor Noggin and SB431542 induce highly efficient neural ectoderm lineage differentiation [35].

Following the guideline of developmental signaling pathways learned from model organism, terminally differentiated human cells can be differentiated from hPSC in vitro [36]. Understanding and improving in vitro directed differentiation offer two unique advantages. From a regenerative medicine perspective, terminally differentiated cells or lineage progenitor cells are great resources for cell replacement therapy. For example, hESC derived pancreatic β-cells[37,38] can potentially be used to treat Type 1 diabetes. From a developmental biology perspective, in vitro directed differentiation of hESC or hiPSC has become a novel model to study the mechanisms of human genetic diseases. For example, the mechanisms of *GATA6* haploinsufficiency associated pancreatic agenesis in

human have been elucidated by many different groups using gene editing technology and in vitro hPSC differentiation [39] [40-43].

Figure 1.3 In vitro directed differentiation of hPSC



The roadmap of endoderm, mesoderm and ectoderm lineage differentiation. Recombinant proteins and chemical inhibitors are used to activate or inhibit signaling pathways transduction. Reprinted from[44] Copyright 2015, with permission from Dove Medical Press Ltd.

1.2 Definitive endoderm differentiation

1.2.1 Signaling pathways involved with DE differentiation

In the TGF-β pathway, Nodal as a ligand transduces its signaling via Type 1 (ACVR1B/ALK4) and Type 2 (ACVR2A/2B) receptors and a co-receptor (Cripto/EGF-CFC)[45]. Upon ligand receptor binding, Type 1 receptor kinases phosphorylate the c-terminal of cytosolic receptor SMAD (SMAD2/3). Phosphorylated SMAD2 and SMAD3 can form homodimer or heterodimer first, and then interact with co-SMAD (SMAD4) as a heterotrimer complex[46]. Then, the SMAD complexes (SMAD2/3-SMAD4) can bind DNA with lineage determining transcription factors and/or chromatin modifying enzymes to regulate gene transcription [47].

In the canonical WNT/β-catenin pathway, the downstream of transcription factor cytosolic β-catenin is normally inhibited by the APC/Axin/GSK-3β complex (β-catenin destruction complex) in the absence of WNT ligands[48]. In the presence of WNT ligands, WNT3a proteins interact with Frizzled and LRP5/6 receptors, which destabilize β-catenin destruction complex and in turn derepress proteasome degradation of β-catenin[48]. Therefore, accumulated nuclear β-catenin interacts with TCF/LEF transcription factors and activates downstream gene expression.

To induce efficient in vitro definitive endoderm differentiation, WNT3a ligand can be substituted by small molecule GSK-3β inhibitor (CHIR99021).

Nodal is normally replaced by Activin A[34,49], another member of TGF-β growth factor that binds to ALK4 receptor without the requirement of Cripto[50]. Definitive endoderm differentiation is a multistep process. The first step of definitive endoderm differentiation is the mesendoderm lineage specification, which is the progenitor cell type in the primitive streak for mesoderm and endoderm differentiation. One day treatment of CHIR99021 and Activin A can induce hESC differentiate to EOMES+ mesendoderm cells. Mesendoderm is the in vitro counter part of primitive streak, which has the potential of definitive endoderm and mesoderm lineage differentiation[51]. Next, two days treatment of Activin A promotes mesendoderm cells differentiate to $CXCR4^{+}SOX17^{+}FOXA2^{+}$ definitive endoderm cells [49,52].

1.2.2. Genetic logic of DE differentiation

Genetic knockout studies demonstrate the important roles of Nodal and Wnt signaling pathways in primitive streak formation. *Nodal, Acvr1b, Smad2, Smad4, Ctnnb1*[53-61]knockout mice were all embryonic lethal and had different defective phenotypes in gastrulation. Transcription factors *Eomes and Brachyury* (*T*) are prominent marker genes of primitive streak[22]. Motif and ChIP-seq analysis reveal that Smad2/3 and Tcf3 motifs and binding sites are located upstream of *Eomes* transcription starts sites[62]. Inhibition of Nodal or Wnt signaling decreases *Eomes* and other mesendoderm gene expression[62]. Therefore, signaling transduction

genes in the Nodal and Wnt pathways are required for the induction of primitive streak transcription factors.

After mesendoderm lineage is established, Eomes cooperates with Nodal signaling transcription factor Smad2-Smad4 complex to drive key definitive endoderm gene expression such as *Gata6/4*, *Foxa2* and *Sox17*[63,64] (Fig. 1.4). ChIP-seq analysis shows a wide range of co-localization of EOMES, SMAD2/3 and GATA6 binding at endoderm enhancers [43,49]. Meanwhile, EOMES suppresses gene expression related to pluripotency maintenance and mesoderm development in the presence of strong Nodal signaling[64]. In contrast, Brachyury is the mesendoderm transcription factor for inducing mesoderm lineage specification[65,66]. In the presence of strong BMP4 signaling, Brachyury interacts with BMP signaling transcription factor SMAD1/5/8 to drive key mesoderm gene expression [67].

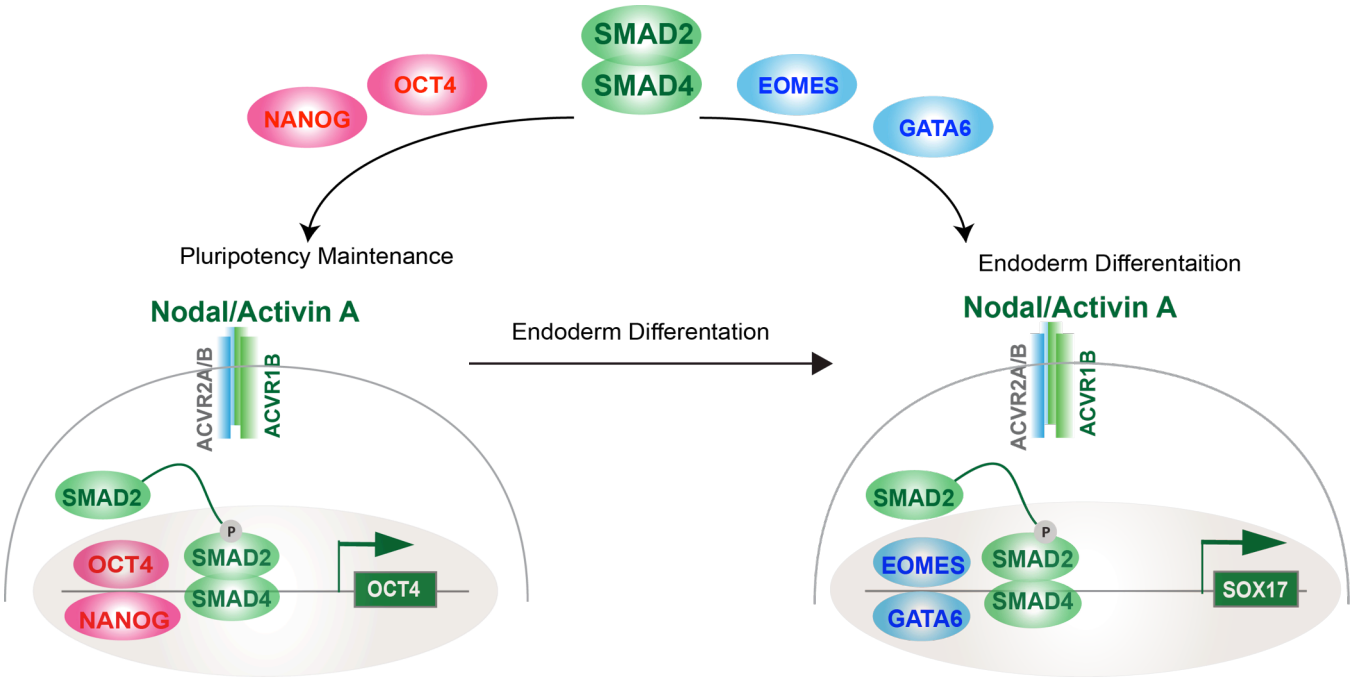
The sequential differentiation events of endoderm and mesoderm differentiation are governed by a network of signaling and lineage transcription facotrs [68]. Although the roles of these transcription factors were validated in mouse and other animal models, their exact functions are still not completely understood. For example, *Nodal* KO mice lacked *Eomes* and *Wnt3a* expression at E6.5 compared to wild type embryos, while *Smad2* KO mice maintained *Eomes* and *Wnt3a* expression[69]. *Lefty2* as a Nodal signaling regulated downstream gene was not expressed in both *Nodal* and *Smad2* KO mice, which demonstrated that Nodal signaling pathway was indeed inactivated in *Smad2* KO mice. This study concluded that primitive streak gene *Eomes* was regulated by

Smad2-independent Nodal signaling pathway in vivo. Interestingly, *Smad3* KO mice are viable and does not have an embryonic development defect [70-72]. These studies demonstrate that Smad2 is the essential Nodal signaling transcription factors during gastrulation. To further investing the role of Smad3 in embryonic development, *Smad2* and *Smad3* compound knockout mice were generated to evaluate the impact of *Smad3* in the context of *Smad2* deficiency [73]. *Smad3*-/-*Smad2*-/+ mice have impaired production of anterior axial mesendoderm. *Smad2* and *Smad3* double knockout mice lacked entire mesoderm induction, which exhibit a much severer phenotype compared to *Smad2* single knockout mice[73]. Therefore, although Smad3 is not absolutely required for embryonic development, but gastrulation is regulated by the combinatorial activities of Smad2 and Smad3.

Given the critical role of Nodal signaling and Smad2/3 in gastrulation, it would be reasonable to assume *Smad4* KO mice would have similar phenotypes as *Nodal* or *Smad2/3* mutant, because Smad4 is the common co-factor for Smad2/3 transcription factor[46]. *Smad4* KO mice have reduced size of egg cylinder and lack mesoderm differentiation, which is similar to the phenotypes of *Smad2* KO mice[59,60]. In order to exclude the impact of *Smad4* deletion in extraembryonic tissue, $Sox2^{Cre}Smad4^{fl/fl}$ mice were generated to conditionally delete *Smad4* in the epiblast. Deletion of *Smad4* in the epiblast does not prevent mesoderm development, although anterior primitive streak and definitive endoderm development are impaired as expected [74]. This study argues against the role of Smad4 as the central effector molecule in the Nodal-Smad2/3 signaling axis. Because Smad4 is selectively participating Nodal signaling during

mouse embryonic development, the mechanism of Smad4-indentpent Nodal signaling pathway remains to be further investigated. In contrast to human, SMAD4 is indeed required for cardiac mesoderm precursor formation[75], as *SMAD4* knockout hESC failed to respond to Activin A and BMP4 mediated cardiac mesoderm differentiation. It will be of interest to determine the roles of these transcription factors in human embryonic development. Forward genetic screens in model organisms provide a powerful approach for uncovering previously unsuspected regulators of development[21]. However, this approach is not directly applicable to studies of human embryogenesis, posing a challenge for identifying unique regulatory mechanisms underlying the developmental control of the human genome.

Figure 1.4 Nodal signaling and transcription factors in hESC pluripotency maintenance and DE differentiation



Phosphorylated SMAD2/4 complex regulates hESC pluripotency maintenance and endoderm differentiation depends on its co-binding transcription factors. Under self-renewal culture condition, SMAD2/4 complex cooperates with OCT4 and NANOG to drive pluripotency gene expression. Under endoderm differentiation culture condition, SMAD2/4 complex cooperates with mesendoderm/endoderm transcription factors EOMES and GATA6 to drive definitive endoderm gene expression.

1.3 Functional genomics approach to study pluripotency and differentiation

1.3.1 Forward genetic screen

Forward genetics is powerful approach to study well-defined biological phenomena, because unbiased genetic screens would eventually identify the causative genes for these phenotypes[76]. Random mutagenesis by retroviral insertion or chemical mutagen was extremely valuable in the pre-genomic era for loss-of-function genetic screens[77,78], as it does not require a reference genome. For example, a retroviral integration screen was performed in mESCs to generate mutant mice, which led to the identification of a key developmental regulator gene *Nodal*[53]. Chemical mutagenesis screens in mice also uncovered a critical requirement of the cilium, a cellular organelle, in spinal cord development and Hedgehog signal transduction[79]. Gain-of-function screens were also used to identify novel genes. Most famously, the transcription factor *Nanog* was identified from a cDNA library over-expression screen based on its ability to sustain mESC self-renewal without LIF activation of JAK/STAT signaling[80]. This essential role of Nanog in pluripotency maintenance is further supported by an independent study showing that the loss of Nanog expression caused mESCs to lose pluripotency and spontaneously differentiate[81]. After completion of human genome project in the post-genomic era, forward genetic screen has transited from a random mutagenesis approach to a genome-wide perturbation approach. Functional genomics utilizes systems loss- and gain-of-function perturbations to interrogate

genes for their roles in various biological processes. The development of RNAi and CRISPR screen technologies greatly improve the efficiency and application of forward genetic screen in cell-based assays.

1.3.2 RNAi and CRISPR screen

RNA interference (RNAi) technology has greatly accelerated the speed of functional genomics studies, as RNAi based genetic screens can be easily integrated with high-throughput sequencing and automation technology to facilitate the process of target identification. In RNAi, synthetic small interfering RNA (siRNA) or short hairpin RNA (shRNA) are typically used to silence gene expression at the post-transcriptional level by forming double strand RNA with complementary mRNA transcripts. As a result, the double stranded RNA-complex is degraded by endogenous RNA interference machinery and the expression level of mRNA for protein translation is downregulated[82].

In contrast to RNAi knockdown, CRISPR/Cas9 gene editing technology alters gene expression at the pre-transcriptional level. The Cas9 protein is an RNA guided DNA nuclease, repurposed from the prokaryotic immune system, that must be exogenously introduced to eukaryotic cells. A guide RNA (gRNA) can target Cas9 to a complementary 20bp genomic DNA sequence with a 5'-NGG-3' motif at the end (PAM sequence). The Cas9-gRNA duplex introduces a DNA double-strand break at the site 3 base pairs upstream of PAM sequence. This DNA double stand break is repaired by either non-homologous end joining (NHEJ) or homology-directed repair. The result from repairing double strand

breaks often produces insertion or deletion (indel) mutations that alter the genomic DNA sequences permanently. CRISPR-Cas9 mediated knockout is achieved by an out-of-frame mutation or an in-frame mutation at a key functional domain of the coding sequence of a gene [83,84].

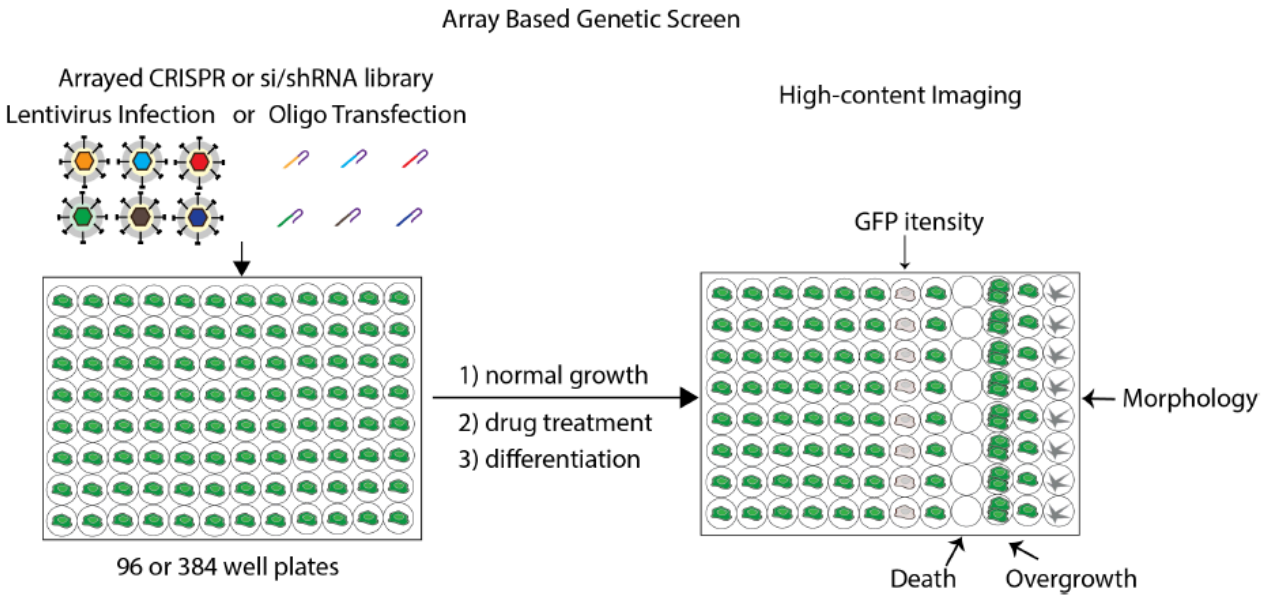
RNAi or CRISPR screens can be conducted in an arrayed or pooled format. Array based screens take advantage of high-throughput robotic automation platforms to perform cell culture and analysis in multi-well plates (e.g. 96- or 384-well plates). In an arrayed screen, cells of interested are matched to a unique short interfering RNA (siRNA), short hairpin RNA (shRNA) or guided RNA in each well. After a period of cell culture with the assay of interest and genetic perturbation, the cellular phenotypic readouts such as cell proliferation, cell death, morphology or marker gene expression can be analyzed by high-throughput imaging platforms such as luminescence plate reader or high-content automatic microscopy (Fig. 1.5).

Pooled based screen takes advantage of high-thought sequencing technology to deconvolute target genes from a large-scale pool of mixed cells cultured in large tissue culture vessels. Cells of interests are infected with shRNA or CRISPR lentiviral libraries at a low multiplicity of infection (MOI) to ensure most of the cells only harbor one shRNA expression cassette (Fig. 1.6). The integrated expression cassette containing the sequence of shRNA or gRNA can be used as a barcode to identify the targeted genes by high-throughput sequencing. Given that RNAi is inherently transient, integration into the genome by retroviral transduction also offers the advantage that siRNAs are consistently

expressed through the course of a long screen. In a pooled screen, the cellular phenotypes of proliferation and death are identified through a selection process. In a positive (enrichment) selection screen, cells with a proliferation or survival advantage will be enriched after a period of cell culture and genetic perturbation. The remaining cells can be collected and sequenced to identify the enriched shRNA of gRNA compared to the non-selected group of cells. In a negative (drop-out) selection screen, cells with a proliferation or survival defect will be depleted after a period of cell culture and genetic perturbation. The identities of missing cells are uncovered by comparing the surviving selected cells to the non-selected group of cells.

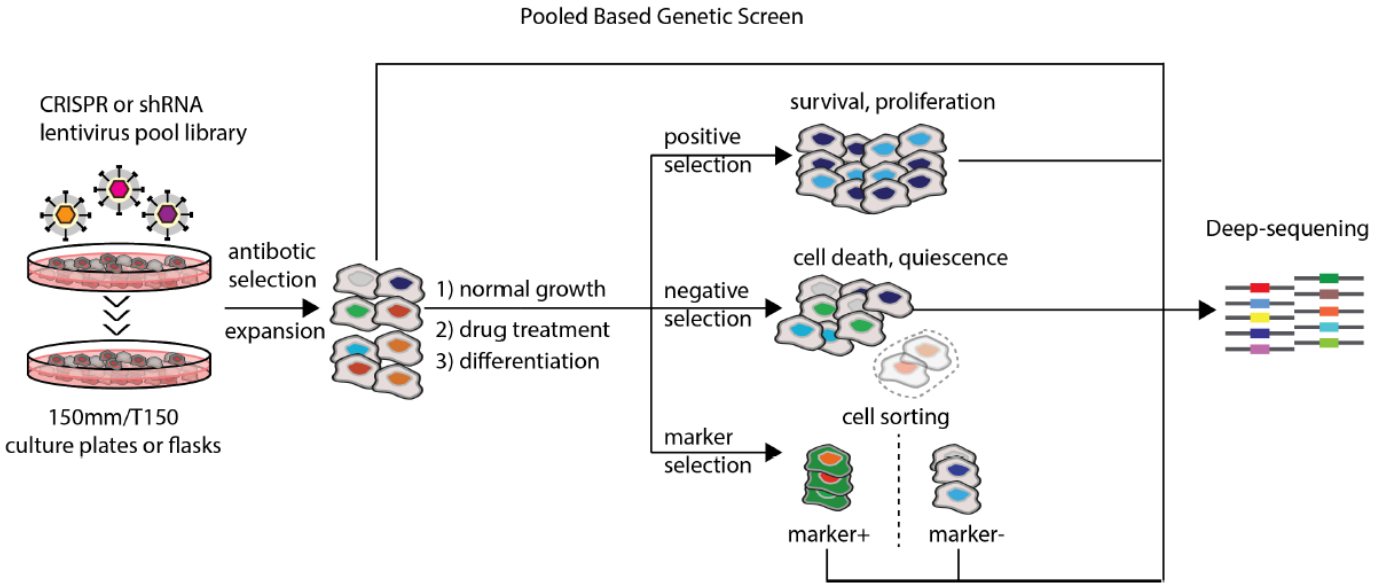
While most pooled screens use proliferation and survival/death phenotype to select cells, it is also possible to select cells based on marker gene expression. Fluorescence or magnetic activated cell sorting can be used to enrich cell populations based on the presence or absence of a selective marker. A selective marker can be a fluorescence-reporter gene or a surface or intracellular antigen can be stained by antibody. Marker positive and negative cells are sorted and sequenced in order to identify the relative enrichment of shRNA or gRNA in each population (Fig. 1.6). Using this strategy, we can identify genes what play roles in promoting or inhibiting the expression of this marker gene during a biological process.

Figure 1.5 Schematic of array based genetic screen



Most arrayed based screens use a reverse transfection/infection approach. Synthetic siRNA, gRNA or lentiviral particles are pre-distributed in 96/384-well plates. After reagent reconstitution according to manufacturer's instruction, cells are distributed to each well by using an automatic liquid dispensing instrument and mixed with RNAi and gRNA reagents. A forward transfection/infection approach can be used as an alternative solution if cells need to be pre-seeded one day earlier before transfection/infection. After the initial transfection/infection step, cells are cultured with media to promote normal cell growth, drug selection or differentiation. The cellular phenotype of interest (overgrowth, death, differentiation) can be evaluated by high-content imaging based on the quantitative signals from of reporter cell line (GFP) or immunostaining.

Figure 1.6 Schematic of pool based genetic screens.



In a typical pooled screen, a lentivirus library is first prepared to express an array of shRNAs or CRISPR gRNAs. With a low multiplicity of infection, most of the infected cells will carry one shRNA or gRNA integration. The infected cells will then be subject to positive selection, negative cell selection, or marker selection. For positive and negative cell selection, the target population is the population that survives the selection, and the control population can be the infected cells either before selection or cultured in parallel without selection. For marker selection, the target and control populations can be the population with or without the marker expression, respectively. Deep sequencing and sequence deconvolution will be performed to quantify the degree of enrichment or depletion of each shRNA or gRNA in the target population compared to the control population.

1.3.3 Genetic screens to interrogate the maintenance of pluripotency

Ivanova et al. designed a focused shRNA library targeting 70 genes encoding DNA binding proteins that were downregulated after retinoic acid (RA) induced differentiation in mESCs [85]. They used mESC proliferation as a readout for pluripotency maintenance. shRNA targeted mESCs were labeled with GFP reporter and co-cultured with unlabeled wild-type mESCs. The proliferation phenotype was evaluated based on the percentage of GFP+ cells by flow cytometry. This competition screen represents the first RNAi screen in mESC and identified *Tbx3*, *Esrrb*, *Tcl1*, and *Dapp4* in addition to known pluripotency regulator like *Oct4* and *Nanog*. The success of this mini-scale screen has inspired larger scale screening efforts. An arrayed siRNA screen targeting 1,008 chromatin proteins was performed to identify the genes required for mESC growth [86]. The knockdown of 68 genes caused varying degrees of phenotypes in mESC viability and morphology as scored by microscopy imaging. These assays have identified genes that regulate cell viability or morphology in mESCs as well as other cell types. It would be of interest to determine which genes identified from the screen are specifically required for maintaining pluripotency, which may be determined by examining the effects of the genetic perturbation on the expression of pluripotency genes (e.g. *Oct4* and *Nanog*) and the differentiation potential of mESCs. Only a couple of screens have been performed in hESCs so far. Ng and colleagues reported the first genome-wide siRNA screen using a transgenic *OCT4* reporter and identified *PRDM14* as a human pluripotency

regulator [87]. Unlike in hESCs, *Prdm14* knockdown in mESCs had no effect on *Oct4* and *Sox2* expression, highlighting the necessity of performing such screens in hESCs to identify unique regulators of pluripotency in hESCs.

1.3.4 Genetic screens to interrogate the exit of pluripotency and differentiation

Most early studies have focused on untangling the mechanisms of pluripotency maintenance, however the natural fate of a pluripotent cell in the developing embryo is not to divide repeatedly in a pluripotent state but to differentiate into somatic lineages when exposed to differentiation cues. To study this complex process, ESCs can be cultured in conditions devoid of signaling molecules necessary for ESC self-renewal or stimulated with strong differentiation signals. To identify regulators of the exit of pluripotency, Paddison and colleagues performed a shRNA screen targeting 312 gene involved in chromatin regulation[88]. A transgenic *Nanog-GFP* reporter was used as a readout, because of the rapid downregulation of *Nanog* expression after LIF withdrawal and RA treatment for neural differentiation. A collection of Baf complex genes (*Brg1*, *Baf47*, *Baf155*, and *Baf57*) was identified as repressors of pluripotency exit, as mESCs with these genes downregulated retained a higher level of *Nanog-GFP* expression after RA treatment compared to wild-type control.

hESC pluripotency maintenance is dependent on FGF2 and TGF-β signaling, which is distinct from the JAK/STAT and BMP signaling that maintains

mESC pluripotency. The difference in signaling requirements suggests that the route of exit of pluripotency and lineage specification may also be different between mouse and human development. Ng and colleagues performed a siRNA screen in a transgenic *Nanog-GFP* hESC reporter line to investigate different routes of exit of pluripotency using five differentiation conditions (withdrawal of FGF2/TGFβ, TGF-β inhibition, MEK inhibition, PI3K inhibition, and RA treatment) [89] Genes involved with DNA replication during S phase or G2/M transiton also stood out in multiple contexts. Depletion of DNA replication genes or G2 phase progression genes, or treatment with S/G2 phase cell cycle inhibitors induced hESCs to extend S and G2 phases and delay the exit of pluripotency after the withdrawal of FGF2 and TGF-β signaling. In turn, extension of S and G2 phases activates ATM/ATR mediated checkpoint signaling pathway to upregulate p53 activity and TGF-β agonist gene expression. As a result, the pluripotent state of hESC is sustained by the upregulated TGF-β agonist, which allows hESCs to have extra time to perform DNA damage repair in the priviledged pluripotency state when homologous recombination and damage repair proteins are highly expressed. By examining the exit of pluripotency in multiple contexts, this study was able to demonstrate the important role of S/G2 cell cycle progression in the regulation of pluripotency and differentiation.

The above screens all used RA treatment and/or withdrawal of self-renewing signals as a general strategy to induce exit of pluripotency and differentiation. However, these methods do not exactly recapitulate the signaling process by which cells exit pluripotency in vivo. Differentiation of definitive

endoderm, mesoderm, and ectoderm require distinct signaling pathways [30,31]. To fully appreciate the orchestration of pluripotency exit and lineage-specific differentiation, it would be necessary to conduct screens in lineage-specific differentiation conditions, and to interrogate the expression of both pluripotency genes and differentiation genes. This will help us to understand the control of pluripotency exit as well as lineage differentiation, which could be explored to improve in vitro differentiation to somatic lineage cells.

1.3.5 The emergence of CRISPR screen.

CRISPR screens have been performed successfully to identify drug resistance genes, tumor suppressors, essential genes, host factors of viral replications[90-93]. CRISPR-based perturbations screen strategy is similar to the arrayed and pooled screening methods established for RNAi screens (Fig. 1.6)[94]. CRISPR screen and RNAi screen have their unique pros and cons[95], therefore choosing the best screen method suitable for the experiments will achieve the best outcome. For instance, while CRISPR/Cas9 can be used to generate complete knockout phenotypes, it is not reversible. On the other hand, RNAi typically generates hypomorphic phenotypes that are reversible and tunable. CRISPR/Cas9 has considerably less off-target effects compared to RNAi [96], though the specificity of RNAi has also been improved in recent years [97].

CRISPR/Cas technology has numerous potential applications in the stem cell field. The advent of CRISPR/Cas9 gene editing technology has inspired enthusiasm for CRISPR-based screens to interrogate pluripotency and

differentiation. Yilmaz et al. recently performed a CRISPR screen to define essential genes of hPSC[98]. Taking advantage of the efficient CRISPR gene editing activity in haploid cells[99], they used genome-wide CRISPR library to generate a pool of mutant haploid hPSCs. hPSCs with mutated genes important for hPSC pluripotency and survival like *PRDM14*, *OCT4* were gradually lost during long-term culturing, while hPSCs with mutated genes restricting cellular growth in the p53-mTOR pathways were gradually enriched. This study provides a rich resource of the functional genome in hPSC pluripotency. CRISPR technology has already proven useful for fast generation of knock-in reporters to study pluripotency and differentiation[100]. These reporter cell lines will open the door to identify positive and negative regulators of lineage differentiation, and integrate functional coding and non-coding genome together to better understand the mechanism of development[101].

1.4 Thesis Aims

Despite the progress, much remains to be learned about mammalian embryonic development, especially human development. Notably, endoderm differentiation and pluripotency maintenance both involve the Nodal/TGF-β pathway [10,13,14,33]. It is unclear how hESCs interpret TGF-β signaling in two opposing ways: promoting self-renewal and promoting endoderm differentiation; thus, an unknown inhibitory mechanism has been postulated as a way to prevent hESCs from precociously differentiating into endoderm cells [102]. In addition, differentiation efficiencies vary among hESC/hiPSC lines [103,104], and homogeneous differentiation remains a major challenge, highlighting the need for discovering additional regulatory mechanisms controlling DE differentiation.

Aim 1: Establish a CRISPR platform to identify regulators of definitive endoderm differentiation

Combining in vitro directed differentiation of hESC and CRISPR technology, we intent to demonstrate the proof of concept of using forward genetics screen approach to identify unique genes related to human embryonic development. Previously, our lab has generated an inducible Cas9 hESC line for versatile gene editing activities. We planned to use this cell line to generate a SOX17-GFP reporter and conduct a pooled based CRISPR screen to identify the known and unknown regulators of DE differentiation.

Aim 2: Investigate the mechanism of definitive endoderm differentiation

Positive regulators are required for efficient DE differentiation; negative regulators impede DE differentiation. From the CRISPR screen, we aim to identify positive and negative regulators of DE differentiation. Next, we plan to validate these genes by generating clonal KO cell lines and assess their phenotypes. Given the important roles of SMAD2/3 and GATA6 in endoderm transcription factor network, we plan to perform ChIP-seq and ATAC-seq assays to study their functional impacts on transcription factors and chromatin landscapes during ESC-DE lineage transition.

CHAPTER 2: Identification of regulator of definitive endoderm differentiation through CRISPR screens

2.1 Introduction

CRISPR/Cas9 gene editing technology is a revolutionary genetic tool that accelerates the pace of discovery-driven research. Due to limited transfection and electroporation efficiency in hPSC, the challenge of efficiently introducing plasmids carrying Cas9 and guided RNA (gRNA) into hPSC hinders CRISPR's gene editing applications and potentials in hPSC. Therefore, our lab developed the iCRISPR genome-engineer platform in hPSC to overcome this obstacle[105]. Using a pair of highly efficient Transcription activator-like effector nucleases (TALENs) targeting the AAVS1 locus, we inserted a doxycycline-inducible Cas9 expression cassette to this locus. After clonal line selection and expansion, we successfully generated many iCas9 hPSC lines that express Cas9 upon doxycycline selection. gRNA can be introduced into hESCs either by synthesized gRNA transfection or gRNA lentivirus infection. This platform allows us to efficiently and rapidly create indel or premise mutations at desired locus.

Taking advantage of this iCRISPR platform, we developed an one-step selection-free knock-in strategy to make reporter cell line[100]. With the elimination of floxed allele selection cassette normally found in a reporter construct, the efficiency of plasmid transfection and homology directed repair is vastly improved. Reporter hESC line can be generated in a shorter time frame compared to the traditional two-step floxed allele method. We have generated *OCT4-2A-GFP* and

PDX1-2A-GFP hESC lines based on this method and validated their activities of reporting lineage differentiation.

We plan to use iCas9 reporter hESC lines and CRISPR screening technology to establish a platform to screen regulators of human embryonic development. Lentivirus CRISPR based genome-wide libraries such as GeCKO and Brunello can be easily adapted into our iCRISPR platform[90,106]. Lentiviral integrated gRNA expression cassette can be used as a barcode for Hi-Seq to identify the corresponding mutation. A low multiplicity of infection (MOI) ratio is needed to ensure the majority of infected cells to receive lentiviral genome integration from one virus. Definitive endoderm formation is the first step of endoderm lineage differentiation, however, no genome-wide screen has ever been performed to specially identify regulators of DE differentiation in human. Here, we conducted genome-scale CRISPR/Cas screens for high-throughput discovery of regulators of DE differentiation. Identification of known regulators indicates the success of our screen. Moreover, we identify several novel genes including five essential JNK/JUN pathway genes that inhibit DE differentiation.

2.2 Results and figures

2.2.1 Generation a SOX17-GFP Knock-in reporter

To screen for regulators of DE differentiation, we used HUES8 iCas9 cells, which express Cas9 upon doxycycline treatment [105] (Fig. 2.1a). We further employed a selection-free knock-in strategy [100] to integrate a GFP transgene into the *SOX17* locus to report the endoderm fate [107] (Fig. 2.1b-c). Both Activin A and CHIR99021 were required to induce $CXCR4^{+}SOX17^{+}$ (~80%) DE cells (Fig. 2.2a-b), and faithful GFP reporter expression was confirmed by immunostaining and flow cytometry (Fig. 2.2 c-d). As a positive control, the requirement for *EOMES* in DE differentiation[63,64] was confirmed by performing differentiation on iCas9 $SOX17^{GFP/+}$ hESCs infected with a lentivirus expressing an *EOMES*-targeting gRNA (Fig. 2.2e).

2.2.2 Genome-scales CRISPR screens

We first performed the screen using the pooled lentiviral human GeCKO v2 library consisting of 58,028 gRNAs targeting 19,009 genes (3 gRNAs per gene) [108]. To improve the confidence of screening hits, we then repeated the screen using a serum-free differentiation condition with the Brunello library [106] consisting of 76,441 gRNAs targeting 19,114 genes (4 gRNAs per gene). After infection of iCas9 $SOX17^{GFP/+}$ hESCs with the lentiviral gRNA libraries, DE differentiation was performed. SOX17-GFP^{+} endoderm and SOX17-GFP^{-} non-

endoderm cells were isolated through fluorescence-activated cell sorting (FACS) (Fig. 2.3a). The abundance of individual gRNAs in each population was determined by high-throughput sequencing.

2.2.3 Analysis of screen results

Two approaches were used to identify and rank hits. We first calculated the Z-score [109] for each gRNA based on the ratio of gRNA reads in the SOX17-GFP^{-} versus SOX17-GFP^{+} populations (Fig. 2.3b-c). 37 positive regulators and 28 negative regulators reached the Z-score cutoff (Z>2.0 or Z<-2.0) in both screens (Fig. 2.4) In the second approach, we used the MAGeCK robust ranking aggregation (RRA) algorithm to identify screening hits [110] (Fig. 2.5a-b). RRA values from the two screens correlated well. All 27 genes with a false discovery rate (FDR) lower than 0.05 in both screens met the Z-score cutoff (>2.0, or <-2.0) (Fig. 2.5a-b). The theoretical FDR predicted by MAGeCK from the GeCKO library screen is consistent with the experimental FDR replicated in the Brunello library screen (Fig. 2.5c-d). The top ranked positive regulator hits based on both methods included almost all non-redundant, cell-autonomously required genes in the Nodal pathway (*ACVR1B*, *SMAD2*, and *FOXH1*) as well as established endoderm TFs (*EOMES* and *MIXL1*) (Fig. 2.4) [26,68].

2.2.4 Validation of hits from the screen

We performed validation experiments on 33 high ranking hits based on the Z-score and FDR (Fig. 2.7a). Lentiviral vectors expressing gene-specific gRNAs were used to infect H1 iCas9 hESCs followed by individual differentiation assays, and two gRNAs were tested for each gene (Fig. 2.6a and 2.7b). We were able to verify endoderm TF genes including *EOMES*, *MIXL1* and genes in Nodal and Wnt signaling pathways, consistent with their known requirements in mice [57,59,61,63,111-113] (Fig. 2.6b). In addition, we verified four Hippo pathway genes (*NF2, TAOK1, PTPN14*, and *PPP2R4*) known to negatively regulate YAP1 nuclear activity [114,115] (Fig. 2.6b). Previous studies showed that transient siRNA knockdown of *YAP1* expression in hESCs led to upregulation of mesendoderm genes *MIXL1*, *EOMES* and *T* [116,117], but a specific requirement of YAP1 or other Hippo pathway members in DE formation has not been established. Supporting the high confidence in the top screening hits, we validated 24 genes out of 33 hits tested (Fig. 2.6c). Screening hits that were not verified are likely false positives, as also found in other large-scale screening studies, though more thorough examinations may be necessary to confirm hits with relatively subtle phenotypes (Fig. 2.7c). Only 4 of the 24 genes showed greater than 2-fold gene expression changes during the ESC-DE transition by RNA-seq analysis (Fig. 2.7d). This highlights the necessity for comprehensive functional analysis rather than relying on transcriptome analysis alone.

Figure 2.1 Generation of HUES8 *SOX17*$^{GFP/+}$ iCas9 cell line

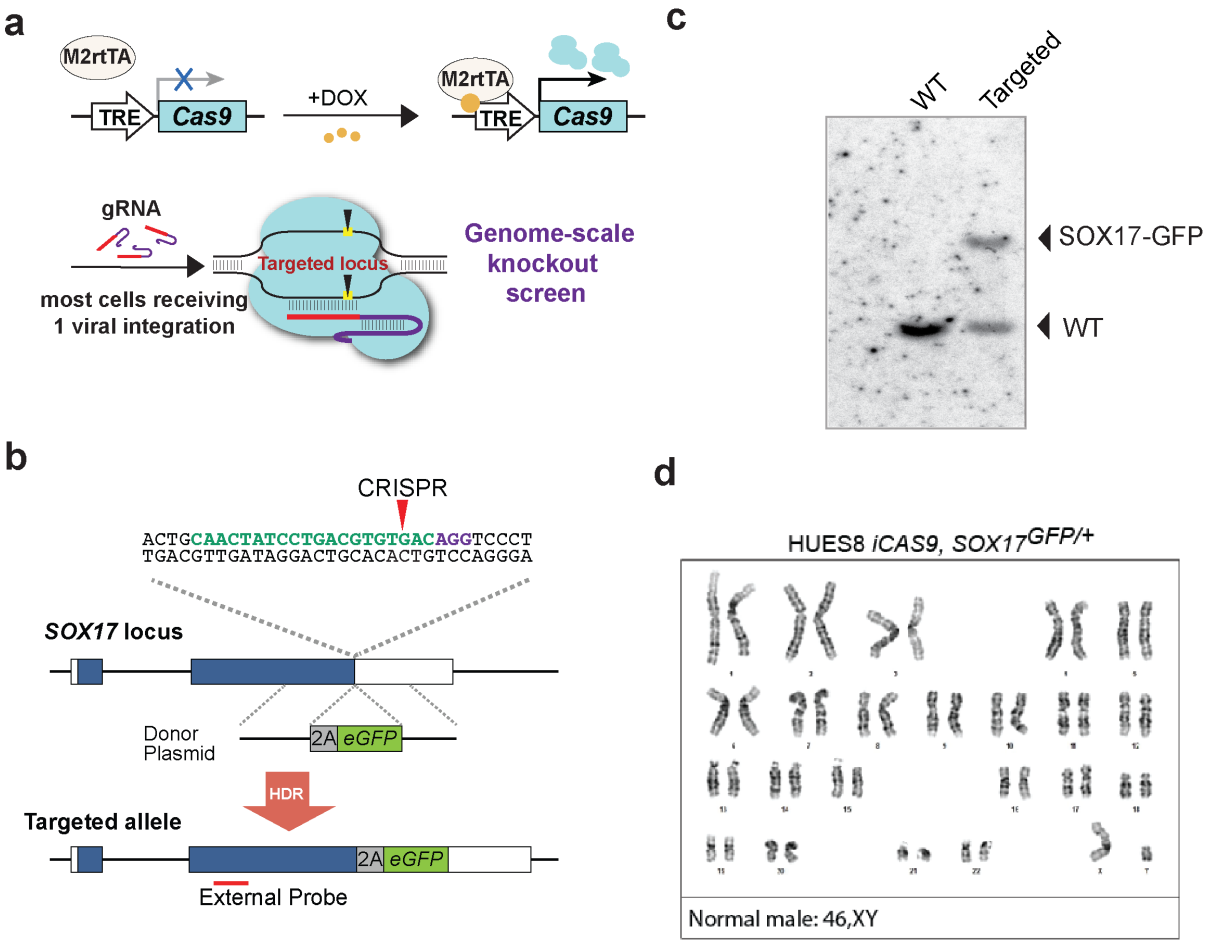


Figure 2.1 Generation of HUES8 *SOX17*$^{GFP/+}$ iCas9 cell line

a. Schematic shows the iCRISPR platform used for efficient genome editing. Doxycycline inducible Cas9 cells were transduced with lentiviruses expressing gRNAs for genome editing.
b. Knock-in strategy for generating the *SOX17*$^{GFP/+}$ reporter cell line. The PAM sequence is labeled in purple, the gRNA sequence is labeled in green.
c. Southern blotting analysis of *SOX17*$^{GFP/+}$ cell line. 5' external probe was used to detect the integration of the GFP cassette at one of the *SOX17* alleles.
d. Karyotyping result of HUES8 *SOX17*$^{GFP/+}$ cell line.

Figure 2.2 Validation of SOX17-GFP reporter cell line

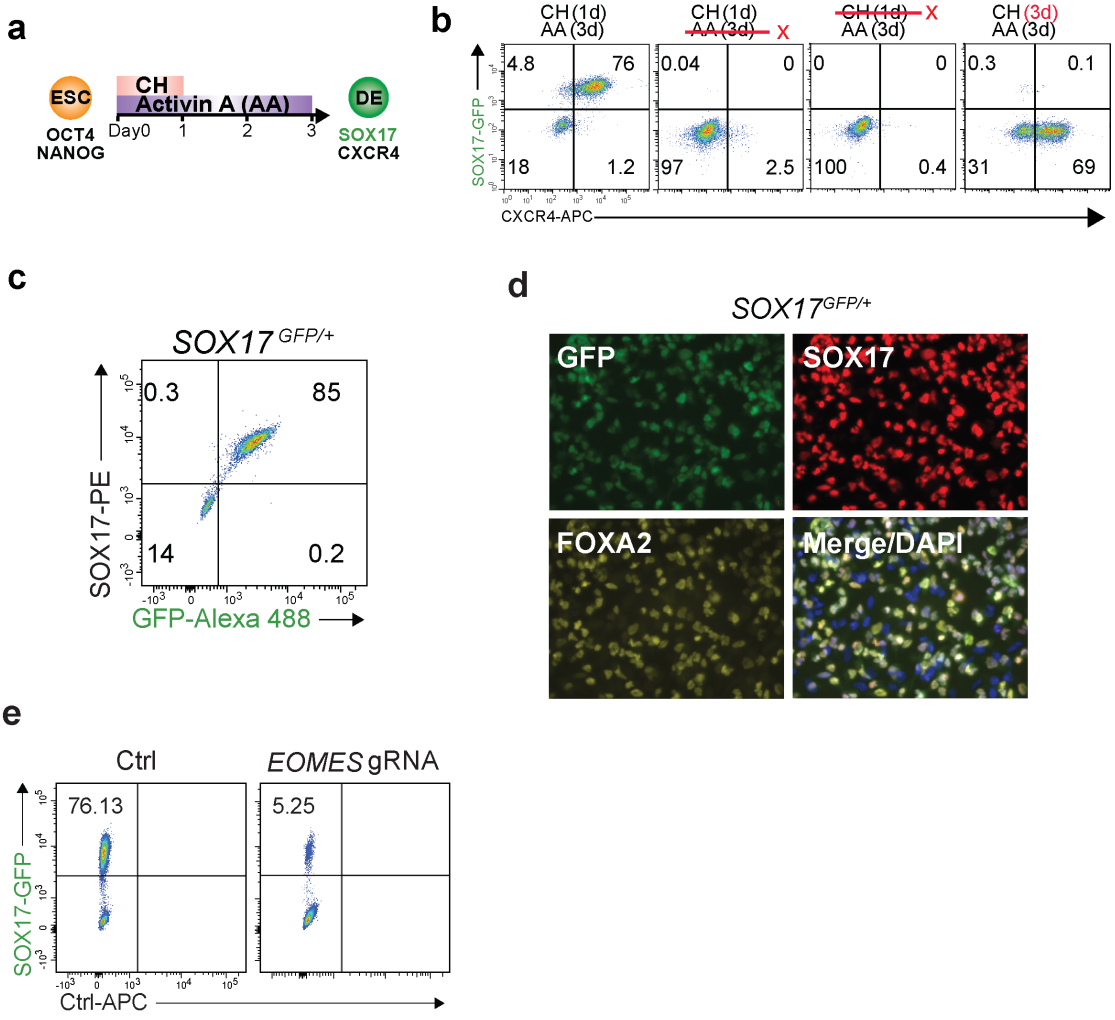


Figure 2.2 Validation of SOX17-GFP reporter cell line

a. DE differentiation protocol. CH: CHIR99021 from Day (D) 0 to D1; AA: Activin A from D0 to D3.
b. Representative FACS dot plots of DE cells stained for CXCR4. The FITC channel was used to detect live GFP fluorescence. Treatment and duration are indicated at the top of each FACS plot.
c. Representative FACS dot plots of DE cells differentiated from the *SOX17*$^{GFP/+}$ cell line co-stained for SOX17 and GFP.
d. Immunostaining of DE cells differentiated from the *SOX17*$^{GFP/+}$ cell line for GFP, SOX17, and FOXA2 expression. Scale bar, 100 μm.
e. Representative FACS dot plots of DE cells differentiated from *SOX17*$^{GFP/+}$ cells that were infected with the *EOMES*-gRNA lentivirus (at MOI 0.36) with puromycin selection. The control sample (Ctrl) represents DE cells differentiated from uninfected WT *SOX17*$^{GFP/+}$ cells.

Figure 2.3 Genome-scale screens identify positive and negative regulators of DE differentiation

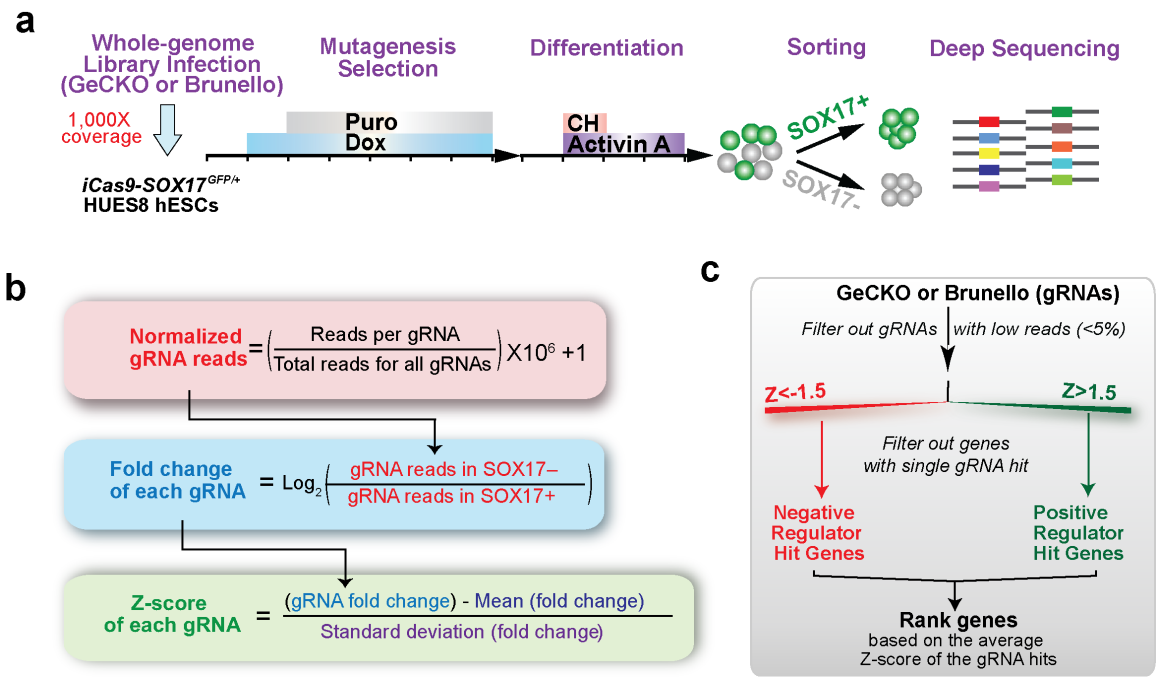


Figure 2.3 Genome-scale screens identify positive and negative regulators of DE differentiation

- **a.** Genome-wide CRISPR screen schematic. Each line segment on the horizontal arrows indicates 1 day of media and chemical treatment.
- **b.** Method of calculating the Z-score of each gRNA from raw read counts.
- **c.** Method of calculating and ranking the Z-score of each hit gene.

Figure 2.4 Z-scores analysis of CRISPR screen results

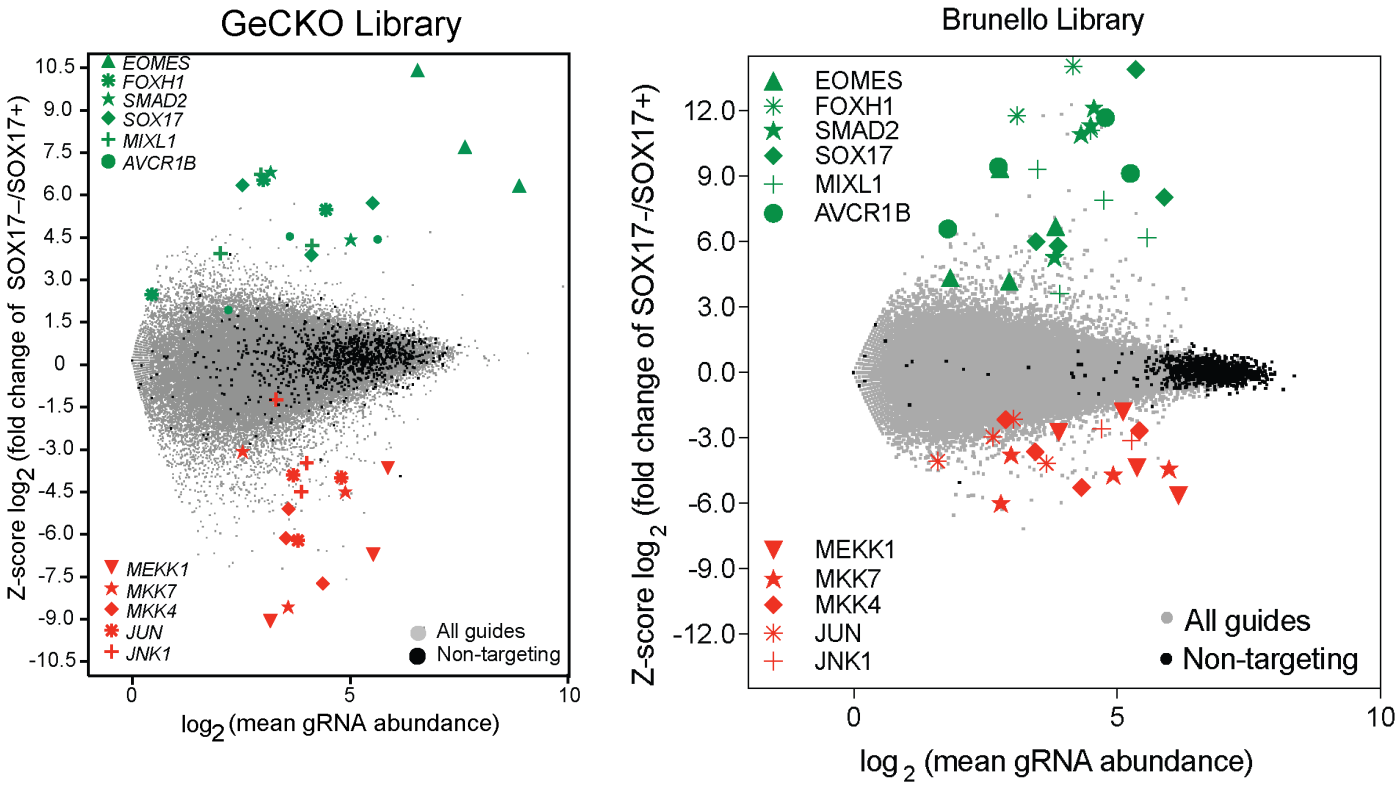


Figure 2.4 Z-scores analysis of CRISPR screen results

Scatter plots of the gRNA distribution from GeCKO and Brunello screen. Y-axis: Z-score of $\log_2$ fold change of $SOX17^-$ vs $SOX17^+$. X-axis: the mean abundance of gRNA reads in the $SOX17^-$ and $SOX17^+$ populations. Each grey dot represents an individual targeting gRNA. Each black dot represents a non-targeting control gRNA (1,000 total). Selected positive and negative regulator hits are labeled in green and red, respectively.

Figure 2.5 Robust Ranking Aggregation analysis of CRISPR screen results

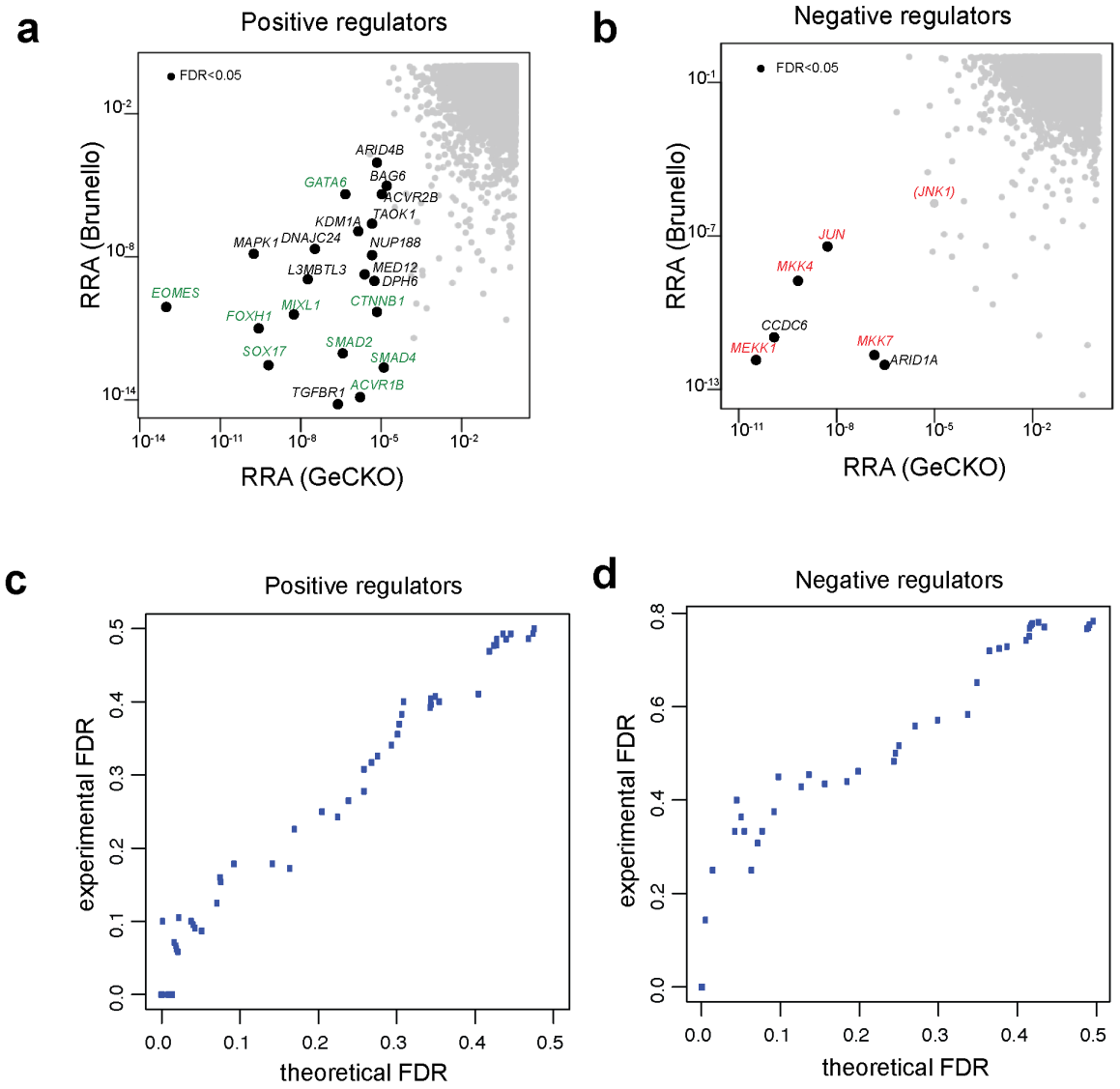


Figure 2.5 Robust Ranking Aggregation analysis of CRISPR screen results

a. Robust Ranking Aggregation (RRA) score for positive regulators from both screens. Genes with FDR<0.05 from both screen are indicated by black dots. Known positive regulators are labeled as green texts.

b. Robust Ranking Aggregation (RRA) score for negative regulators from both screens. Genes with FDR<0.05 from both screen are indicated by black dots. JNK pathway genes are labeled as red texts. JNK1 is shown in parenthesis as it just missed the FDR cutoff (JNK1 Brunello screen FDR<0.05, GeCKO screen FDR=0.056).

c.d Theoretical FDR calculated for GeCKO screen based on MAGeCK v.0.5.7. For each theoretical FDR, the replicated fraction (f_{repl}), is the fraction of genes below that FDR which are also below that FDR in the Brunello screen. Experimental FDR is 1-f_{repl}.

Figures 2.5a-d are contributed by Mike Beer.

Figure 2.6 Validation of gene hits identified from the genome-scale screens

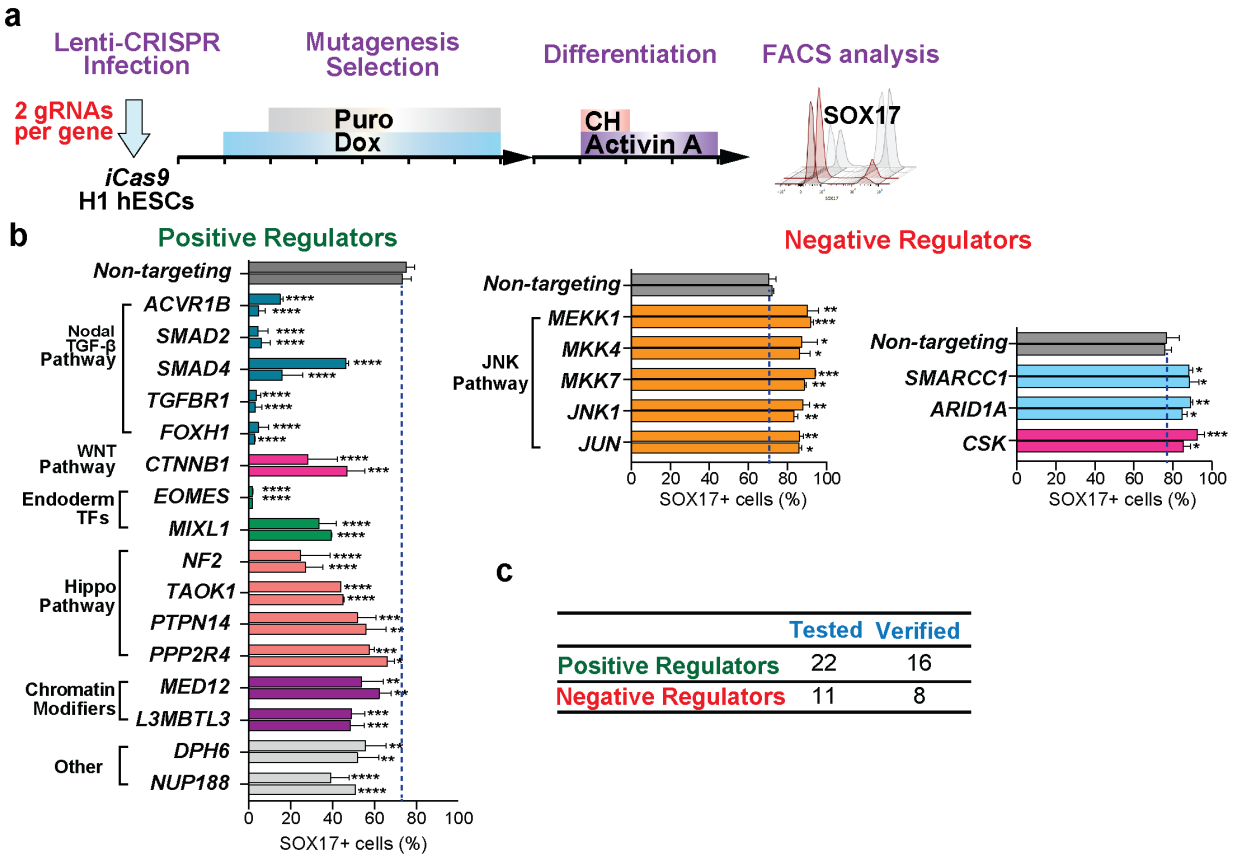


Figure 2.6 Validation of gene hits identified from the genome-scale screens

- **a.** Schematic showing validation of top hit genes using individual lentiviruses expressing different gRNAs.
- **b.** Bar graphs show the percentage of $SOX17^+$ cells obtained after DE differentiation following gRNA targeting. n=2 independent experiments.
- **c.** A summary of the number of tested and verified hits.

Figure 2.7 Information of identified positive and negative regulators

a

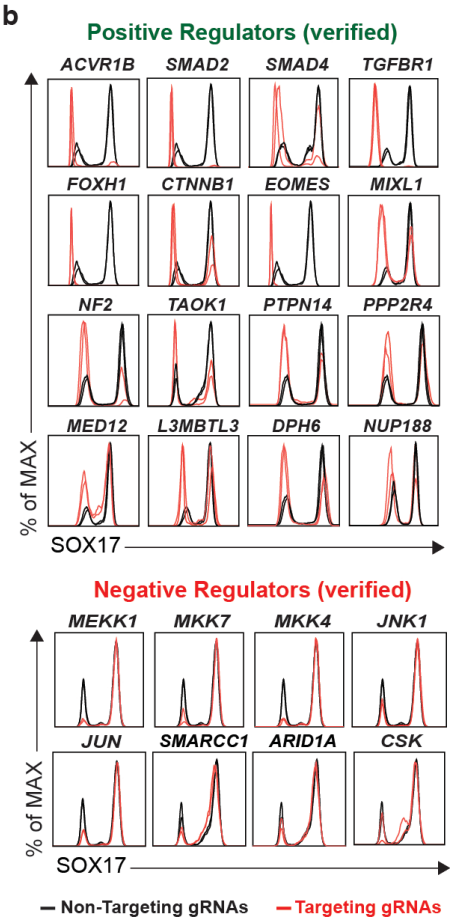
Positive Regulators (tested)

	GeCKO			Brunello		
Gene	Z-score	RRA	FDR	Z-score	RRA	FDR
EOMES	8.15	8.1E-14	0.0007	6.14	7.5E-11	0.0001
FOXH1	6.01	2.2E-10	0.0007	12.31	1.0E-11	0.0001
SMAD2	5.62	3.8E-07	0.0025	9.91	8.1E-13	0.0001
MIXL1	4.96	4.9E-09	0.0007	6.75	3.8E-11	0.0001
MED12	4.17	2.6E-06	0.0149	5.38	1.9E-09	0.0001
DPH6	3.81	5.3E-06	0.0218	3.56	8.9E-10	0.0001
AVCR1B	3.64	1.5E-06	0.0095	9.21	1.2E-14	0.0001
TGFBR1	3.49	2.1E-07	0.0019	10.58	6.3E-15	0.0001
L3MBTL3	3.39	1.7E-08	0.0007	3.51	1.1E-09	0.0001
NF2	2.79	2.6E-04	0.3045	1.79	6.8E-03	0.7158
SMAD4	2.77	1.3E-05	0.0413	8.08	2.1E-13	0.0001
TAOK1	2.71	4.3E-06	0.0195	3.43	2.7E-07	0.0005
PPP2R4	2.67	4.2E-03	0.8306	2.17	3.3E-04	0.1283
PTPN14	2.64	3.1E-05	0.0767	3.32	1.2E-06	0.0012
NUP188	2.34	4.6E-06	0.0201	3.07	1.0E-08	0.0001
CTNNB1	2.26	6.1E-06	0.0232	6.78	4.9E-11	0.0001
MAPK1	5.12	1.6E-10	0.0007	4.13	1.3E-08	0.0001
BAG6	3.46	1.5E-05	0.0439	3.03	8.4E-06	0.0082
MLL2	2.93	1.7E-04	0.2595	4.29	2.3E-08	0.0001
KDM1A	2.89	1.3E-04	0.2258	5.38	1.5E-09	0.0001
ARID4B	2.74	6.2E-06	0.0232	2.16	8.5E-05	0.0492
DUSP8	2.63	9.0E-05	0.1706	0.24	5.2E-02	1.0000

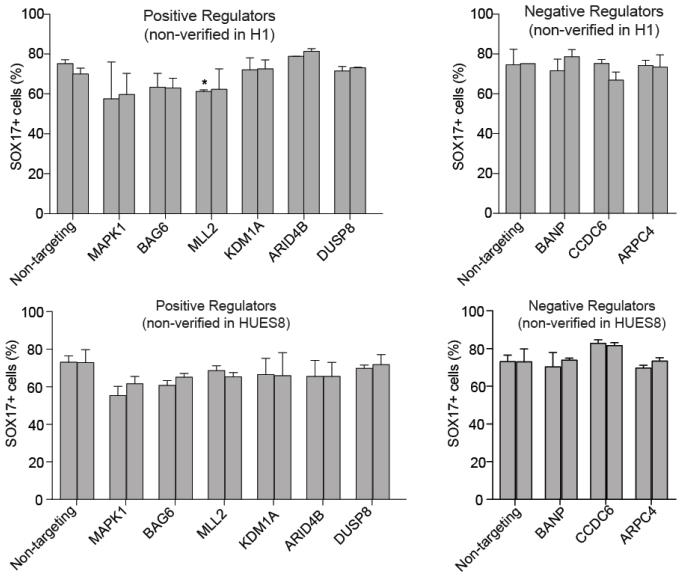
Negative Regulators (tested)

	GeCKO			Brunello		
Gene	Z-score	RRA	FDR	Z-score	RRA	FDR
MEKK1	-6.48	3.3E-11	0.0012	-3.64	1.7E-12	0.0003
MKK4	-6.32	6.3E-10	0.0012	-4.29	1.0E-11	0.0003
MKK7	-5.38	1.4E-07	0.0025	-3.64	1.7E-12	0.0003
JUN	-4.70	5.3E-09	0.0012	-3.44	1.7E-09	0.0003
SMARCC1	-4.42	6.5E-07	0.0064	-2.84	1.8E-06	0.0043
ARID1A	-4.37	2.9E-07	0.0025	-3.34	3.8E-08	0.0003
JNK1	-3.99	9.2E-06	0.0557	-4.94	8.1E-13	0.0003
CSK	-2.95	1.6E-06	0.0155	0.85	9.9E-01	1.0000
BANP	-5.60	1.8E-04	0.3002	0.49	8.7E-01	1.0000
CCDC6	-4.47	1.2E-10	0.0012	-3.34	2.0E-12	0.0003
ARPC4	-4.06	4.5E-04	0.4123	-2.37	6.8E-03	0.8555

b



c



d

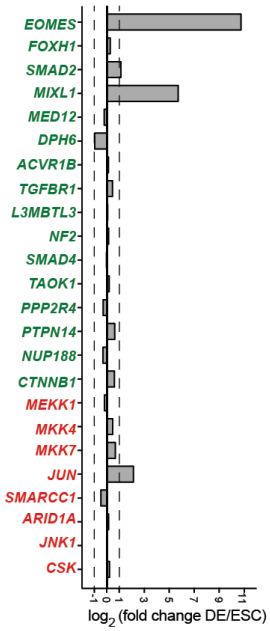


Figure 2.7 Information of identified positive and negative regulators

a. A table of the Z-score, RRA and FDR of selected tested genes.

b. Histograms showing SOX17 expression from FACS analysis. The black lines represent results from control cells infected with non-targeting gRNA lentiviruses. The red lines represent results from cells infected with targeting gRNA lentiviruses. Each plot shows a representative result from two non-targeting gRNAs and two different targeting gRNAs.

c. FACS quantification of differentiation efficiency based on the percentage $SOX17^+$ DE cells in H1 and HUES8 reporter line after gRNA targeting (2 gRNAs/gene). *p<0.05 for one of the *MLL2* gRNAs, while results from all other gRNAs were not significantly different from the control non-targeting gRNAs. n=2 independent experiments.

d. The gene expression $\log_2$ (fold change DE/ESC) of the 24 verified screen hits. Positive regulators are labeled in green and negative regulators are labeled in red. Dashed lines indicate the 2-fold change.

Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05).

2.3 Discussion

Successful completion of this CRISPR screen project demonstrates the proof of concept of using CRISPR screen and hESC to identify unique development regulators involved with human embryonic development. The main objective of this screen is to identify the genes as positive and negative regulators for the process of turning on *SOX17* expression during DE differentiation. After one week of selection and mutagenesis, we not only eliminated non-infected cells, but also eliminated cells carrying mutations with genes involved with pluripotency maintenance, self-renewal and cellular machinery genes required for survival and proliferation. Certain pluripotency maintenance genes such as *OCT4* and *NANOG* may play a role in lineage differentiation[118]; however, their exact contributions to DE lineage specification cannot be completely uncoupled with their pre-existing roles in pluripotency maintenance. Therefore, all of the identified genes from this screen do not play roles in hESC survival or pluripotency maintenance, which emphasize their distinct roles in DE lineage specification.

Among the identified positive regulators, we uncovered several known Nodal pathway related signaling genes and transcription factors. From the upstream receptor *ACVR1B* to downstream transcription factors *SMAD2, SMAD4* and *FOXH1*, their KO phonotypes are consistent with previous mouse KO studies[56,57,59,112,119]. Interestingly, the identification of *SMAD2* and *SMAD4* from the DE screen indicates SMAD2 or SMAD4 alone is not essential for hESC pluripotency and self-renewal, although TGF-β signaling is still required for hESC

pluripotency maintenance[14,120]. This finding also suggests that signaling transcription factor SMAD2 or SMAD4 have different roles in pluripotency and DE differentiation. The divergent functions of SMAD2 and SMAD4 in embryonic development remain to be further investigated.

We also identified *TGFBR1* as a strong positive regulator for DE differentiation. TGFBR1 is the essential Type 1 TGF-β family receptor that transduces signaling from TGFβ1/2/3. However, TGF-β1 proteins are not used for DE differentiation in vitro, and Activin A binds to ACVR1B and does not rely on TGFBR1 for signaling transduction[46]. In addition, *Tgfbr1* knockout mice are embryonic lethal due to abnormal vascular development at E9.5[107]. *Tgfbr1* knockout mice do not have early gastrulation phenotype as *Acvr1b* knockout mice[56,121]. The exact mechanism of TGFBR1 dependent DE differentiation is still unclear, but the possibility that TGFBR1 dependent signaling pathway may play a unique role in human definitive endoderm differentiation cannot be ruled out.

Overall, the results from known positive regulators are consistent with previous knowledge learned from reverse genetic studies in mice. The robustness of the CRISPR screen in identifying all the essential genes in the Nodal signaling pathway from one single experiment demonstrates the power and ease of using forward genetic screen to study a development process that had been investigated in other animal models for the past three decades. Identification of novel genes from human definitive endoderm differentiation also encourages us to use this platform to perform genetic screens in other lineage differentiation, such as neuroectoderm and mesoderm.

2.4 Materials and Methods

Culture of hESCs and hiPSCs

This study used three hESC lines: HUES6, HUES8 and H1 (NIHhESC-09-0019, NIHhESC-09-0021 and NIHhESC-10-0043); and two hiPSC lines (BJ and CV). The generation of the iCas9 lines from HUES8 and H1 cells was previously described [40,105]. Undifferentiated hESC and hiPSC cells were maintained as previously described [52] in the chemically defined feeder-free, serum-free E8 condition (Thermo Fisher Scientific, A1517001) on polystyrene plates coated with human recombinant vitronectin (Thermo Fisher Scientific, A14700) at 37^{o}C with 5% CO_2. hESCs/hiPSCs were dissociated with 0.5 mM EDTA and split every 3-4 days at a 1:10-1:15 ratio. Cells were routinely confirmed to be mycoplasma-free by the MSKCC Antibody and Bioresource Core Facility. All experiments were approved by the Tri-SCI Embryonic Stem Cell Research Oversight Committee (ESCRO).

Generation of the HUES8 *SOX17$^{GFP/+}$* reporter line

The HUES8 iCas9 cell line carries a puromycin resistance gene [105]. Since the GeCKO v2 and Brunello library also relies on puromycin selection, we first transfected HUES8 iCas9 cells with an *in vitro* transcribed gRNA (Table 4) to knockout the puromycin resistance gene. We then established the *SOX17$^{GFP/+}$* iCas9 line using a selection-free knock-in strategy [100] (Fig. 2.1b). Correct targeting was verified by Southern blotting as previously described [105]. The sequence for the 5' external probe used in the southern blotting is:

AATCGCTAGGCCGATTTCTTAAACCCCAAACTGTTCTTTGCGAGCCTGACGC
CCAAAACCAGGGGTGTGTAGCGGCCACGTCCTTTCTTAAGGCTCTGGGTTC
CCTTCCCGCTTCCCGCCCTCCGACCCTCCAAAGCAGCTTTCCGCCTTGCTC
TCCGGCTCCCGGATTCCCCAGGTGGCCGGGGGCGCGGGTCCAACGGCTC
TGGGAAGGCGACTTCCCGGCACCTCCGGGCGGCGCGAGAGCACCCTTGG
CCCTGAACTGGGCCGGTTGTGTCCATCCCTCGACCCCTTCCCTAGTTAGGT
GTCCTTTTCTGTTTTTCGAACGACCGGGTGATGGGTGAGCGGAAAGCCGCT
TCCAGGAGACCAAAAGAAAGGGGTGCCTTTAGAGGACGGGTGTTCCCCAA
GGGCTCGGACTCAGGAGTCCCAGATCTCCCTCTTTAACTTCACCCCGGTTG
CGCAATTCAAAGTCTGAGGGG. The probe was synthesized by PCR using the PCR DIG probe Synthesis Kit (Roche Applied Sciences, 11636090910). 20 µg genomic DNA was digested with XmnI, which produced a 3,974 bp DNA fragment with the GFP insertion, and a 3,188 bp DNA fragment without the GFP insertion (Fig. 2.1b-c). Southern blotting identified one clonal cell line, in which the stop codon of one *SOX17* allele was replaced by the 2A-GFP sequence, and the other *SOX17* allele remained wild-type. This HUES8 iCas9 *SOX17*$^{GFP/+}$ cell line maintains a normal karyotype (Fig. 2.1d). Genomic DNA genotyping revealed no *de novo* mutation in the *TP53* coding sequences in the HUES8 *SOX17*$^{GFP/+}$ line.

Production of the lentivirus library

Human CRISPR libraries GeCKO v2 and Brunello were obtained from Addgene (1000000049 and 73178) [106,108]. 50 µg CRISPR library plasmid with 20 µg PAX2

and 5 μg VSVG plasmid were transfected with the Jetprime (VMR, 89137972) reagent into 293T cells to pack the lentivirus. Viral supernatant was collected, filtered, aliquoted and stored at -80^{o}C. We used lentivirus infection efficiency to estimate the multiplicity of infection (MOI) according to the formula P (n) = $\frac{m^n e^{-m}}{n!}$, where m is the MOI; n is the occurrence of event that virus enters into cells; P (n) is the probability that a cell will get infected by n viruses. The infection efficiency can be viewed as the probability of being infected which equals to 1-P (0). When MOI equals to 0.36, the infection efficiency [1-P (0)] is 30%, and the probability of a cell getting 2 or more viruses is 16.28%. To determine the viral titer for an infection efficiency of 30%, 0.15 million SOX17$^{GFP/+}$ hESCs per well were infected with different amounts of virus (0-20 μL) in duplicates in six-well plates. 48 hours later, puromycin (0.5 μg/mL) was added into one set of the cells to select infected cells. After 48 hours treatment of puromycin, uninfected control cells were killed by puromycin selection. The ratio of the cell number of the selected set (treated with puromycin) over the unselected set (not treated with puromycin) was calculated to determine the infection efficiency. The amount of virus needed for the 30% infection efficiency in the six-well format is scalable to 150 mm plate by a factor of 15.

Genome wide CRISPR screens

A minimum of 200-fold library coverage is typically recommended for screens based on basic phenotypes such as cell survival and growth. Considering the relatively complex nature of a screen for lineage regulators, we aimed for a

~1,000-fold coverage to maximize sensitivity. 200 million iCas9 *SOX17*$^{GFP/+}$ HUES8 cells were harvested and infected with the lentiviral library at a low MOI of 0.36 at D0 in 150 mm plates (100 plates total). 6 μg/mL protamine sulfate was added on the first day of infection to enhance the infection efficiency. Next, infected cells were treated with 2 μg/mL doxycycline (D1 to 7) and 0.5 μg/mL puromycin (D2 to 7). On D7, cells were treated with TrypLE Select (Thermo Fisher Scientific, 12563029) and 120 million cells were plated into 150mm plates (59 plates total). On D8, cells were switched from the maintenance E8 medium to the DE differentiation medium (described in the "DE differentiation" subsection). After 3 days of DE differentiation (D8 to D11), cells were dissociated using TrypLE Select and sorted using FACSAria according to GFP expression. Sorted GFP positive and negative cells were pelleted and genomic DNA was immediately extracted using the Qiagen blood & cell culture DNA maxi kit (Qiagen, 13362).

Hi-Seq and data analysis

A two-step PCR method was performed to amplify the gRNA sequence for Hi-Seq. For the first step, 380 μg of DNA per sample (6.6 μg of genomic DNA per 1 million cells) was used to perform PCR using Herculase II fusion DNA polymerase (Agilent, 600679) in order to achieve a 1,000-fold converge of the GeCKO library containing 58,028 gRNAs. Primers sequences to amplify lentiCRISPR gRNAs for the first PCR are:

F1: AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG

R1: CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC. In total, we performed 38 separate 100 µL reaction with 10 µg genomic DNA per sample for 18 cycles and combined the resulting amplicons. For the second step, 5 µL of the product from the first PCR was used in a 100 µL PCR reaction for 24 cycles with primers to attach Illumina adaptors for barcoding. Primers used in this reaction are:

F2:AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCT(1-9bp variable length sequence)tcttgtggaaaggacgaaacaccg

R2:CAAGCAGAAGACGGCATACGAGAT(6bp barcode) GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtctactattctttcccctgcactgt

Gel purified amplicons from the second PCR were quantified, mixed and sequenced using Illumina HiSeq 2500 by MSKCC Integrated Genomics Operation (IGO). Raw FASTQ files demultiplexed by MSKCC IGO were further processed to contain only the unique gRNA sequences, and the processed reads were aligned to the designed gRNA sequences from the library using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The read counts were further normalized to total reads of that sample to offset differences in read depth.

Hit validation and the lentiCRISPR approach

The H1 iCas9 line [40] was used for most validation experiments. Some hits were tested again in HUES8 iCas9 cells. The differentiation outcome was evaluated by directly measuring intracellular SOX17 expression through flow cytometry instead of relying on the GFP reporter. For hit validation and all experiments using the

lentiCRISPR approach, gRNAs were cloned into lentiGuide-puro (Addgene, 52963) following published protocols [108]. The lentiCRISPR gRNA construct expresses a puromycin resistance gene, allowing the selection of infected cells through puromycin treatment. 1 µg lentiGuide-puro, 0.1 µg VSVG and 0.4 µg PAX2 plasmids were transfected with the Jetprime (VMR, 89137972) reagent into 293T cells to pack lentiviruses. Viral supernatant was collected, filtered, aliquoted and stored at -80^{o}C. A MOI of 0.30~0.36 was used for the infection of the H1 iCas9 cell line with different lentiCRISPR viruses. Cells were treated with 2 µg/mL doxycycline (D1 to 7) to induce Cas9 expression, and 0.5 µg/mL puromycin to eliminate non-infected cells (D2 to 7). On D7, cells were treated with TrypLE Select and 0.1 million cells per well were plated in duplicate sets in a six-well plate. One set was used for DE differentiation the next day. The other set was cultured in E8 media for maintenance until the second repeat of differentiation. gRNA targeting sequences selected from GeCKO v2 library are listed in Table 4. Two gRNAs per gene were tested for validation. Two non-targeting gRNAs were used as WT controls.

DE differentiation

All DE differentiation experiments were performed on hESCs grown on vitronectin. For the CRISPR screens and screening hit validation, cultured hESCs/hiPSCs were grown to 80-90% confluence, and then treated with TrypLE Select for dissociation into single cells. Typically, 0.15 million hESCs/hiPSCs were plated in one well of the six-well plates with 10 µM Y-27632 in 2 mL E8

media. For the H1 line, we typically plated 0.1 million cells to accommodate the higher proliferation rate compared to other lines. For differentiation in 150mm dishes, we multiply the numbers by a factor of 15 based on the numbers from six-well plates. 24 hours later, cells were washed with PBS once and culture medium was changed to Advanced RPMI (Thermo Fisher Scientific, 12633012) with penicillin/streptomycin (Thermo Fisher Scientific, 15070063), GlutaMAX (Thermo Fisher Scientific, 35050079), 0.003% BSA (Thermo Fisher Scientific, 15260037), 5 μM CHIR99021 (Tocris, 4423) and 100 (high) or 20 (low) ng/mL Activin A (PeproTech, 12014E; Bon Opus Biosciences, BP08001). Over the next 2 days, culture medium was changed to Advanced RPMI with penicillin/streptomycin, GlutaMAX, 0.2% FBS and 100 or 20 ng/mL Activin A. Although FBS was used in the initial GeCKO library screen, we have since found that FBS is not essential for DE differentiation. Thus, we omitted FBS in the Brunello screen. For JNK inhibition, 1 μM JNK-IN-8 (Selleck Chemicals, S4901) was added at the indicated time window, while equal concentration of DMSO was used as a vehicle control.

CHAPTER 3: Investigation of the roles of JNK/JUN pathway in DE differentiation

3.1 Introduction

During gastrulation, Nodal signals via the Smad2-Smad4-Foxh1 axis to promote endoderm formation, whereas the secreted proteins Lefty1, Lefty2, and Cer1 (Cerberus 1) act as Nodal antagonists to restrict endoderm or mesoderm formation in mice [23]. For example, *lefty2*-/- embryos form excess mesoderm[122], and *Cer1*-/-*Lefty*-/- mutant embryos develop ectopic primitive streak[123]. Complete absences of *lefty1* and *lefty2* in zebrafish embryos cause severe gastrulation patterning defects and lethal expansion of mesendoderm[124]. The expressions of these antagonists are also Nodal signaling dependent, therefore the feedback inhibition of Nodal/Lefty system tightly controls the events of gastrulation and body axis formation. Negative regulators are essential for establishing a morphogen gradient to control and fine-tune lineage decision marking, because precise lineage commitment requires negative regulators to counterbalance the inductive signaling[125]. From the CRISPR screen, we identified five genes in the JNK pathway as top negative regulators of human DE differentiation. Less is known about the negative regulators of in vitro directed DE differentiation. Therefore, we decided to investigate the role of JNK pathway in DE differentiation.

JNK pathway is part of the stress-activated MAP kinase signaling pathways[126]. MAP kinase pathway operates signaling transduction in a kinase

cascade manner. The signaling cascade starts from phosphorylation of MAP kinase kinase kinases (including MEKK, TAK1, MLK) then phosphorylation of dual specificity MAP kinase kinases (MKK7 and MKK4), finally to phosphorylation of MAP kinases (JNK1, JNK2, JNK3). *JNK1/2/3* are the core nuclear MAP kinase genes that translate upstream cellular signaling transduction to transcription factors[127]. JNK signaling is also part of the non-smad pathways in TGF-β signaling[128], that activation of TGF-β receptors induces phosphorylation of JNK pathways through TAK1. One major class of the JNK substrates is the AP-1 transcription factor family. AP-1 transcription factors including c-Jun (JUN, c-Fos, Jun-B, Jun-D and others[129] regulate gene transcription through their binding to AP-1 motifs (TGACTCA or TGACGTCA) in the form of homodimers or heterodimers. The leucine zipper domain of JUN is responsible for the dimerization activity[130]. JNK mediated Phosphorylation at serine 63 and serine 73 of JUN positively enhance its transactivation activity[131,132].

The diverse functions of AP-1 mediated transcription is context dependent. AP-1 interacts with many other transcription factors and chromatin-modifying enzymes, therefore, the outcome of transcription can be either activation of inhibition[133]. For example, repression of TGF-β signaling caused by hyperactivation of the JNK pathway contributes to HTLV-1 associated adult T-cell leukemia[134]. In this context, phosphorylated JUN interacts with Smad3 and inhibits Smad3 DNA binding activity. Studies performed in COS-7 and HepG2 cell lines show that JUN also suppresses SMAD2 transcriptional activity by stabilizing a SMAD2 co-repressor complex with SKI or TGIF[135,136]. In another

context, AP1 and SMAD3/4 complex cooperatively mediates TGF-β induced transcription[137].

In this chapter, we used genetic and pharmacological approaches to study the roles of JNK pathway in DE differentiation. We generated *MKK7* and *JUN* KO hESC lines and validated their phenotypes, which was consistent with our findings from the CRISPR screen. Next, we found that inhibition of JNK pathway by using small molecule JNK inhibitor could not only improve DE differentiation efficiency, but also improve subsequent lineage differentiation efficiency to pancreatic and lung lineage progenitors. Finally, we used ATAC-seq and ChIP-seq approaches to study the impacts of JNK inhibition on chromatin accessibility and transcription factor binding during ESC-DE lineage transition.

3.2 Results and figures

3.2.1. The JNK pathway inhibits DE differentiation

Four of the six negative regulator hits with FDR<0.05 in both screens belonged to the mitogen-activated protein kinase (MAPK) JNK pathway, including the MAPK kinase kinase *MEKK1* (*MAP3K1*), the MAPK kinase *MKK4* (*MAP2K4*) and *MKK7* (*MAP2K7*), and the AP1 family TF *JUN* (*C-JUN*) (Fig. 2.5b and 3.1a). A fifth JNK pathway gene *JNK1* (*MAPK8*) just missed the FDR cutoff (Brunello screen FDR<0.05, GeCKO screen FDR=0.056). The JNK/JUN pathway can be activated by a variety of environmental signals including stress, cytokines

and growth factors [138], but it has not been reported to regulate DE differentiation. Using lentiviruses expressing gene-specific gRNAs, we were able to validate the inhibitory effects of all five JNK pathway genes on DE differentiation (Fig. 2.6b). We further generated clonal H1 *MKK7* and *JUN* knockout (KO) lines and verified the knockouts by western blotting (Fig. 3.1b-c and Fig.3.2a). We confirmed that *MKK7* KO cells had greatly reduced JNK phosphorylation, undetectable levels of JUN phosphorylation, and reduced JUN expression due to loss of positive autoregulation [139] (Fig. 3.1c). *MKK7* or *JUN* KO cells did not exhibit a difference in cell growth during hESC self-renewal or DE differentiation (Fig. 3.2f); nor did flow cytometry detect a significant difference in the number of cells expressing proliferation or apoptosis markers (phospho-Histone H3 and cleaved caspase-3, respectively) during differentiation (Fig. 3.2g). *MKK7* and *JUN* KO hESCs formed neuroectoderm cells [35] expressing PAX6 and SOX1 with comparable efficiency and kinetics to the WT control (Fig. 3.3a-d). In contrast, *MKK7* and *JUN* KO cells readily formed DE cells expressing CXCR4, SOX17, GATA6 and GATA4 at a higher efficiency compared to isogenic wild-type (WT) controls (Fig. 3.1d-e). RNA-seq analysis showed significant global transcriptional differences between WT and *MKK7* KO cells at the DE but not the ESC stage (Fig. 3.1f), as verified using a panel of DE and ESC markers for flow cytometry, immunostaining and RT-qPCR analyses (Fig. 3.2b-e). These findings demonstrate that the JNK/JUN pathway inhibits the ESC-DE transition.

3.2.2 Pharmacological JNK inhibition improves DE differentiation

We tested the effects of JNK-IN-8 [140], a selective JNK inhibitor (JNKi) on DE differentiation (Fig. 3.4a). Western blotting analysis confirmed that JUN phosphorylation was barely detectable in JNKi treated H1 hESCs (Fig. 3.4b). Similar to genetic perturbation, JNKi treatment improved the efficiency of DE differentiation to >90% based on flow cytometry analysis for expression of key endoderm genes including CXCR4, SOX17, GATA6, and GATA4, and confirmed by the corresponding RT-qPCR analysis (Fig. 3.4c and 3.5a). JNKi also significantly increased DE differentiation efficiency in additional hESC (HUES8 and HUES6) and hiPSC (BJ and CV) lines (Fig. 3.5b-c). These findings demonstrate that JNK inhibition improves DE differentiation efficiency and reduces differentiation variability.

We performed Drop-seq [141] and conducted unsupervised hierarchical clustering on 1,604 and 1,732 individual DE cells from control and JNKi-treated conditions, respectively, using a panel of DE and ESC marker genes (Fig. 3.4d) or the 1,000 most variably expressed genes (Fig. 3.5d). JNKi treatment during differentiation induced a relatively homogenous cell population, in which DE signature genes were highly expressed, and ESC signature genes were low to undetectable. In comparison, the control differentiation condition resulted in two cell populations, one expressing DE signature genes which clustered closely with cells from the JNKi condition, and the other retaining some expression of ESC signature genes which clustered separately (Fig. 3.4d and 3.5d). This difference

in cell heterogeneity can be further visualized in scatter plots showing the expression of individual DE and ESC genes (Fig. 3.4e-f and 3.5e). Thus, JNKi treatment during differentiation results in a more homogeneous DE population.

3.2.3 JNK inhibitor improve pancreatic and lung progenitor differentiation

We further examined the effects of JNKi using an optimized serum-free DE differentiation protocol [37,38,40] (see Methods). One day of JNKi treatment was sufficient to improve DE differentiation with this improved protocol as evaluated by flow cytometry analysis of CXCR4 and SOX17-GFP expression when compared to the untreated control (Fig. 3.6a-c). We further demonstrated that one day of JNKi treatment improved DE differentiation across different hESC (HUES8, H1, HUES6) and hiPSC (CV, BJ) lines under this optimized differentiation condition (Fig. 3.4g and 3.6d). Different culture media could influence the outcome of DE differentiation. Our work so far shows that JNK inhibition improves DE differentiation in both serum-free and serum-containing conditions.

Furthermore, DE cells from the 1-day JNKi treated condition formed a higher percentage of $PDX1^{+}NKX6.1^{+}$ pancreatic progenitor cells or $NKX2.1^{+}FOXA2^{+}$ lung progenitor cells as shown by immunostaining and quantified by flow cytometry analysis (Fig. 3.4h-k). Additional pancreatic (*NEUROD1*, *MNX1* and *NGN3*) or lung (*FOXP2* and *CPM*) progenitor markers were upregulated in the JNKi treated condition by RT-qPCR analysis (Fig. 3.6e-

f). These results demonstrate that JNK inhibition improves DE differentiation, which leads to more robust subsequent differentiation to the pancreatic and lung lineages.

3.2.4 JUN impedes chromatin landscape remodeling during ESC-DE differentiation

The Nodal/TGF-β-SMAD2/3 signaling axis is required for both hESC pluripotency maintenance and DE differentiation [13,14,33]. SMAD2/3 binding patterns have been shown to be associated with lineage transcription factors [49,142,143]. Our ChIP-seq analysis of SMAD2/3 differentially bound regions at the ESC versus DE stages, identified OCT4/NANOG as the most enriched motif among SMAD2/3 bound regions at the ESC stage, SMAD and GATA as the most enriched motifs at the DE stage (Fig. 3.8a-b). This observation suggests that SMAD2/3 binds cooperatively with OCT4/NANOG at the ESC stage and becomes free to occupy its own binding sites or co-occupy with GATA6 at the DE stage.

To investigate the effect of JNK inhibition on chromatin remodeling during DE differentiation, we performed ATAC-seq on hESCs after one day of DE differentiation (DE D1, the primitive streak stage). Using two *de novo* discriminative motif algorithms, HOMER [144] and MotifSpec [145], we found AP1 and TEAD binding sites were enriched in regions with decreased accessibility upon JNK inhibition (Fig. 3.7a). We also found GATA and SMAD binding sites were

enriched in regions with increased accessibility upon JNK inhibition. Using an alternative gapped kmer (oligomers of length k) based approach, we trained gkmSVM [146] on regions with increased versus decreased accessibility and identified significant feature weights for kmers matching these same binding sites (AP1, TEAD, GATA, SMAD), which in combination could account for almost all the changes in the chromatin accessibility landscape (AUROC=0.923).

Next, we performed SMAD2/3 and GATA6 ChIP-Seq analysis at the DE stage, and found that indeed regions with increased accessibility upon JNK inhibition (DE D1) showed significant GATA6 and SMAD2/3 binding at the DE stage (Fig. 3.7b-c). These results suggest that JNK inhibition promotes the access of SMAD2/3 and GATA6 to DE enhancers during the ESC-DE transition. Importantly, we found enrichment of JUN binding at the ESC stage in regions with decreased accessibility upon JNK inhibition, indicating a direct role of JUN in regions that lost rather than gained chromatin accessibility. We speculate that JNK inhibition may promote ESC-DE transition through accelerating the loss of chromatin accessibility at ESC enhancers. Supporting this notion, regions with decreased accessibility upon JNK inhibition also showed OCT4 and SMAD2/3 enrichment at the ESC stage (Fig. 3.7b-c).

The association of JUN and OCT4 binding in regions with decreased accessibility upon JNK inhibition prompted us to examine the global pattern of JUN and OCT4 binding in undifferentiated hESCs. Using default (MACS2 FDR q=0.05) peak calling for OCT4, we observed significant overlap between OCT4 and JUN binding (Fig. 3.8c) ($p < 3.89e-224$): ~60% of JUN-bound sites were co-

bound by OCT4. Focusing on only the strongest OCT4 bound sites by using very stringent (q=1e-6) peak calling for OCT4, the overlap is still significant (Fig. 3.7d) ($p<1e-300$): ~28% of JUN-bound sites were co-bound by OCT4. Compared to the OCT4+JUN- sites, the OCT4+JUN+ sites showed a prominent increase in SMAD2/3 binding along with a small increase in OCT4 binding (Fig. 3.7e and 3.8c). We further observed an association of JUN binding with a stronger ESC chromatin signature. OCT4+JUN+ sites showed much increased chromatin accessibility and H3K27ac and P300 signals (Fig. 3.7e and 3.8c) [147]. Indeed, at the ESC stage, co-occupancy of JUN, OCT4, NANOG and SMAD2/3 was observed at enhancers of ESC genes (e.g. *OCT4, NANOG* and *SOX2*), and JNK inhibition during DE differentiation decreased the chromatin accessibility at these enhancers (Fig. 3.7f). Meanwhile, we observed co-occupancy of SMAD2/3 and GATA6 at enhancers of DE genes (e.g. *SOX17, GATA6,* and *CXCR4*), and JNK inhibition during DE differentiation increased the chromatin accessibility (Fig. 3.7f). Together these findings support a direct role of JUN binding in impeding the decommissioning of ESC enhancers, which then indirectly impedes the reconfiguration of SMAD2/3 occupancy from ESC to DE enhancers.

3.2.5 JUN impedes SMAD2/3 reconfiguration during ESC-DE differentiation

To better understand the effect of JUN on the enhancer landscape during DE differentiation, we focused on the top OCT4-bound ESC enhancers (9,248, excluding promoter regions) and the top GATA6-bound DE enhancers (9,248) for

further analyses (Fig. 3.9a). NANOG binding was strongly enriched at the ESC enhancers as expected, and there was a dramatic reconfiguration of SMAD2/3 binding during the ESC-DE transition from OCT4-bound ESC enhancers to GATA6-bound DE enhancers (Fig. 3.9a). When comparing OCT4-bound regions with high versus low JUN ChIP-seq signals, we found a strong correlation of SMAD2/3 occupancy with the JUN^{high} group, which also showed stronger OCT4 and NANOG signals compared to the JUN^{low} group (Fig. 3.9a and 3.8f). Together, these findings support a strong association of JUN binding to ESC enhancers, especially those that are also occupied by SMAD2/3.

Next, we analyzed the effects of JNK inhibition on these ESC and DE enhancers. JNK inhibition (DE D1) decreased chromatin accessibility at $OCT4^{+}JUN^{high}$ ESC enhancers, and increased chromatin accessibility at $GATA6^{+}$ DE enhancers (Fig. 3.9b). Such reciprocal effects were also observed on SMAD2/3 binding at ESC and DE enhancers respectively (Fig. 3.9c), indicating that JNK inhibition promotes the reconfiguration process of SMAD2/3 binding pattern from ESC to DE enhancers. Similar reciprocal effects on chromatin accessibility and SMAD2/3 occupancy were also observed when examining ESC and DE enhancers that are also bound by SMAD2/3 or when broadly examining SMAD2/3 occupied regions at the ESC and DE stages irrespective of OCT4 or GATA6 occupancy (Fig. 3.8d). These effects were not due to an effect of JNK inhibition on SMAD2 protein expression or phosphorylation (at Ser465/467 or Ser245/250/255) (Fig. 3.8e).

The tight correlation of JUN and SMAD2/3 binding at the ESC stage prompted us to directly examine the global effect of SMAD2/3 occupancy upon JNK inhibition at DE D1. We found that regions with decreased SMAD2/3 binding showed enrichment of JUN, SMAD2/3 and OCT4 binding at the ESC stage and decreased ATAC-seq signal after JNK inhibition, whereas regions with increased SMAD2/3 binding showed enrichment of SMAD2/3 and GATA6 binding at the DE stage and increased ATAC-seq signal after JNK inhibition (Fig. 3.9e-f and 3.8g). Corresponding changes were observed at many individual ESC and DE enhancers (Fig. 3.7f and 3.8h). In addition, JNKi treatment increased SMAD2/3 and GATA6 co-occupancy (Fig 3.8i). Further supporting a direct causal relationship between decreased JUN binding and decreased SMAD2/3 binding at ESC enhancers, HOMER, MotifSpec and gkmSVM analysis showed that regions with decreased SMAD2/3 binding upon JNK inhibition were enriched with AP1 and TEAD binding sites, while regions of increased SMAD2/3 binding were enriched with GATA and SMAD binding sites, and kmers matching these same binding sites can account for most of the changes in the SMAD2/3 binding (AUROC=0.962) (Fig. 3.9d).

Lastly, we investigated the transcriptional consequences of JUN binding at ESC enhancers. JUN, SMAD2/3, OCT4, and NANOG showed co-occupancy at cis-regulatory regions of many ESC genes including *OCT4*, *NANOG*, *SOX2*, *PRDM14* and others (Fig. 3.9g). The expression levels of these genes at the ESC stage were not affected by genetic inactivation of the JNK pathway (*MKK7* KO) (Fig. 3.9h). However, after DE differentiation, they were further

downregulated in *MKK7* KO DE cells compared to WT DE cells (Fig. 3.9h). Globally, the most decreased JUN binding sites from ESC to DE D1 were enriched near genes with the most decreased expression during the ESC-DE transition (Fig. 3.8j). Taken together, we show that JUN, SMAD2/3 and OCT4/NANOG co-occupy ESC enhancers, and JUN binding impedes the decommissioning of ESC enhancers and the corresponding down-regulation of ESC genes during DE differentiation; inhibition of JUN binding accelerates the reconfiguration of SMAD2/3 binding pattern from ESC to DE signature and promotes the establishment of the DE enhancer landscape (Fig. 3.10a).

3.2.6 Stochastic simulations of ESC-DE transition

We performed stochastic simulations to further test the model that JNK inhibition promotes DE differentiation primarily through weakening ESC enhancers. In this minimal bifurcation model, ESC TFs (e.g. OCT4) activate their own transcription, but negatively regulate DE TFs either directly or indirectly (e.g. GATA6) (Fig. 3.10b-c). Conversely, DE TFs activate their own transcription but negatively regulate ESC TFs. Activin A produces a transient activation of DE TFs, modeled as an impulse. A low Activin A impulse activates GATA6, but does not stimulate GATA6 sufficiently to produce transitions to the stable $GATA6^{high}$ DE state. When the auto-activation of OCT4 is weakened by 15%, the same low Activin A impulse is now sufficient to stimulate 7/10 of the simulations into the high $GATA6^{high}$ DE state. We can then simulate the rate at which cells transition

to the DE state under varying Activin A doses (Fig. 3.10d). If JNK inhibition reduces either transcriptional activation of the ESC genes or the binding of ESC TFs to their auto-regulatory enhancers, the model predicts a reduction in the Activin A dose required for efficient transition to the DE state.

We tested this prediction by performing an Activin A dose titration experiment using the HUES8 *SOX17*$^{GFP/+}$ reporter line. The efficiency of DE differentiation depended on the Activin A concentration. JNK inhibition did not eliminate the requirement for Activin A, but enabled efficient DE differentiation with lower Activin A doses (Fig. 3.10e) exactly as predicted by the model. JNK inhibition also promoted differentiation of H1 hESCs treated with a low Activin A dose (20 ng/mL, instead of the typical 100 ng/mL) (Fig. 3.11a-b). Similar effects were observed with genetic inactivation of the JNK pathway members *MKK7* and *JUN* (Fig. 3.11c-e). Overall, these results together with the mathematical modeling and genomic studies support the concept that JNK inhibition during DE differentiation destabilizes the ESC enhancers that were maintained by the co-binding of JUN, SMAD2/3 and OCT4/NANOG, thereby accelerating the establishment of the DE enhancer landscape and lowering the Activin A dose requirement for efficient DE differentiation.

Figure 3.1 Genetic inactivation of JNK pathway improves DE differentiation

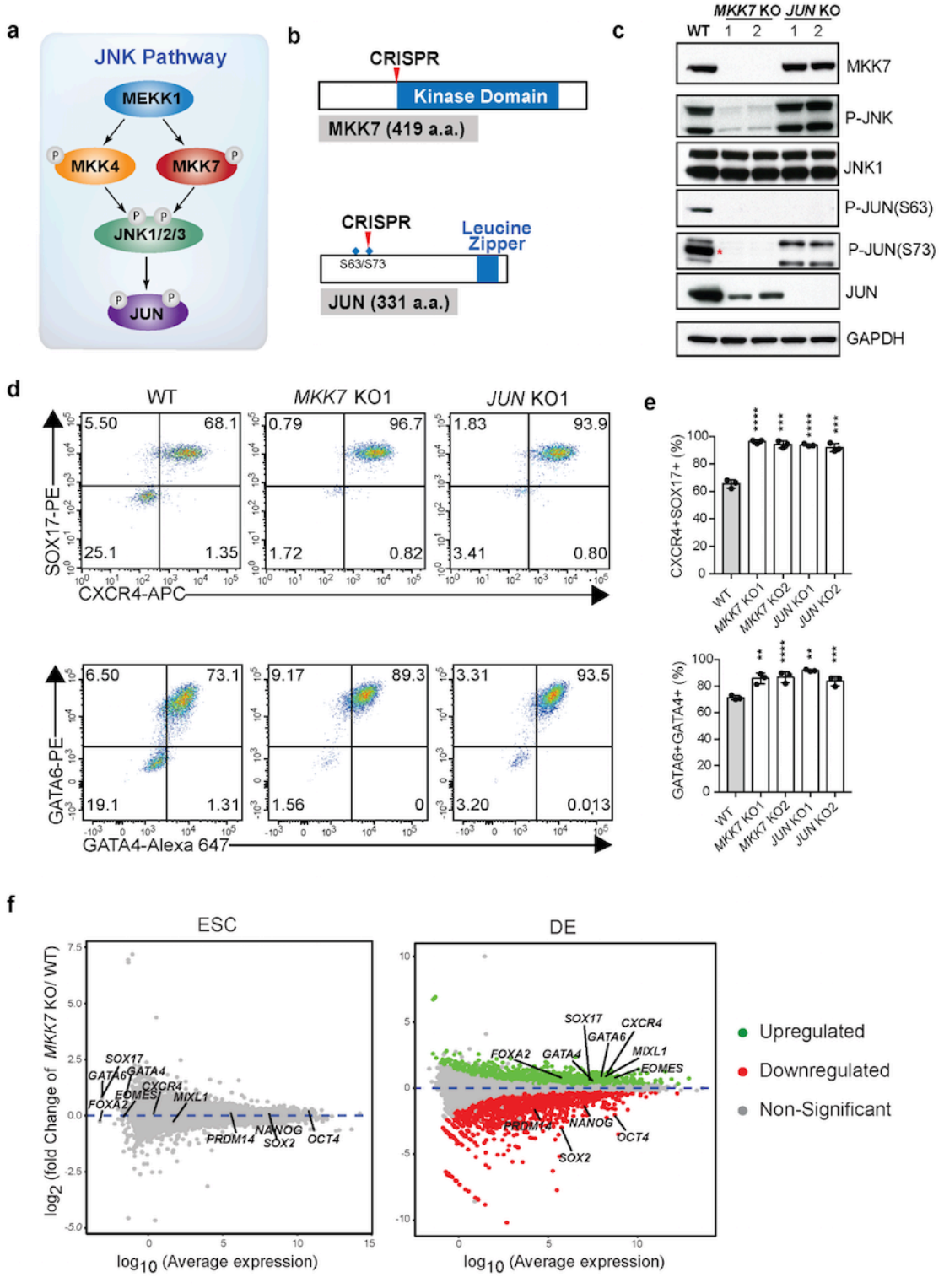


Figure 3.1 Genetic inactivation of JNK pathway improves DE differentiation

a. Illustration of the JNK-JUN pathway.
b. MKK7 and JUN protein structures depicting their important functional domains. The arrows indicate the locations of the gRNA target sequences. The blue diamonds indicate the location of phosphorylation sites.
c. Western blotting analysis for JNK pathway protein expression of DE cells (D3). GAPDH was used as a loading control. * indicates the specific band for P-JUN (S73).
d. Representative FACS dot plots of DE D3 cells stained for CXCR4 and SOX17, GATA6 and GATA4 (H1 hESC line, WT vs *MKK7* and *JUN* KO).
e. FACS quantification of differentiation efficiency from (d). n=3 independent experiments.
f. MA plot of RNA-seq analysis of WT and *MKK7* KO from ESC and DE stages. Each dot represents a gene expression profile with its average expression level on the X-axis and fold change on the Y-axis. Selected genes are indicated. FDR of 0.05 was used to define significant differentially expressed genes.

Figure 3.1f is contributed by Chunlong Xu.
Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05).

Figure 3.2 Phenotypes of *MKK7* and *JUN* KO hESC lines

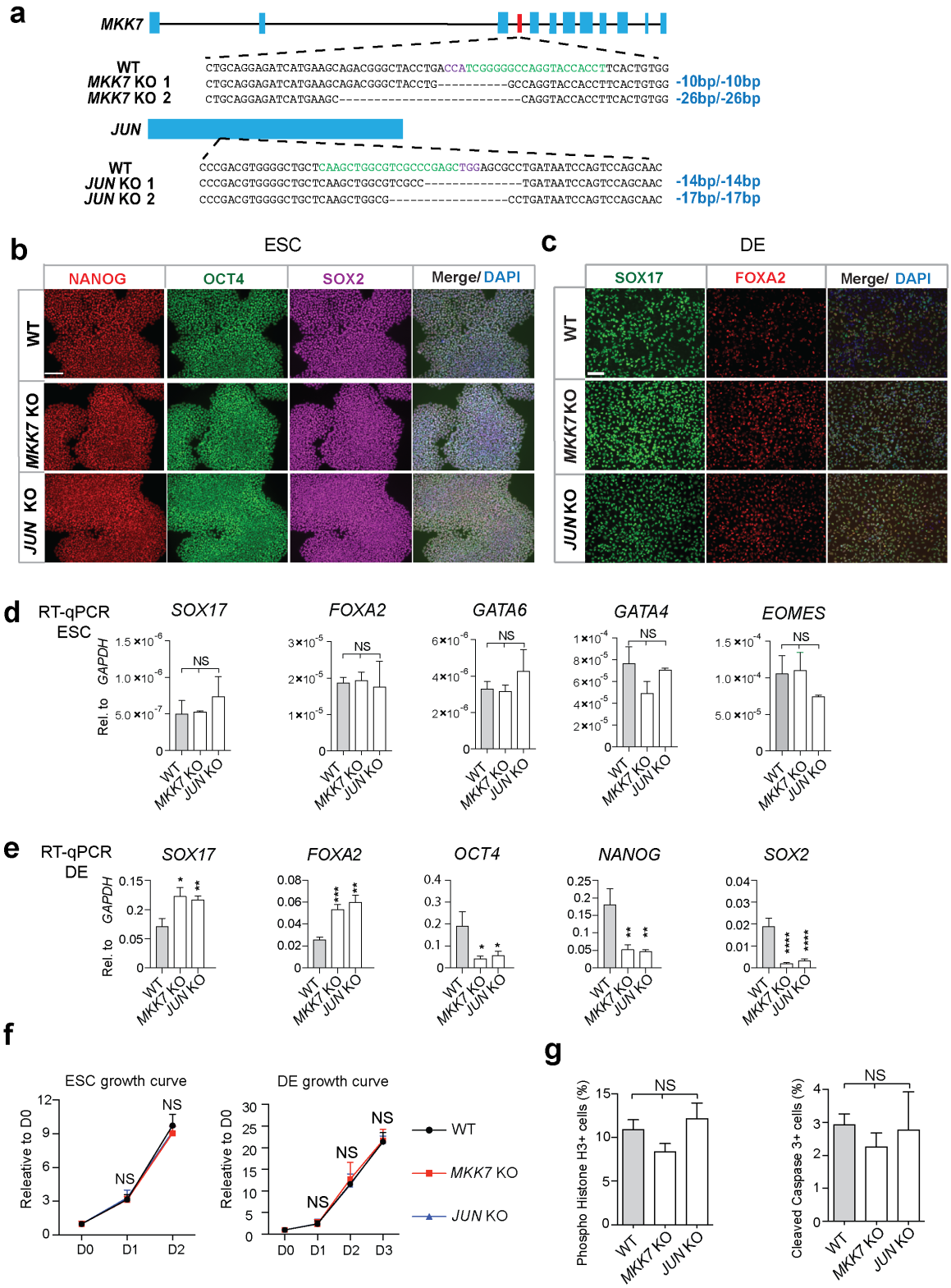


Figure 3.2 Phenotypes of *MKK7* and *JUN* KO hESC lines

a. CRISPR/Cas9 gene targeting strategy used to generate *MKK7* and *JUN* KO lines. Two homozygous KO cells were picked for further analysis. The blue bars indicate the exons of the gene. Note that *JUN* is an intronless gene. CRISPR target and PAM sequences are shown in green and purple, respectively.

b. Immunostaining of ESC markers OCT4, NANOG and SOX2 in WT, *MKK7* KO and *JUN* KO ESCs.

c. Immunostaining with DE markers SOX17 and FOXA2 in WT, *MKK7* KO and *JUN* KO DE D3 cells.

d. RT-qPCR analysis of DE marker genes in undifferentiated WT, *MKK7* KO and *JUN* KO ESCs. The relative expression level is normalized to the housekeeping gene *GAPDH*. n=3 independent experiments.

e. RT-qPCR analysis of DE (*SOX17, FOXA2*) and ESC (*OCT4*, *NANOG*, *SOX2*) marker genes in WT, *MKK7* KO and *JUN* KO DE cells. The relative expression level is normalized to the housekeeping gene *GAPDH*. n=3 independent experiments.

f. Growth curve of WT, *MKK7* KO and *JUN* KO cells cultured in the self-renewing condition (ESC) or DE differentiation condition. For the ESC condition, cell numbers are normalized to D0 (1 day after splitting). For the DE condition, cell numbers are normalized to D0 (1 day after splitting and the day when DE differentiation was initiated). n=3 independent experiments.

g. FACS quantification of the proliferation marker phospho-Histone H3 and the apoptosis marker cleaved caspase-3 from WT, *MKK7* KO and *JUN* KO DE D3 cells. n=3 independent experiments.

Error bars indicate SD, and significance is indicated as *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$; ****$p< 0.0001$; NS, not significant ($p \geq 0.05$). Scale bar, 100 μm.

Figure 3.3 Neuroectoderm (NE) differentiation of *MKK7* and *JUN* KO hESC lines

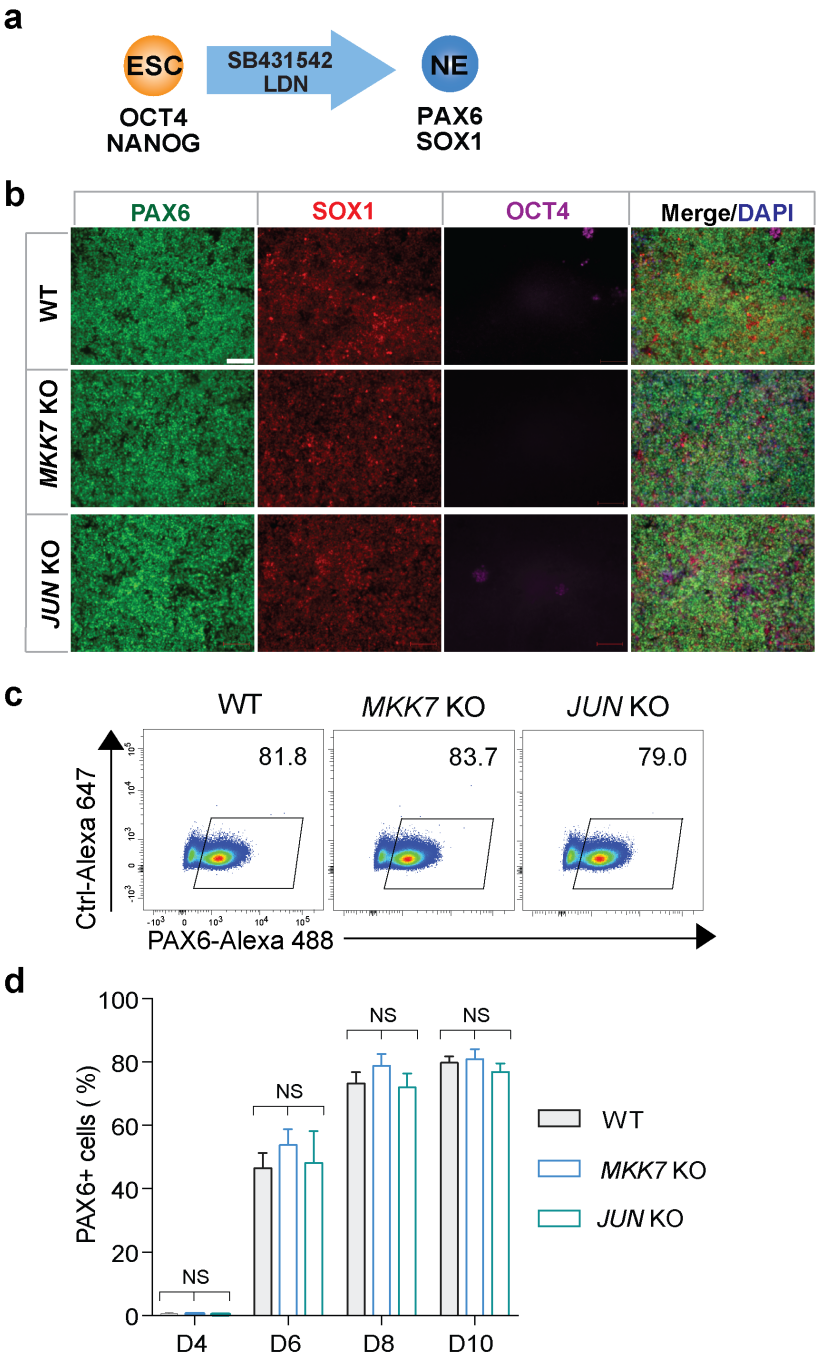


Figure 3.3 Neuroectoderm (NE) differentiation of *MKK7* and *JUN* KO hESC lines

a. NE differentiation schematic. The day when NE differentiation was initiated is designated as D0. Cells were examined on D4, 6, 8 and 10 of NE differentiation.
b. Immunostaining of PAX6, SOX1 and OCT4 on D10 of NE differentiation from WT, *MKK7* KO and *JUN* KO cell lines.
c. Representative FACS dot plots of D10 NE cells stained for PAX6 (WT, *MKK7* KO and *JUN* KO cell lines).
d. FACS quantification of differentiation efficiency based on the percentage of $PAX6^+$ NE cells from D4 to D10 (WT, *MKK7* KO and *JUN* KO cell lines). . n=3 independent experiments.

Figures 3.3c-d are contributed by Nipun Verma.
Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05). Scale bar, 100 μm.

Figure 3.4 Pharmacological JNK inhibition improves endoderm differentiation

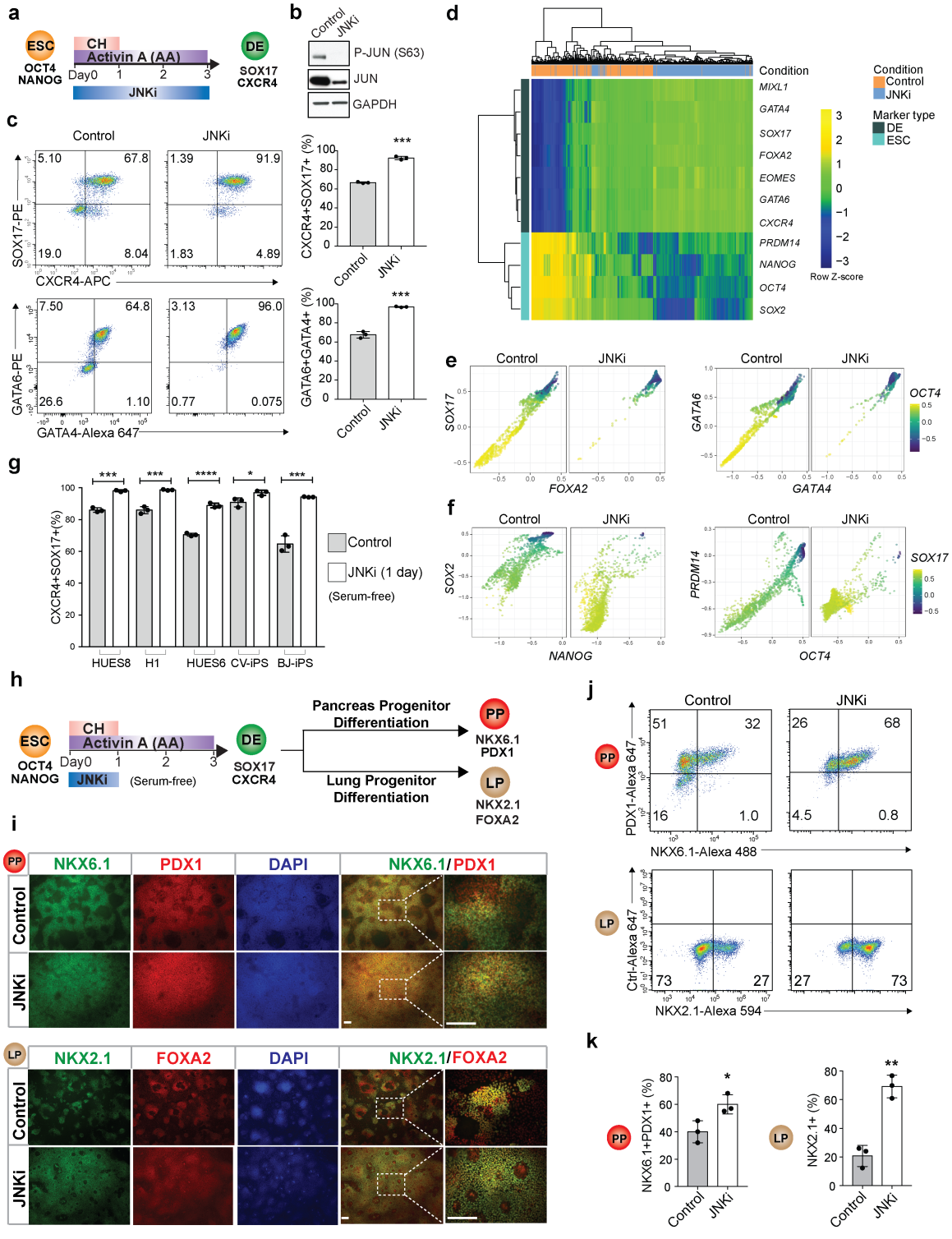


Figure 3.4 Pharmacological JNK inhibition improves endoderm differentiation

a. Illustration of JNK pathway inhibition during DE differentiation. JNKi indicates the JNK inhibitor (JNK-IN-8).
b. Western blotting analysis for JUN and P-JUN (S63) expression of H1-derived DE D3 cells from control and JNKi treated condition. DMSO was used as a vehicle control.
c. Representative FACS dot plots (left) and quantification (right) for DE D3 cells co-stained for CXCR4 and SOX17, GATA6 and GATA4 from control and JNKi treated condition. n=3 independent experiments. Hierarchical clustering of single cell expression values for known marker genes of DE and ESC. Analysis was performed following DE differentiation. Each row represents an individual gene, with DE signature genes annotated with teal-colored bars and ESC signature genes marked in cyan. Each column represents an individual cell, where JNKi treated cells are labeled blue and control cells are labeled orange. The color scheme of the heatmap corresponds to the expression levels of the genes in the cells (row-based z-scores) where dark blue reflects the lowest expression value across all cells for a given gene.
d. Expression patterns of selected DE genes in individual cells following DE differentiation. Every dot represents a single cell; its position being determined by the expression levels for the genes indicated on the x and y-axis. The color gradient reflects *OCT4* expression.
e. Expression patterns of selected ESC genes in individual cells following DE differentiation. Every dot represents a single cell; its position being determined by the expression levels for the genes indicated on the x and y-axis. The color gradient reflects *SOX17* expression.
f. FACS quantification of differentiation efficiency based on the percentage of $CXCR4^{+}SOX17^{+}$ DE D3 cells from untreated control and JNKi (1 Day) treated condition, including HUES8, H1, HUES6 hESC lines, CV,BJ-hiPSC lines using the serum free protocol. n=3 independent experiments.
g. Schematic of pancreatic and lung progenitor differentiation (PP: Pancreatic Progenitor, LP: Lung Progenitor). Representative PP differentiation data shown below were differentiated from DE cells treated with 20 ng/mL Activin A. Similar results were observed with 100 ng/ml Activin A treatment.
h. Immunostaining with PP markers PDX1 and NKX6.1 and LP markers NKX2.1 and FOXA2. Scale bar, 100 μm.
i. Representative FACS dot plots for PP cells co-stained for NKX6.1 and PDX1, or LP cells stained for NKX2.1.

j. FACS quantification of differentiation efficiency of PP or LP cells. n=3 independent experiments.

Figures 3.4 d-f are contributed by Friederike Dundar.
Figures 3.4 h-k about lung differentiation are contributed by Miriam Gordillo.
Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05). Scale bar, 100 μm.

Figure 3.5 Transcriptional profiling of JNKi treated DE cells

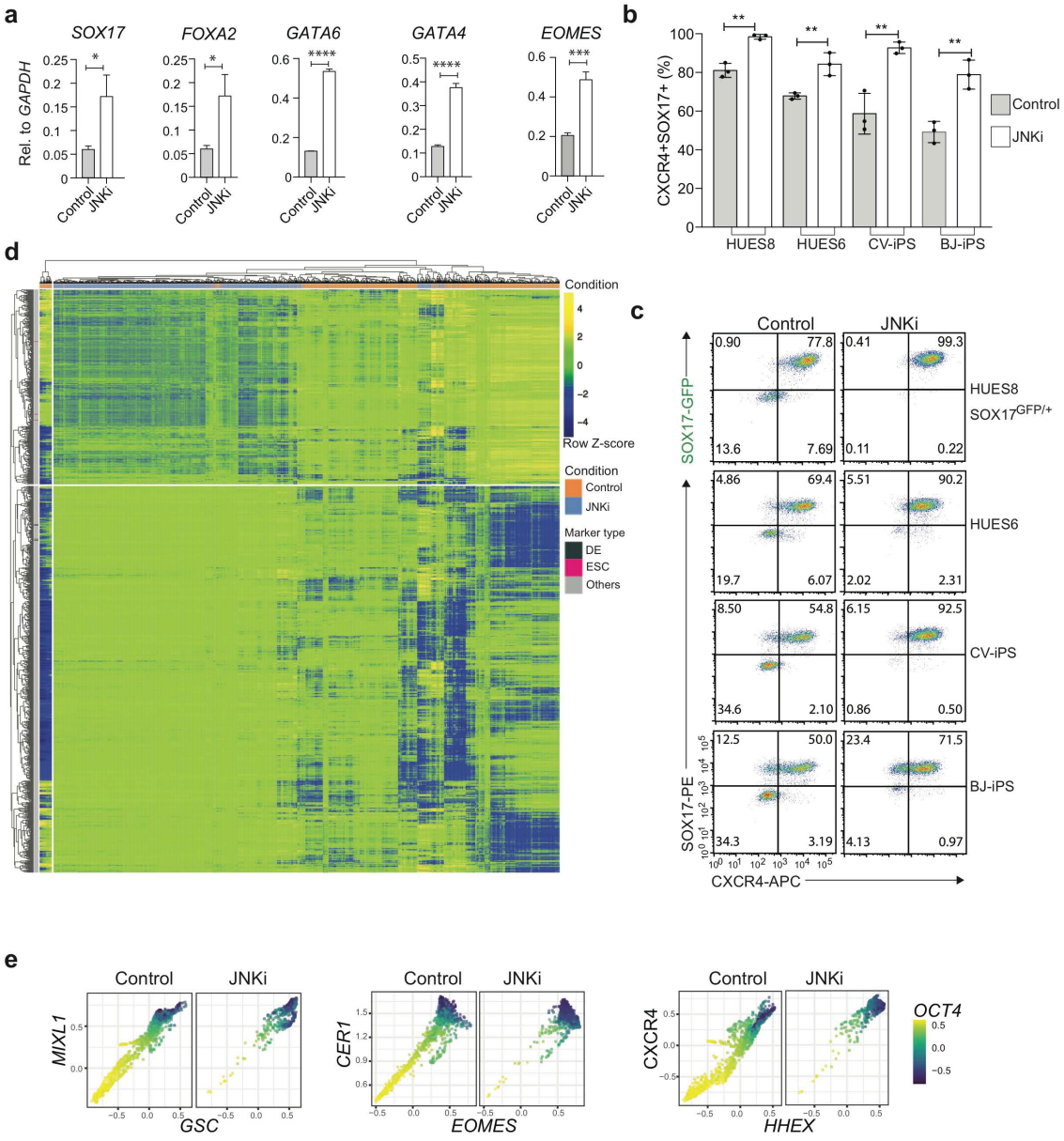


Figure 3.5 Transcriptional profiling of JNKi treated DE cells

a. Gene expression analysis by RT-qPCR of DE marker genes in DE cells differentiated from control and JNKi treated condition, n=3 independent experiments.

b. FACS quantification of differentiation efficiency based on the percentage of $CXCR4^{+}SOX17^{+}$ DE D3 cells from untreated control and JNKi treated condition, including HUES8-$SOX17^{GFP/+}$ line, HUES6 hESC line, CV,BJ-hiPSC lines. n=3 independent experiments.

c. Representative FACS dot plots of DE D3 cells co-stained for SOX17 and CXCR4 from untreated control and JNKi treated condition, including HUES8-$SOX17^{GFP/+}$ line, HUES6 hESC line, CV,BJ-hiPSC lines.

d. Unsupervised hierarchical clustering of the expression values from individual JNKi treated and control DE cells for the 1,000 most variably expressed genes. Each vertical line represents the gene expression profile for each single cell (orange bars label control DE cells; blue bars label JNKi treated DE cells). The ESC and DE signature genes that were shown in Fig. 3d are indicated by the dark grey (DE) and pink (ESC) bars. The color gradient of the heatmap reflects the row-normalized z-score of expression where dark blue indicates the lowest expression value for a particular gene. The gaps in the heatmap serve as a visual separator of the top two branches of the dendrograms for the clustering of both cells and genes, respectively.

e. Expression patterns of selected DE genes in individual cells following DE differentiation. Every dot represents a single cell. Their positions are determined by the expression levels for the genes indicated on the x and y-axis. The color gradient reflects *OCT4* expression.

Figures 3.5 d-e are contributed by Friederike Dundar.
Error bars indicate SD, and significance is indicated as *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$; ****$p< 0.0001$; NS, not significant ($p \geq 0.05$).

Figure 3.6 JNKi treatment improves DE and subsequent endoderm lineage differentiation

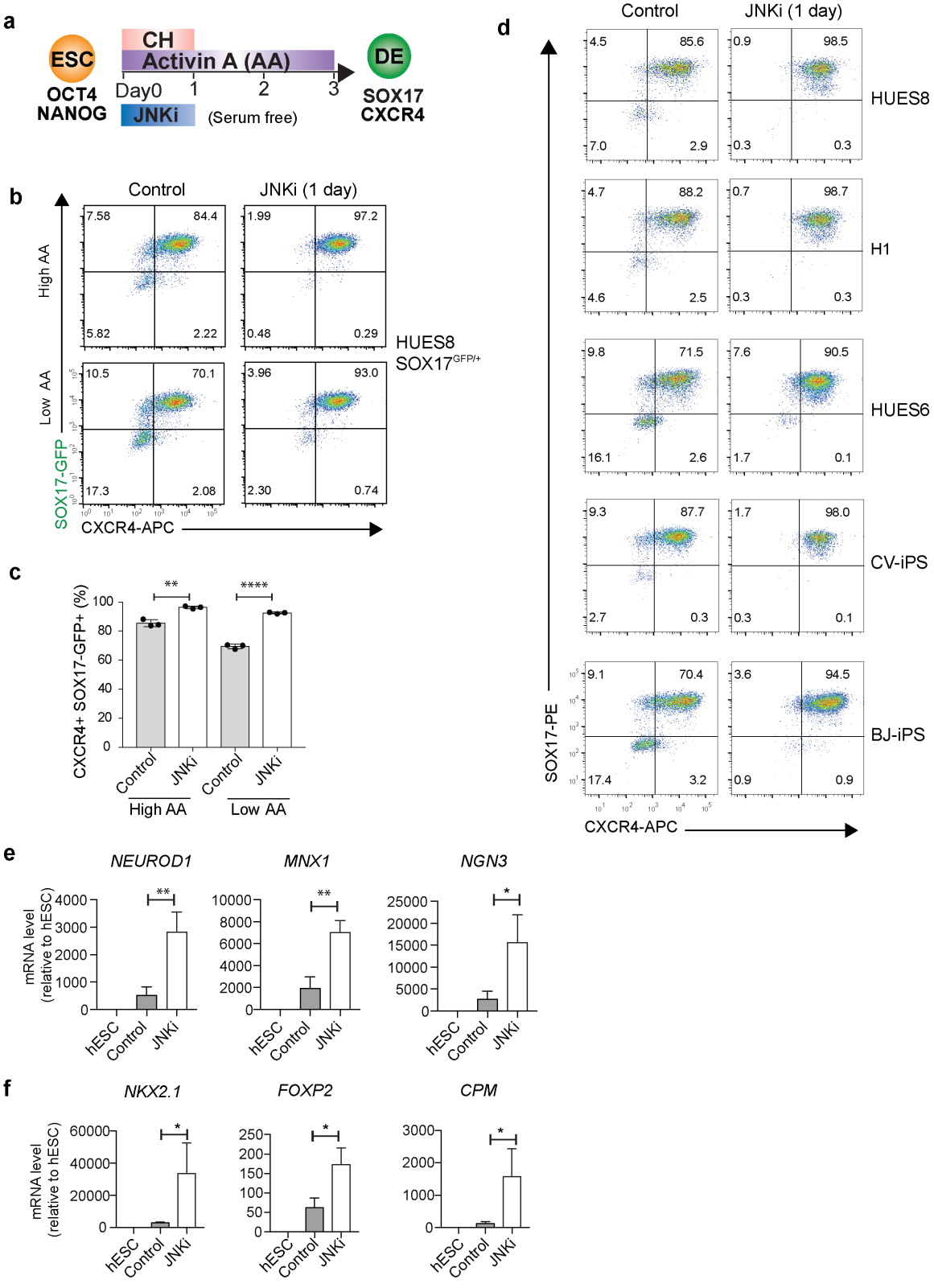


Figure 3.6 JNKi treatment improves DE and subsequent endoderm lineage differentiation

a. Schematic of DE differentiation for endoderm derivative differentiation. JNKi was only added during the first day of differentiation.
b. Representative FACS dot plots of DE D3 cells stained for CXCR4 and SOX17-GFP (FITC-gating) from untreated control and JNKi (1 day) treated condition, HUES8-SOX17$^{GFP/+}$ line differentiated from high (100 ng/ml) and low (20 ng/ml) Activin A.
c. FACS quantification of differentiation efficiency based on the percentage of CXCR4^{+}SOX17-GFP^{+} DE D3 cells from (b). n=3 independent experiments.
d. Representative FACS dot plots of DE D3 cells stained for CXCR4 and SOX17-GFP from untreated control and JNKi (1 day) treated condition, including HUES8 (non-reporter), H1, HUES6 hESC lines, CV,BJ-hiPSC lines.
e. RT-qPCR analysis of pancreatic progenitor markers gene expression. The relative expression levels were normalized to the housekeeping gene *GAPDH*, and further normalized to the expression level in hESCs. n=3 independent experiments.
f. RT-qPCR analysis of lung progenitor markers gene expression. The relative expression levels were normalized to the housekeeping gene *GAPDH*, and further normalized to the expression level in hESCs. n=3 independent experiments.

Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05).

Figure 3.7 JUN impedes chromatin landscape remodeling during ESC-DE differentiation

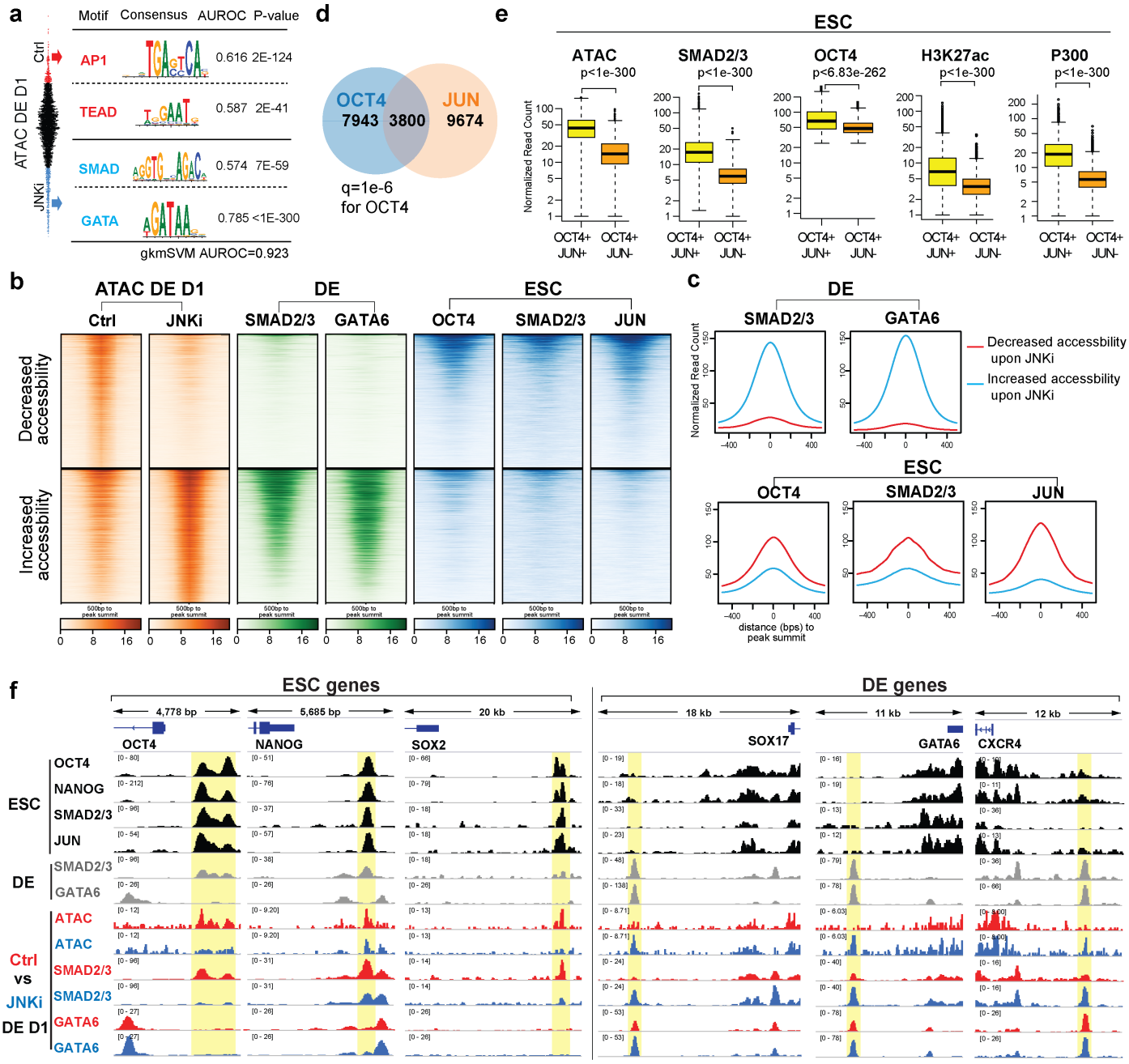


Figure 3.7 JUN impedes chromatin landscape remodeling during ESC-DE differentiation

a. Motif enrichment in ATAC-seq peaks with decreased or increased accessibility upon JNKi treatment in DE D1 cells compared to WT (Ctrl). Beeswarm plot shows change in ATAC DE D1 signal WT vs JNKi, motifs and gkmSVM are trained on genomic intervals in the tails of this distribution.

b. Heatmaps of ATAC-seq signals at peaks with decreased/increased accessibility upon JNKi treatment (orange). Increased accessibility regions are enriched in ChIP-seq signal of SMAD2/3 and GATA6 in DE cells (green). Decreased accessibility regions are enriched in ChIP-seq signals of JUN, OCT4 and SMAD2/3 in ESC cells (blue).

c. Average ChIP-seq signals at ATAC-seq peaks with decreased accessibility (red) or increased accessibility (blue) upon JNKi treatment.

d. The Venn diagram shows overlapping OCT4 and JUN binding sites at the ESC stage using a more stringent peak calling criterion for OCT4 (q=1e-6) ($p<1e-300$).

e. The boxplots show the average ATAC-seq, SMAD2/3, OCT4, H3K27ac[48], and p300[48] ChIP-seq signals over the OCT4+JUN+ and OCT4+JUN- regions at the ESC stage. On these boxplots, boxes show interquartile range, whiskers show fixed multiples of interquartile range, and solid line shows median.

f. Relevant ATAC-seq and ChIP-seq tracks of ESC genes (*OCT4*, *NANOG* and *SOX2*) and DE genes (*SOX17*, *GATA6* and *CXCR4*)

Figures 3.7a-e are contributed by Mike Beer in collaboration with Danwei Huangfu.

Figure 3.8 The effect of JNK inhibition on chromatin accessibility and transcription factors binding in ESC-DE lineage transition

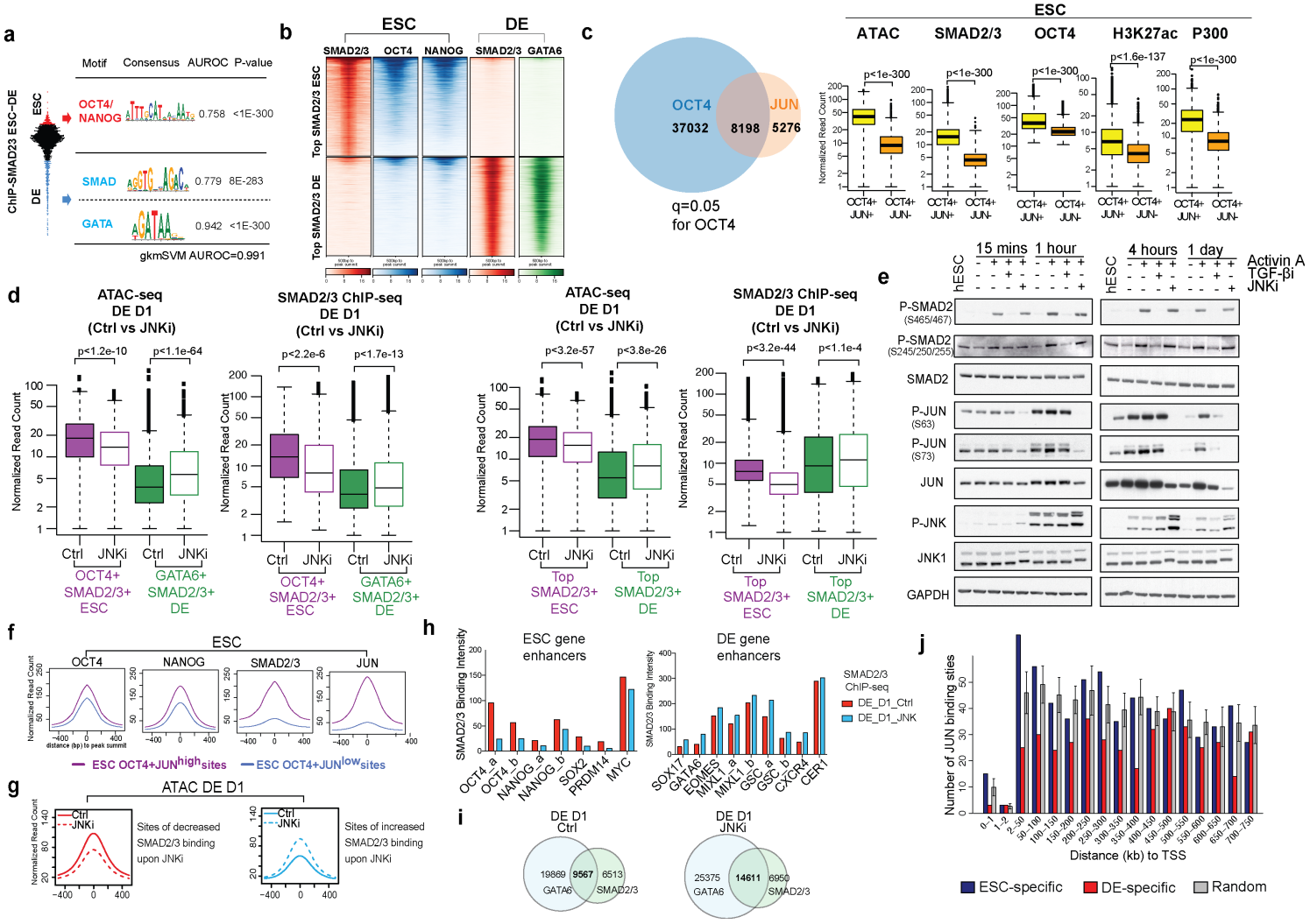


Figure 3.8 The effect of JNK inhibition on chromatin accessibility and transcription factors binding in ESC-DE lineage transition

a. Motif enrichment at regions of differential SMAD2/3 ChIP-seq binding signals in ESC and DE cells. Beeswarm plot shows change in SMAD2/3 binding signal ESC vs DE, motifs and gkmSVM are trained on genomic intervals in the tails of this distribution.

b. Heatmaps of TF binding intensities at differentially bound SMAD2/3 sites in ESC and DE D3 stage.

c. The Venn diagram shows overlapping OCT4 and JUN binding sites at the ESC stage using the standard peak calling criteria (q=0.05) (p<3.89e-224) for OCT4. The boxplots show the average ATAC-seq, SMAD2/3, OCT4, H3K27ac[48], and p300[48] ChIP-seq signals over OCT4+JUN+ and OCT4+JUN- regions at the ESC stage. On these and following boxplots, boxes show interquartile range, whiskers show fixed multiples of interquartile range, and solid line shows median.

d. Boxplots show the normalized read counts of ATAC-seq and SMAD2/3 ChIP-seq of DE D1 Ctrl and JNKi condition in OCT4+SMAD2/3+ ESC sites and GATA6+SMAD2/3+ DE sites (left 2 panels). Boxplots show the normalized read counts of ATAC-seq and SMAD2/3 ChIP-seq of DE D1 Ctrl and JNKi condition in the strongest SMAD2/3+ESC and strongest SMAD2/3+ DE sites (right 2 panels).

e. Western blotting analysis of protein expression at 15 minutes, 1 hour, 4 hours and 1 day after initiating DE differentiation showed that Activin A treatment induces C-terminal SMAD2 phosphorylation (at Ser465/467) as expected, an effect blocked by SB431542, a selective inhibitor of ACVR1B/ALK4, TGFBR1/ALK5 and ACVR1C/ALK7. JNKi treatment inhibited JUN phosphorylation, but did not change the level of C-terminal SMAD2 phosphorylation (S465/467) and linker SMAD2 phosphorylation (S245/250/255). Conversely SB431542 treatment also did not affect the level of JUN phosphorylation.Average ChIP-seq signals of OCT4, NANOG, SMAD2/3 and JUN at regions of OCT4+JUN^{high} ESC enhancers and OCT4+JUN^{low} ESC enhancers.

f. Average ATAC-seq signals at regions of decreased (red) or increased (blue) SMAD2/3 binding upon JNKi treatment at DE D1.

g. SMAD2/3 binding intensity (number of reads per 300bp genomic interval) at ESC and DE associated gene enhancers based on SMAD2/3 ChIP-seq analysis of DE D1 control and DE D1 JNKi.

h. Venn diagram of GATA6 and SMAD2/3 unique and common binding sites in DE D1 Ctrl and DE D1 JNKi.

i. ChIP-seq peaks of increased JUN binding in ESC relative to DE D1 are preferentially located near the top 200 ESC-expressed genes (blue) and farther from the top 200 DE-expressed genes (red) compared to random gene sets (grey). 142 are located within 100kb of the transcription start site (TSS) of one of the 200 genes with the greatest fold decrease in expression during DE differentiation, which is a significant enrichment compared to the expected overlap of 107.6±11.3 binding sites by chance from sampling ten random sets of 200 genes ($p<0.0012$).

Figures 3.7 a-d,f-l are contributed by Mike Beer in collaboration with Danwei Huangfu.

Figure 3.9 JUN impedes SMAD2/3 reconfiguration during ESC-DE differentiation

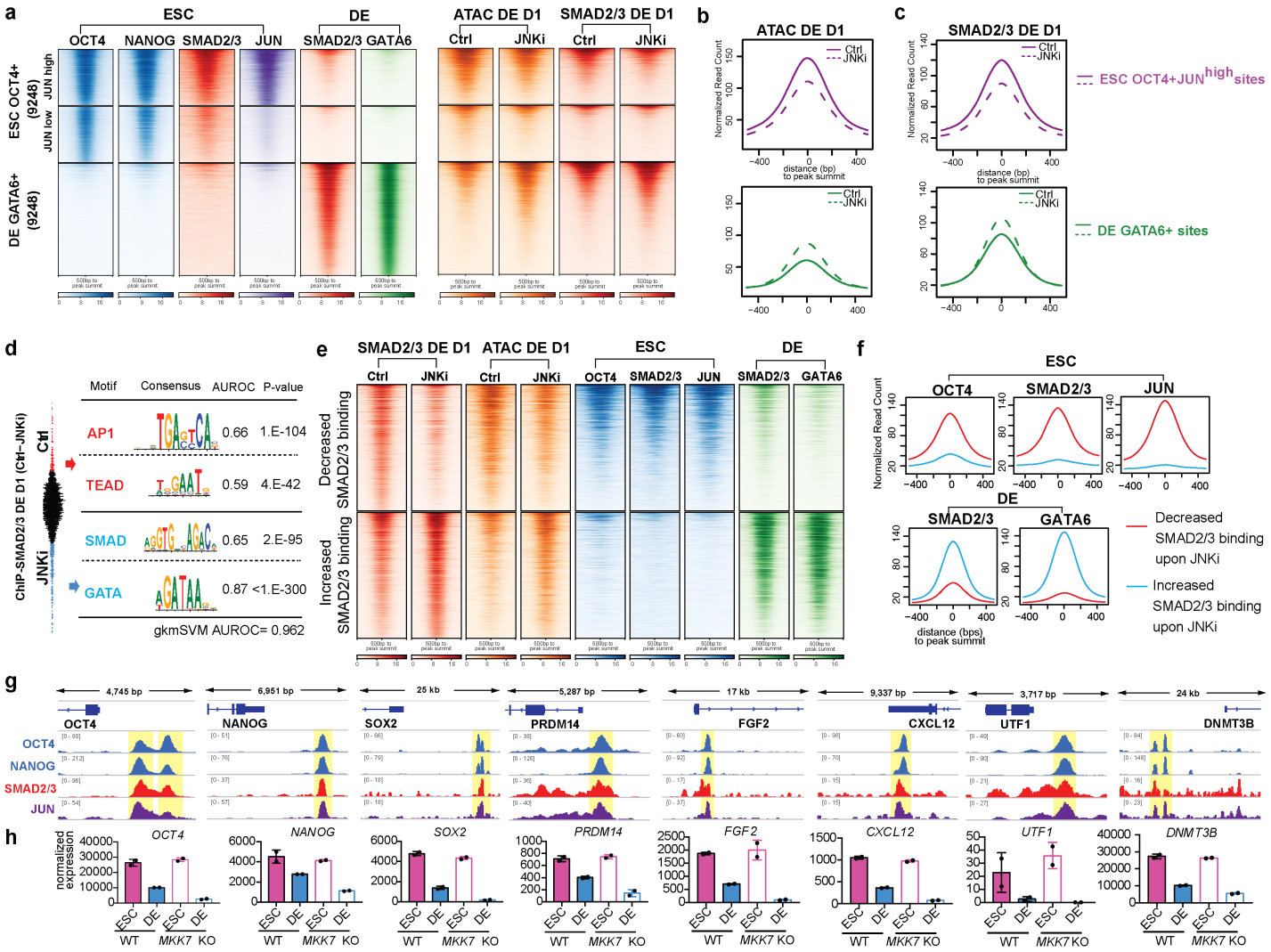


Figure 3.9 JUN impedes SMAD2/3 reconfiguration during ESC-DE differentiation

a. Heatmaps of ChIP-seq and ATAC-seq signals enrichment at top OCT4 bound enhancers in ESC (9248) and top GATA6 bound enhancers in DE (9248). OCT4+ enhancers were divided into JUN^{high} and JUN^{low} groups based on JUN ChIP-seq signals. Similar results were obtained when examining OCT4+ ChIP-seq peaks based on standard peak calling parameter (q=0.05).

b. Average ATAC-seq signals of DE D1 (Ctrl vs JNKi) at regions of OCT4+JUN^{high} ESC enhancers and GATA6+ DE enhancers.

c. Average SMAD2/3 ChIP-seq signals of DE D1 (Ctrl vs JNKi) at regions of OCT4+JUN^{high} ESC enhancers and GATA6+ DE enhancers.

d. Motif enrichment in SMAD2/3 ChIP-seq peaks with decreased SMAD2/3 binding (Ctrl) and increased SMAD2/3 binding (JNKi) upon JNKi treatment in DE D1 cells. Beeswarm plot shows change in SMAD2/3 binding signal WT vs JNKi, motifs and gkmSVM are trained on genomic intervals in the tails of this distribution.

e. Heatmaps of SMAD2/3 ChIP-seq signals at peaks with decreased/increased SMAD2/3 binding upon JNKi treatment. Decreased SMAD2/3 binding sites are enriched in ChIP-seq signals of OCT4, SMAD2/3 and JUN in ESC (blue). Increased SMAD2/3 binding sites are enriched in ChIP-seq signals of SMAD2/3 and GATA6 in DE (green). Heatmaps of ATAC-seq signals at peaks with decreased/increased SMAD2/3 binding were shown in orange.

f. Average ChIP-seq signals at regions of decreased (red) or increased (blue) SMAD2/3 binding upon JNKi treatment.

g. Representative ChIP-seq track images of OCT4, NANOG, SMAD2/3 and JUN binding at ESC gene enhancers at the ESC stage.

h. RNA-seq analysis of the corresponding genes (g) in ESC and DE cells from WT and *MKK7* KO cell lines. n=2 independent experiments. Error bars indicate SD.

Figures 3.9a-f are contributed by Mike Beer in collaboration with Danwei Huangfu.

Figure 3.10 Modeling the effect of JNK inhibition on ESC-DE transition

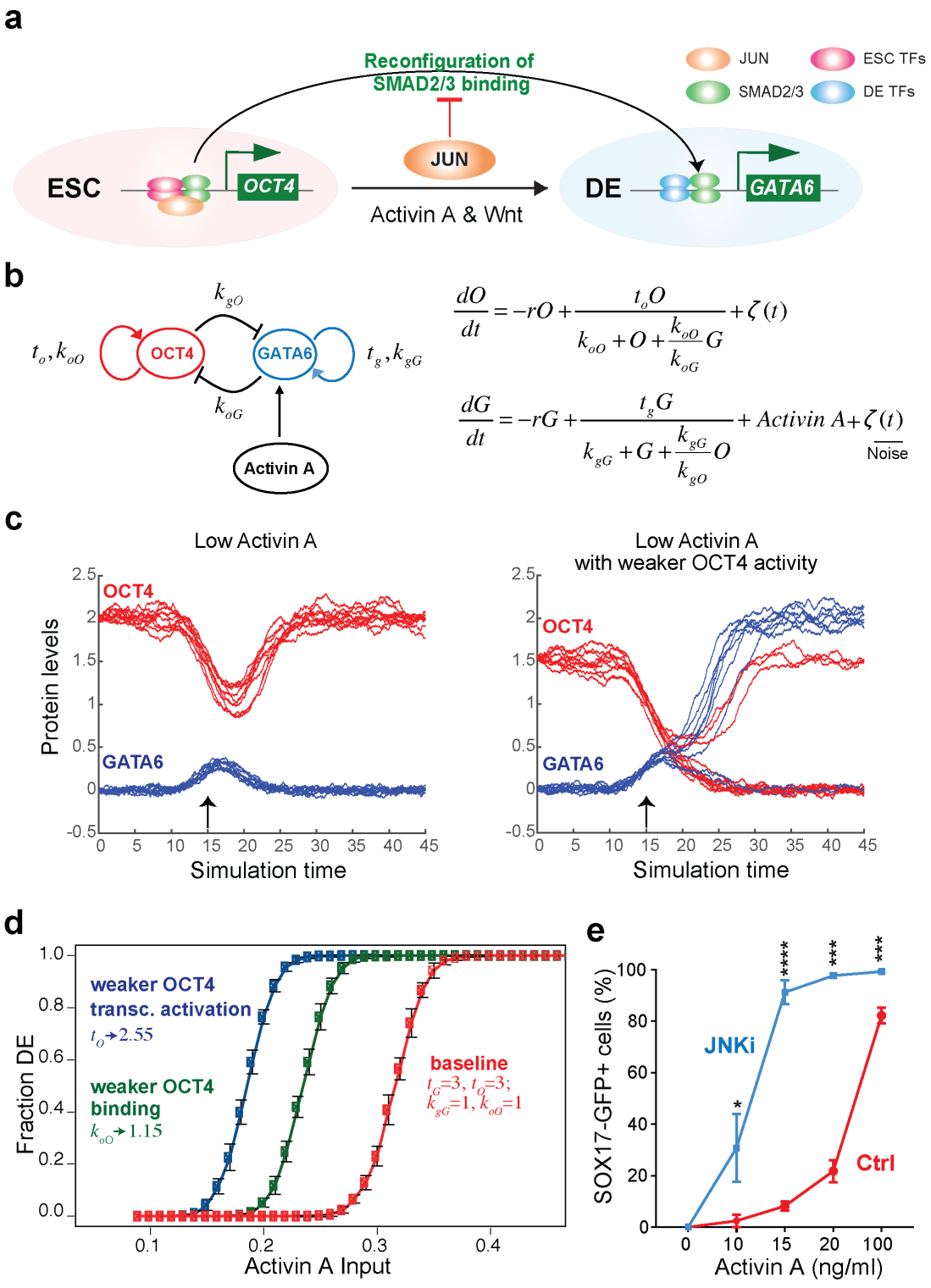


Figure 3.10 Modeling the effect of JNK inhibition on ESC-DE transition

a. Schematic for the role of JUN in reconfiguration of SMAD2/3 binding during ESC-DE transition. Different types of transcription factors (TFs) are labeled with different colors at top right corner.

b. Stochastic bifurcation model: O=OCT4 and G=GATA6 represent ESC and DE TF protein concentrations, r: degradation rate, t: transcription rate, k: dissociation constants for binding of each TF (upper case) to either enhancer (lower case). Binding of a TF to its own enhancer is transcriptionally activating, binding to the other enhancer is repressive.

c. Response of OCT4 (red) and GATA6 (blue) to a weak Activin A impulse at time t=15 (pseudo time) with baseline parameters (left), and with reduced auto-activation of OCT4 (right) via reduced transcription, t_o, or reduced binding at OCT4 enhancer (increased k_{oO}). Reduced auto-activation of OCT4 enables transition to the DE (high GATA6) state.

d. Modeling the effect of OCT4 auto-activation on endoderm differentiation efficiency in response to varying Activin A doses. Red: baseline OCT4 activity. Green line: weaker OCT4 binding at its enhancer. Blue line: weaker OCT4 transcriptional activation. 20 simulations.

e. FACS quantification of differentiation efficiency based on the percentage of SOX17-GFP$^+$ DE cells at different Activin A doses. n=3 independent experiments.

Figures 3.10b-d are contributed by Mike Beer in collaboration with Danwei Huangfu.

Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05).

Figure 3.11 JNK inhibition lowers the threshold of Activin A doses for efficient endoderm differentiation

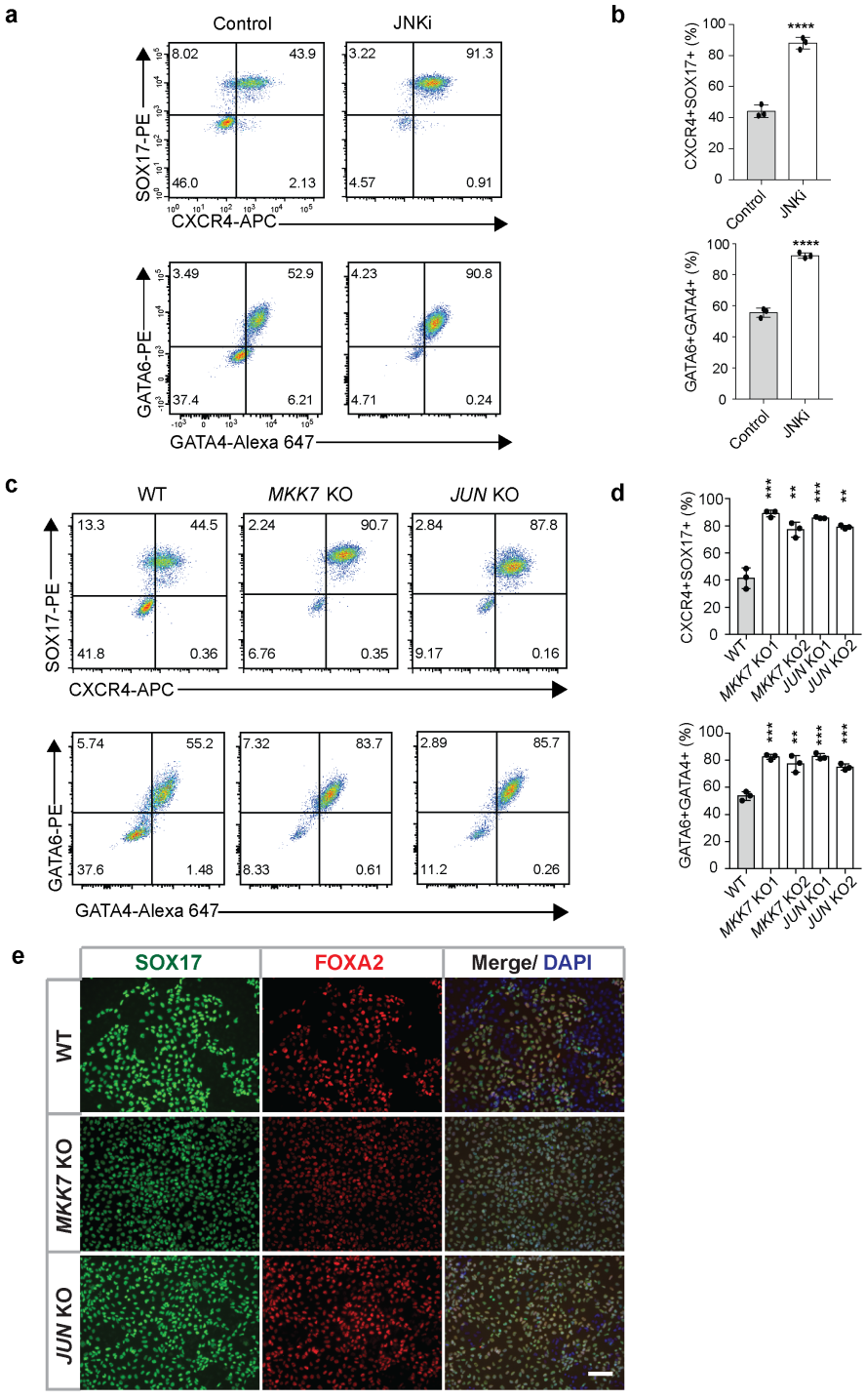


Figure 3.11 JNK inhibition lowers the threshold of Activin A doses for efficient endoderm differentiation

a. Representative FACS dot plots of DE D3 cells stained for CXCR4 and SOX17, GATA4 and GATA6 from untreated control and JNKi treated condition (H1 hESC line, 20ng/ml Activin A).
b. FACS quantification of differentiation efficiency based on the percentage of $CXCR4^{+}SOX17^{+}$, $GATA6^{+}GATA4^{+}$ DE D3 cells from (a). n=3 independent experiments.
c. Representative FACS dot plots of DE D3 cells stained for CXCR4 and SOX17, GATA6 and GATA4 (H1 hESC line, WT vs *MKK7* and *JUN* KO, 20ng/ml Activin A).
d. FACS quantification of differentiation efficiency based on the percentage of $CXCR4^{+}SOX17^{+}$, $GATA6^{+}GATA4^{+}$ DE D3 cells from (c). n=3 independent experiments.
e. Immunostaining of DE markers SOX17 and FOXA2 in WT, *MKK7* KO and *JUN* KO DE cells differentiated from 20 ng/ml Activin A condition. Scale bar, 100 μm.

Error bars indicate SD, and significance is indicated as *p < 0.05, **p < 0.01, and ***p < 0.001; ****p< 0.0001; NS, not significant (p ≥ 0.05). Scale bar, 100 μm.

3.3 Discussion

The role of AP-1 in ESC-DE transition

Both hESC self-renewal and DE differentiation involve Nodal/TGF-β signaling via the SMAD2/3 complex. During differentiation, SMAD2/3 chromatin binding undergoes a dramatic reconfiguration from ESC to DE enhancers. Our comprehensive screens using two independent CRISPR libraries uncovered the JNK/JUN pathway as a key barrier from pluripotency to DE differentiation. Unlike a typical differentiation inhibitory pathway, the JNK/JUN pathway does not act through directly inhibiting the DE enhancers. Instead it primarily acts on the ESC enhancers. JUN and OCT4 co-occupy many enhancers at the ESC stage, and JUN binding correlates with higher enhancer activities as indicated by increased chromatin accessibility as well as H3K27ac and P300 signals. JNK inhibition during DE differentiation accelerates the destabilization of the pluripotency network, as supported by the decreased chromatin accessibility at ESC enhancers, which is accompanied by decreased SMAD2/3 occupancy at ESC enhancers thus facilitating the reconfiguration of SMAD2/3 chromatin binding to DE enhancers. Our mathematical modeling further supports that the JNK/JUN pathway safeguards the pluripotency gene regulatory network, and also predicts that JNK inhibition during DE differentiation would lower the Activin A dose requirement, which was experimentally verified.

The similar phenotypes observed in *JUN* and *MKK7* knockout hESCs argue that JUN is the predominant AP-1 TF involved in DE differentiation, though potential overlapping requirements of additional AP-1 TFs cannot be excluded. It

is of great interest to determine whether the inhibitory role of the JNK/JUN pathway in DE differentiation is conserved in mice or human-specific. It is not feasible to directly examine the roles of the JNK/JUN pathway in gastrulating human embryos, and the exact role of the JNK/JUN pathway in mouse DE differentiation is unclear. *Jun* knockout mouse embryos showed developmental arrest at around embryonic day 12, but earlier phenotypes during DE differentiation have not been examined[148,149]. Deletion of *Jnk1/2* or *Jun* in mouse ESCs (mESCs) did not have a significant effect on early endoderm gene expression in spontaneous embryoid body and retinoic acid-induced differentiation assays, but a specific effect on DE differentiation has not been examined [150,151]. More quantitative experiments are needed to characterize the exact involvement of the JNK/JUN pathway in DE differentiation from mouse embryos and mESCs.

The potential function of AP-1 in guarding lineage identity

Recent studies on somatic cell reprogramming show that the AP1 motif is strongly associated with chromatin regions undergoing closing during the initial phase of reprogramming mouse fibroblasts to iPSCs [151-153]. Further over-expression and loss-of-function studies show that AP1 TFs such as Jun and Fra1 act as barriers to the acquisition of the pluripotent state. Here we show that in pluripotent hESCs, JUN binding strongly associates with OCT4-bound ESC enhancers with higher enhancer activities. Instead of inhibiting the acquisition of the pluripotent state as shown in the context of somatic cell reprogramming, JUN

impedes the exit from the pluripotent state. It may seem counterintuitive that JUN inhibits both the acquisition and the dissolution of the pluripotent state. This paradox can be explained by a unifying theme that AP1 TF JUN is a guardian of lineage identity, which is the ESC identity in the context of pluripotency dissolution, or the initial fibroblast identity in the context of pluripotency acquisition. This concept of AP1 in safeguarding lineage identity may also apply to other lineages as has been suggested recently for macrophages, in which the AP-1 motif is found through Hi-C genomic data analysis to be highly enriched in the activation hubs that connect multiple enhancers to promoters[154].

The role of JUN and SMAD factors in hESC pluripotency maintenance

Although I provided a mechanistic explanation for JNK-JUN mediated inhibition of DE differentiation, the role of JNK pathway in pluripotency maintenance still need to be investigated. ChIP-seq analysis revealed that ESC enhancers are co-occupied by OCT4, NANOG, SMAD2/3 and JUN, however *JUN*-/- and *MKK7*-/- hESCs can self-renew and express all pluripotency genes. These results demonstrate that JUN is not required for pluripotency maintenance for undifferentiated hESCs. The screen results also suggest *SMAD2* and *SMAD4* are not required for pluripotency maintenance for undifferentiated hESCs. Clonal *SMAD4*-/- hESC generated by Chien and colleagues generated express normal level of *OCT4* and *NANOG*, and shows no sign of loss of pluripotency maintenance and self-renewal[75]. Our lab also successfully derived clonal *SMAD4* and *SMAD2* knockout hESC and confirmed the phenotype that loss of

SMAD4 or *SMAD2* alone did not affect pluripotency maintenance (unpublished data).

However, Nodal signaling is indeed required for hESC pluripotency maintenance[155] , as hESC went through rapid differentiation after TGFβ SB431542 inhibitor treatment. SB431542 is a potent inhibitor for TGFβ, Activin A and Nodal signaling, because it inhibits the kinase activity of ALK4, ALK5 and ALK7[156]. As a result, SMAD2 or SMAD3 dependent TGFβ pathway should still be required for hESC pluripotency maintenance. To explore the exact contribution of each SMAD factor in pluripotency maintenance, I will need to examine the phenotype of *SMAD2*-/- *SMAD3*-/- double knockout hESC line. It is possible SMAD3 may compensate the role of SMAD2 in maintaining pluripotency. If *SMAD2*-/- *SMAD3*-/- double knockout hESC line shows loss of pluripotency phenotype, then it will demonstrate the important role of SMAD2 and SMAD3 in Nodal signaling mediated pluripotency maintenance. However, SMAD4 is the common transcription co-factor for SMAD2/3 in the canonical TGFβ-SMAD2/3-SMAD4 signaling axis, and SMAD4 is not required for pluripotency maintenance. In this case, it is possible other SMAD2/3 co-factors may exist to compensate the loss of *SMAD4*. Therefore, it will be informative to perform a CRISPR screen in *SMAD2*-/- hESC line and identify the co-factors that are required to maintain hESC pluripotency when *SMAD2* is lost.

Specific function of SMAD2 and SMAD4 in DE differentiation

Interestingly, loss of *SMAD2* or *SMAD4* alone is sufficient to block DE differentiation in hESC, which is a phenotype similar to previous mouse studies[57,59]. *SMAD3* is not found to be a positive regulator of DE differentiation, and the phonotype will need to be confirmed from clonal line study in future. It seems that Nodal signaling is acting through the conical SMAD2-SMAD4 axis in the context of DE differentiation, which is slightly different from the SMAD4 independent Nodal signaling pathway in the context of hESC pluripotency maintenance. Increasing evidence suggests substantial biochemical differences between SMAD2 and SMAD3 and their ability to translate TGFβ signaling[157]. It will be of interest to perform specific SMAD2, SMAD3, and SMAD4 ChIP-seq experiments in ESC and DE cells to study their specific function in future.

TEAD motif is co-regulated with AP-1 motif

The abundance of TEAD motif in addition to AP1 motif is decreased upon JNK inhibition in DE D1 cells from ATAC-seq and SMAD2/3-ChIP analysis. The strong association of AP-1 motif and TEAD motif is also commonly observed in many other recent studies [158] [159-162]. ChIP-seq analysis of YAP1, TEAD and AP1 transcription factors in cancer cells show a strong degree of overlapping binding. Co-IP studies also demonstrate protein-protein interaction between TEAD and AP1 TFs[158,160]. The strong co-binding pattern of AP1 and TEAD is also supported by identifying a defined co-occurring AP1 (TGA[GC]TCA) and TEAD(ATTCC)

motifs connected by a 7bp spacer [159]. This suggests that the binding activity of TEAD can be regulated by JNK inhibition.

TEAD proteins are a group of transcription factors (TEAD1/2/3/4) in the Hippo signaling pathway. As a co-factor of YAP1, TEADs bind to DNA and regulate YAP1-mediated gene expression. *TEAD1* is identified as a negative regulator from the Brunello screen (ranked 56th), and *YAP1* is identified as a negative regulator from the GeCKO screen (ranked 40th). Therefore, inhibition of TEAD or YAP1 binding may have similar effects as JNK inhibition on DE differentiation. The involvement of Hippo pathway in DE differentiation is also supported by the validation results of *NF2, TAOK1, PTPN14*, and *PPP2R4* as positive regulators of DE differentiation. These four genes are in the upstream components of Hippo signaling pathway, which negatively regulate YAP1 activity. A recent paper shows *YAP*-/- hESC line can self-renew and express normal level ESC genes [163]. Loss of *YAP1* in hESCs promotes Activin A induced mesendoderm and cardiac lineage differentiation. The phenotype of DE differentiation has not been examined, but it is reasonable to predict that loss of *YAP1* may promote hESC exit ESC pluripotency state, which in turn allow for efficiency differentiation to either mesoderm or endoderm lineage. In order to investigate the function of TEAD in ESC-DE transition, I performed TEAD4 ChIP-seq in hESCs (unpublished data). TEAD4 indeed binds to many enhancers of ESC genes (such as *OCT4* and *NANOG*), supporting the idea that TEAD is also part of the ESC TF network. Because loss of *YAP1* does not affect hESC pluripotency, additional factors may need to be deleted in order to completely

disrupt the pluripotency network. Considering that SMAD2, JUN, TEAD4 all bind to ESC gene enhancers, it is interesting that loss of a single gene alone by itself does not affect pluripotency network. Future studies should employ a double or compound knockout approach to dissect the functions of these genes. Alternatively, a sensitized genetic screen can be performed in *SMAD2*-/-, *JUN*-/- or *YAP1*-/- hESC line to identify the co-factors required for pluripotency maintenance in the special genetic background.

3.4 Materials and Methods

Generation of clonal KO hESC lines

Clonal KO lines were created as previously described [105] with some modifications. H1 iCas9 cells were infected with lentiviruses expressing *MKK7* or *JUN* targeting gRNAs made from the lentiCRISPR v2 construct (Table 4) on D0. Next, cells were treated with 2 μg/mL doxycycline (D1 to 7) and 0.5 μg/mL puromycin (D2 to 7). On D7, infected ES cells were dissociated to single cells using TrypLE Select. 500 cells were plated into one 100 mm tissue culture dish with 10 mL E8 media supplemented with 10 μM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals, S1049) for colony formation. After 10 days of expansion, 50 colonies were picked into individual wells in a 96-well plate. Genomic DNA was extracted for PCR genotyping. Primers used for PCR and sequencing are listed in Table 3.

Serum-free DE differentiation and subsequent pancreatic and lung differentiation

We performed DE differentiation using an optimized serum-free protocol[37,52] followed by pancreatic or lung lineage differentiation. For serum-free DE differentiation, 0.3 million HUES8, H1, HUES6, BJ-iPSC cells were plated in E8 media in one well of a six-well plate coated with human recombinant vitronectin (Thermo Fisher Scientific, A14700). 24 hours later, DE differentiation was initiated. For CV-iPSC line, 0.8 million cells were plated and cultured in E8 media for two days before DE differentiation due to its slower growth rate. Cells were

washed with PBS once and culture media was changed to MCDB131 (Life Technologies, 10372-019) with 1x GlutaMax, 0.5% BSA (Lampire Biological Products, 7500804), 1.5 g/L NaHCO3, 10 mM glucose (Thermo Fisher Scientific, S22060), supplemented with 5 μM CHIR99021 and 100 or 20 ng/mL Activin A for 1 day. On the second day, the media was supplemented with 0.5 μM CHIR99021 and 100 or 20 ng/mL Activin A as specified. On the third day, the media was supplemented with 100 or 20 ng/mL Activin A only.

For pancreatic lineage differentiation[37,38,40,52], ~1 million HUES8 hESCs were plated for DE differentiation as described above. DE cells were washed with PBS and cultured in MCDB131 with 1x GlutaMax, 0.5% BSA, 1.5 g/L NaHCO3, 10 mM glucose supplemented with 0.25 mM Vitamin C (Sigma, A4544), 50 ng/mL FGF7 (R&D, 251-KG) and 1.25 μM IWP2 (Tocris, 3533) for 2 days. Next, cells were cultured in MCDB131 with 1x Glutamax, 2% BSA, 2.5 g/L NaHCO3, and 10mM glucose supplemented with 0.25 mM Vitamin C, 50 ng/mL FGF7, 0.25 μM SANT-1 (Sigma, S4572), 1 μM all−trans−Retinoic acid (ATRA) (Sigma, R2625), 100 nM BMP inhibitor LDN193189 (Stemgent, 04-0019), 200 nM TPB (EMD Millipore, 565740) and 1X ITS-X (Life Technologies, 51500-056) for 2 days. Then, cells were cultured in MCDB131 with 1x Glutamax, 2% BSA, 2.5 g/L NaHCO3, and 10mM glucose supplemented with 0.25 mM Vitamin C, 50 ng/mL FGF7, 0.25 μM SANT-1, 0.1 μM ATRA, 200 nM LDN193189, 100 nM TPB and 1X ITS-X for 3 days. Media was changed every day.

For lung lineage differentiation[164,165], ~0.3 million HUES8 hESCs were plated for DE differentiation as described above. After 3 days of DE differentiation, further differentiation used base media consisting of 75% IMDM (Life Technologies) and 25% Ham's Modified F12 medium (Cellgro) supplemented with N2 and B27 (Life Technologies), 0.05% BSA Fraction V (Life Technologies), 200 mM L-glutamine (Life Technologies), 0.05 mg/mL ascorbic acid (Sigma) and $4.5x10^{-4}$ M MTG (Sigma). DE cells were differentiated in lung base media containing 10 μM SB431542 (Stemgent, 04-0010) and 2 μM Dorsomorphin (Tocris, 3093)[165] from D3 to D6. Then, from D6 to D9, cells were grown in lung base media containing 100 ng/mL of human recombinant BMP4 (R&D Systems, 314-BP-010), 0.5 μM ATRA (Sigma) and 3 μM CHIR99021[164]. Media was changed every day.

Neuroectoderm Differentiation

hESCs cultured in E8 were disaggregated using TrypLE Select for 5 minutes and washed using E8 media. The cells were plated on Matrigel (BD, 354234) coated dishes in E8 media with 10 μM Y-27632 at a density of 180,000-200,000 cells/cm^2. After 1 day of culture in E8 media, differentiation was initiated by switching to knockout serum replacement (KSR) media with 10 μM TGF-β receptor inhibitor SB431542 (Tocris, 161410) and 100 nM BMP inhibitor LDN193189. On D1 and D2 of differentiation, the media was removed and fresh KSR with 10 μM SB431542 and 100 nM LDN193189 was added. Starting on D4 of differentiation an increasing amount of N2 media was added to the KSR media every two days, while maintaining 10 μM SB431542 and 100 nM LDN193189. On

D4 a 3:1 mixture of KSR/N2 media was added. On D6 a 1:1 mixture of KSR/N2 media was added and on D8, a 1:3 mixture of KSR/N2 media was added. The cells were isolated for flow cytometry analysis on D4, 6, 8 and 10 of differentiation and for immunostaining on D10 of differentiation. KSR media contains Knockout DMEM (Thermo Fisher Scientific, 10829018), Knockout Serum Replacement (Thermo Fisher Scientific, 10828028), 1X MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140050), 1X GlutaMAX (Thermo Fisher Scientific, 35050079), and 2-mercaptoethanol (Thermo Fisher Scientific, 21985023). N2 media contains DMEM/F12 medium (Thermo Fisher Scientific, 12500-062), glucose (Sigma, G8270), sodium bicarbonate (Sigma, S5761), putrescine (Sigma, P5780), progesterone (Sigma, P8783), sodium selenite (Sigma, S5261), apo-transferrin (Sigma, T1147), and insulin (Sigma, I2643).

Flow cytometry

Cells were dissociated using TrypLE Select and resuspended in FACS buffer (5% FBS, 5 mM EDTA in PBS). First, cells were stained with surface antibody (CXCR4-APC) with LIVE/DEAD violet dye (Molecular Probe, L34955, 1:1,000) for 30 minutes at 4^{o}C. After washing with FACS buffer, cells were fixed and permeabilized in 1X fixation/permeabilization buffer (eBioscience, 00-5523-00) for 30 minutes at RT. After fixation and permeabilization, cells were stained with intracellular antibody (SOX17-PE, GATA6-PE, GATA4-Alexa-647, NKX6.1, PDX1, NKX2.1) in permeabilization buffer (eBioscience, 00-5523-00) for 30

minutes at R.T. After washing with permeabilization buffer, cells were resuspended in permeabilization buffer with fluorescence conjugated secondary antibodies for 30 minutes at R.T. After staining, cells were washed, suspended in FACS buffer and analyzed using BD LSRFortessa or BD LSRII. Flow cytometry analysis and figures were generated in FlowJo V10. Detailed information of antibodies used is listed in Table 1.

RNA isolation and RT-qPCR

Cell pellets were lysed in TRIzol (Thermo Fisher Scientific, 15596018). RNA was extracted from TRIzol lysate using the RNeasy Mini Kit (Qiagen, 74106). cDNA was produced using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 43814). Quantitative real-time PCR was performed in triplicate using Absolute Blue QPCR SYBR Green Mix with low ROX (Thermo Fisher Scientific, AB4322B) on the ABI PRISM® 7500 Real Time PCR System (Applied Biosystems) using the following protocol: 15 minutes at 95°C followed by 40 cycles of 15 seconds at 95°C, 30 seconds at 58°C, and 30 seconds at 72°C. The signal was detected at 72°C. All primers used for RT-qPCR are listed in Table 2.

RNA-seq analysis

ESC and DE stage RNA-seq data from both WT and *MKK7* KO cells were analyzed by Tophat [166], featureCounts [167] and DESeq2 [168]. First, reads were mapped to the human genome (hg19) by Tophat (v.2.1.1) and read counts per gene were summarized by featureCounts from subread package. DESeq2 was

used to analyze gene differential expression by comparing transcriptomes of WT and *MKK7* KO cells from ESC and DE stages. MA plot was generated by ggplot2 [169]. ChIP-seq peaks of increased JUN binding in ESC relative to DE D1 (Fig.3.8j) are preferentially located closer to the TSS of ESC expressed genes farther from DE expressed genes, following the similar analysis in Fig. 5 supplement 1 of reference [170].

Single Cell RNA-seq (Drop-seq) and analysis

DE cells were washed with PBS once and treated with TrypLE for 5 minutes. E8 media was added to stop the digestion. Dissociated single cells were resuspended in PBS with 0.01% BSA, passed through a 40 μm filter and counted. Cell concentration was adjusted to 100-120 cells/μL with >90% viability (in 1 mL volume). Cell suspension were kept on and transferred to the Weill Cornell Genomics and Epigenomics Core Facility for drop-seq. Drop-seq experiment procedure follows the protocol published by the McCarroll lab (version 3.1, 12/18/2015) [141] http://mccarrolllab.com/download/905/. The sequenced reads were processed and turned into a digital expression matrix using the Drop-Seq computational pipeline (v.1.2) of the McCarroll lab (http://mccarrolllab.com/dropseq/). The reference genome was GRCh38, the gene annotation was from Ensembl release 76. The resulting count matrix was filtered to remove lowly covered cells and rarely captured transcripts, normalized for differences in sequencing depth, and de-noised using the MAGIC algorithm [171]. The scatter plots and heatmaps are based on $\log_{10}$-transformed expression

values. Heatmaps were generated with the R package pheatmap (https://CRAN.R-project.org/package=pheatmap) including default settings for hierarchical clustering of the columns (representing cells). The values that are shown are z-scores after row-based scaling (scale = “row”). A detailed version of data processing is available for downloading.

Immunofluorescence Staining

Cells were fixed in 4% paraformaldehyde (Thermo Fisher Scientific, 50980495) for 10 minutes at room temperature (R.T.). After washing with PBST (PBS with 0.2% Triton X-100) three times for 5 minutes each, cells were blocked in 5% donkey serum in PBST buffer for 5 minutes. Cells were incubated with primary antibodies for 1 hour at R.T. After washing with PBST three times for 5 minutes each, cells were incubated with fluorescence conjugated secondary antibodies and 0.2 µg/mL DAPI (Sigma, 32670-5mg-F) for 1 hour at R.T. After washing with PBST three times for 5 minutes each, images were taken with equal exposure for the same field. All primary and secondary antibodies used are listed in Table 1.

Western Blot

Cell pellets were snap frozen in liquid nitrogen and lysed in cell lysis buffer (Cell Signaling Technology, 9803) with proteinase/phosphatase inhibitors (Cell Signaling Technology, 5872) and 1mm PMSF (MP Biomedicals, ICN19538105). Proteins were pre-cleared by centrifugation at 14,000g 4^{o}C for 10 minutes. Protein concentration was determined by the Bradford Protein Assay (Bio-Rad,

500-0202). Equal amounts of protein were loaded into Bis-Tris 10% gel (Novex, NP0301BOX) and transferred to nitrocellulose membranes (Novex, LC2001). Membranes were blocked with 5% milk (for non-phosphorylated proteins) or 5% BSA (for phosphorylated proteins). Primary antibody was incubated overnight at 4^{o}C. Membranes were washed with TBST 3 times for 10 minutes each and incubated with secondary antibody for 1 hour at R.T. Membranes were washed with TBST 3 times for 10 minutes each. ECL western blotting detection reagents (Amersham, RPN2236 and Thermo Fisher Scientific, 32106) were used to visualize the protein bands. All antibodies and dilution factors are listed in Table 1.

ChIP-seq

ChIP was performed as previously described [62]. Cells were cross-linked with 1% formaldehyde (Sigma, F1635) at 37°C for 15 minutes and quenched with 0.125 M glycine for 5 minutes at room temperature. Fixed cells were scraped off the plates in cold PBS buffer and washed twice in cold PBS buffer. Cell pellets were obtained by centrifugation at 3000 RPM for 5 minutes at 4 ^{o}C and frozen in liquid nitrogen immediately before transferring to the -80C freezer. ~25 million cells were used for one ChIP reaction. Cell pellets were thawed on ice, resuspended in 1 mL SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8) containing proteinase and phosphatase inhibitor for one reaction in an Eppendorf tube and incubated on ice for 10 minutes. Sonication was performed on a Branson Sonifier 250 with a 20% amplitude setting for 5.5 minutes (10 seconds

on/off pulsing). Sonication products were spun down at 14,000 RPM at 4°C for 10 minutes and 1 mL supernatant containing chromatin and DNA were transferred to a falcon tube containing 9 mL ChIP dilution buffer (0.01% SDS, 1.1% TritonX-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8, 167 mM NaCl) with proteinase and phosphatase inhibitor. 50 µL of Dynabeads (Life Technologies, 10009D) were added to samples and incubated at 4°C with rotation for 1 hour. After pre-clearing, Dynabeads beads were removed and 200 µL of sample were collected as 2% input separately. 5 µg antibodies were added to the pre-cleared samples for overnight incubation at 4°C with rotation. 200 µL Dynabeads were added into one ChIP reaction and incubated for 4-6 hours at 4 °C with rotation. Dynabeads were collected by centrifugation with 3,000 RPM at 4 °C for 5 minutes and washed in 1 mL low salt buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl) for 5 minutes at 4°C with rotations. Then beads were washed in 1 mL high salt buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 500 mM NaCl) twice and TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA) twice for 5 minutes at 4°C with rotation. After the last wash, beads were resuspended in 250 µL elution buffer (1% SDS, 0.1 M NaHCO3) and incubated in a thermomixer: 850 RPM for 15 minutes at 60 °C. Supernatant were collected and added with 5 M NaCl for overnight decrosslinking at 65 °C. 10 µL 0.5 M EDTA, 20 µL 1 M Tris-HCl pH6.5 and 1 µL proteinase K (20 mg/mL) were added to decrosslinked product and incubated for 1 hour at 45 °C. DNA were isolated by using QIAquick PCR purification kit (Qiagen, 28104). Libraries were prepared using the NEBNext® ChIP-seq Library

Prep Master Mix Set for Illumina® (NEB, E6240L) and quality controlled using Agilent Technologies 2200 TapeStation to determine fragment size and PicoGreen (Life Technologies, P7589) to quantify the concentration. Samples were pooled and submitted to New York Genome Center for SE50 sequencing using Illumina HiSeq 2500.

ATAC-seq

50,000 cells were pelleted by centrifugation at 1500 RPM for 5 minutes at 4 ^{o}C in 1 mL cold PBS buffer. Cell pellets were washed with 1 mL ice-cold ATAC buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2) and lysed in 50 µl of ATAC lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40 or IGEPAL-Ca630) for 2 minutes on ice. Lysis was terminated by adding 1 ml of cold ATAC buffer to the mixture. Nuclear pellets were extracted by centrifugation at 1500 RPM for 10 minutes at 4 ^{o}C and aspiration of supernatant to leave behind 22.5 µL sample liquid. 2.5 µL Tagmentation Enzyme (transposase) and 25 µL of Tagmentation Buffer (Illumina Nextera DNA Sample Preparation Kit) were added to the resuspended nuclear pellets. Reaction mixtures were transferred to PCR tubes and incubated for 30 minutes at 37^{o}C. Next, 0.2% SDS was added to the reaction mixtures and incubated for 5 minutes at room temperature. 2X Agencourt AMPure XP beads (Beckman Coulter A63881) were used for sample purification. 50 µL of purified sample were mixed with 55 µL of NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB, catalogue M0543L) and 5 µL Nextera primers[172] (the universal primer Ad1 mixed with the unique index primer 2.X at a

final concentration of 25 μM). Amplification was performed as the following PCR program for 12 times: 65 ºC, 5 minutes; 98 ºC, 30 seconds; 98 ºC, 10 seconds, 65 ºC 30 seconds. 1.5X AMPure XP beads were used for sample purification. The concentration and median fragment size were quality controlled by PicoGreen and Agilent D1000 screentape on the Agilent Technologies 2200 TapeStation respectively. Samples were sequenced PE50 using Illumina HiSeq 2500.

ATAC-seq and ChIP-seq analysis

We mapped reads to hg19 using bowtie2 [173] (paired reads for ATAC, local for ChIP) and ran MACS2 [174] to call peaks, and defined 300 bp bound regions by extending 150bp from each summit. Replicated ChIP-seq peaks were defined using MACS2 default parameters (default MACS2 FDR q-value parameter q=0.05) and were filtered against hg19 ENCODE blacklisted regions. To demonstrate that our conclusions did not depend on the large number of MACS2 OCT4 default peaks (n=45320), we also defined a very stringent set of the strongest OCT4 peaks using q=1e-6 (n=11743) for OCT4 only. In Fig. 3.7d and Fig. 3.8c, the OCT4+JUN+ sets were defined by the intersection of the OCT4 and JUN replicated peaks (n=8198 for Fig. 3.8c, n=3800 for Fig. 3.7d), and the OCT4+JUN- sets were defined to be equal size sets of OCT4+ peaks with the lowest JUN ChIP-seq signal (n=8198 or n=3800). Both sets give consistent results.

When comparing signals across experiments we normalized ATAC and ChIP-seq signals so columns in Fig. 3.7b, 3.9a and 3.9e had constant average signal across all bound regions shown, to correct for read depth and other technical variation. We evaluated the average signal (reads/bp) within peak regions in all experiments using bigWigAverageOverBed [175]. We defined fixed size sets of differentially bound (or accessible) regions (N=5,000 for ATAC (Fig. 3.7a-b), N=2500 for SMAD2/3 DE WT vs JNKi (Fig. 3.9d-e) and SMAD2/3 ESC vs DE (Fig.3.8a-b) by their distance from y=x or the regression line for imbalanced read depth.

To show the direct effect of JUN binding on changes in the enhancer landscape during the ESC to DE transition, in Fig. 3.9a we defined an OCT4+ ESC enhancer set using the stringent set of OCT4 bound replicated non-blacklisted peaks defined above (using q=1e-6, n=11743) further filtered to remove promoters (<2kb from TSS) (n=9248). We defined an equal size set of GATA6 bound distal DE enhancers consisting of the strongest GATA6 signal replicated MACS2 peaks filtered to remove promoters and blacklisted regions (n=9248). We then split the ESC enhancers into equal JUN high and JUN low sets based on their JUN ESC ChIP-seq signal averaged from both replicates. Similarly, to show the direct effect of JNK inhibition on changes in chromatin accessibility and SMAD2/3 binding at DED1, we defined an OCT4+SMAD2/3+ ESC enhancer set from the overlap of the stringent q=1e-6 OCT4+ ESC enhancers defined above with default q=0.05 MACS2 SMAD2/3 peaks (n=682), a GATA6+SMAD2/3+ DE

enhancer set from the overlap of the strongest GATA6+ DE enhancers with SMAD2/3 default DE peaks (n=8500) (Fig. 3.8d); a SMAD2/3 bound ESC set from replicated default peaks (n=5511) (Fig. 3.8d), and a SMAD2/3 bound DE set from the strongest SMAD2/3 DE peaks (n=8500) (Fig. 3.8d), and calculated p-values by Welch two-sample t-test.

Significance of Co-binding

We can estimate the p-value for the significance of the co-binding of OCT4 and JUN at ESC by comparing to the null hypothesis that the two factors bind randomly among n=105269 ESC accessible genomic loci (MACS2 peaks on ESC ATAC). For default peak calling (Fig. 3.8c, nOCT4=45230, nJUN=13474) the observed p-value is p<3.89e-224 for the overlap of OCT4 and JUN. For more stringent (q=1e-6) peak calling for OCT4 (Fig. 3.7c, nOCT4=11743) the observed p-values are p<1e-300 for the overlap of OCT4 and JUN.

We can calculate the significance of the increase in overlap of GATA6 and SMAD2/3 binding upon JNK inhibition shown in Fig. 3.8i with the following model. Since there is a large overlap of GATA and SMAD2/3 binding compared to the available genome, we can treat the overlap by modeling that GATA and SMAD2/3 are detected as binding independently at a small number of loci, N, and estimated this number of loci to make the overlap $N_{GS}=Np_Sp_G$ match the observed N_G, N_S, and overlap, N_{GS}. The observed overlap N_{GS}=9,567 between GATA6 and SMAD2/3 in DE D1 WT in Fig. 3.9c is consistent with sampling from

a population size of N=49,475 genomic loci with $p_G = N_G/N = 29,436/N$ and $p_S = N_S/N = 16,080/N$. Then with JNKi treatment, the number of total GATA and SMAD2/3 binding sites increases, but the overlap also increases. We needed to calculate the significance of the new overlap given that N_G and N_S both increase slightly. Under the hypothesis that p_S and p_G are constant and the binding is still independent at a new number of loci N', new denominator is N'=66,773 loci yielding expected binding N_G=39,728 and N_S=21,702, which is very close to the observed JNKi binding 39,986 and 21,5561 (by construction). This is a way of correcting for sequencing depth, technical variation, etc. If the binding were independent in these loci as before, the expected overlap would be N_{GS}=12,912±102. We use that variance to estimate the significance of the larger than expected observed overlap of N_{GS}=14,611, yielding p<2e-62.

Sequence Determinants of Regulatory Reconfiguration

We used complementary approaches to identify sequence features mediating changes in accessibility and binding. We used two discriminative motif identification algorithms, HOMER [144] and MotifSpec [145] trained on WT vs JNKi increased vs decreased accessibility and SMAD2/3 binding and vice versa. We also used a gapped kmer based approach, gkm-SVM [146], trained on chromatin accessible regions as described [176], which can describe how the full set of kmers can in combination describe the changes in binding and accessibility. We reported in Fig. 3.7a, Fig. 3.9d and Fig.3.8a the highest AUC single motifs found by either HOMER or MotifSpec and the hypergeometric p-value associated with

its differential enrichment. These motifs match the significant kmers detected by gkm-SVM. That the CV test-set AUROC from gkm-SVM is high (0.923, 0.962, 0.991), and the large positive and negative weight kmers match a relatively small set of motifs (AP1, TEAD, GATA, SMAD) suggests that the differential binding and activity of these factors can directly account for the majority of the changes in the chromatin accessibility and binding landscape. In addition to test-set cross-validation, the gapped kmer vocabulary detected by gkm-SVM has been shown to be predictive of chromatin accessibility and massively parallel expression reporter assays in blind experimental comparisons [177,178]. While gkm-SVM trained on SMAD2/3 ESC vs DE detected SMAD2/3 and GATA as the top kmer weights in DE and OCT4 kmers in ESCs, standard motif enrichment methods identified OCT4 and SMAD2/3 motifs, but they were not the most highly scoring motifs (which were poly-A and GC rich).

Bifurcation Model

We modeled the transition from ESC-DE as a minimal two gene auto-regulatory system with negative feedback between antagonistic representative ESC and DE TFs. This model is derived based on two interacting genes, their enhancers, and their protein products, which we will call O=OCT4 and G=GATA6. However, the protein concentrations could represent a set of ESC TFs and DE TFs, and the interactions described as direct interactions here could be direct or indirect. Many of these interactions are supported by our ChIP-seq data and prior results. We denote protein concentrations with uppercase O=[O] and G=[G], and genes and

enhancers with lower case, e.g., [oO] is the concentration of O bound to the *OCT4* gene enhancer, o, and [oG] is the concentration of G bound to the *OCT4* gene enhancer, etc. Each protein (O,G) is degraded at a fixed rate, r, and is produced by a combined transcription and translation rate proportional to the occupancy of the gene's enhancer (lower case) by its own TF product t_o[oO] and t_g[gG] (auto-activation). As transcription changes protein concentrations, the enhancer's occupancy changes as determined by fast Michelis-Menten binding kinetics (relative to transcription and translation), so in equilibrium the concentration of transcriptionally active occupied enhancer [oO]=[O]/(k_{oO} +[O]+(k_{oO}/k_{oG})[G]). k_{oO} and k_{oG} are the dissociation constants for O and G at gene o's enhancer, respectively, so weaker binding means larger k. The last term in the denominator reflects the fact that when gene o's promoter is occupied by G ([oG]) it is transcriptionally inactive. Similarly, the active occupied enhancer concentration, [gG], is given by [gG] = [G]/(kgG +[G]+(kgG/kgO)[O]). This bifurcation model is similar to previous models[179,180] and our simplifying approximations are consistent with and supported by the results of more complete simulations[181]. We chose baseline parameters r=1, t_o=3, t_g=3, k_{oO}=1, k_{gG}=1, k_{oG}=0.25, and k_{gO}=0.25 which yield stable fixed points at high O and low G (ESC-like) or high G and low O (DE-like): (O=t_o/r-k_{oO},G=0) and (O=0,G=t_g/r-k_{gG}). To simulate ESC-DE transition within cell populations, we start all simulations in the high O state, add stochastic noise, and modeled Activin A as a Gaussian impulse at t=15 with width=3: Activin = c_A*exp(-(t-15)^2/3^2). To model the effects of JNK inhibition we performed additional simulations reducing either the

transcriptional activation of O by 15% (t_o=2.55) or by reducing the binding strength of O to its own enhancer by 15%, k_{oO}=1.15. We performed N=100 simulations at each Activin A concentration c_A=0.0 to 0.5 and report the fraction of simulations which transition to the high G state in Fig. 3.10c. We in turn repeated this 20 times for each concentration to determine the standard deviation of the transition rate. These simulations show that weakening the auto-regulation of ESC-like genes either by reducing binding or transcriptional activation are both consistent with JNK inhibition increasing the rate of transition from ESC-DE.

Statistical Analysis

Quantification of flow cytometry and qPCR data are shown as mean ± standard deviation (SD) unless otherwise indicated. Unpaired two-tailed Student's t-test was used for comparison between two groups. p-values of ≥0.05 were considered not significant and p-values <0.05 (*), <0.01 (**), <0.001 (***), <0.0001 (****) were considered significant.

Table 1. Antibody information

Protein	Application	Species	Vendor	Cat. Number	Dilution
MKK7	WB	rabbit	Cell Signaling Technology	4172S	1:1,000
C-JUN	WB	rabbit	Cell Signaling Technology	9165S	1:1,000
P-JUN S63	WB	rabbit	Cell Signaling Technology	9261S	1:500
P-JUN S73	WB	rabbit	Cell Signaling Technology	3270S	1:1,000
P-SMAD2 (s465/467)	WB	rabbit	Cell Signaling Technology	3108S	1:500
P-SMAD2 (S245/250/255)	WB	rabbit	Cell Signaling Technology	3104	1:1000
SMAD2	WB	rabbit	Cell Signaling Technology	5339S	1:1,000
JNK1	WB	mouse	Cell Signaling Technology	3708S	1:1,000
P-JNK	WB	mouse	Cell Signaling Technology	9255S	1:500
GAPDH	WB	rabbit	Cell Signaling Technology	5174S	1:25,000
anti-rabbit HRP	WB		Cell Signaling Technology	7074s	1:5,000
anti-mouse HRP	WB		Cell Signaling Technology	7076s	1:5,000
SOX17	Flow	mouse	BD BIOSCIENCES	561591	1:50
CXCR4	Flow	mouse	R&D Systems	FAB170A	1:25
PAX6	Flow/IF	rabbit	Covance	PRB278P100	1:50
GATA4	Flow	mouse	BD Biosciences	560400	1:50
GATA6	Flow	rabbit	Cell Signaling Technology	26452s	1:50
Phospho Histone H3	Flow	rabbit	Cell Signaling Technology	3465S	1:50
Cleaved Caspase 3	Flow	rabbit	Cell Signaling Technology	9602s	1:50
SOX17	IF	goat	R&D Systems	AF1924	1:500
FOXA2	IF	rabbit	Millipore	7633	1:500
OCT4	IF	goat	Santa Cruz	sc8628	1:500
NKX6.1	Flow/IF	mouse	DSHB	FGGA12-C	1:250
NKX2.1	Flow/IF	rabbit	Abcam	Ab76013	1:250
PDX1	Flow/IF	goat	R&D Systems	AF2419	1:250
OCT4	ChIP	rabbit	Cell Signaling Technology	5677	5 μg
NANOG	IF	rabbit	Cosmobio Japan	RECRCAB0004PF	1:500
NANOG	ChIP	rabbit	Cell Signaling Technology	5232	5 μg
GFP	Flow/IF	rat	Biolegend	338008	1:100
SMAD2/3	ChIP	rabbit	Cell Signaling Technology	8685s	5 μg
JUN	ChIP	rabbit	Cell Signaling Technology	9165S	5 μg
GATA6	ChIP	rabbit	Cell Signaling Technology	5851S	5 μg
Alexa Fluor 488	IF	goat	Thermo Fisher Scientific	A11055	1:500
Alexa Fluor 488	IF	rat	Thermo Fisher Scientific	A21208	1:500
Alexa Fluor 488	IF/Flow	mouse	Thermo Fisher Scientific	A21202	1:500
Alexa Fluor 594	IF/Flow	rabbit	Thermo Fisher Scientific	A21207	1:500
Alexa Fluor 594	IF	goat	Thermo Fisher Scientific	A11058	1:500
Alexa Fluor 647	IF	rabbit	Thermo Fisher Scientific	A31573	1:500
Alexa Fluor 647	IF/Flow	goat	Thermo Fisher Scientific	A21447	1:500

WB: western blotting; IF: immunofluorescence staining; Flow: flow cytometry.

Table 2. qPCR primers

Gene	Forward Primer (5' to 3')	Reverse primer (5' to 3')
OCT4	AGTGAGAGGCAACCTGGAGA	ACACTCGGACCACATCCTTC
NANOG	CATGAGTGTGGATCCAGCTTG	CCTGAATAAGCAGATCCATGG
SOX2	TGGACAGTTACGCGCACAT	CGAGTAGGACATGCTGTAGGT
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
EOMES	CAACATAAACGGACTCAATCCCA	ACCACCTCTACGAACACATTGT
GATA6	CCCACAACACAACCTACAGC	GCGAGACTGACGCCTATGTA
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
NEUROD1	GGATGACGATCAAAAGCCCAA	GCGTCTTAGAATAGCAAGGCA
MNX1	TCCACCGCGGGCATGATCCT	GCGCTTGGGCCGCGACAGGTA
NGN3	CTATTCTTTTGCGCCGGTAG	ACTTCGTCTTCCGAGGCTCT
NKX2.1	CGGCATGAACATGAGCGGCAT	GCCGACAGGTACTTCTGTTGCTTG
FOXP2	ACAGCCTGCTGTTGTTGGAGAAG	CGGCATGAACATGAGCGGCAT
CPM	TCCAAGGTGGAATGCAAGAT	TCAAAAACTTGACCCTTTACACC
GAPDH	GGAGCCAAACGGGTCATCATCTC	GAGGGGCCATCCACAGTCTTCT

Table 3. Genotyping primer information

Gene	Forward PCR Primer (5' to 3')	Reverse PCR Primer (5' to 3')	Sequencing Primer (5' to 3')
MKK7	AGCCTCCTCCATCTCTTTCC	ATGAGGATGCGCTTGTTCTC	GATGAGACAAGGGACCCTGA
JUN	ACTTTTCAAAGCCGGGTAGC	CACTGTCTGAGGCTCCTCCT	TGACTGCAAAGATGGAAACG

Table 4. gRNA targeting sequence

Gene	gRNA targeting sequence
ACVR1B	AGCCTGAGCACCCGTCCATG
ACVR1B	ACACCTGCACATGGAGATCG
ARID1A	TTCAATAGATGACCTCCCCA
ARID1A	TGAGCGAGACTGAGCAACAC
ARID4B	TAGACATGATTCTTCAACAG
ARID4B	AGTTCAGGATGACCACATAA
ARPC4	TACAACCTGTGACCATCAGC
ARPC4	GTTTCACAGCAATGCTGACC
BAG6	CATCTTGCAGAACTCGTCCC
BAG6	GCAAGATGATAAGAAGCTTC
BANP	CAACCTCCAGATCCATCACG
BANP	GACCGTCCTGCCCCACGTGA
CCDC6	GATTGACCTTGAAAATACAT
CCDC6	CTTTCTCATAATTTACAGCA
CSK	TCTCCTCCACGATCACGCCC
CSK	CGGCTGGGCCCTGAACATGA
CTNNB1	GAAACAGCTCGTTGTACCGC
CTNNB1	AGAACGCATGATAGCGTGTC
DPH6	CTAAGACCAGCTGAAAACCA
DPH6	TCAGGAAAAAGAAGAAGTAG
EOMES	GGACACTCACATCGGTGTTT
EOMES	CCACTACAATGTGTTCGTAG
FOXH1	TGGCAGAACGGAGGTGCGCG
FOXH1	TGGCCGGCCGTGCAGCACGT
JUN	CAAGCTGGCGTCGCCCGAGC
JUN	GATTATCAGGCGCTCCAGCT
KDM1A	TATAAGGTGCTTCTAATTGT
KDM1A	TGTGGTCCACTGATAATATC
L3MBTL3	ATGGTACCAACTGCTCAAGA
L3MBTL3	TGTGAGAACTGTTGTCAGTA
MAPK1	GCCTACAGACCAAATATCAA
MAPK1	GCAGTAGGTCTGGTGCTCAA

Gene	gRNA targeting sequence
JNK1	CTCATAAAGTTACATAGTCA
JNK1	AGAATCAGACTCATGCCAAG
MED12	ACAGGTCATCTTAATGAGCC
MED12	CTCAGAGATTGCTGCATAGT
MEKK1	AAGGTTCTAGTTGAAACACC
MEKK1	AACCTTCAGACCCAATGTTA
MIXL1	AGCTCGTCTTCCGCCGGACC
MIXL1	GCGCCGCGTTTCCAGCGTAC
MKK7	AGGTGGTACCTGGCCCCCGA
MKK7	CCTGTTCACACCCCGCAGCA
MLL2	TTCCCCTGCGCGACTGCCAG
MLL2	CTCACCATTGGTGTGCTGCA
NF2	TGAGCCTACCTTGGCCTGGA
NF2	ATTCCACGGGAAGGAGATCT
Non-targeting	ACGGAGGCTAAGCGTCGCAA
Non-targeting	CGCTTCCGCGGCCCGTTCAA
NUP188	CGATATCCATGCCCTCCACC
NUP188	CTTAAACCCAGTTCCTTCAA
PPP2R4	ACCCTGGTTCACTTTGGACC
PPP2R4	TTCATCCTTACCCTCAACGA
PTPN14	CTAGCCGGCCTAGCTGTGCA
PTPN14	GAAATAGCACATACTCTCTG
Puromycin resistence gene	CACGCGCCACACCGTCGATC
SMAD2	ATGTTATATATTGCCGATTA
SMAD2	CTCCAGGTATCCCATCGAAA
SMAD4	CAGGTGCCTTAGTGACCACG
SMAD4	GGATTAACACTGCAGAGTAA
SMARCC1	TGCTCCTACCAATAAAACAC
SMARCC1	TGGGAAGCATGTCACCAACC
TAOK1	TCTGCTTCGGATTTACTAGA
TAOK1	ATTTACGTGAACACACAGCA
TGFBR1	CCACGAACGTTCTTCTCTAG
TGFBR1	CCATCGAGTGCCAAATGAAG

CONCLUDING REMARKS

Here we conducted genome-scale CRISPR/Cas screens for high-throughput discovery of regulators of DE differentiation. In addition to known regulators of DE differentiation, we identified novel genes including five essential JNK/JUN pathway genes that inhibit DE differentiation. The JNK/JUN pathway is not required for the maintenance of the pluripotent state. Instead, JNK inhibition specifically accelerates the decommissioning of ESC enhancers during DE differentiation, and promotes the reconfiguration of SMAD2/3 binding to DE enhancers co-bound by GATA6. Thus, the JNK/JUN pathway constitutes a key barrier from pluripotency to DE differentiation. Our findings demonstrate the power of large-scale forward genetic screens for uncovering genes that regulate hESC/hiPSC differentiation and human development.

Our work also demonstrates the power of using unbiased CRISPR screen to identify the barriers of hESC differentiation. Targeting negative regulators of DE differentiation represents a rational strategy to improve broad endoderm lineage differentiation. JNK inhibitor treatment not only improved the differentiation efficiency of DE, but also improved DE-derived pancreatic and lung lineage cells. Other groups have successfully improved pancreatic and lung differentiation by focusing on differentiation at later stages[165,182]. For instance, the Kotton[165] and Snoeck[183] groups have previously reported ~10-40% NKX2.1+ cells at Day 13-15 of differentiation, and optimization at later stages improves the efficiency to ~70% at Day 15 based on NKX2.1 and FOXA2 immunostaining[184].

The use of JNKi during DE differentiation improves NKX2.1+ cells from ~20% to ~70% by flow cytometry after only 9 days of differentiation, demonstrating the importance of improving DE differentiation for generating DE-derived cell types.

Our screening strategy can be extended to systematically interrogate sequential lineage decisions leading to the formation of mature cell types such as pancreatic β cells. Such screens may also benefit from inducibly expressing Cas9[105] or dCas9-KRAB[185] during differentiation to interrogate the temporal control of human development. Knowledge gained from unbiased screening efforts such as ours can contribute to the understanding of human development and improve hESC/hiPSC directed differentiation for disease modeling and regenerative medicine.

REFERENCES

1. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154 (1981).
2. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7634-7638 (1981).
3. Thomson, J.A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**, 1145 (1998).
4. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews Molecular Cell Biology* **17**, 183 (2016).
5. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663-676.
6. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150 (2004).
7. Smith, A.G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690 (1988).
8. Williams, R.L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-687 (1988).
9. Ying, Q.-L., Nichols, J., Chambers, I. & Smith, A. BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. *Cell* **115**, 281-292 (2003).
10. Amit, M., Shariki, C., Margulets, V. & Itskovitz-Eldor, J. Feeder Layer- and Serum-Free Culture of Human Embryonic Stem Cells1. *Biology of Reproduction* **70**, 837-845 (2004).
11. Ginis, I. *et al.* Differences between human and mouse embryonic stem cells. *Developmental Biology* **269**, 360-380 (2004).
12. Beattie, G.M. *et al.* Activin A Maintains Pluripotency of Human Embryonic Stem Cells in the Absence of Feeder Layers. *STEM CELLS* **23**, 489-495 (2005).
13. James, D., Levine, A.J., Besser, D. & Hemmati-Brivanlou, A. TGF β /activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**, 1273 (2005).
14. Vallier, L., Alexander, M. & Pedersen, R.A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *Journal of Cell Science* **118**, 4495 (2005).
15. Xu, C. *et al.* Basic Fibroblast Growth Factor Supports Undifferentiated Human Embryonic Stem Cell Growth Without Conditioned Medium. *STEM CELLS* **23**, 315-323 (2005).
16. Xu, R.-H. *et al.* Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature Methods* **2**, 185 (2005).

17. Brons, I.G.M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191 (2007).
18. Tesar, P.J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196 (2007).
19. Davidson, K.C., Mason, E.A. & Pera, M.F. The pluripotent state in mouse and human. *Development* **142**, 3090 (2015).
20. Han, G., Wang, H. & Hao, J. Molecular Mechanisms of Embryonic Stem Cell Pluripotency. (eds. Bhartiya, D. & Lenka, N.) (IntechOpen, Rijeka, 2013).
21. Anderson, K.V. & Ingham, P.W. The transformation of the model organism: a decade of developmental genetics. *Nature Genetics* **33**, 285 (2003).
22. Tam, P.P.L. & Loebel, D.A.F. Gene function in mouse embryogenesis: get set for gastrulation. *Nature Reviews Genetics* **8**, 368 (2007).
23. Shen, M.M. Nodal signaling: developmental roles and regulation. *Development* **134**, 1023 (2007).
24. van Amerongen, R. & Berns, A. Knockout mouse models to study Wnt signal transduction. *Trends in Genetics* **22**, 678-689 (2006).
25. Tam, P.P.L. & Behringer, R.R. Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of Development* **68**, 3-25 (1997).
26. Robertson, E.J. Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Seminars in Cell & Developmental Biology* **32**, 73-79 (2014).
27. Wang, R.N. *et al.* Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes & Diseases* **1**, 87-105 (2014).
28. Camus, A., Perea-Gomez, A., Moreau, A. & Collignon, J. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Developmental Biology* **295**, 743-755 (2006).
29. Loebel, D.A.F., Watson, C.M., De Young, R.A. & Tam, P.P.L. Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Developmental Biology* **264**, 1-14 (2003).
30. Murry, C.E. & Keller, G. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell* **132**, 661-680 (2008).
31. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes & Development* **19**, 1129-1155 (2005).
32. Zhu, Z. & Huangfu, D. Human pluripotent stem cells: an emerging model in developmental biology. *Development* **140**, 705-717 (2013).
33. D'Amour, K.A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotech* **23**, 1534-1541 (2005).
34. Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nature Biotechnology* **26**, 443 (2008).
35. Chambers, S.M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotech* **27**, 275-280 (2009).

36. Tabar, V. & Studer, L. Pluripotent stem cells in regenerative medicine: challenges and recent progress. *Nature Reviews Genetics* **15**, 82 (2014).
37. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat Biotech* **32**, 1121-1133 (2014).
38. Pagliuca, Felicia W. *et al.* Generation of Functional Human Pancreatic β Cells In Vitro. *Cell* **159**, 428-439 (2014).
39. Allen, H.L. *et al.* GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nature Genetics* **44**, 20 (2011).
40. Shi, Z.-D. *et al.* Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development. *Cell Stem Cell* **20**, 675-688.e6 (2017).
41. Fisher, J.B., Pulakanti, K., Rao, S. & Duncan, S.A. GATA6 is essential for endoderm formation from human pluripotent stem cells. *Biology Open* **6**, 1084 (2017).
42. Tiyaboonchai, A. *et al.* GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic β Cells. *Stem Cell Reports* **8**, 589-604 (2017).
43. Chia, C.Y. *et al.* GATA6 Cooperates with EOMES/SMAD2/3 to Deploy the Gene Regulatory Network Governing Human Definitive Endoderm and Pancreas Formation. *Stem Cell Reports* **12**, 57-70 (2019).
44. Thomas F Allison, N.S.P.-G., Veronica Biga, Peter W Andrews,Ivana Barbaric. Human pluripotent stem cells as tools for high-throughput and high-content screening in drug discovery. *International Journal of High Throughput Screening* **5**, 1-13 (2014).
45. Heldin, C.-H. & Moustakas, A. Signaling Receptors for TGF-β Family Members. *Cold Spring Harbor Perspectives in Biology* **8**(2016).
46. Shi, Y. & Massagué, J. Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. *Cell* **113**, 685-700 (2003).
47. Gaarenstroom, T. & Hill, C.S. TGF-β signaling to chromatin: How Smads regulate transcription during self-renewal and differentiation. *Seminars in Cell & Developmental Biology* **32**, 107-118 (2014).
48. MacDonald, B.T., Tamai, K. & He, X. Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell* **17**, 9-26 (2009).
49. Loh, Kyle M. *et al.* Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. *Cell Stem Cell* **14**, 237-252 (2014).
50. Gray, P.C., Harrison, C.A. & Vale, W. Cripto forms a complex with activin and type II activin receptors and can block activin signaling. *Proceedings of the National Academy of Sciences* **100**, 5193 (2003).
51. Tada, S. *et al.* Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. *Development* **132**, 4363 (2005).

52. Zhu, Z. *et al.* Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes. *Cell Stem Cell* **18**, 755-768 (2016).
53. Conlon, F.L., Barth, K.S. & Robertson, E.J. 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 (1991).
54. Conlon, F.L. *et al.* A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919 (1994).
55. Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L.M. & Kuehn, M.R. Nodal is a novel TGF-β-like gene expressed in the mouse node during gastrulation. *Nature* **361**, 543-547 (1993).
56. Gu, Z. *et al.* The type I activin receptor ActRIB is required for egg cylinder organization and gastrulation in the mouse. *Genes & Development* **12**, 844-857 (1998).
57. Nomura, M. & Li, E. Smad2 role in mesoderm formation, left-right patterning and craniofacial development. *Nature* **393**, 786-790 (1998).
58. Waldrip, W.R., Bikoff, E.K., Hoodless, P.A., Wrana, J.L. & Robertson, E.J. Smad2 Signaling in Extraembryonic Tissues Determines Anterior-Posterior Polarity of the Early Mouse Embryo. *Cell* **92**, 797-808 (1998).
59. Sirard, C. *et al.* The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes & Development* **12**, 107-119 (1998).
60. Yang, X., Li, C., Xu, X. & Deng, C. The tumor suppressor SMAD4/DPC4 is essential for epiblast proliferation and mesoderm induction in mice. *Proceedings of the National Academy of Sciences* **95**, 3667 (1998).
61. Haegel, H. *et al.* Lack of beta-catenin affects mouse development at gastrulation. *Development* **121**, 3529 (1995).
62. Wang, Q. *et al.* The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* **20**, 70-86 (2017).
63. Arnold, S.J., Hofmann, U.K., Bikoff, E.K. & Robertson, E.J. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501 (2008).
64. Teo, A.K.K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes & Development* **25**, 238-250 (2011).
65. Showell, C., Binder, O. & Conlon, F.L. T-box genes in early embryogenesis. *Developmental Dynamics* **229**, 201-218 (2004).
66. Papaioannou, V.E. The T-box gene family: emerging roles in development, stem cells and cancer. *Development* **141**, 3819 (2014).
67. Faial, T. *et al.* Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in

differentiating human embryonic stem cells. *Development* **142**, 2121 (2015).
68. Zorn, A.M. & Wells, J.M. Vertebrate Endoderm Development and Organ Formation. *Annual Review of Cell and Developmental Biology* **25**, 221-251 (2009).
69. Brennan, J. *et al.* Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965 (2001).
70. Datto, M.B. *et al.* Targeted Disruption of Smad3 Reveals an Essential Role in Transforming Growth Factor β-Mediated Signal Transduction. *Molecular and Cellular Biology* **19**, 2495 (1999).
71. Zhu, Y., Richardson, J.A., Parada, L.F. & Graff, J.M. Smad3 Mutant Mice Develop Metastatic Colorectal Cancer. *Cell* **94**, 703-714 (1998).
72. Yang, X. *et al.* Targeted disruption of SMAD3 results in impaired mucosal immunity and diminished T cell responsiveness to TGF‐β. *The EMBO Journal* **18**, 1280 (1999).
73. Dunn, N.R., Vincent, S.D., Oxburgh, L., Robertson, E.J. & Bikoff, E.K. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development* **131**, 1717 (2004).
74. Chu, G.C., Dunn, N.R., Anderson, D.C., Oxburgh, L. & Robertson, E.J. Differential requirements for Smad4 in TGFβ-dependent patterning of the early mouse embryo. *Development* **131**, 3501 (2004).
75. Xu, J., Gruber, P.J. & Chien, K.R. SMAD4 Is Essential for Human Cardiac Mesodermal Precursor Cell Formation. *STEM CELLS* **37**, 216-225 (2019).
76. Moresco, E.M.Y., Li, X. & Beutler, B. Going forward with genetics: recent technological advances and forward genetics in mice. *The American journal of pathology* **182**, 1462-1473 (2013).
77. Acevedo-Arozena, A. *et al.* ENU Mutagenesis, a Way Forward to Understand Gene Function. *Annual Review of Genomics and Human Genetics* **9**, 49-69 (2008).
78. Ranzani, M., Annunziato, S., Adams, D.J. & Montini, E. Cancer Gene Discovery: Exploiting Insertional Mutagenesis. *Molecular Cancer Research* **11**, 1141 (2013).
79. Huangfu, D. *et al.* Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature* **426**, 83 (2003).
80. Chambers, I. *et al.* Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* **113**, 643-655 (2003).
81. Mitsui, K. *et al.* The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113**, 631-642 (2003).
82. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).
83. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819 (2013).
84. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816 (2012).

85. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533 (2006).
86. Fazzio, T.G., Huff, J.T. & Panning, B. An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. *Cell* **134**, 162-174 (2008).
87. Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316-320 (2010).
88. Schaniel, C. *et al.* Smarcc1/Baf155 Couples Self-Renewal Gene Repression with Changes in Chromatin Structure in Mouse Embryonic Stem Cells. *STEM CELLS* **27**, 2979-2991 (2009).
89. Gonzales, Kevin Andrew U. *et al.* Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell* **162**, 564-579 (2015).
90. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84 (2014).
91. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343**, 80 (2014).
92. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096 (2015).
93. Marceau, C.D. *et al.* Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **535**, 159 (2016).
94. Shalem, O., Sanjana, N.E. & Zhang, F. High-throughput functional genomics using CRISPR–Cas9. *Nature Reviews Genetics* **16**, 299 (2015).
95. Boettcher, M. & McManus, Michael T. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular Cell* **58**, 575-585 (2015).
96. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nature Biotechnology* **34**, 631 (2016).
97. Fellmann, C. & Lowe, S.W. Stable RNA interference rules for silencing. *Nature Cell Biology* **16**, 10 (2013).
98. Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat Cell Biol* **20**, 610-619 (2018).
99. Wutz, A. Haploid Mouse Embryonic Stem Cells: Rapid Genetic Screening and Germline Transmission. *Annual Review of Cell and Developmental Biology* **30**, 705-722 (2014).
100. Zhu, Z., Verma, N., González, F., Shi, Z.-D. & Huangfu, D. A CRISPR/Cas-Mediated Selection-free Knockin Strategy in Human Embryonic Stem Cells. *Stem Cell Reports* **4**, 1103-1111 (2015).
101. Shukla, A. & Huangfu, D. Decoding the noncoding genome via large-scale CRISPR screens. *Current Opinion in Genetics & Development* **52**, 70-76 (2018).
102. Brown, S. *et al.* Activin/Nodal Signaling Controls Divergent Transcriptional Networks in Human Embryonic Stem Cells and in Endoderm Progenitors. *STEM CELLS* **29**, 1176-1185 (2011).

103. Bock, C. *et al.* Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. *Cell* **144**, 439-452 (2011).
104. Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotech* **26**, 313-315 (2008).
105. González, F. *et al.* An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Cell Stem Cell* **15**, 215-226 (2014).
106. Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34**, 184 (2016).
107. Kanai-Azuma, M. *et al.* Depletion of definitive gut endoderm in Sox17 null mutant mice. *Development* **129**, 2367 (2002).
108. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Meth* **11**, 783-784 (2014).
109. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675-686 (2015).
110. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology* **15**, 554 (2014).
111. Gu, Z. *et al.* The type I serine/threonine kinase receptor ActRIA (ALK2) is required for gastrulation of the mouse embryo. *Development* **126**, 2551 (1999).
112. Yamamoto, M. *et al.* The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes & Development* **15**, 1242-1256 (2001).
113. Hart, A.H. *et al.* Mixl is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development* **129**, 3597 (2002).
114. Plouffe, S.W. *et al.* Characterization of Hippo Pathway Components by Gene Inactivation. *Molecular Cell* **64**, 993-1008 (2016).
115. Ribeiro, P.S. *et al.* Combined Functional Genomic and Proteomic Approaches Identify a PP2A Complex as a Negative Regulator of Hippo Signaling. *Molecular Cell* **39**, 521-534 (2010).
116. Estarás, C., Benner, C. & Jones, Katherine A. SMADs and YAP Compete to Control Elongation of β-Catenin:LEF-1-Recruited RNAPII during hESC Differentiation. *Molecular Cell* **58**, 780-793 (2015).
117. Beyer, Tobias A. *et al.* Switch Enhancers Interpret TGF-β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells. *Cell Reports* **5**, 1611-1624 (2013).
118. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell* **10**, 440-454 (2012).
119. Hoodless, P.A. *et al.* FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes & Development* **15**, 1257-1271 (2001).

120. Vallier, L. *et al.* Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* **136**, 1339 (2009).
121. Larsson, J. *et al.* Abnormal angiogenesis but intact hematopoietic potential in TGF‐β type I receptor‐deficient mice. *The EMBO Journal* **20**, 1663 (2001).
122. Meno, C. *et al.* Mouse Lefty2 and Zebrafish Antivin Are Feedback Inhibitors of Nodal Signaling during Vertebrate Gastrulation. *Molecular Cell* **4**, 287-298 (1999).
123. Perea-Gomez, A. *et al.* Nodal Antagonists in the Anterior Visceral Endoderm Prevent the Formation of Multiple Primitive Streaks. *Developmental Cell* **3**, 745-756 (2002).
124. Rogers, K.W. *et al.* Nodal patterning without Lefty inhibitory feedback is functional but fragile. *eLife* **6**, e28785 (2017).
125. Ashe, H.L. & Briscoe, J. The interpretation of morphogen gradients. *Development* **133**, 385 (2006).
126. Paul, A. *et al.* Stress-activated Protein Kinases: Activation, Regulation and Function. *Cellular Signalling* **9**, 403-410 (1997).
127. Plotnikov, A., Zehorai, E., Procaccia, S. & Seger, R. The MAPK cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1813**, 1619-1633 (2011).
128. Zhang, Y.E. Non-Smad pathways in TGF-β signaling. *Cell Research* **19**, 128 (2008).
129. Eferl, R. & Wagner, E.F. AP-1: a double-edged sword in tumorigenesis. *Nature Reviews Cancer* **3**, 859 (2003).
130. Kouzarides, T. & Ziff, E. The role of the leucine zipper in the fos–jun interaction. *Nature* **336**, 646-651 (1988).
131. Pulverer, B.J., Kyriakis, J.M., Avruch, J., Nikolakaki, E. & Woodgett, J.R. Phosphorylation of c-jun mediated by MAP kinases. *Nature* **353**, 670-674 (1991).
132. Smeal, T., Binetruy, B., Mercola, D.A., Birrer, M. & Karin, M. Oncogenic and transcriptional cooperation with Ha-Ras requires phosphorylation of c-Jun on serines 63 and 73. *Nature* **354**, 494-496 (1991).
133. Foletta, V., H, Segal, D. & Cohen, D. *Transcriptional regulation in the immune system: All roads lead to AP-1*, 139-52 (1998).
134. Arnulf, B. *et al.* Human T-cell lymphotropic virus oncoprotein Tax represses TGF-β1 signaling in human T cells via c-Jun activation: a potential mechanism of HTLV-I leukemogenesis. *Blood* **100**, 4129 (2002).
135. Pessah, M. *et al.* c-Jun Associates with the Oncoprotein Ski and Suppresses Smad2 Transcriptional Activity. *Journal of Biological Chemistry* **277**, 29094-29100 (2002).
136. Pessah, M. *et al.* c-Jun interacts with the corepressor TG-interacting factor (TGIF) to suppress Smad2 transcriptional activity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6198-6203 (2001).

137. Zhang, Y., Feng, X.-H. & Derynck, R. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-[beta]-induced transcription. *Nature* **394**, 909-913 (1998).
138. Davis, R.J. Signal Transduction by the JNK Group of MAP Kinases. *Cell* **103**, 239-252 (2000).
139. Angel, P., Hattori, K., Smeal, T. & Karin, M. The jun proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell* **55**, 875-885 (1988).
140. Zhang, T. *et al.* Discovery of potent and selective covalent inhibitors of JNK. *Chemistry & biology* **19**, 140-154 (2012).
141. Macosko, Evan Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
142. Mullen, A.C. *et al.* Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling. *Cell* **147**, 565-576 (2011).
143. Tsankov, A.M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344-349 (2015).
144. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576-589 (2010).
145. Karnik, R. & Beer, M.A. Identification of Predictive Cis-Regulatory Elements Using a Discriminative Objective Function and a Dynamic Search Space. *PLOS ONE* **10**, e0140557 (2015).
146. Ghandi, M., Lee, D., Mohammad-Noori, M. & Beer, M.A. Enhanced Regulatory Sequence Prediction Using Gapped k-mer Features. *PLOS Computational Biology* **10**, e1003711 (2014).
147. The, E.P.C. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).
148. Hilberg, F., Aguzzi, A., Howells, N. & Wagner, E.F. c-Jun is essential for normal mouse development and hepatogenesis. *Nature* **365**, 179 (1993).
149. Johnson, R.S., van Lingen, B., Papaioannou, V.E. & Spiegelman, B.M. A null mutation at the c-jun locus causes embryonic lethality and retarded cell growth in culture. *Genes & Development* **7**, 1309-1317 (1993).
150. Xu, P. & Davis, R.J. c-Jun NH2-Terminal Kinase Is Required for Lineage-Specific Differentiation but Not Stem Cell Self-Renewal. *Molecular and Cellular Biology* **30**, 1329-1340 (2010).
151. Liu, J. *et al.* The oncogene c-Jun impedes somatic cell reprogramming. *Nature Cell Biology* **17**, 856 (2015).
152. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459.e20 (2017).
153. Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* **21**, 819-833.e6 (2017).
154. Phanstiel, D.H. *et al.* Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Molecular Cell* **67**, 1037-1048.e6 (2017).
155. Xu, R.-H. *et al.* NANOG Is a Direct Target of TGF β /Activin-Mediated SMAD Signaling in Human ESCs. *Cell Stem Cell* **3**, 196-206 (2008).

156. Inman, G.J. *et al.* SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor-β Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. *Molecular Pharmacology* **62**, 65 (2002).
157. Liu, L. *et al.* Smad2 and Smad3 have differential sensitivity in relaying TGFβ signaling and inversely regulate early lineage specification. *Scientific Reports* **6**, 21602 (2016).
158. Liu, X. *et al.* Tead and AP1 Coordinate Transcription and Motility. *Cell Reports* **14**, 1169-1180 (2016).
159. Obier, N. *et al.* Cooperative binding of AP-1 and TEAD4 modulates the balance between vascular smooth muscle and hemogenic cell fate. *Development* **143**, 4324 (2016).
160. Zanconato, F. *et al.* Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nature Cell Biology* **17**, 1218 (2015).
161. Verfaillie, A. *et al.* Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nature Communications* **6**, 6683 (2015).
162. Vierbuchen, T. *et al.* AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Molecular cell* **68**, 1067-1082.e12 (2017).
163. Estarás, C., Hsu, H.-T., Huang, L. & Jones, K.A. YAP repression of the WNT3 gene controls hESC differentiation along the cardiac mesoderm lineage. *Genes & Development* **31**, 2250-2263 (2017).
164. Gotoh, S. *et al.* Generation of Alveolar Epithelial Spheroids via Isolated Progenitor Cells from Human Pluripotent Stem Cells. *Stem Cell Reports* **3**, 394-403 (2014).
165. McCauley, K.B. *et al.* Efficient Derivation of Functional Human Airway Epithelium from Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. *Cell Stem Cell* **20**, 844-857.e6 (2017).
166. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
167. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
168. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
169. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer-Verlag New York, 2016).
170. Mo, A. *et al.* Epigenomic landscapes of retinal rods and cones. *eLife* **5**, e11613 (2016).
171. van Dijk, D. *et al.* MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. *bioRxiv* (2017).

172. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. in *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., 2001).
173. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357 (2012).
174. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
175. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207 (2010).
176. Lee, D. *et al.* A method to predict the impact of regulatory variants from DNA sequence. *Nature Genetics* **47**, 955 (2015).
177. Beer, M.A. Predicting enhancer activity and variant impact using gkm-SVM. *Human Mutation* **38**, 1251-1258 (2017).
178. Kreimer, A. *et al.* Predicting gene expression in massively parallel reporter assays: A comparative study. *Human Mutation* **38**, 1240-1250 (2017).
179. Moris, N., Pina, C. & Arias, A.M. Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics* **17**, 693 (2016).
180. Huang, S., Guo, Y.-P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology* **305**, 695-713 (2007).
181. François, P. & Hakim, V. Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 580-585 (2004).
182. Nostro, M.C. *et al.* Efficient Generation of NKX6-1+ Pancreatic Progenitors from Multiple Human Pluripotent Stem Cell Lines. *Stem Cell Reports* **4**, 591-604 (2015).
183. Green, M.D. *et al.* Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nature Biotechnology* **29**, 267 (2011).
184. Huang, S.X.L. *et al.* Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nature Biotechnology* **32**, 84 (2013).
185. Mandegar, Mohammad A. *et al.* CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541-553 (2016).