Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

*Nat Genet*. Author manuscript; available in PMC 2019 November 20.

Published in final edited form as:

*Nat Genet*. 2019 June ; 51(6): 999–1010. doi:10.1038/s41588-019-0408-9.

# Genome-scale screens identify JNK/JUN signaling as a barrier for pluripotency exit and endoderm differentiation

**Qing V. Li**[1,2], **Gary Dixon**[1,3], **Nipun Verma**[1,4], **Bess P. Rosen**[1,3], **Miriam Gordillo**[5], **Renhe Luo**[1,2], **Chunlong Xu**[1], **Qiong Wang**[1,6], **Chew-Li Soh**[1], **Dapeng Yang**[1], **Miguel Crespo**[5], **Abhijit Shukla**[1], **Qing Xiang**[1], **Friederike Dündar**[7], **Paul Zumbo**[7], **Matthew Witkin**[1], **Richard Koche**[1], **Doron Betel**[8], **Shuibing Chen**[5], **Joan Massagué**[1], **Ralph Garippa**[1], **Todd Evans**[5], **Michael A. Beer**[9,*], and **Danwei Huangfu**[1,*]

[1]Sloan Kettering Institute, 1275 York Avenue, New York, New York 10065, USA

[2]Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA

[3]Weill Graduate School of Medical Sciences at Cornell University, 1300 York Avenue, New York, New York 10065, USA

[4]Weill Graduate School of Medical Sciences at Cornell University/The Rockefeller, University/ Sloan Kettering Institute Tri-Institutional M.D.-Ph.D. Program, 1300 York Avenue, New York, New York 10065, USA

[5]Department of Surgery, Weill Cornell Medical College, New York, New York 10065, USA

[6]Department of Histo-Embryology, Genetics and Developmental Biology, and Shanghai Key Laboratory of Reproductive Medicine, Shanghai Jiaotong University, School of Medicine, Shanghai 200025, China.

[7]Applied Bioinformatics Core, Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York 10065, USA

[8]Department of Medicine and Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York 10065, USA

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

[*]Correspondence to: huangfud@mskcc.org (D.H.), mbeer@jhu.edu (M.A.B.).

Author Contributions

Q.V.L. and D.H. designed experiments, analyzed and interpreted results. Q.V.L. performed most experiments. M.A.B. performed the mathematical modeling and computational analysis. G.D., B.P.R., R.L. and A.S. assisted with the screen and the validation experiments. N.V. performed neuroectoderm differentiation experiments. Q.X. and R.G. provided assistance related to the CRISPR library construction and sequencing. Q.W. and J.M. provided protocols, reagents and technical support on experiments related to ChIP-seq and TGF-β signaling. M.W. and R.K. assisted with ATAC-seq and ChIP-seq experiments. C.-L.S., D.Y., M.G., C.X., M.C., S.C. and T.E. tested JNK inhibitors and assisted with additional experiments and data analyses. F.D., P.Z. and D.B. performed Drop-seq data analysis. Q.V.L., M.A.B. and D.H. wrote the manuscript and all other authors provided editorial advice.

Competing Interests Statement

D.H. and Q.V.L. are listed as inventors for a patent application (Pub. No: WO/2018/035454) relating to this work for the application of JNK inhibitor in endoderm differentiation.

Data Availability Statement

Sequencing data are available under NCBI GEO accession number GSE109524.

H3K27ac ChIP-seq data are from GSM733718. p300 ChIP-seq data are from GSM803542. Codes used in this study are available from the corresponding authors upon request.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[9]Department of Biomedical Engineering and McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.

## Abstract

Human embryonic and induced pluripotent stem cells (hESCs/hiPSCs) hold great promise for cell-based therapies and drug discovery. However, homogeneous differentiation remains a major challenge, highlighting the need for understanding developmental mechanisms. We performed genome-scale CRISPR screens to uncover regulators of definitive endoderm (DE) differentiation, which unexpectedly uncovered five JNK/JUN family genes as key barriers of DE differentiation. The JNK/JUN pathway does not act through directly inhibiting the DE enhancers. Instead JUN co-occupies ESC enhancers with OCT4, NANOG and SMAD2/3, and specifically inhibits the exit from the pluripotent state by impeding the decommissioning of ESC enhancers and inhibiting the reconfiguration of SMAD2/3 chromatin binding from ESC to DE enhancers. Therefore, the JNK/JUN pathway safeguards pluripotency from precocious DE differentiation. Direct pharmacological inhibition of JNK significantly improves the efficiencies of generating DE and DE-derived pancreatic and lung progenitor cells, highlighting the potential of harnessing the knowledge from developmental studies for regenerative medicine.

## Introduction

By 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[1]. These findings form the basis for guided differentiation of hESCs or hiPSCs into the three embryonic lineages and their derivatives[2]. On the other hand, CRISPR/Cas-mediated genome editing in hESCs/hiPSCs combined with guided differentiation offers a valuable platform to investigate human development. For instance, the guided differentiation of hESCs/hiPSCs into DE is induced by Activin A[3]. This mimics the action of Nodal, a member of the TGF-β superfamily known to signal via the SMAD2/3-SMAD4-FOXH1 complex to promote endoderm differentiation in gastrulating mouse embryos[4]. Using Activin A-induced DE differentiation and reverse genetics in hESCs, we and others have previously demonstrated critical requirements for transcription factors (TFs) EOMES and GATA6 in the formation of human endoderm[5–8].

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[3,9–11]. 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[12]. In addition, differentiation efficiencies vary among hESC/hiPSC lines[13,14], and homogeneous differentiation remains a major challenge, highlighting the need for discovering additional regulatory mechanisms controlling DE differentiation.

Forward genetic screens in model organisms provide a powerful approach for uncovering previously unsuspected regulators of development[1]. The key endoderm regulator *Nodal* was

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

itself first identified in a genetic screen performed in mice[15]. 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. 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. Moreover, JNK inhibitor treatment improves DE and DE-derived pancreatic and lung lineage differentiation and reduces the dose requirement for Activin A, highlighting the potential of harnessing the knowledge gained from developmental studies for regenerative medicine.

## Results

### Genome-scale knockout screens using pooled CRISPR libraries

To screen for regulators of DE differentiation, we used the HUES8 iCas9 hESC line, which expresses Cas9 upon doxycycline treatment[16] (Supplementary Fig. 1a). We further employed a selection-free knock-in strategy[17] to integrate a *GFP* transgene into the *SOX17* locus to report endoderm fate[18] (Supplementary Fig. 1b–c). Both Activin A and CHIR99021 were required to induce $CXCR4^+SOX17^+$ (~80%) DE cells (Fig. 1a–b), and faithful GFP reporter expression was confirmed by immunostaining and flow cytometry (Supplementary Fig. 1d–e). As a positive control, the requirement for *EOMES* in DE differentiation[5,19] was confirmed by performing differentiation on HUES8 $SOX17^{GFP/+}$ iCas9 hESCs infected with a lentivirus expressing an *EOMES*-targeting gRNA (Supplementary Fig. 1f).

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)[20]. To improve the confidence of screening hits, we then repeated the screen using a serum-free differentiation condition with the Brunello library[21] consisting of 76,441 gRNAs targeting 19,114 genes (4 gRNAs per gene). After infection of HUES8 $SOX17^{GFP/+}$ iCas9 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. 1c). The abundance of individual gRNAs in each population was determined by high-throughput sequencing.

Two approaches were used to identify and rank hits. We first calculated the Z-score[22] for each gRNA based on the ratio of gRNA reads in the SOX17-$GFP^-$ versus SOX17-$GFP^+$ populations (Supplementary Fig. 2a–b). 37 positive regulators and 28 negative regulators exceeded the absolute Z-score cutoff ($|Z\text{-score}| > 2$) in both screens (Fig. 1d, Supplementary Fig. 2c and Supplementary Data Set 1). In the second approach, we used the MAGeCK robust ranking aggregation (RRA) algorithm to identify screening hits[23]. RRA values from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the two screens correlated well (Fig. 1e–f). All 27 genes with a false discovery rate (FDR) lower than 0.05 in both screens met the Z-score cutoff (|Z-score| > 2.0, Fig. 1e–f and Supplementary Data Set 1). The theoretical FDR predicted by MAGeCK from the GeCKO library screen is consistent with the experimental FDR replicated in the Brunello library screen (Supplementary Fig. 2d). 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*, *SMAD4* and *FOXH1*) as well as established endoderm TFs (*EOMES* and *MIXL1*) (Fig. 1d and Supplementary Fig. 2c)[4,24].

## Validation of screening hits

We performed validation experiments on 33 high-ranking hits based on the Z-score and FDR (Supplementary Table 1). 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. 1g and Supplementary Fig. 3a). 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[19,25–30] (Fig. 1h). In addition, we verified four Hippo pathway genes (*NF2, TAOK1, PTPN14*, and *PPP2R4*) known to negatively regulate YAP1 nuclear activity[31,32] (Fig. 1h). Previous studies showed that transient siRNA knockdown of *YAP1* expression in hESCs led to upregulation of mesendoderm genes *MIXL1*, *EOMES* and *T*[33,34], 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. 1i). 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 (Supplementary Fig. 3b). Only 4 of the 24 genes showed greater than 2-fold gene expression changes during the ESC-DE transition by RNA-seq analysis (Supplementary Fig. 3c and Supplementary Data Set 2). This highlights the necessity for comprehensive functional analysis rather than relying on transcriptome analysis alone.

## 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. 1f and 2a). 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[35], 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. 1h and 2a). We further generated clonal H1 *MKK7* and *JUN* knockout (KO) lines and verified the knockouts by western blotting (Fig. 2b–c and Supplementary Fig. 4a). 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[36] (Fig. 2c). *MKK7* or *JUN* KO cells did not exhibit a difference in cell growth during hESC self-renewal or DE differentiation (Supplementary Fig. 4b); nor did flow cytometry detect a significant

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

difference in the number of cells expressing proliferation or apoptosis markers (phospho-histone H3 and cleaved caspase-3, respectively) during differentiation (Supplementary Fig. 4c). *MKK7* and *JUN* KO hESCs formed neuroectoderm cells[37] expressing PAX6 and SOX1 with comparable efficiency and kinetics to the WT control (Supplementary Fig. 5a–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. 2d–e). RNA-seq analysis showed significant global transcriptional differences between WT and *MKK7* KO cells at the DE but not the ESC stage (Fig. 2f and Supplementary Data Set 3), as verified using a panel of DE and ESC markers for flow cytometry, immunostaining and RT-qPCR analyses (Supplementary Fig. 4d–g). These findings demonstrate that the JNK/JUN pathway inhibits the ESC-DE transition but does not affect pluripotency maintenance.

## Pharmacological JNK inhibition improves DE differentiation

We tested the effects of JNK-IN-8[38], a selective JNK inhibitor (JNKi) on DE differentiation (Fig. 3a). Western blotting analysis confirmed that JUN phosphorylation was barely detectable in JNKi treated H1 hESCs (Fig. 3b). 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. 3c and Supplementary Fig. 6a). JNKi also significantly increased DE differentiation efficiency in additional hESC (HUES8 and HUES6) and hiPSC (BJ and CV) lines (Supplementary Fig. 6b–c). These findings demonstrate that JNK inhibition improves DE differentiation efficiency and reduces differentiation variability.

We performed Drop-seq[39] 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. 3d) or the 1,000 most variably expressed genes (Supplementary Fig. 6d). 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. 3d and Supplementary Fig. 6d). This difference in cell heterogeneity can be further visualized in scatter plots showing the expression of individual DE and ESC genes (Fig. 3e–f and Supplementary Fig. 6e). Thus, JNKi treatment during differentiation results in a more homogeneous DE population.

We further examined the effects of JNKi using an optimized serum-free DE differentiation protocol[6,40,41] (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 (Supplementary Fig. 7a–c). We further demonstrated that one day of JNKi treatment improved DE differentiation across different hESC (HUES8, H1, HUES6) and hiPSC (CV,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

BJ) lines under this optimized differentiation condition (Fig. 3g and Supplementary Fig. 7d). Different culture media could influence the outcome of DE differentiation. Our work 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. 3h–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 (Supplementary Fig. 7e–f). These results demonstrate that JNK inhibition improves DE differentiation, which leads to more robust subsequent differentiation to the pancreatic and lung lineages.

## 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[3,10,11]. SMAD2/3 binding patterns have been shown to be associated with lineage transcription factors[42–44]. We performed SMAD2/3 ChIP-seq at the ESC and DE stages, and analysis of differentially bound regions 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 (Supplementary Fig. 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, corresponding to the primitive streak stage). Using two *de novo* discriminative motif algorithms, HOMER[45] and MotifSpec[46], we found AP1 and TEAD binding sites were enriched in regions with decreased accessibility upon JNK inhibition (Fig. 4a). 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 gkm-SVM[47] 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) (Supplementary Data Set 4).

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. 4b–c). These results suggest that JNK inhibition promotes the access of SMAD2/3 and GATA6 to DE enhancers during the ESC-DE transition. Importantly, as suggested by the motif analysis, regions with decreased accessibility upon JNK inhibition showed enrichment of JUN binding at the ESC stage, 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,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

regions with decreased accessibility upon JNK inhibition also showed enriched OCT4 and SMAD2/3 binding at the ESC stage (Fig. 4b–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. We observed significant overlap ($P < 3.89 \times 10^{-224}$) between OCT4 and JUN binding (Supplementary Fig. 8c): ~60% of JUN-bound sites were co-bound by OCT4. Focusing on the strongest OCT4 bound sites (using MACS2 FDR $q = 1 \times 10^{-6}$ instead of the default $q = 0.05$ for peak calling), the overlap is still significant ($P < 1 \times 10^{-300}$): ~28% of JUN-bound sites were co-bound by OCT4 (Fig. 4d). 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. 4e and Supplementary Fig. 8d). 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. 4e and Supplementary Fig. 8d). 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. 4f). 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. 4f). Together these findings suggest 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.

## 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. 5a). 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. 5a). 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. 5a and Supplementary Fig. 8e). Together, these analyses uncovered 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. 5a–b). Such reciprocal effects were also observed on SMAD2/3 binding at ESC and DE enhancers respectively (Fig. 5a and 5c), 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

occupied regions (including promoters) at the ESC and DE stages irrespective of OCT4 or GATA6 occupancy (Supplementary Fig. 8f). These effects were not due to an effect of JNK inhibition on SMAD2 protein expression or phosphorylation at Ser465/467 or Ser245/250/255 (Supplementary Fig. 8g).

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 upon JNK inhibition 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. 5d–e and Supplementary Fig. 8h). Corresponding changes were observed at many individual ESC and DE enhancers (Fig. 4f and Supplementary Fig. 8i). In addition, JNKi treatment significantly increased SMAD2/3 and GATA6 co-occupancy ($P < 2 \times 10^{-62}$, Supplementary Fig. 8j). Further supporting a direct causal relationship between decreased JUN binding and decreased SMAD2/3 binding at ESC enhancers, HOMER, MotifSpec and gkm-SVM 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. 5f and Supplementary Data Set 4).

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. 5g). The expression levels of these genes at the ESC stage were not affected by genetic inactivation of the JNK pathway (*MKK7* KO) (Fig. 5h). However, after DE differentiation, they were further downregulated in *MKK7* KO DE cells compared to WT DE cells (Fig. 5h). 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 (Supplementary Fig. 8k). 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 an ESC to DE signature and promotes the establishment of the DE enhancer landscape (Fig. 6a).

## 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 (e.g. GATA6) either directly or indirectly (Fig. 6b–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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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. 6d). 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. 6e). 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) (Supplementary Fig. 9a–b). Similar effects were observed with genetic inactivation of the JNK pathway members *MKK7* and *JUN* (Supplementary Fig. 9c–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.

## Discussion

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 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. This is accompanied by decreased SMAD2/3 occupancy at ESC enhancers, thus JNK inhibition also facilitates the reconfiguration of SMAD2/3 chromatin binding. Our mathematical modeling further supports the notion 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 AP1 TF involved in DE differentiation, though potential overlapping requirements of additional AP1 TFs cannot be excluded. It is of great interest to determine whether the JNK/JUN pathway also inhibits DE *in vivo*. 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 show developmental arrest at around embryonic day 12, but earlier phenotypes during DE differentiation have not been examined[49,50]. Deletion of *Jnk1/2* or *Jun* in mouse ESCs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(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[51,52]. More quantitative experiments are needed to characterize the exact involvement of the JNK/JUN pathway in mouse embryos during the crucial time window of pluripotency exit and DE differentiation. The mouse knockout phenotypes may also differ from those observed in hESC directed differentiation due to potential compensatory mechanisms *in vivo* or species-specific gene requirements.

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[52–54]. 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 AP1 motif is found through Hi-C genomic data analysis to be highly enriched in the activation hubs that connect multiple enhancers to promoters[55].

Our work demonstrates the power of using unbiased CRISPR screens to identify the barriers of hESC differentiation. JNK inhibitor treatment not only improved the differentiation efficiency of DE, but also improved DE-derived pancreatic and lung lineage cells. Thus, targeting negative regulators of DE differentiation represents a rational strategy to improve differentiation to endoderm-derived lineages. Other groups have successfully improved pancreatic and lung differentiation by focusing on differentiation at later stages[56,57]. For instance, the Kotton[57] and Snoeck[58] 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[59]. The use of JNKi during DE differentiation increases the frequency of 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 inducible expression of Cas9[16] or dCas9-KRAB[60] 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

# 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[6,16]. Undifferentiated hESCs and hiPSCs were maintained as previously described[61] 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°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[16]. Since the GeCKO v2 and Brunello library also relies on puromycin selection, we first transfected HUES8 iCas9 cells with an *in vitro* transcribed gRNA (Supplementary Table 2) to knockout the puromycin resistance gene. We then established the HUES8 *SOX17$^{GFP/+}$* iCas9 line using a selection-free knock-in strategy[17] (Supplementary Fig. 1b). Correct targeting was verified by Southern blotting as previously described[16]. The sequence for the 5' external probe is shown in Supplementary Table 3. 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 (Supplementary Fig. 1c). 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 *SOX17$^{GFP/+}$* iCas9 cell line maintains a normal karyotype (Supplementary Fig. 1g). Genomic DNA genotyping revealed no *de novo* mutation in the *TP53* coding sequences in the HUES8 *SOX17$^{GFP/+}$* iCas9 cell line.

## Production of the lentiviral library

Human CRISPR libraries GeCKO v2 and Brunello were obtained from Addgene (1000000049 and 73178)[20,21]. 50 μg CRISPR library plasmids with 20 μg PAX2 and 5 μg VSVG plasmids 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°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 1-P (0). When MOI is equal 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 HUES8 *SOX17$^{GFP/+}$* hESCs per well were infected

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 HUES8 *SOX17*$^{GFP/+}$ iCas9 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 150 mm 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 from GeCKO screen and ~504 μg of DNA per sample from Brunello screen (6.6 μg of genomic DNA per 1 million cells) were used to perform PCR using Herculase II fusion DNA polymerase (Agilent, 600679) in order to achieve a 1,000-fold converge of the GeCKO and Brunello libraries. Primer sequences to amplify lentiGuide-puro for the first PCR are shown in Supplementary Table 3 (F1 and R1). 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 shown in Supplementary Table 3 (F2 and R2). 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. Z-score of each gRNA and gene was calculated as illustrated in Supplementary Figure 2a–b.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

### Hit validation and the lentiCRISPR approach

The H1 iCas9 line[6] was used for most validation experiments. Some hits were tested again in HUES8 *SOX17*$^{GFP/+}$ 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[20]. The lentiGuide-puro construct expresses a puromycin resistance gene, allowing the selection of infected cells through puromycin treatment. 1 μg lentiGuide, 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 °C. A MOI of 0.30~0.36 was used for the infection of the 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 Supplementary Table 2. Two gRNAs per gene were tested for validation. Two non-targeting gRNAs were used as WT controls.

### Generation of clonal KO hESC lines

Clonal KO lines were created as previously described[16] with some modifications. H1 iCas9 cells were infected with lentiviruses expressing *MKK7* or *JUN* targeting gRNAs made from the lentiGuide-puro construct (Supplementary Table 2) 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 hESCs were dissociated to single cells using TrypLE Select. 500 cells were plated into one 100 mm tissue culture dish with 10 ml E8 medium 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 Supplementary Table 3.

### 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 medium. For the H1 line, we typically plated 0.1 million cells to accommodate the higher proliferation rate compared to other lines. For differentiation in 150 mm 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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.

### Serum-free DE differentiation and subsequent pancreatic and lung differentiation

We performed DE differentiation using an optimized serum-free protocol[40,61] followed by pancreatic or lung lineage differentiation. For serum-free DE differentiation, 0.3 million HUES8, H1, HUES6, and BJ iPSCs were plated in E8 medium 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 medium for two days before DE differentiation due to its slower growth rate. Cells were washed with PBS once and culture medium was changed to MCDB131 (Life Technologies, 10372–019) with 1× GlutaMax, 0.5% BSA (Lampire Biological Products, 7500804), 1.5 g/l $NaHCO_3$, 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 medium was supplemented with 0.5 μM CHIR99021 and 100 or 20 ng/ml Activin A as specified. On the third day, the medium was supplemented with 100 or 20 ng/ml Activin A only.

For pancreatic lineage differentiation[6,40,41,61], ~1 million HUES8 hESCs were plated for DE differentiation as described above. DE cells were washed with PBS and cultured in MCDB131 with 1× GlutaMax, 0.5% BSA, 1.5 g/l $NaHCO_3$, 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 1× Glutamax, 2% BSA, 2.5 g/l $NaHCO_3$, and 10 mM 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 1× ITS-X (Life Technologies, 51500–056) for 2 days. Then, cells were cultured in MCDB131 with 1× Glutamax, 2% BSA, 2.5 g/l $NaHCO_3$, and 10 mM 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 1× ITS-X for 3 days. Medium was changed every day.

For lung lineage differentiation[57,62], ~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 $4.5 \times 10^{-4}$ M MTG (Sigma). DE cells were differentiated in lung base medium containing 10 μM SB431542 (Stemgent, 04–0010) and 2 μM Dorsomorphin (Tocris, 3093)[57] from D3 to D6. Then, from D6 to D9, cells were grown in lung base

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

medium containing 100 ng/ml of human recombinant BMP4 (R&D Systems, 314-BP-010), 0.5 μM ATRA (Sigma) and 3 μM CHIR99021[62]. Medium was changed every day.

### Neuroectoderm Differentiation

hESCs cultured in E8 were disaggregated using TrypLE Select for 5 minutes and washed using E8 medium. The cells were plated on Matrigel (BD, 354234) coated dishes in E8 medium with 10 μM Y-27632 at a density of 180,000–200,000 cells/cm$^2$. After 1 day of culture in E8 medium, 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 medium 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 medium 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 medium 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 medium contains Knockout DMEM (Thermo Fisher Scientific, 10829018), Knockout Serum Replacement (Thermo Fisher Scientific, 10828028), 1× MEM Non-Essential Amino Acids (Thermo Fisher Scientific, 11140050), 1× GlutaMAX (Thermo Fisher Scientific, 35050079), and 2-mercaptoethanol (Thermo Fisher Scientific, 21985023). N2 medium 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 °C. After washing with FACS buffer, cells were fixed and permeabilized in 1× 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 Supplementary Table 4. FACS gating strategy is shown in Supplementary Figure 1h.

### 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 Supplementary Table 5.

## RNA-seq analysis

ESC and DE stage RNA-seq data from both WT and *MKK7* KO cells were analyzed by Tophat[63], featureCounts[64] and DESeq2[65]. 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[66]. ChIP-seq peaks of increased JUN binding in ESC relative to DE D1 (Supplementary Fig. 8k) are preferentially located closer to the TSS of ESC expressed genes farther from DE expressed genes, following the similar analysis in Figure 5 supplement 1 of reference[67].

## Single-cell RNA-seq (Drop-seq) and analysis

DE cells were washed with PBS once and treated with TrypLE for 5 minutes. E8 medium 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 ice 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 laboratory (v. 3.1, 12/18/2015)[39] 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 laboratory (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[68]. 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 download (Supplementary Note 1).

## 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

exposure for the same field. All primary and secondary antibodies used are listed in Supplementary Table 4.

## 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 1 mM PMSF (MP Biomedicals, ICN19538105). Proteins were pre-cleared by centrifugation at 14,000 g for 10 minutes at 4 °C. 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 °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 Supplementary Table 4.

## ChIP-seq

ChIP was performed as previously described[69]. 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 3,000 RPM for 5 minutes at 4 °C and frozen in liquid nitrogen immediately before transferring to the −80 °C 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 was 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 1,500 RPM for 5 minutes at 4 °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 $MgCl_2$) and lysed in 50 μl of ATAC lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 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 1,500 RPM for 10 minutes at 4 °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 °C. Next, 0.2% SDS was added to the reaction mixtures and incubated for 5 minutes at room temperature. 2× 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[70] (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 repeated 12 times: 65 °C, 5 minutes; 98 °C, 30 seconds; 98 °C, 10 seconds, 65 °C 30 seconds. 1.5× 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[71] (paired reads for ATAC, local for ChIP) and ran MACS2[72] to call peaks, and defined 300 bp bound regions by extending 150 bp 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 = 45,320), we also defined a very stringent set of the strongest OCT4 peaks using $q = 1 \times 10^{-6}$ (n = 11,743) for OCT4 only. In Figure 4d and Supplementary Figure 8c, the $OCT4^+JUN^+$ sets were defined by the intersection of the OCT4 and JUN replicated peaks (n = 8,198 for Supplementary Fig. 8c, n = 3,800 for Fig. 4d), and the $OCT4^+JUN^-$ sets were defined to be equal size sets of $OCT4^+$ peaks with the lowest JUN ChIP-seq signal (n = 8,198 or n = 3,800). Both sets give consistent results.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

When comparing signals across experiments we normalized ATAC and ChIP-seq signals so columns in Figures 4b, 5a and 5e 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[73]. We defined fixed size sets of differentially bound or accessible regions by their distance from y = x or the regression line for imbalanced read depth: N = 5,000 for ATAC (Fig. 4a), N = 2500 for SMAD2/3 DE WT vs JNKi (Fig. 5f) and SMAD2/3 ESC vs DE (Supplementary Fig. 8a).

To show the direct effect of JUN binding on changes in the enhancer landscape during the ESC to DE transition, in Figure 5a we defined an $OCT4^+$ ESC enhancer set using the stringent set of OCT4-bound replicated non-blacklisted peaks defined above (using $q = 1 \times 10^{-6}$, n = 11,743) further filtered to remove promoters (< 2 kb from TSS) (n = 9,248). 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 = 9,248). 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 DE D1, we defined an $OCT4^+SMAD2/3^+$ ESC enhancer set from the overlap of the stringent $q = 1 \times 10^{-6}$ $OCT4^+$ 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 =8,500) (Supplementary Fig. 8f); a SMAD2/3-bound ESC set from replicated default peaks (n = 5,511) (Supplementary Fig. 8f), and a SMAD2/3-bound DE set from the strongest SMAD2/3 DE peaks (n = 8,500) (Supplementary Fig. 8f), 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 = 105,269 ESC accessible genomic loci (MACS2 peaks on ESC ATAC). For default peak calling (Supplementary Fig. 8c, nOCT4 = 45,230, nJUN = 13,474) the observed *P* value is $P < 3.89 \times 10^{-224}$ for the overlap of OCT4 and JUN. For more stringent ($q = 1 \times 10^{-6}$) peak calling for OCT4 (Fig. 4c, nOCT4 = 11,743) the observed *P* values are $P < 1 \times 10^{-300}$ 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 Supplementary Figure 8j 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 Supplementary Figure 8j 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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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', the 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 21,561 (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 $P < 2 \times 10^{-62}$.

## 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[45] and MotifSpec[46] 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, trained on chromatin accessible regions as described[74], which can describe how the full set of kmers can in combination describe the changes in binding and accessibility. We reported in Figures 4a, 5f and Supplementary Figure 8a the highest AUROC 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 (Supplementary Data Set 4). 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[75,76]. 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 (Supplementary Data Set 4), 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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[77,78] and our simplifying approximations are consistent with and supported by the results of more complete simulations[79]. 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 Figure 6c. 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. The MATLAB source code for the model is provided in Supplementary Note 2.

### 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 (NS) and *P* values < 0.05 (*), < 0.01 (**), < 0.001 (***), < 0.0001 (****) were considered significant.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The CV iPSC line was a gift from Larry Goldstein. The 293T cell line was a gift from Robert A. Weinberg's laboratory. We thank Z. Zhu, Z.-D. Shi, I. Caspi and T. Leonardo for technical assistance, S. Mehta for assistance with CRISPR screen Hi-seq, Y. Zou for assistance with ChIP-seq analysis, A.-K. Hadjantonakis and S. Tao for assisting with additional experiments not included in the manuscript, and K.V. Anderson, L. Studer, W. Guo, L. Dow and members of D.H.'s laboratory for insightful discussions. This study was supported in part by New York State Stem Cell Science (NYSTEM, C029156 and C32593GG), NIH/NIDDK (R01DK096239), the MSKCC Cancer Center Support Grant (P30CA008748), U01HG009380 (M.A.B.), R01HG007348 (M.A.B.), B.P.R., N.V. and Q.W. were supported by an NIH T32 Training Grant in Developmental and Stem Cell Biology (T32HD060600, G.D.), an NIH T32 Training Grant in Molecular and Cellular Biology (T32GM008539, B.P.R.), a Howard Hughes Medical Institute (HHMI) Medical Research Fellowship (N.V.), and a postdoctoral fellowship from a NYSTEM

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

grant (C026879, Q.W.) to the Center for Stem Cell Biology of the Sloan Kettering Institute and the National Natural Science Foundation of China (31771512, Q.W.).

## References


1. Anderson KV & Ingham PW The transformation of the model organism: a decade of developmental genetics. Nature Genetics 33, 285 (2003). [PubMed: 12610538]
2. Zhu Z & Huangfu D Human pluripotent stem cells: an emerging model in developmental biology. Development 140, 705–717 (2013). [PubMed: 23362344]
3. D'Amour KA et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotech 23, 1534–1541 (2005).
4. Robertson EJ Dose-dependent Nodal/Smad signals pattern the early mouse embryo. Seminars in Cell & Developmental Biology 32, 73–79 (2014). [PubMed: 24704361]
5. Teo AKK et al. Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes & Development 25, 238–250 (2011). [PubMed: 21245162]
6. 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). [PubMed: 28196600]
7. Fisher JB, Pulakanti K, Rao S & Duncan SA GATA6 is essential for endoderm formation from human pluripotent stem cells. Biology Open 6, 1084 (2017). [PubMed: 28606935]
8. 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). [PubMed: 28196690]
9. 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). [PubMed: 14627547]
10. Vallier L, Alexander M & Pedersen RA Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. Journal of Cell Science 118, 4495 (2005). [PubMed: 16179608]
11. James D, Levine AJ, Besser D & Hemmati-Brivanlou A TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273 (2005). [PubMed: 15703277]
12. 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). [PubMed: 21630377]
13. 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). [PubMed: 21295703]
14. Osafune K et al. Marked differences in differentiation propensity among human embryonic stem cell lines. Nat Biotech 26, 313–315 (2008).
15. Conlon FL, Barth KS & Robertson EJ 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). [PubMed: 1879365]
16. 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). [PubMed: 24931489]
17. 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). [PubMed: 26028531]
18. Kanai-Azuma M et al. Depletion of definitive gut endoderm in Sox17 null mutant mice. Development 129, 2367 (2002). [PubMed: 11973269]
19. Arnold SJ, Hofmann UK, Bikoff EK & Robertson EJ Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. Development 135, 501 (2008). [PubMed: 18171685]
20. Sanjana NE, Shalem O & Zhang F Improved vectors and genome-wide libraries for CRISPR screening. Nat Meth 11, 783–784 (2014).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript


21. Doench JG et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology 34, 184 (2016).
22. Parnas O et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. Cell 162, 675–686 (2015). [PubMed: 26189680]
23. Li W et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biology 15, 554 (2014). [PubMed: 25476604]
24. Zorn AM & Wells JM Vertebrate Endoderm Development and Organ Formation. Annual Review of Cell and Developmental Biology 25, 221–251 (2009).
25. 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). [PubMed: 10226013]
26. Nomura M & Li E Smad2 role in mesoderm formation, left-right patterning and craniofacial development. Nature 393, 786–790 (1998). [PubMed: 9655392]
27. 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). [PubMed: 9420335]
28. 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). [PubMed: 11358868]
29. Haegel H et al. Lack of beta-catenin affects mouse development at gastrulation. Development 121, 3529 (1995). [PubMed: 8582267]
30. Hart AH et al. Mixl is required for axial mesendoderm morphogenesis and patterning in the murine embryo. Development 129, 3597 (2002). [PubMed: 12117810]
31. Plouffe SW et al. Characterization of Hippo Pathway Components by Gene Inactivation. Molecular Cell 64, 993–1008 (2016). [PubMed: 27912098]
32. Ribeiro PS 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). [PubMed: 20797625]
33. 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). [PubMed: 25936800]
34. 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). [PubMed: 24332857]
35. Davis RJ Signal Transduction by the JNK Group of MAP Kinases. Cell 103, 239–252 (2000). [PubMed: 11057897]
36. 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). [PubMed: 3142689]
37. Chambers SM et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotech 27, 275–280 (2009).
38. Zhang T et al. Discovery of potent and selective covalent inhibitors of JNK. Chemistry & biology 19, 140–154 (2012). [PubMed: 22284361]
39. Macosko Evan Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214 (2015). [PubMed: 26000488]
40. 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).
41. Pagliuca Felicia W. et al. Generation of Functional Human Pancreatic β Cells In Vitro. Cell 159, 428–439 (2014). [PubMed: 25303535]
42. Mullen AC et al. Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling. Cell 147, 565–576 (2011). [PubMed: 22036565]
43. Tsankov AM et al. Transcription factor binding dynamics during human ES cell differentiation. Nature 518, 344–349 (2015). [PubMed: 25693565]
44. 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). [PubMed: 24412311]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

45. 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). [PubMed: 20513432]
46. Karnik R & Beer MA Identification of Predictive Cis-Regulatory Elements Using a Discriminative Objective Function and a Dynamic Search Space. PLOS ONE 10, e0140557 (2015). [PubMed: 26465884]
47. Ghandi M, Lee D, Mohammad-Noori M & Beer MA Enhanced Regulatory Sequence Prediction Using Gapped k-mer Features. PLOS Computational Biology 10, e1003711 (2014). [PubMed: 25033408]
48. The EPC et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57 (2012). [PubMed: 22955616]
49. Hilberg F, Aguzzi A, Howells N & Wagner EF c-Jun is essential for normal mouse development and hepatogenesis. Nature 365, 179 (1993). [PubMed: 8371760]
50. Johnson RS, van Lingen B, Papaioannou VE & Spiegelman BM A null mutation at the c-jun locus causes embryonic lethality and retarded cell growth in culture. Genes & Development 7, 1309–1317 (1993). [PubMed: 8330736]
51. Xu P & Davis RJ 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). [PubMed: 20065035]
52. Liu J et al. The oncogene c-Jun impedes somatic cell reprogramming. Nature Cell Biology 17, 856 (2015). [PubMed: 26098572]
53. Chronis C et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell 168, 442–459.e20 (2017). [PubMed: 28111071]
54. Li D et al. Chromatin Accessibility Dynamics during iPSC Reprogramming. Cell Stem Cell 21, 819–833.e6 (2017). [PubMed: 29220666]
55. Phanstiel DH et al. Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. Molecular Cell 67, 1037–1048.e6 (2017). [PubMed: 28890333]
56. Nostro MC et al. Efficient Generation of NKX6–1+ Pancreatic Progenitors from Multiple Human Pluripotent Stem Cell Lines. Stem Cell Reports 4, 591–604 (2015). [PubMed: 25843049]
57. McCauley KB 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). [PubMed: 28366587]
58. Green MD et al. Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. Nature Biotechnology 29, 267 (2011).
59. Huang SXL et al. Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. Nature Biotechnology 32, 84 (2013).
60. Mandegar Mohammad A. et al. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. Cell Stem Cell 18, 541–553 (2016). [PubMed: 26971820]

## Method-only references

61. 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). [PubMed: 27133796]
62. 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). [PubMed: 25241738]
63. Trapnell C, Pachter L & Salzberg SL TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–11 (2009). [PubMed: 19289445]
64. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–30 (2014). [PubMed: 24227677]
65. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]
66. Wickham H ggplot2: Elegant Graphics for Data Analysis, (Springer-Verlag New York, 2016).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript


67. Mo A et al. Epigenomic landscapes of retinal rods and cones. eLife 5, e11613 (2016). [PubMed: 26949250]
68. van Dijk D et al. MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. bioRxiv (2017).
69. 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). [PubMed: 27889317]
70. Buenrostro JD, Wu B, Chang HY & Greenleaf WJ ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide in Current Protocols in Molecular Biology (John Wiley & Sons, Inc., 2001).
71. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357 (2012). [PubMed: 22388286]
72. Zhang Y et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137 (2008). [PubMed: 18798982]
73. Kent WJ, Zweig AS, Barber G, Hinrichs AS & Karolchik D BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26, 2204–2207 (2010). [PubMed: 20639541]
74. Lee D et al. A method to predict the impact of regulatory variants from DNA sequence. Nature Genetics 47, 955 (2015). [PubMed: 26075791]
75. Beer MA Predicting enhancer activity and variant impact using gkm-SVM. Human Mutation 38, 1251–1258 (2017). [PubMed: 28120510]
76. Kreimer A et al. Predicting gene expression in massively parallel reporter assays: A comparative study. Human Mutation 38, 1240–1250 (2017). [PubMed: 28220625]
77. Moris N, Pina C & Arias AM Transition states and cell fate decisions in epigenetic landscapes. Nature Reviews Genetics 17, 693 (2016).
78. Huang S, Guo Y-P, May G & Enver T Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental Biology 305, 695–713 (2007). [PubMed: 17412320]
79. 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). [PubMed: 14704282]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

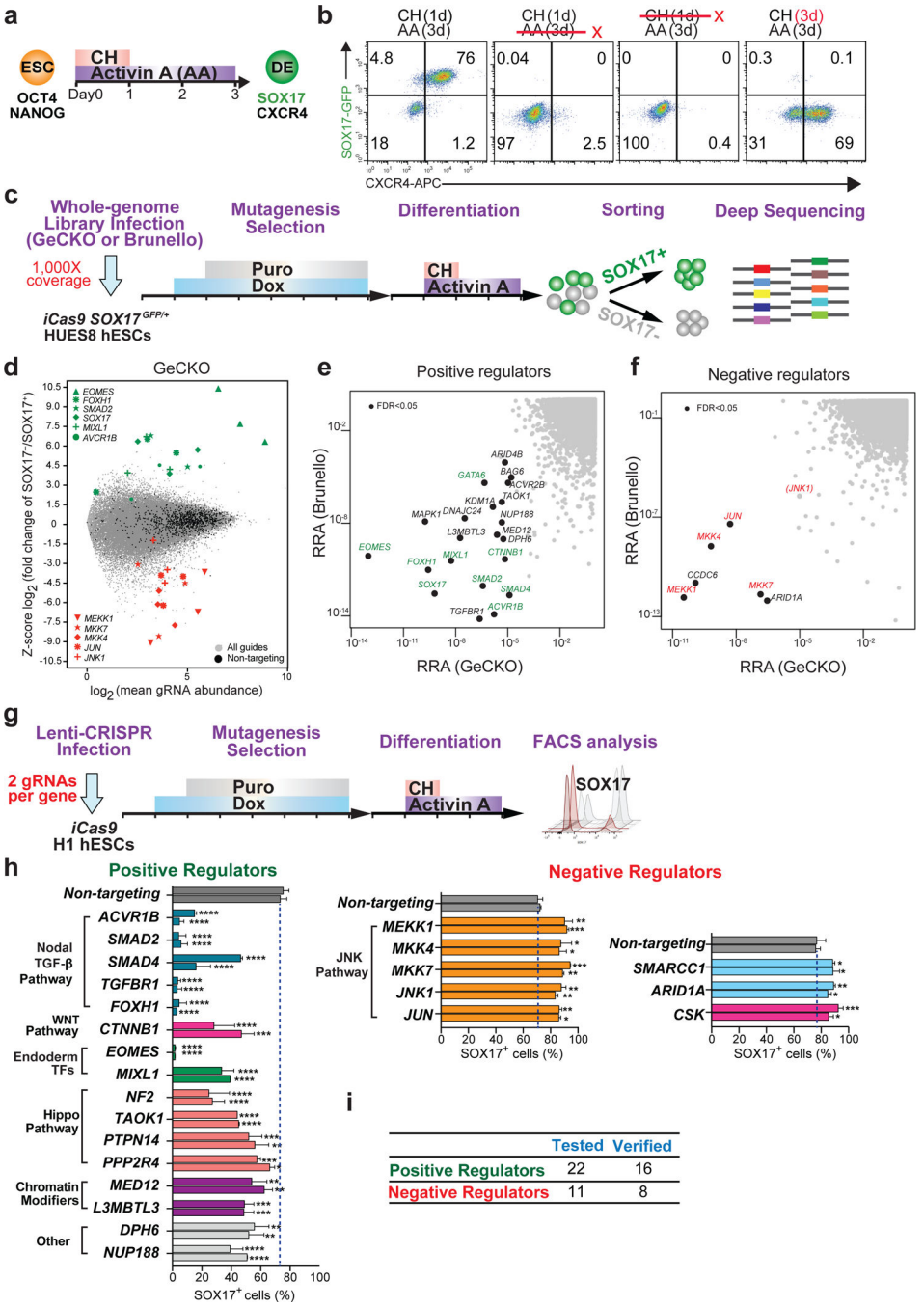


**Fig.1 |. Genome-scale screens identify regulators of DE differentiation.**

**a,** DE differentiation protocol. CH: CHIR99021 from Day (D) 0 to D1; AA: Activin A from D0 to D3. **b**, Representative flow cytometry 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 flow cytometry plot. Representative plots are from 3 independent experiments. **c**, Genome-wide CRISPR screen schematic. Each line segment on the horizontal arrows indicates 1 day of media and chemical treatment. **d,** A scatter plot of the gRNA distribution from GeCKO screen. Y-axis: Z-score of $\log_2$ fold-change of $SOX17^{-}$ vs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$SOX17^+$. 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. **e-f,** Robust Ranking Aggregation (RRA) score for positive regulators **(e)** and negative regulators **(f)** from the two screens. Known positive regulators are labeled as green texts. JNK pathway genes are labeled as red texts. Genes with FDR < 0.05 from both screens are indicated by black dots. JNK1 is shown in parenthesis as it just missed the stringent FDR cutoff (Brunello screen FDR < 0.05, GeCKO screen FDR = 0.056). **g,** Schematic of validation of top hit genes using by lentiCRISPR infection. **h**, Bar graphs show the percentage of $SOX17^+$ cells obtained after DE differentiation following gRNA targeting. n = 2 independent experiments. Error bars indicate SD. Statistical analysis was performed by unpaired two-tailed Student's *t* test. Significance is indicated as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$. **i**, A summary of the number of tested and verified hits.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

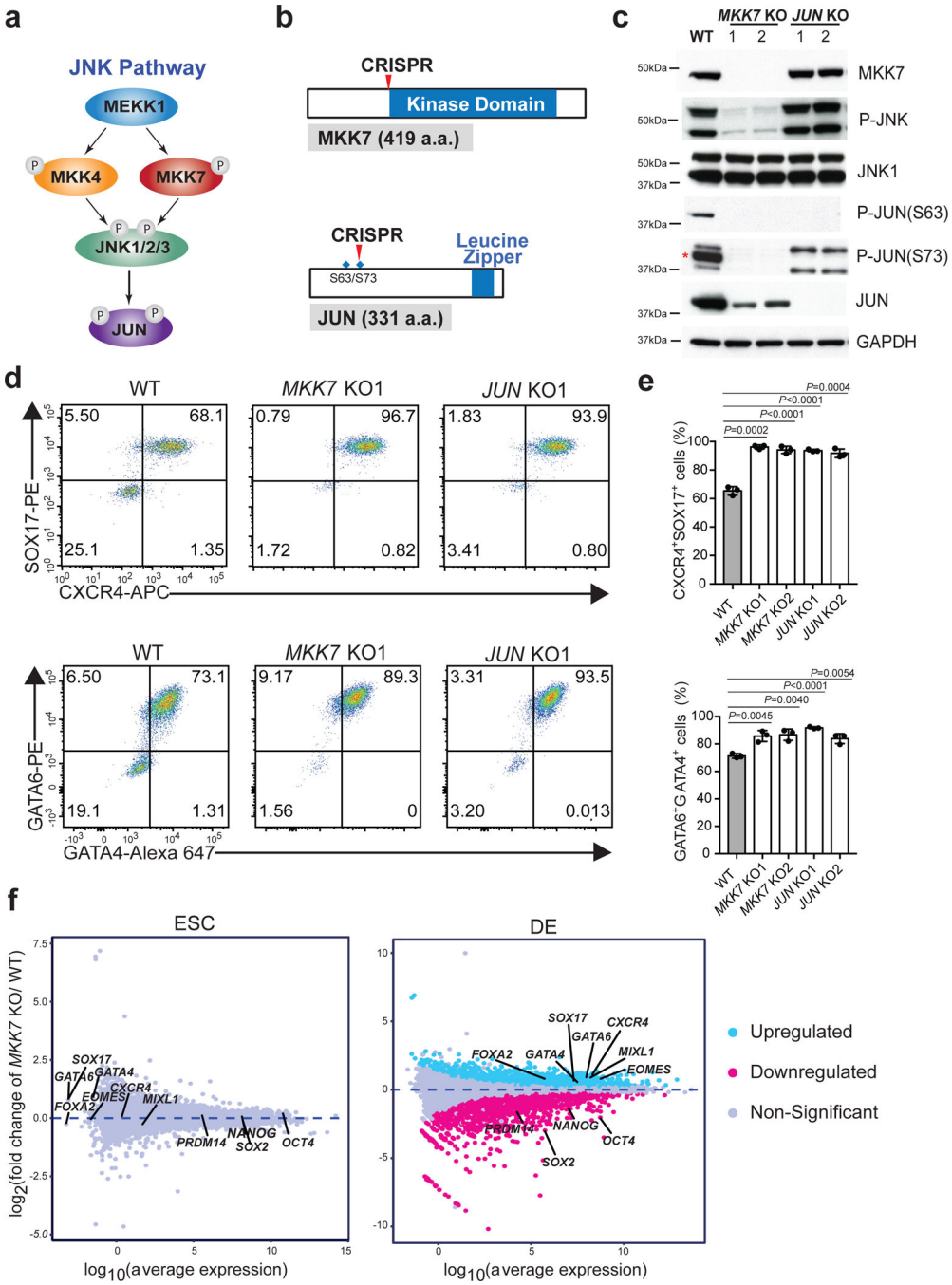


**Fig.2 |. Genetic inactivation of JNK pathway improves DE differentiation**

**a,** Illustration of the JNK pathway. **b,** MKK7 and JUN protein structures and 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 proteins in DE cells (Day 3). GAPDH was used as a loading control. * indicates the specific band for P-JUN (S73). Uncropped images are shown in Supplementary Figure 10. Representative images are from 2 independent experiments. **d,** Representative flow cytometry dot plots of DE D3 cells stained for CXCR4 and SOX17, GATA6 and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

GATA4 (H1 hESC line, WT vs *MKK7* and *JUN* KO). **e,** flow cytometry quantification of differentiation efficiency from **(d).** n = 3 independent experiments. Error bars indicate SD. Statistical analysis was performed by unpaired two-tailed Student's *t* test. Exact *P* values are shown. **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.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

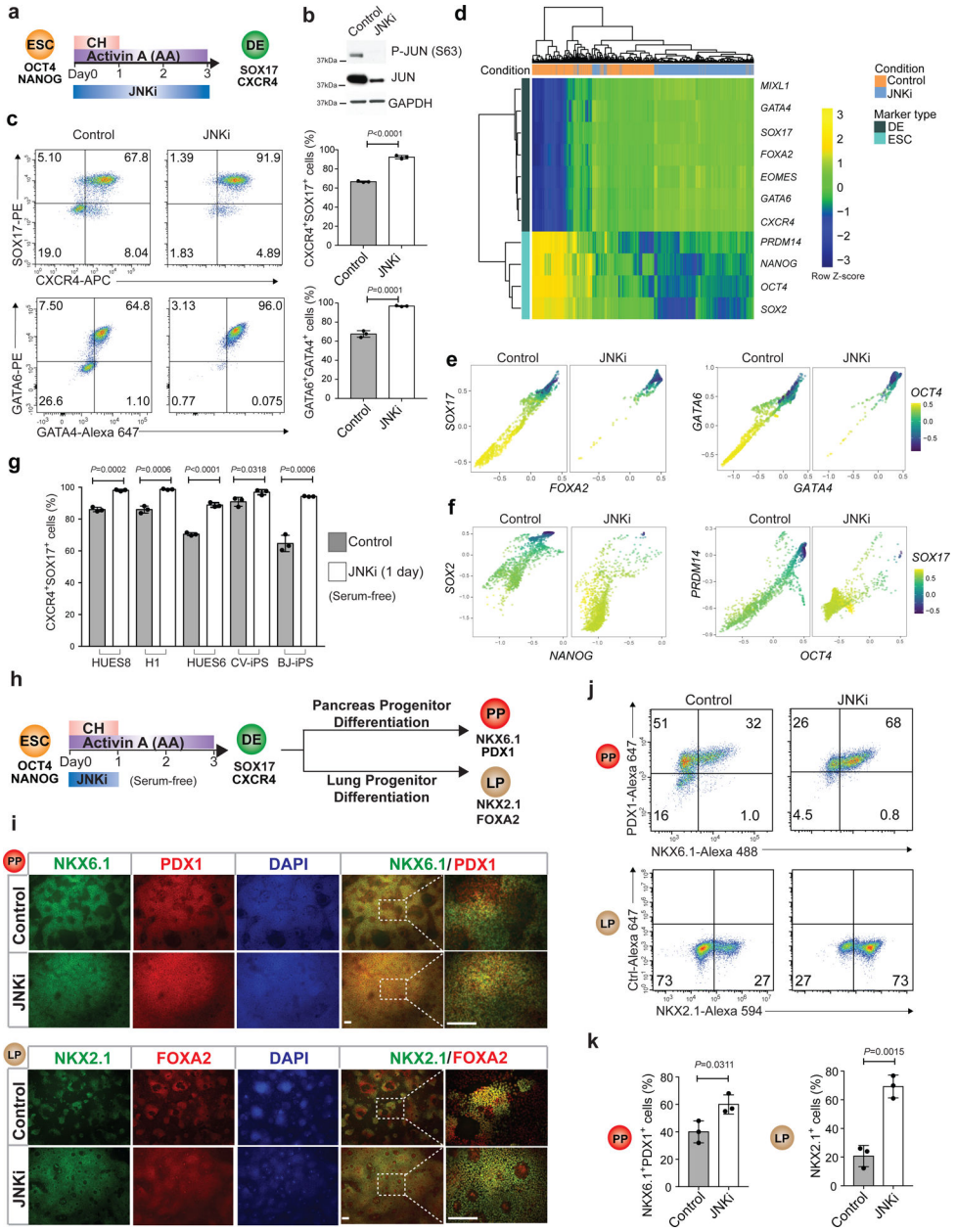


**Fig.3 |. Pharmacological JNK inhibition improves endoderm differentiation**

**a,** Illustration of JNK pathway inhibition during DE differentiation. **b**, Western blotting analysis for JUN and P-JUN (S63) expression of H1-derived DE D3 cells from control and JNKi treated condition. Uncropped images are shown in Supplementary Figure 11. Representative images are from 2 independent experiments. **c**, Representative flow cytometry 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. **d**, Hierarchical clustering of single cell expression values for known marker genes of DE and ESCs. 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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. **e-f**, Expression patterns of selected DE genes **(e)** and ESC genes **(f)** 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* **(e)** *or SOX17* **(f)** expression. **g**, Flow cytometry 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 hiPSC and BJ hiPSC lines using the serum free protocol. n = 3 independent experiments. **h**, 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. **i**, Immunostaining with PP markers PDX1 and NKX6.1 and LP markers NKX2.1 and FOXA2. Scale bar, 100 μm. Representative images are from 2 independent experiments. **j.** Representative flow cytometry dot plots for PP cells co-stained for NKX6.1 and PDX1, or LP cells stained for NKX2.1. **k,** Flow cytometry quantification of differentiation efficiency of PP or LP cells. n = 3 independent experiments. Error bars indicate SD. Statistical analysis was performed by unpaired two-tailed Student's *t* test. Exact *P* values are shown.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

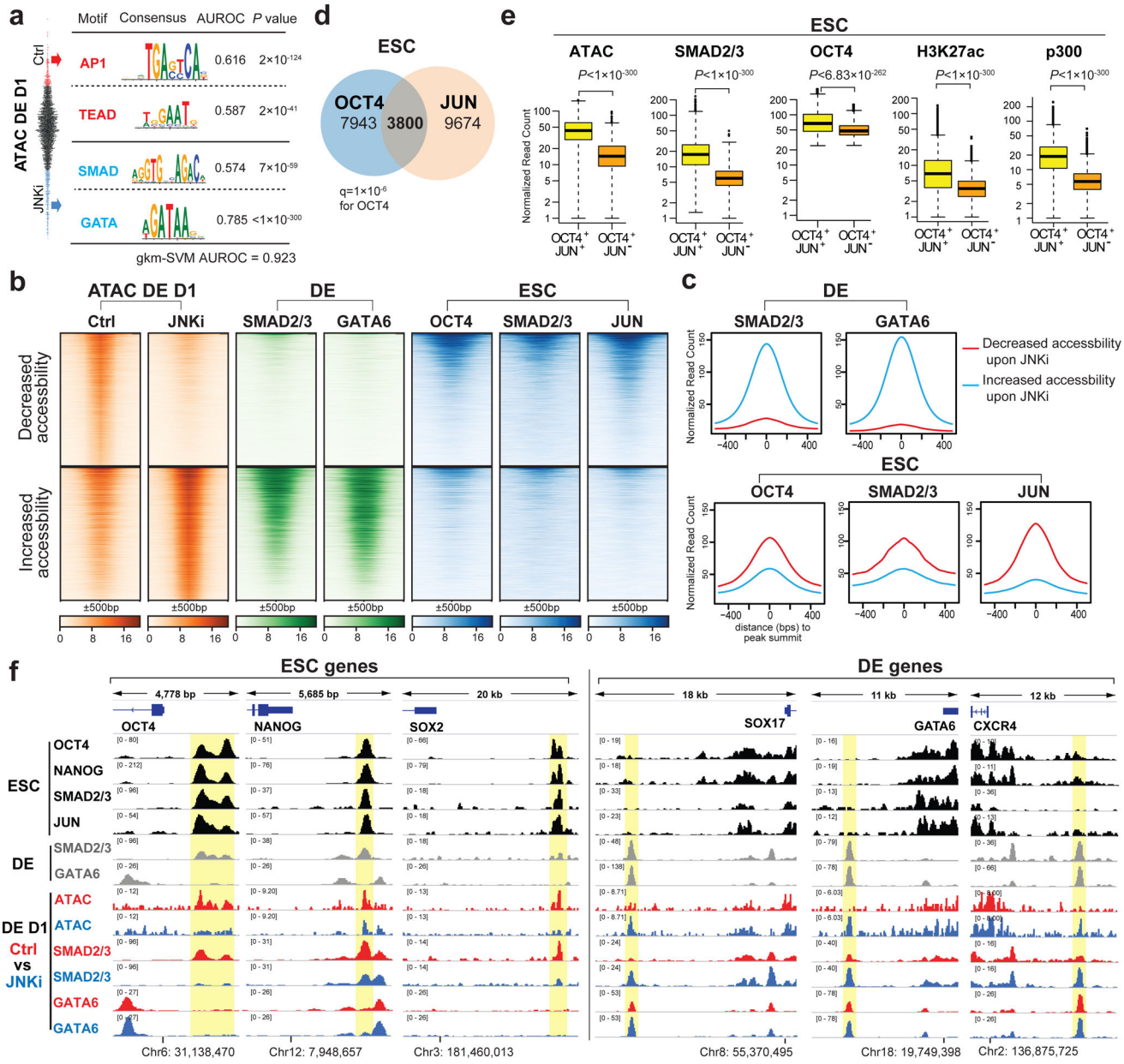


**Fig.4 |. JUN impedes chromatin landscape remodeling during ESC-DE differentiation**

**a,** Motif analysis for ATAC-seq peaks of DE D1 cells with decreased accessibility (ctrl) or increased accessibility (JNKi) from JNKi treatment compared to untreated control. Beeswarm plot shows change in ATAC-seq DE D1 signal Ctrl vs JNKi, motifs and gkm-SVM 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 ESCs (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 in ESCs by using a more

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

stringent peak calling criterion for OCT4 ($q = 1 \times 10^{-6}$) ($P < 1 \times 10^{-300}$). **e**, The boxplots show the average ATAC-seq, SMAD2/3, OCT4, H3K27ac (GSM733718) and p300 (GSM803542) ChIP-seq signals[48] over the $OCT4^{+}JUN^{+}$ and $OCT4^{+}JUN^{-}$ regions in ESCs. 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*). Genomic coordinate from GRCh37 (Human hg19) for each gene is labeled. Representative images are from 2 independent experiments.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

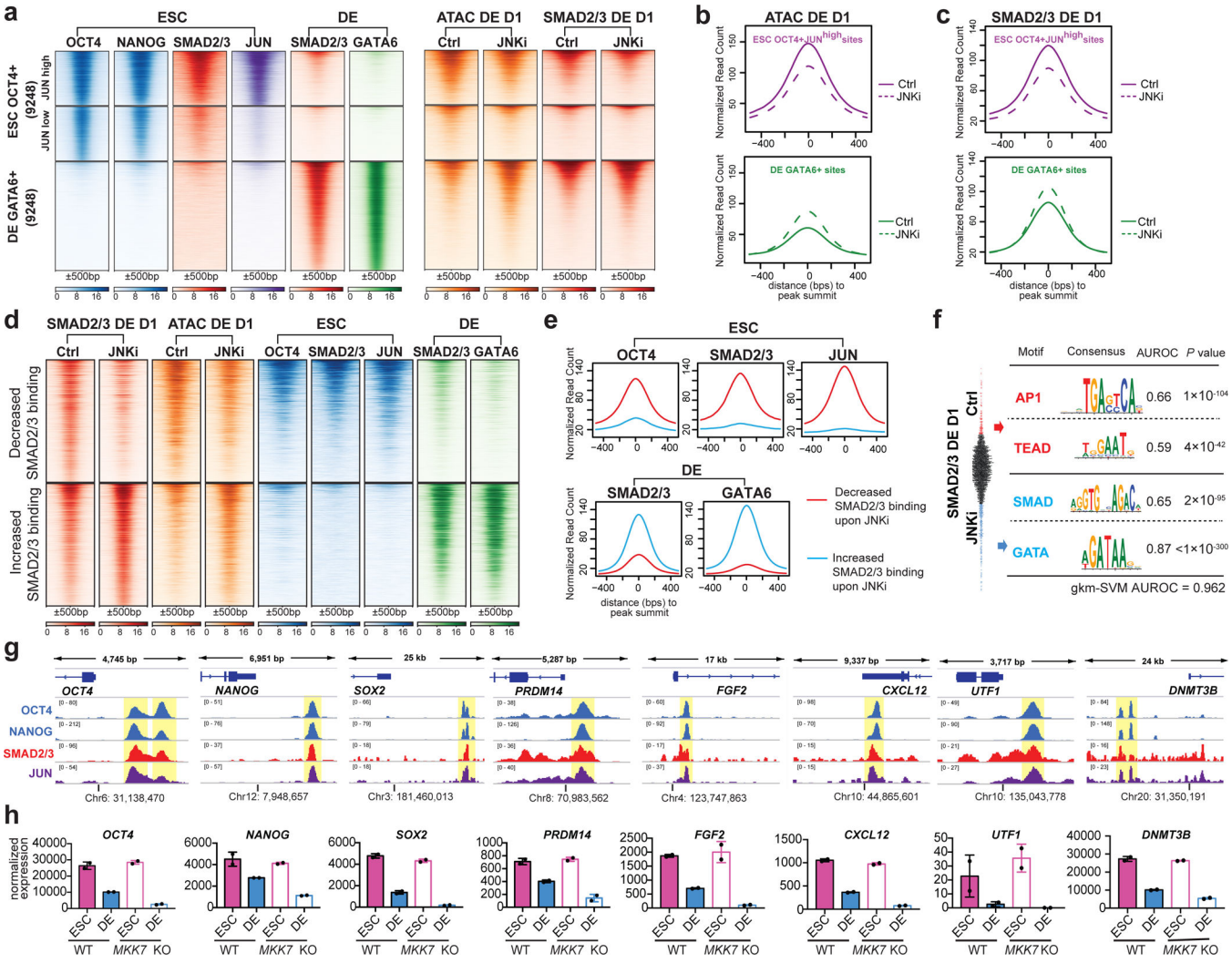


**Fig.5 |. JUN impedes SMAD2/3 reconfiguration during ESC-DE differentiation**

**a,** Heatmaps of ChIP-seq and ATAC-seq signals enrichment at top OCT4-bound enhancers in ESCs (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 in ESCs. 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,** 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 ESCs (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. **e,** Average ChIP-seq signals at regions of decreased (red) or increased (blue) SMAD2/3 binding upon JNKi treatment. **f,** Motif analysis for SMAD2/3 ChIP-seq peaks of DE D1 cells with decreased SMAD2/3 binding

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Ctrl) and increased SMAD2/3 binding (JNKi) from JNKi treatment compared to untreated control. Beeswarm plot shows change in SMAD2/3 binding signal Ctrl vs JNKi, motifs and gkm-SVM are trained on genomic intervals in the tails of this distribution. **g,** Representative ChIP-seq track images of OCT4, NANOG, SMAD2/3 and JUN binding at ESC gene enhancers in ESCs. Genomic coordinates from GRCh37 (Human hg19) for each gene are labeled. Representative images are from 2 independent experiments. **h**, RNA-seq analysis of the corresponding genes **(g)** in ESCs and DE cells from WT and *MKK7* KO cell lines. n = 2 independent experiments. Error bars indicate SD.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

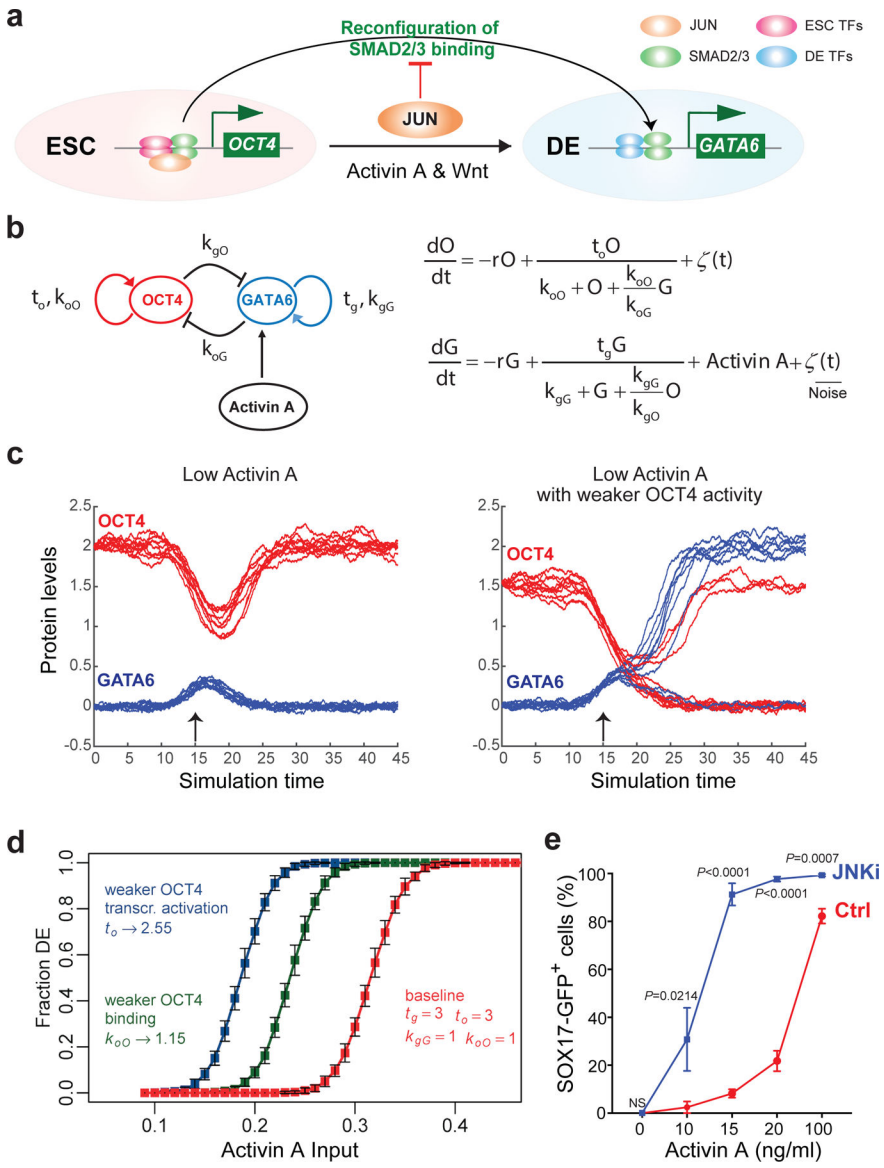


**Fig. 6 |. Modeling the effect of JNK inhibition on ESC-DE transition**

**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:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

weaker OCT4 binding at its enhancer. Blue line: weaker OCT4 transcriptional activation. 20 simulations. **e**, Flow cytometry quantification of differentiation efficiency based on the percentage of SOX17-GFP$^{+}$ DE cells at different Activin A doses. n = 3 independent experiments. Error bars indicate SD. Statistical analysis was performed by unpaired two-tailed Student's *t* test. Exact *P* values of comparison between JNKi vs Ctrl at each AA concentration are shown in the figure. NS, not significant ($P \geq 0.05$).