THE ROLES OF SOX2 AND R-SMADS IN HUMAN EMBRYONIC STEM CELL DIFFERENTIATION TOWARDS NEUROECTODERM

by

YI QI

(Under the Direction of Stephen Dalton)

ABSTRACT

Sox2 protein is necessary in maintaining the pluripotency state and self-renewal capacity of human embryonic stem cells; it also plays a role in human pluripotent stem cell's differentiation towards neuroectodermal lineages. The Smad proteins are involved in signaling pathways that impact pluripotent cell fates, with phosphorylated Smad2,3 involved in Activin/Nodal-mediated self-renewal maintenance and definitive endoderm differentiation and phosphorylated Smad1,5,8 involved in BMP-induced mesoderm differentiation, both forming complexes with Smad4 to exert their functions. In this study, I show that Sox2 and Smad complexes act in an antagonistic way in determining the stem cell fate by competitively binding to the transcriptionally active open chromatin regions and triggering the activation of signaling networks governing neuroectodermal or mesendodermal cell fates.

INDEX WORDS: Stem cells, pluripotency, neuroectoderm, R-Smads, Sox2, enhancers

THE ROLES OF SOX2 AND R-SMADS IN HUMAN EMBRYONIC STEM CELL DIFFERENTIATION TOWARDS NEUROECTODERM

by

YI QI

BS, The Hong Kong University of Science and Technology, Hong Kong China, 2015

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

MASTER OF SCIENCE

ATHENS, GEORGIA

2019

THE ROLES OF SOX2 AND R-SMADS IN HUMAN EMBRYONIC STEM CELL DIFFERENTIATION TOWARDS NEUROECTODERM

by

YI QI

Major Professor:	Stephen Dalton
Committee:	Nadja Zeltner
	Shaying Zhao

Electronic Version Approved:

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

DEDICATION

This thesis is dedicated to my parents. Without their love and support, this work would not have been accomplished.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Dr. Stephen Dalton for accepting me as a laboratory member and offering me the opportunities to perform graduate researches under his guidance. I am grateful to Dr. Amar Singh for the useful advice that he provided me with on the practical wet-lab experiments. I would like to thank my committee members, Dr. Nadja Zeltner and Dr. Shaying Zhao, for their help and support throughout the years of my graduate studies and their comments on my master thesis. I am also thankful to all the Dalton lab members for the technical and moral support they provided me with whenever I am in need. I would also like to specifically acknowledge Dr. Lance Wells for the help, support and guidance as the BCMB Graduate Coordinator.

Last but not least, I would like to thank my boyfriend and my family for their continuing love, care, patience, moral support, and encouragement throughout my graduate career. I probably would not have been able to make through the difficult days without their presence in my life.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

PLURIPOTENT STEM CELLS

Human embryonic stem cells are defined as pluripotent owing to their capability to propagate and self-renew indefinitely *in vitro* while being able to differentiate into the three germ layers, namely endoderm, mesoderm and ectoderm, and any type of tissues that are further derived from the germ layers under specified culture conditions (Figure 1.1) (De Los Angeles et al. 2015). Embryonic stem cells are derived from the inner cell mass of the human blastocyst, which raises ethical issues while considering them as a potential tool in regenerative medicine. In 2006, a new type of pluripotent stem cell called "induced pluripotent stem cell" (iPS cell) was generated by reprogramming differentiated mouse somatic cells via expressing exogenous pluripotency factors Oct4, Sox2, cMyc and Klf4 (Takahashi and Yamanaka, 2006). In 2007, human induced pluripotent stem cells were also produced via reprogramming adult human fibroblasts (Takahashi et al. 2007). The achievement of human induced pluripotent stem cell shed light on human stem cell application in regenerative medicine by making it possible to generate patient-specific stem cell lines by converting their somatic cells back to the pluripotency state and subsequently differentiating them into target cell or tissue types, without ethical concerns upon using human embryos as cell source.

SIGNALING PATHWAYS IN PLURIPOTENT STEM CELL AND GERM LAYER SPECIFICATION

The indefinite self-renewal and proliferation capacity as well as the potential of differentiating towards various somatic cell lineages as the characteristics of pluripotency state in ES cells are maintained through a tightly-regulated signaling network, with the transcription factors NANOG, octamer-binding transcription factor 4 (OCT4) and Sex determining region Y-box2 (SOX2) serving as the central participants. The three core pluripotency factors co-bind to regulatory DNA elements such as promoters and enhancers of genes active in ES cells or developmental processes, activating or repressing their expressions to maintain pluripotency in ES cells while they also bind to their own gene promoters, forming an interconnected autoregulatory and feedforward circuitry (Boyer et al. 2005). The pluripotency-impacting signaling pathways involving genes that interact with the core pluripotency factors include PI3K/Akt, Activin/NODAL, Wnt/Gsk3β and BMP pathways, with TGF-β/Activin A signaling and basic fibroblast growth factor (bFGF)-mediated repression of BMP signaling playing essential roles in human ES cell self-renewal (Singh et al. 2012; Vallier et al. 2009a; Davidson et al. 2012; Xu et al. 2005). These pathways regulate the levels of pluripotency factors Oct4, Nanog and Sox2 (ONS factors) through the downstream Smad proteins to maintain pluripotency in human ES cells, for example the TGF-β responsive Smad2,3 serving as the effector of Activin A/NODAL bind directly to the promoter of Nanog to sustain its activity in human ES cells (Xu et al. 2008). Whether ES cells remain pluripotent or start differentiation towards a certain germ layer depends on the levels of the growth factors, which are controlled by the complex crosstalk between these pathways. With PI3K/Akt being active, Erk and Wnt signaling pathways are suppressed and Activin A/NODAL maintains pluripotency by Smad2,3 to activate gene sets required for self-

renwal; when PI3K/Akt pathway is shut down, Erk and Wnt pathways are active and their effectors such as β-catenin can let Smad2,3 activate genes involved in mesendodermal differentiation (Singh et al. 2012).

In embryogenesis, the primitive streak (PS) forms starting the gastrulation where the pluripotent ES cells from the inner cell mass of the blastocyst ingress through PS during the epithelial to mesenchymal transition (EMT) process and are then differentiated into mesoderm and definitive endoderm, based on their spatial distribution determined by the site and sequential order of the epiblast cell ingression (Kinder et al. 1999). The ectoderm is derived from the epiblast cells that do not ingress through the PS (Lawson et al. 1991). The germ layer specification is controlled by the Transforming Growth Factor Beta (TGF- β) superfamily members, including Activin A, NODAL and bone morphogenetic proteins (BMPs). Activin/NODAL signaling, together with FGF and WNT pathways, drive the differentiation of human ES cell towards definitive endoderm lineages by activating the downstream signal transducers SMAD2/3 and β-catenin, with Activin/NODAL being necessary in that WNT and FGF pathways cannot induce mesendoderm specification in its absence (Vallier et al. 2009b). BMP4 can function to modulate mesoderm specification through phosphorylation of SMAD1/5/8 (Zhang et al. 2008). The receptor-regulated Smads (R-Smads) SMAD2/3 and SMAD1/5/8, when phosphorylated upon activation of Activin/NODAL or BMP signaling, interact with the co-SMAD4 and are translocated into the nucleus to bind to their target gene sequences and regulate their expressions (Figure 1.2) (Gaarenstroom and Hill, 2014). The differentiation towards neuroectoderm, on the other hand, can be achieved by blocking the BMP and Activin/NODAL pathways via dual inhibition of SMAD signaling using NOGGIN/LDN193189 and SB431542 *in vitro* (Chambers et al. 2009; Kriks et al. 2011).

It has been suggested that the ONS factors also participate in the lineage specification of the ES cells. NANOG complexes with SMAD2/3 to initiate EOMES expression and definitive endoderm differentiation, which is inhibited by SOX2 and OCT4 expression (Teo et al. 2011). ChIP-seq data revealed that OCT4 binding level at regulatory DNA elements was increased in mesoderm compared to ES cells (Tsankov et al. 2015; Thomson et al.2011), whereas SOX2 protein level was induced in neuroectoderm (Thomson et al.2011). OCT4 has been shown to be necessary for mesendodermal differentiation and has a role in inhibition of differentiation towards neuroectoderm (Thomson et al.2011), while SOX2 plays an asymmetric role with OCT4 in lineage specification that will be discussed in the next part.

ROLES OF SOX2 PROTEIN IN PLURIPOTENCY MAINTENANCE AND ECTODERM SPECIFICATION

SOX2 is a member of the SOXB1 family of transcription factors with a highly conserved HMG (high-mobility group) box DNA binding domain (Kamachi et al. 2000). It plays an important role in the maintenance of pluripotency state in ES cells, and deficiency in SOX2 leads to mouse embryo lethality (Avilion et al. 2003). In pluripotent stem cells, OCT4 preferentially forms heterodimers with SOX2 to bind regulatory DNA elements to activate pluripotency genes in an autoregulatory manner while repressing differentiation-promoting genes (Figure 1.3) (Boyer et al. 2005; Mistri et al. 2015). Knockdown of SOX2 and its functional equivalent SOX3 in human embryonic stem cells leads to differentiation with the expression of mesendodermal markers GATA6, GATA4, FOXA2 and SOX17 (Wang et al. 2012), suggesting that SOX2 is necessary in repressing embryonic stem cells differentiation towards mesendodermal lineages. SOX2 is also involved in reprogramming somatic cells to iPS cells through exogenous co-expression with

Oct4, cMyc and Klf4 (Takahashi and Yamanaka, 2006), indicating its critical role in pluripotency establishment.

SOX2 also promotes neural ectoderm differentiation. During gastrulation, SOX2 expression is mainly restricted to neuroectoderm (Avilion et al. 2003), with ChIP-qPCR results showing that SOX2 binding is enriched at the neuroectoderm regulatory regions (Thomson et al. 2011). In mouse ES cells, modest increase (two-fold or less) in SOX2 expression results in cells exiting pluripotency state and differentiating into multiple lineages including neuroectoderm (Kopp et al. 2008). Constitutive SOX2 expression leads to maintenance of neural progenitor state and inhibition of differentiation towards downstream neural lineages (Graham et al. 2003). In human ES cells, SOX2 overexpression suppresses definitive endoderm differentiation and enhances neuroectoderm differentiation under the specified differentiation-inducing culture conditions *in vitro* (Wang et al. 2012). Genome-wide analysis shows that SOX2 binds to different sets of gene promoters and enhancers in ES cells and neuroectoderm complexing with distinct POU transcription factors, with Oct4 in ESCs and Brn2 in neuroectoderm (Figure 1.3) (Lotado et al. 2013), thereby regulating cell type-specific gene expressions and functioning in cell fate determination. SOX2 also complexes with other transcription factors to regulate downstream lineage-specific development, including Pax6 in lens development initiation (Aota et al. 2003) and Oct1 in olfactory placode development (Donner et al. 2007).

REGULATORY DNA ELEMENTS AND EPIGENETIC HISTONE MODIFICATIONS

The DNA-level regulation of gene expression is achieved by cis-regulatory DNA elements consisting of promoters and enhancers. Eukaryotic promoters are non-coding DNA sequences consisting of TATA boxes and transcription factor-specific binding motifs around 35 base pairs

upstream of the transcription start sites (TSS) of the target genes (Butler and Kadonaga, 2002). Gene-specific transcription factors bind to the promoter sequences to form transcriptional machinery that activates or represses the expression of the gene. Enhancers are DNA sequences that contain motifs for specific transcription factor binding to recruit co-activators or co-repressors and form loops to interact with the transcriptional machinery at promoters to regulate gene expression (Deng et al. 2012). Enhancers can be located hundreds of kilobases from its target gene and function independently of orientation and distance. Active enhancers and promoters are exposed DNA sequences that are in an open chromatin state. (Shlyueva et al. 2014). Therefore, identification of the open chromatin region can be utilized to determine the regulatory DNA elements. Genome-wide open chromatin map was generated by high-throughput sequencing and tiled microarray following enzymatic cleavage of accessible DNA by using micrococcal nuclease (MNase) in *Saccharomyces cerevisiae* and DNase I in CD4+ T cells, respectively (Yuan et al. 2005; Boyle et al. 2008). In 2013, a new technology termed assay for transposase-accessible chromatin using sequencing (ATAC-seq) involving the use of Tn5 transposase to integrate into accessible chromatin regions was introduced and it could generate comparable signal-to-noise ratio to DNase-seq while using ~3-5 orders of magnitude fewer cells and shorter time (Figure 1.4) (Buenrostro et al. 2013).

The term “epigenetics” refers to the study of the heritable changes in gene function that do not involve changes in DNA sequence (Dupont et al. 2009). Post-translational covalent histone modification is an important component of epigenetic control in eukaryotic cells. The basic DNA packaging unit, nucleosome, comprises of eight subunits representing different histone variants, namely H2A, H2B, H3 and H4 (Felsenfeld and Groudine, 2003). Acetylation of lysine residues on H3 and H4 are associated with gene activation and open chromatin structure while

methylation can either be related to gene activation with accessible chromatin (H3K4, H3K36 and H3K79) or repression with packed chromatin (H3K9, H3K27, and H4K20) (Dupont et al. 2009; Sims et al. 2003). Active enhancers are characterized by nucleosome depletion and are flanked by nucleosomes with the typical post-translational histone modifications H3 lysine 27 acetylation (H3K27ac) and H3 lysine 4 monomethylation (H3K4me1) (Figure 1.5), allowing the DNA to be exposed for interaction with transcription factors (Creyghton et al. 2010). In contrast, inactive enhancers are enriched in the repressive H3 lysine 27 trimethylation (H3K27me3) mark (Figure 1.5) (Shlyueva et al. 2014).

In ES cells, the enhancers for developmental genes are in a "poised" state bearing H3K27me3 and H3K4me1 (Rada-Iglesias et al. 2011). These genes are silenced in ES cells and are induced to express when the lineage-specific differentiation involving them is triggered, and the corresponding chromatin landscape of their enhancers switches from the "poised" state (H3K27me3 + H3K4me1) to the active state (H3K27ac + H3K4me1), thus activating the enhancers to allow binding of other developmental transcription factors (Gifford et al. 2013).

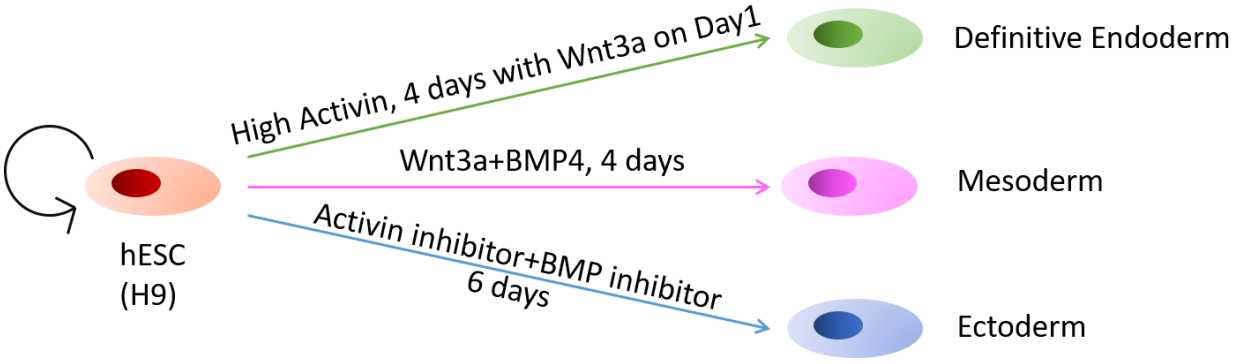


Figure 1.1 | Human Embryonic Stem Cells can be Differentiated towards the Three Germ Layers Human embryonic stem cells (hESCs) can propagate and self-renew indefinitely when cultured *in vitro*. They can be induced to differentiate towards endoderm, mesoderm and endoderm under specified *in vitro* culture conditions.

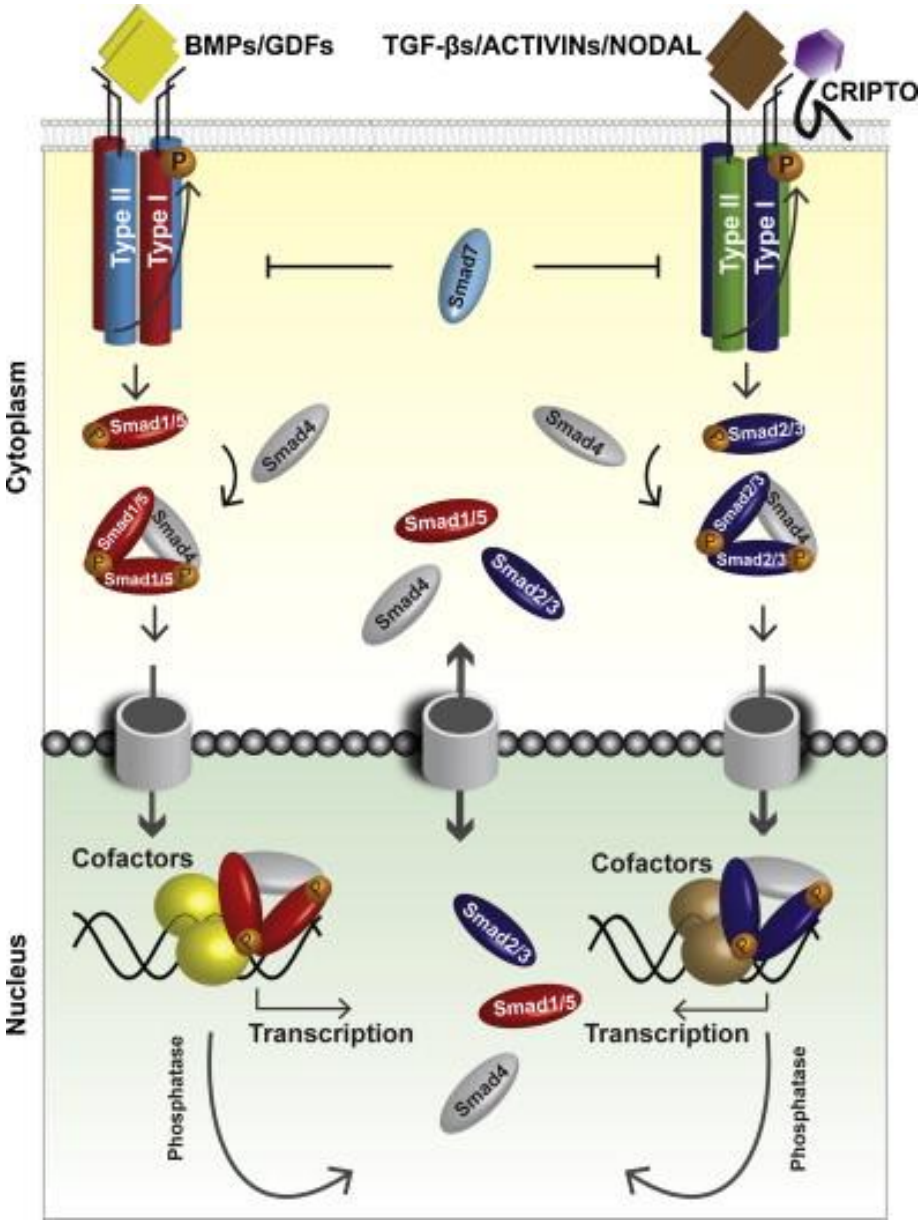


Figure 1.2 | Overview of TGFβ/Smad signaling pathways TGFβ ligands including BMPs and Activins bind to cell-surface receptors. The phosphorylated receptors phosphorylate R-Smads, allowing them to complex with the co-Smad Smad4 to be shuttled into and accumulated in the nucleus for regulation of gene transcription through interaction with other cofactors. Adapted from (Gaarenstroom and Hill, 2014).

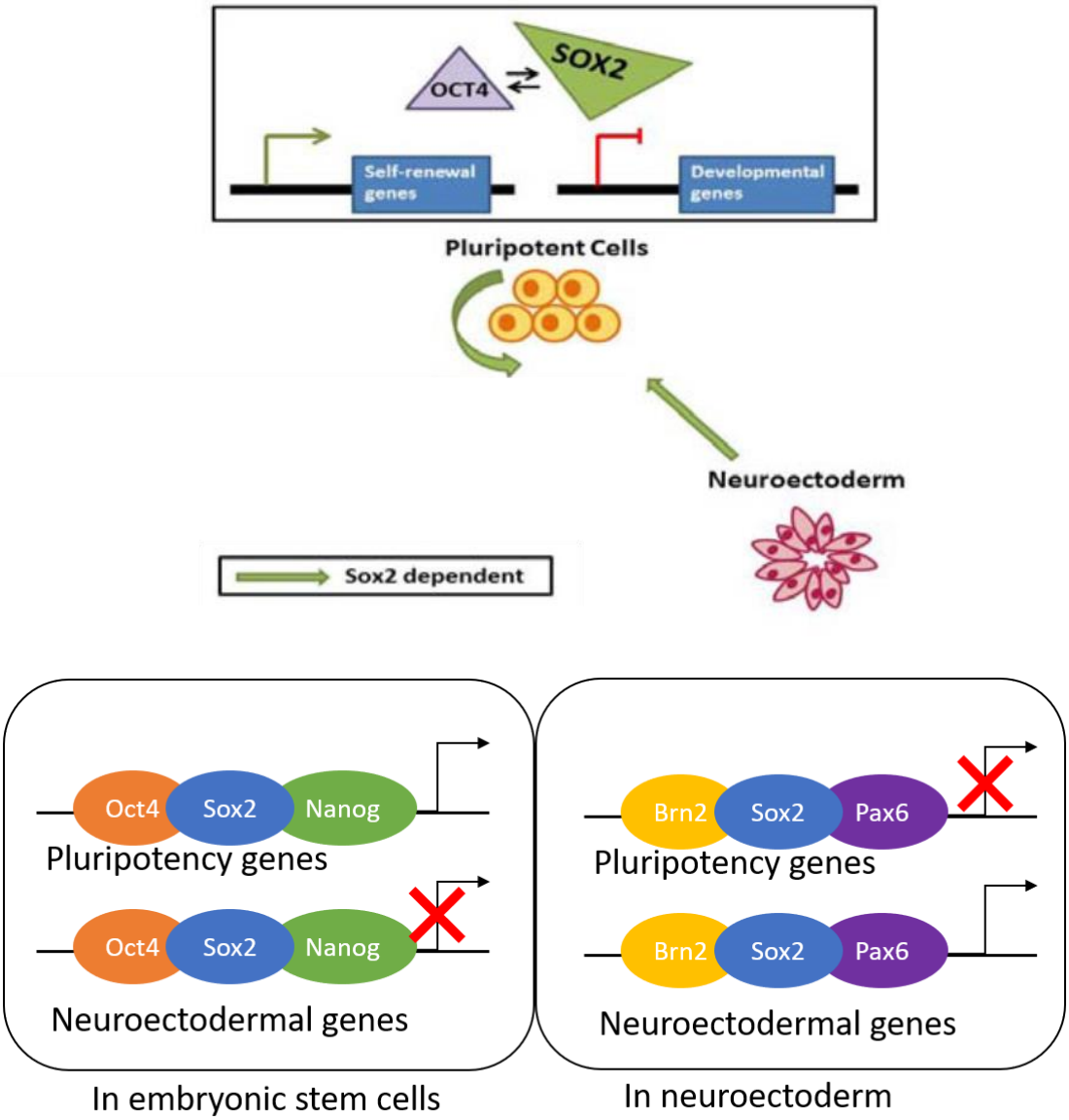


Figure 1.3 | Sox2 activates and represses different sets of genes in embryonic stem cells and neuroectoderm Sox2 is required in pluripotency maintenance and neuroectoderm development. In embryonic stem cells, Sox2 forms a heterodimer with Oct4 to activate self-renewal genes and repress neuroectoderm differentiation-related genes. During neuroectoderm differentiation, Sox2 interacts with neural lineage transcription factors Pax6 and Brn2 to activate neuroectodermal gene expression and repress pluripotency genes. Adapted from (Chanoumidou et al. 2017)

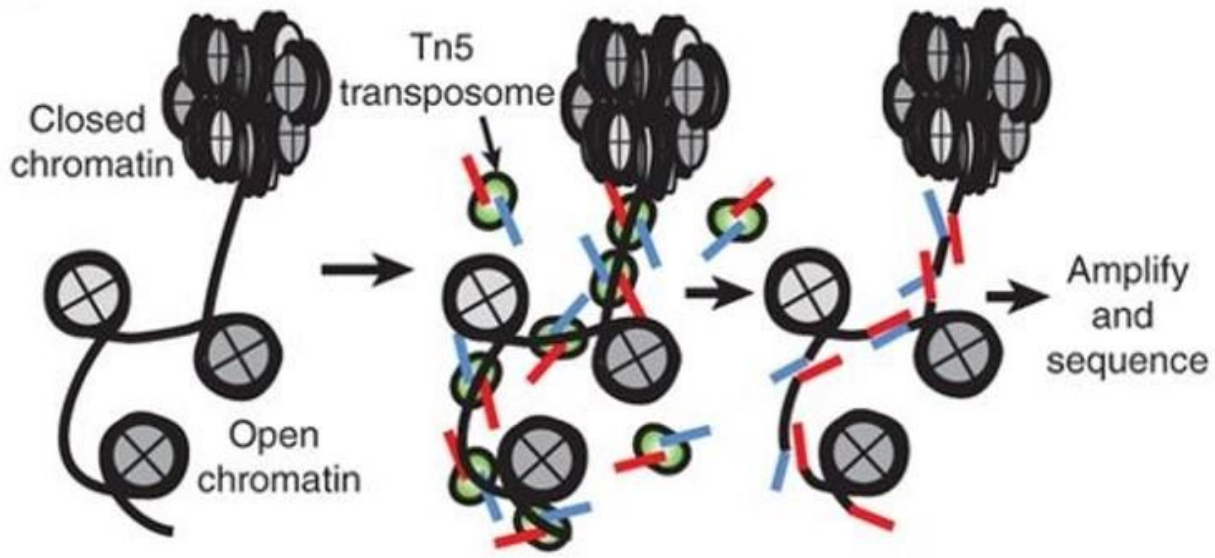


Figure 1.4 | Working Principle of ATAC-seq The hyperactive Tn5 transposase has its adaptor payload integrated into the DNA sequences in the accessible chromatin regions and cleaves DNA into short fragments. The steric hindrance in the less accessible heterochromatin prevents Tn5 transposase integration. The DNA fragments are then amplified and used for high-throughput sequencing. Adapted from (Buenrostro et al. 2013).

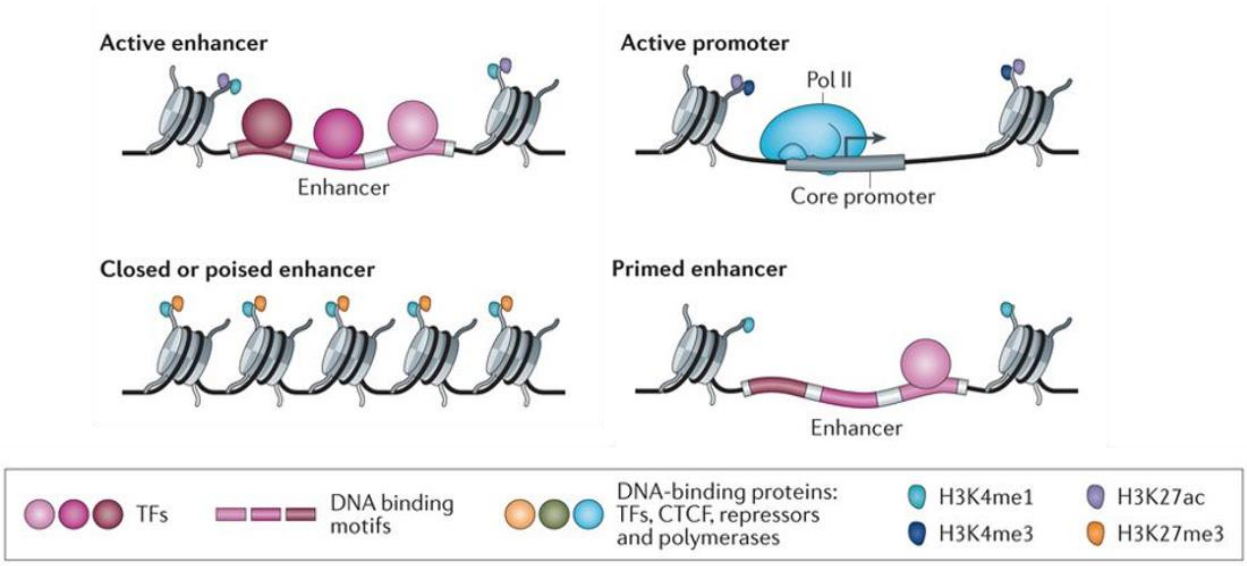


Figure 1.5 | Histone Modification Marks and Chromatin Accessibility at Cis-Regulatory DNA Elements Accessible regulatory elements (active enhancers or promoters) are free of nucleosomes and allow binding of transcription factor or other DNA-binding proteins. Active enhancers are characterized by histone H3 lysine 27 acetylation (H3K27ac) marks and H3 lysine 4 monomethylation (H3K4me1) marks on the flanking nucleosomes. Active promoters are characterized by flanking nucleosomes with H3K27ac and H3K4me1 marks. Heterochromatin contains closed or poised enhancers that are inaccessible for protein binding. The closed or poised enhancers are characterized with densely packed nucleosomes bearing H3K27me3 and H3K4me1 modifications. Adapted from (Shlyueva et al. 2014).

REFERENCES

1. Aota, S. I., Nakajima, N., Sakamoto, R., Watanabe, S., Ibaraki, N., & Okazaki, K. (2003). Pax6 autoregulation mediated by direct interaction of Pax6 protein with the head surface ectoderm-specific enhancer of the mouse Pax6 gene. Developmental biology, 257(1), 1-13.
2. Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development, 17(1), 126-140.
3. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., ... & Gifford, D. K. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. cell, 122(6), 947-956.
4. Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., ... & Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell, 132(2), 311-322.
5. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods, 10(12), 1213.
6. Butler, J. E., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & development, 16(20), 2583-2592.

7. Chambers, S. M., Fasano, C. A., Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology, 27(3), 275.
8. Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., ... & Boyer, L. A. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences, 107(50), 21931-21936.
9. Davidson, K. C., Adams, A. M., Goodson, J. M., McDonald, C. E., Potter, J. C., Berndt, J. D., ... & Moon, R. T. (2012). Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. Proceedings of the National Academy of Sciences, 201118777.
10. De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., ... & Lensch, M. W. (2015). Hallmarks of pluripotency. Nature, 525(7570), 469.
11. Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., ... & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell, 149(6), 1233-1244.
12. Donner, A. L., Episkopou, V., & Maas, R. L. (2007). Sox2 and Pou2f1 interact to control lens and olfactory placode development. Developmental biology, 303(2), 784-799.
13. Dupont, C., Armant, D. R., & Brenner, C. A. (2009, September). Epigenetics: definition, mechanisms and clinical perspective. In Seminars in reproductive medicine (Vol. 27, No. 5, p. 351). NIH Public Access.
14. Felsenfeld, G., & Groudine, M. (2003). Controlling the double helix. Nature, 421(6921), 448.

15. Gaarenstroom, T., & Hill, C. S. (2014, August). TGF-β signaling to chromatin: how Smads regulate transcription during self-renewal and differentiation. In Seminars in cell & developmental biology (Vol. 32, pp. 107-118). Academic Press.
16. Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., ... & Zhang, X. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell, 153(5), 1149-1163.
17. Graham, V., Khudyakov, J., Ellis, P., & Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. Neuron, 39(5), 749-765.
18. Kamachi, Y., Uchikawa, M., & Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends in Genetics, 16(4), 182-187.
19. Kinder, S. J., Tsang, T. E., Quinlan, G. A., Hadjantonakis, A. K., Nagy, A., & Tam, P. P. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. Development, 126(21), 4691-4701.
20. Kopp, J. L., Ormsbee, B. D., Desler, M., & Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem cells, 26(4), 903-911.
21. Kriks, S., Shim, J. W., Piao, J., Ganat, Y. M., Wakeman, D. R., Xie, Z., ... & Yang, L. (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature, 480(7378), 547.
22. Lawson, K. A., Meneses, J. J., & Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development, 113(3), 891-911.

23. Lodato, M. A., Ng, C. W., Wamstad, J. A., Cheng, A. W., Thai, K. K., Fraenkel, E., ... & Boyer, L. A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS genetics, 9(2), e1003288.
24. Mistri, T. K., Devasia, A. G., Chu, L. T., Ng, W. P., Halbritter, F., Colby, D., ... & Wohland, T. (2015). Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. EMBO reports, e201540467.
25. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature, 470(7333), 279.
26. Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nature Reviews Genetics, 15(4), 272.
27. Sims III, R. J., Nishioka, K., & Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. TRENDS in Genetics, 19(11), 629-639.
28. Singh, A. M., Reynolds, D., Cliff, T., Ohtsuka, S., Mattheyses, A. L., Sun, Y., ... & Dalton, S. (2012). Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance between self-renewal and differentiation. Cell stem cell, 10(3), 312-326.
29. Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, *126*(4), 663-676.
30. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, *131*(5), 861-872.

31. Teo, A. K. K., Arnold, S. J., Trotter, M. W., Brown, S., Ang, L. T., Chng, Z., ... & Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes & development.

32. Thomson, M., Liu, S. J., Zou, L. N., Smith, Z., Meissner, A., & Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell, 145(6), 875-889.

33. Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., ... & Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. Nature, 518(7539), 344.

34. Wang, Z., Oron, E., Nelson, B., Razis, S., & Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell stem cell, 10(4), 440-454.

35. Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L. E., ... & Brons, G. (2009a). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development, 136(8), 1339-1349.

36. Vallier, L., Touboul, T., Chng, Z., Brimpari, M., Hannan, N., Millan, E., ... & Pedersen, R. A. (2009b). Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. PloS one, 4(6), e6082.

37. Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T., & Thomson, J. A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nature methods, 2(3), 185.

38. Xu, R. H., Sampsell-Barron, T. L., Gu, F., Root, S., Peck, R. M., Pan, G., ... & Thomson, J. A. (2008). NANOG is a direct target of TGFβ/activin-mediated SMAD signaling in human ESCs. Cell stem cell, 3(2), 196-206.

39. Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science, 309(5734), 626-630.

40. Zhang, P., Li, J., Tan, Z., Wang, C., Liu, T., Chen, L., ... & Ding, M. (2008). Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells. Blood, 111(4), 1933-1941.

CHAPTER 2

SMAD SIGNALING SUPPRESSES NEUROECTODERM DIFFERENTIATION BY INHIBITING SOX2 RELOCATION

INTRODUCTION

Human embryonic stem cells (hESCs) have the capacity to self-renew or differentiate towards mesoderm, endoderm and ectoderm *in vitro* by responding to the different culture environment through several signaling pathways. Transforming Growth Factor Beta (TGF- β) signaling, as has been mentioned in Chapter 1, is one of the signaling pathways that are involved in pluripotency maintenance and germ layer specifications. The intracellular TGF- β signaling cascade is activated upon binding of TGF- β factors, such as BMP, Activin and Nodal, to the cell-surface TGF- β Type I and Type II serine/threonine kinase transmembrane receptors (Figure 1.2, Chapter 1). Type I receptor is activated upon heterodimer formation with and trans-phosphorylation at multiple threonine and serine residues by Type II receptor upon ligand binding (Shi and Massagué, 2003). The phosphorylated Type I receptor exhibits high affinity for R-Smads (Smad1/2/3/5/8) binding and specifically phosphorylates R-Smads at the serine residues of the C-terminal MH2 domain (Huse et al. 2001). The phosphorylated R-Smads, after heterodimerization with the co-Smad (Smad4) and translocation into the nucleus, interact with other transcription factors at regulatory DNA elements to regulate target gene expressions. Genome-wide ChIP-seq analysis revealed that Smad3 co-occupied enhancer sites with Oct4 and Nanog in hESC, while it co-occupied sites of different genes with MyoD in myotubes (Mullen et

al. 2011). Smad1 was also shown to bind to regulatory DNA with Gata2 hematopoietic cells and interacts with C/EBPα in erythroid cells (Trompouki et al. 2011), indicating that Smads participate in gene regulation by interaction with transcription factors at regulatory DNA sites in a cell type-specific manner.

Genome-wide histone H3 ChIP-analysis showed that the regulatory elements bound by Smads were mostly depleted of nucleosomes (Mullen et al. 2011), suggesting that these are accessible DNA sequences in an open chromatin state. Active regulatory elements in open chromatin state are also characterized by the histone H3 lysine 27 acetylation (H3K27ac) level at the corresponding DNA sequences. ChIP-seq data of H3K27ac in hESCs and the three germ layers showed distinct signal profiles (Tsankov et al. 2015), indicating the occurrence of chromatin remodeling during lineage specification. It has been established that chromatin remodeling plays a critical role in differentiation, with the recruitment of different transcription factor sets to distinct accessible regulatory DNA sequences during lineage specification (Voss and Hager, 2014; Du et al. 2017).

Sox2 is involved in pluripotency maintenance and neuroectoderm induction. It has been shown to retain high protein expression levels in hESC and neuroectoderm and bind to regulatory elements of distinct gene sets while interacting with different transcription factors in the two cell types, respectively (Lotado et al. 2013; Du et al. 2017). Meanwhile, neuroectoderm can be differentiated *in vitro* from hESCs by dual-Smad inhibition, i.e. inhibition of BMP/Smad1/5 signaling using LDN193189 and inhibition of Activin/Smad2/3 signaling using SB431542 (Fathi et al. 2015), which implies removal of R-Smads from the regulatory DNA elements for pluripotency maintenance. The mechanism underlying the binding dynamics of Sox2 during hESC differentiation to neuroectoderm still remains to be elucidated, and whether there is a

correlation between Smad depletion and Sox2 relocation is still unclear. This study looks into the binding patterns of R-Smads and Sox2 at the active regulatory elements in lineage-specific open chromatin regions in neuroectoderm and hESCs, particularly focusing on the potential interaction between these two transcription factors and the impact on hESC-derived neuroectoderm differentiation.

RESULTS

Open Chromatin Profile and Transcription Factor Binding Pattern in Human Embryonic Stem Cells and the Three Germ Layers

To gain insight into the distribution of open chromatin sites in human embryonic stem cells (hESCs), mesoderm, endoderm and neuroectoderm, we differentiated hESCs *in vitro* towards the three cell lineages and performed ATAC-seq analyses on the harvested cells. Open chromatin regions were indicated by the pooled ATAC-seq peak signals of hESC, definitive endoderm, mesoderm and neuroectoderm. K-means clustering algorithm was adopted for categorizing the ATAC-seq peaks corresponding to the cell-type specific open chromatin regions, with Group 7 on the bottom representing the neuroectoderm-specific sites (Figure 2.1.1). The neuroectoderm-specific open chromatin regions were defined as the increased ATAC-seq peak signals in neuroectoderm compared to other cell types. To find out whether the neuroectoderm-specific open chromatin profile was related to the transcription pattern, we performed RNA-seq on hESCs and neuroectoderm. The genes with a more than 2-fold increase in transcript level in neuroectoderm compared to hESCs were categorized as the up-regulated genes in neuroectoderm differentiation, and those with a more than 2-fold decrease in transcript level were categorized as the down-regulated genes. Among the 2315 up-regulated genes, 882 (38.10%) had sites with

increased ATAC-seq peak signals, and the percentage was higher than the 24.20% in the genes with stable expression levels in neuroectoderm differentiation (2048 out of 8769) and the 19.95% in the down-regulated genes (804 out of 4030), indicating that open chromatin regions were more associated with the actively expressed genes. In neuroectoderm, sites with increased ATAC-seq signals compared to hESC could be found in the lineage marker genes PAX6, SOX1 and SOX2, all of which were up-regulated after neuroectoderm differentiation from hESCs (Figure 2.1.2), suggesting that these genes could have gone through chromatin state change and become more accessible for transcript factor binding during the differentiation process.

We performed ChIP-seq to study the transcription factor binding pattern in hESCs and neuroectoderm at the regions with lineage-specific ATAC-seq signals. Smad2/3 and Sox2 were involved in the study, and H3K27ac ChIP-seq from public database was used to identify whether the distant open chromatin regions were associated with the lineage-specific active enhancers that participated in gene regulation. Since Activin A/Smad2/3 signaling was required in pluripotency maintenance in hESC and was inhibited in the *in vitro* differentiation of neuroectoderm, we expected that Smad2/3 binding should be enriched at hESC-specific open chromatin regions. On the other hand, Sox2 had a role in maintaining pluripotency in hESCs and was also essential in neuroectoderm differentiation. Therefore, Sox2 binding was expected to be enriched at the lineage-specific open chromatin regions in both hESCs and neuroectoderm. ChIP-seq profile of H3K27ac at the neuroectoderm-specific open chromatin regions in genes up-regulated in neuroectoderm showed strongly elevated signal in neuroectoderm compared to hESCs, indicating that these regions represented the regulatory DNA elements that were activated in neuroectoderm (Figure 2.2.1A). Similarly, the H3K27ac ChIP-seq pattern at the hESC-specific open chromatin regions in genes up-regulated in hESCs suggested that these

regions corresponded to the hESC-specific regulatory DNA elements (Figure 2.2.1B). At the neuroectoderm-specific regulatory regions, Smad2/3 showed weak binding signals in hESCs and no signal in neuroectoderm, while Sox2 showed increased binding signals at the distal elements (enhancers) (Figure 2.2.1A). Meanwhile, at the hESC-specific regulatory regions, both Smad2/3 and Sox2 showed stronger binding signals in hESCs at the enhancer sites (Figure 2.2.1B). These results suggested that Smad2/3 bound to DNA regulatory elements in hESCs but not in neuroectoderm, while Sox2 binding was enriched at hESC-specific enhancers in hESCs and was enriched at neuroectoderm-specific enhancers in neuroectoderm. Tracks in Integrative Genomics Viewer (IGV) displaying the ChIP-seq signals at the neuroectoderm marker genes PAX6, SOX1, and SOX2 also showed enhanced Sox2 binding and reduced Smad2/3 binding at the neuroectoderm-specific active enhancers, where ATAC-seq signals were up-regulated in neuroectoderm compared to hESCs (Figure 2.2.2), supporting the previous results.

Phosphorylated Smad2/3 Compete with Sox2 to Suppress Neuroectoderm Differentiation

Inhibition of Smad signaling has been shown to disrupt the pluripotency state of hESCs and lead to neuroectodermal lineage differentiation (Fathi et al. 2015). The binding enrichment profile of Smad2/3 and Sox2 proteins at the hESC-specific and the neuroectoderm-specific active enhancers suggested a competitive binding pattern of the two proteins in hESCs and neuroectoderm, which has not been addressed in previous studies. To investigate whether phosphorylated Smad2/3 competes with Sox2 to bind at neuroectoderm enhancers and suppress neuroectoderm differentiation, we differentiated hESCs towards neuroectoderm using SB431542 and LDN193189 to inhibit Activin/Smad2/3 and BMP/Smad1/5 signaling. On Day 2 and Day 4 of the differentiation, respectively, we removed SB431542 and LDN193189 from the culture

media and added BMP4 and Activin to restore the Smad signaling pathways (Figure 2.3.1). The cells were harvested on Day 6 of differentiation for RT-qPCR and ChIP-qPCR assays to measure the neuroectodermal marker gene expression levels and binding of Sox2 or Smad2/3 at the neuroectodermal enhancers. Cells with Smad signaling restored on Day 2 and Day 4 displayed morphological difference from the Day 6 neuroectoderm cells differentiated from hESCs in vitro adopting the dual-Smad inhibition method (Figure 2.3.2). RT-qPCR results showed that the expression levels of PAX6, SOX1 and SOX2 in cells with Smad signaling restored on Day 2 and Day 4 were markedly reduced compared to the Day 6 neuroectoderm, with the cells having Smad signaling restored on Day 2 demonstrating more decreased expression than those having Smad restoration on Day 4 (Figure 2.3.3A), suggesting that neuroectoderm differentiation was suppressed upon restoration of Smad signaling. ChIP-qPCR tested Smad2/3 and Sox2 binding at a 140bp sequence of the active enhancer sites that were also neuroectoderm-specific open chromatin regions of the three neuroectoderm marker genes, PAX6, SOX1 and SOX2. Sox2 binding increased after hESCs were induced to differentiate towards neuroectoderm for 6 days where Smad signaling was inhibited, and the binding was drastically reduced in the case that Smad signaling was restored on Day 2 of differentiation (Figure 2.3.3B). On the contrary, Smad2/3 binding peaked when Smad signaling was restored on Day 2 of differentiation, and the levels were markedly higher than those in Day 6 neuroectoderm (Figure 2.3.3B). The target DNA sequences did not contain Sox2 or Smad2/3 binding motifs, suggesting that Sox2 or Smad2/3 binding to these sequences might not be motif-dependent. Taken together, these findings implied that Smad signaling suppressed hESC differentiation towards neuroectoderm, possibly by preventing Sox2 binding to the neuroectoderm-specific enhancer sequences and subsequently activating neuroectodermal gene expressions.

Smad Signaling Suppression via Smad4 Knockdown Leads to hESC Spontaneous Differentiation towards Neuroectoderm Lineage

Since Smad signaling is involved in pluripotency maintenance and suppresses neuroectoderm differentiation, we then assessed whether Smad signaling disruption via Smad4 knockdown led to spontaneous differentiation of hESCs towards neuroectoderm *in vitro*. Smad4 was selected as the knockdown target in that it served as the co-Smad that was required for complexing with the phosphorylated R-Smads (Smad2/3 or Smad1/5) to shuttle the R-Smads into the nucleus and regulate target gene expressions. Knockdown of Smad4 was postulated to remove the Smad inhibition on neuroectodermal differentiation. To test this, lentivirus-delivered Smad4 shRNA was transduced into hESCs and the cells were maintained in hESC culture media under puromycin selection until a stable knockdown cell line was generated (Figure 2.4.1). The Smad4 knockdown efficiency reached 68% by Day 6 and further achieved 73% by Day 11 compared to the control hESCs transfected with lentivirus-delivered scrambled shRNA (Figure 2.4.2A). Gene expression levels of pluripotency, mesendodermal lineage and neuroectoderm markers were measured by RT-qPCR to assay whether hESCs underwent non-directed differentiation upon Smad4 knockdown. The expression level of the pluripotency marker NANOG, which was also a target of Smad2/3 signaling in pluripotency maintenance, was significantly declined by Day 11, suggesting suppression of Smad2/3 signaling and potential disruption of hESC integrity upon Smad4 knockdown (Figure 2.4.2B). The expression levels of EOMES and BRACHYURY, two markers of the mesendodermal lineages, were comparable between the Smad4-knockdown cells and the control cells transfected with scrambled shRNA, indicating that Smad4 knockdown did not induce differentiation towards mesendodermal lineage under hESC culture conditions

(Figure 2.4.2C). The expression levels of the neuroectodermal marker genes, namely OTX2, PAX6, SOX1, and ZIC2, were increased in Smad4-knockdown cells compared to the control cells (Figure 2.4.2D), suggesting that absence of Smad4 led to hESCs spontaneous differentiation favoring a neuroectodermal fate in pluripotency-maintaining culture conditions. Interestingly, SOX2 expression was not significantly elevated, which might imply that the switch in the role of Sox2 during early-stage neuroectoderm differentiation did not occur on the transcription level (Figure 2.4.2D).

Constitutively Active Smad Overexpression in hESCs Impedes Induced Neuroectodermal Differentiation

The *in vitro* differentiation of neuroectoderm from hESCs involves using the cell surface TGF-β receptor inhibitors SB431542 and LDN193189 to inhibit Activin/Smad2/3 and BMP/Smad1/5 signaling pathways, thereby disrupting pluripotency state and suppressing mesodermal and endodermal differentiations. To test whether overexpression of exogenous constitutively active (ca) Smad proteins can impede the neuroectoderm differentiation from hESCs, we cloned the caSmad1, caSmad2 and caSmad3 open reading frames into a vector driven by the constitutive CAG promoter. The open reading frames were linked to a neomycin (neoR) selection marker by an internal ribosomal entry site (IRES) sequence (Figure 2.5.1A). The caSmads were designed by mutating the C-terminus serine sites that could be phosphorylated into aspartic acid or glutamic acid residues, mimicking the phosphorylated serine residues in the active R-Smad proteins (Figure 2.5.1B). Wild-type (wt) Smad1, Smad2 and Smad3 open reading frames were cloned into the same type of vector as controls. The hESCs transfected with the Smad-overexpression plasmids went through neomycin selection until stable resistant cell lines were

generated. The Smad-overexpressing cells were then induced to differentiate towards neuroectoderm *in vitro* for 6 days and RT-qPCR was performed to monitor the differentiation efficiency by Day 2, Day 4 and Day 6, respectively. The NANOG expression levels decreased drastically during the differentiation, indicating that the Smad2/3 overexpression was not sufficient to maintain NANOG expression upon inhibition of Smad signaling via SB431542 and LDN193189 (Figure 2.5.2A). The expression levels of the neuroectodermal markers NEUROD1, PAX6 and SOX1 were comparable between caSmad1-expressing cells and the hESCs throughout the 6-day differentiation process, indicating that the designed caSmad1 did not affect the neuroectoderm differentiation (Figure 2.5.2B). Overexpression of caSmad2 and caSmad3 greatly decreased the expression of NEUROD1, PAX6 and SOX1 by Day 6, suggesting that the designed caSmad2 and caSmad3 could impede the neuroectoderm differentiation without Activin or BMP to activate the intracellular Smad signaling cascade (Figure 2.5.2B). Interestingly, the exogenous wild-type R-Smads also seemed to suppress neuroectoderm differentiation by reducing NEUROD1, PAX6 and SOX1 expressions (Figure 2.5.2B). Marker gene expression levels were comparable between the Smad-overexpressing cells and the hESCs by Day 2 and Day 4 but were greatly reduced on Day 6 in the Smad-overexpressing cells, implying that Smad overexpression negatively impacts neuroectoderm differentiation after 4 days of induction, where the markers NEUROD1, PAX6 and SOX1 should have undergone drastic increases in expression levels.

DISCUSSION

In this study, we looked into the binding patterns of Smads and Sox2 at the neuroectoderm-specific enhancers. Neuroectoderm-specific open chromatin regions addressed by ATAC-seq

data in genes upregulated in neuroectoderm were proposed as the neuroectoderm-specific regulatory DNA elements due to the strong increase in H3K27ac modifications in neuroectoderm compared to hESC. We found out that Smad2/3 bound to regulatory DNA in hESCs but not in neuroectoderm, while Sox2 binding relocated from the hESC-specific enhancers in hESCs to neuroectoderm-specific enhancers in neuroectoderm. Active Smad signaling suppressed hESC-derived *in vitro* neuroectoderm differentiation, with the increased binding of Smad2/3 reduced Sox2 binding to the neuroectoderm-specific enhancers, suggesting that Smads might suppress neuroectoderm differentiation in pluripotent stem cells by blocking Sox2 translocation to the neuroectoderm-specific enhancers sites. This hypothesis was also supported by the suppressed neuroectoderm differentiation from the caSmad-overexpressing hESCs. Repression of Smad signaling by Smad4 knockdown led to hESC spontaneous differentiation towards neuroectoderm, suggesting that Sox2 relocation to neuroectoderm-specific enhancers was allowed in the absence of R-Smads. Based on all these results, we proposed a model that when Smad signaling was active, Sox2 was hindered from the neuroectoderm enhancers and hESC-derived neuroectoderm differentiation was suppressed; when Smad signaling was inhibited, Sox2 could relocate to the neuroectoderm-specific enhancers and trigger neuroectoderm differentiation (Figure 2.7). Future experiments adopting ChIP-qPCR of R-Smads and Sox2 binding at the neuroectoderm-specific enhancer sites in the Smad4-knockdown hESCs as well as during neuroectoderm differentiation of Smad-overexpressing hESCs would be required to further test this model.

Sox2 occupying neuroectodermal enhancers has been shown to be critical in neuroectoderm specification (Lotado et al. 2013). It is still not clear, however, at which specific time point is Sox2 binding to regulatory DNA elements necessary in hESC-derived neuroectoderm

differentiation. To address this question, we set up a Sox2 conditional knockout hESC line adopting the Auxin-inducible degron (AID) technology. The AID tag-fused protein can be polyubiquitinated and degraded rapidly in cells with Skp1-Cullin-F-box (SCF) ligase upon auxin induction; when auxin is removed, degradation is stopped and target protein level can be restored, which allows testing protein functions in a time-sensitive manner (Figure 2.6.1A). To set up the Sox2 homozygous AID tag knockin hESC line, we made the plasmid for tagging endogenous SOX2 gene by fusing minimal AID (mAID) tag with an mNectarine fluorescence and a blasticidin selection marker to construct the insert cassette and cloned the cassette into a vector harboring 200-bp Sox2 homology donor arms (Figure 2.6.1B). Cells were transfected with the insertion plasmid and an CRISPR/Cas9 plasmid for accurate knockin of the mAID-mNectarine-blasticidin cassette into the SOX2 locus. The drug-selected cells exhibiting red fluorescence were sorted and grown into single cell-derived colonies for genotyping (Figure 2.6.2). The cell colony with homozygous insertion confirmed by genomic PCR was used as the hESC line for experiments (Figure 2.6.3). By using the AID-tagged cells, we can explore the exact time window in which Sox2 affects neuroectoderm differentiation by conditionally depleting Sox2 proteins at different time points. It is expected that during early stages of neuroectoderm specification, conditional depletion of Sox2 by Auxin-induced degradation will result in Smad2/3 staying bound to enhancers and delay of differentiation, which can be rescued upon Sox2 restoration by Auxin removal that leads to Sox2 replacing Smad2/3 at the enhancer regions. This method may help provide deeper insight into the signaling network involving Sox2 and other key transcription factors governing hESC-derived neuroectoderm differentiation.

EXPERIMENTAL PROCEDURES

hESC maintenance The human embryonic stem cell line used in this study was WA09. Cells were cultured on Geltrex (ThermoFisher)-coated plates in ES cell maintenance media made up of base media supplemented with 10 ng/ml recombinant human heregulinβ-1 (Peprotech), 10 ng/ml Activin A (R&D Systems), 200 ng/ml Insulin-like Growth Factor (IGF)-1 (Sigma) and 8 ng/ml Fibroblast Growth Factor (FGF)-2 (R&D Systems). The base media consists of DMEM/F12 (Fisher Scientific), 2% Probumin, 1x Corning™ MEM Nonessential amino acids (Fisher Scientific), 1x Corning™ Cellgro™ Antibiotic-Antimycotic (Fisher Scientific), 1x Corning™ cellgro™ glutaGRO Supplement (200mM L-alanyl-L-Glutamine) (Fisher Scientific), 1x Corning™ Cellgro™ Trace elements A, B and C (Fisher Scientific), 50µg/ml ascorbic acid, 10 µg/ml human plasma transferrin (Athens Research Technology) and 0.1 mM beta-mercaptoethanol (Life Technologies).

Induced neuroectoderm differentiation WA09 cells were seeded onto Geltrex (ThermoFisher)-coated plates at 90,000 cells/cm^2 and were cultured in neuroectoderm differentiation media consisting of base media (as has been described in *hESC maintenance*) supplemented with 10 ng/ml recombinant human heregulinβ-1 (Peprotech), 200 ng/ml Insulin-like Growth Factor (IGF)-1 (Sigma), 10 µg/ml SB431542 (Tocris) and 500 µM LDN193189 (Sigma). Cells were harvested 6 days after differentiation induction for further RT-qPCR assays.

Induced definitive mesoderm and definitive differentiation For mesoderm differentiation, WA09 cells were seeded onto Geltrex (ThermoFisher)-coated plates at 50,000 cells/cm^2 and were cultured in mesoderm differentiation media consisting of base media (as has been described in *hESC maintenance*) supplemented with 10 ng/ml recombinant human heregulinβ-1 (Peprotech), 10 ng/ml Activin A (R&D Systems), 200 ng/ml IGF-1 (Sigma), 8 ng/ml FGF-2 (R&D Systems),

25 ng/ml recombinant human Wnt-3a (Bio-Techne) and 100 ng/ml recombinant human Bone Morphogenic Protein (BMP)-4 (R&D Systems). Cells were harvested 4 days after differentiation induction for ATAC-seq assays. For definitive endoderm differentiation, WA09 cells were seeded onto Geltrex (ThermoFisher)-coated plates at 50,000 cells/cm^2 and were cultured in definitive endoderm differentiation media consisting of base media (as has been described in *hESC maintenance*) supplemented with 100 ng/ml Activin A (R&D Systems) and 8 ng/ml FGF-2 (R&D Systems). 25 ng/ml Wnt-3a (Bio-Techne) was added to the media on the first day of differentiation only. Cells were also harvested 4 days after differentiation induction for ATAC-seq assays.

RT-qPCR Total RNA was isolated from 1 million lysed cells using E.Z.N.A total RNA kit (Omega Bio-tek) followed by cDNA conversion using iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad). RT-qPCR was done on ViiA™ 7 Real-Time PCR System (Applied Biosystems) following manufacturer instructions applying TaqMan Assay (ThermoFisher) to cDNA fractions. 18s ribosomal RNA (18s rRNA) is used as the standard reference to normalize the RT-qPCR data output. The RT-qPCR results are analyzed by the ΔΔCT method.

ATAC-seq 50,000 cells were harvested and washed with cold PBS buffer. Transposition reaction mix containing Tn5 Transposase (Nextera) was added to the cells for a 30-minute reaction at 37°C followed by purification of transposed DNA using DNA clean and concentrator kit (Zymo Research). Eluted DNA fragments were amplified via a 5-cycle PCR reaction with Custom Nextera PCR primers (Nextera), followed by a 20-cycle qPCR reaction using Custom Nextera PCR primers (Nextera) and SYBR Green I reagent on 10% of the amplified DNA. Additional number of PCR cycles needed were determined as the cycle number corresponding to 1/3 of the maximum fluorescent intensity (Buenrostro et al. 2015). A second PCR reaction was performed

on the remaining DNA with the cycle number determined by the qPCR reaction. Size selection was performed on the amplified DNA library to eliminate fragments larger than 1000bp or smaller than 100bp using Mag-Bind® RxnPure Plus beads (Omega). Sequencing was performed on Illumina NextSeq High Output platform generating 400 million reads per 75nt single-end sequencing. ATAC-seq reads were aligned to human genome assembly GRCh37 (hg19) by Bowtie2 version 2.2.9 and sorted by Samtools version 1.3.1. Peak calling was performed on the aligned ATAC-seq reads using MACS2 version 2.1.1 and differential peak analyses were done between the peak files of distinct cell types generated via MACS2 algorithms. The differential peaks were then annotated by their proximity to the transcription start sites (TSS) of genes using the ChIPseek website (http://chipseek.cgu.edu.tw/analysis_form.php).

RNA-seq Total RNA was extracted from 1 million lysed cells using E.Z.N.A total RNA kit (Omega Bio-tek). Sequencing was performed on Illumina NextSeq High Output platform generating 50 million reads per 75nt paired-end sequencing. RNA-seq reads were aligned to human genome assembly GRCh37 (hg19) by Tophat version 2.0.13 (Trapnell et al. 2010). Transcript assembly and expression level differential analyses were performed by Cufflinks version 2.2.1 (Trapnell et al. 2010).

ChIP-seq 2 million cells were harvested, washed with cold PBS and crosslinked with 1% formaldehyde for 8 minutes at room temperature followed by 5-minute quenching with 0.2M glycine and cold PBS wash. Cells were lysed in buffer containing 10 mM pH=8.0 Tris-HCl, 100 mM NaCl, 1 mM EDTA, 0.1% Na-Deoxycholate and 1x protease inhibitors (Sigma), followed by a 5-minute sonication in Covaris S220 until the fragment sizes were within 200-700bp. The sonicated cell lysates were then centrifuged at 20,000rcf at 4°C for 15 minutes. Dynabeads™ Protein G (ThermoFisher) were blocked in washing buffer consisting of PBS, 0.05% Tween-20

and 0.5% bovine serum albumin (BSA) and then conjugated to either mouse anti-SOX2 (MAB2018, R&D Systems), goat anti-SMAD2/3 (AF3797, R&D Systems), or rabbit anti-SMAD1 (9743S, Cell Signaling Tech) by rotating at 4°C for 4 hours. Supernatant from the centrifuged sonicated cell lysate was added to the conjugated beads and rotated overnight at 4°C. Beads were then washed with TF-WBI buffer (20mM pH=7.4 Tris-HCl, 0.1% SDS, 150mM NaCl, 2mM EDTA and 1% Triton-X-100) twice and TF-WBII buffer (10mM pH=7.4 Tris-HCl, 0.7% sodium deoxycholate, 250mM LiCl, 1mM EDTA and 1% Triton-X-100) twice. ChIPmentation was performed by adding reaction mix containing Tn5 Transposase (Nextera) to the beads followed by 1-minute incubation at 37°C. Beads were then washed with TF-WBI twice and TET buffer (10 mM pH=8.0 Tris-HCl, 1mM EDTA and 0.2% Tween-20) twice, followed by incubation with 70 µl elution buffer supplemented with 2 µl Proteinase K (Roche) at 55°C for 1 hour and 65°C overnight for reversion of formaldehyde crosslinking. The eluted DNA was purified using DNA clean and concentrator kit (Zymo). Sequencing was performed on Illumina NextSeq High Output platform generating 400 million reads per 75nt single-end sequencing. ChIP-seq reads were aligned to human genome assembly GRCh37 (hg19) by Bowtie2 version 2.2.9 and sorted by Samtools version 1.3.1.

Heatmap generation DNA regions where the corresponding gene expression is upregulated in neuroectoderm compared to WA09 cells with the ATAC-seq signal increased in neuroectoderm compared to WA09 cells were determined by intersecting the RNA-seq differential expression analysis result with the ATAC-seq differential peak analysis result in RStudio version 1.0.136. Region files in BED format were generated by further classifying the intersect DNA sequences into promoter (located within 0-2 kb from TSS), intermediate (located 2-5 kb from TSS) and enhancer (located more than 5 kb from TSS) regions. SOX2, SMAD1, SMAD2 and SMAD3

ChIP-seq coverage tracks (bigWig files) used in the matrix computation for heatmap plotting were generated by converting the aligned and sorted ChIP reads (BAM files) generated as has been described in *ChIP-seq* at bin size of 100 bp using Deeptools version 2.3.1. The ChIP-seq coverage tracks of other transcription factors and histone modifications used in the study were obtained from the public online Gene Expression Omnibus with the accession numbers GSE61475 and GSE62193. Heatmaps were plotted using the matrices generated with the ChIP signal distributed relative to the center of the intersect DNA regions at the distance of ±3 kb via Deeptools version 2.3.1.

ChIP-qPCR Cells were harvested, crosslinked, lysed and sonicated following the same procedures as has been described in *ChIP-seq*. Dynabeads™ Protein G (ThermoFisher) were blocked in washing buffer, conjugated to either mouse anti-SOX2 (MAB2018, R&D Systems), goat anti-SMAD2/3 (AF3797, R&D Systems), or rabbit anti-SMAD1 (9743S, Cell Signaling Tech) by rotating at 4°C for 4 hours, and combined with supernatant from the centrifuged cell lysate followed by overnight rotation at 4°C as has been described in *ChIP-seq*. Beads were then washed with TF-WBI buffer twice, TF-WBII buffer twice, and TET buffer twice. 100 µl elution buffer and 6 µl Proteinase K (Roche) were then added to the beads. The beads were incubated at 55°C for 1 hour and 65°C overnight to reverse the formaldehyde crosslinking. The eluted DNA was purified by DNA clean and concentrator kit (Zymo). Eluted DNA was then mixed with primers specifically designed for detection at target gene sites, ROX Low dye and KAPA SYBR® FAST qPCR Master Mix for the SYBR Green dye-based qPCR reactions. The primers used for ChIP-qPCR are listed in Table 1.

Table 1 | List of ChIP-qPCR primers	
Primer name	Sequence
Pax6	TTCCTTGACATGCAACATCC
	AGAGGCATCATTTCCCATTG
SOX2	GGCTTTGTTTGACTCCGTGT
	ATTTTAGCCGCTCTCCCATT
SOX1	CAATGGCTTCACAAAGCTGA
	GCCTTGGACTTGTGTGGTCT

SMAD4 shRNA knockdown Lentiviruses containing pLKO.1-puro vector-based MISSION® shRNA (Sigma) against human SMAD4 gene (target sequences: 5'-TACCATACAGAGAACATTGGA-3', 5'-GTACTTCATACCATGCCGATT-3', 5'-CGAGTTGTATCACCTGGAATT-3') were produced using HEK293T cells. The lentivirus containing pLKO.1-puro vector-based scramble shRNA (Addgene, catalog No. 1864) was used as negative control. WA09 cells were transfected with the shRNA lentivirus at MOI=2. Puromycin selection was performed on the transfected WA09 cells cultured in ES cell maintenance media 24 hours after initial transfection at 0.2 μg/ml, and the concentration was increased to 1 μg/ml after 48 hours. Cells were harvested 6 days after transfection.

SMAD overexpression Wild type SMAD1, SMAD2 and SMAD3 open reading frame DNA sequences were amplified using the primers listed in Table 2 and cloned into a CAG-neo expression vector driven by a CAG promoter. DNA sequences that can be transcribed into constitutively active SMAD1, SMAD2 and SMAD3 proteins were generated by mutating the phosphorylation sites Ser462, Ser463 and Ser465 to S462D, S463D and S465D in SMAD1, Ser465 and Ser467 to S465E and S467E in SMAD2, and Ser422, Ser423 and Ser425 to S422D, S423D and S425D in SMAD3 using the primers listed in Table 2. The constitutively active SMAD1, SMAD2 and SMAD3 were also cloned into the CAG-neo expression vector driven by a CAG promoter. The wild-type and constitutively active SMAD expression plasmids were transformed into WA09 cells via electroporation. The transformed cells were cultured in ES cell maintenance media for 48 hours before neomycin selection at the concentration of 150 μg/ml.

Table 2 | List of SMAD amplification primers	
Primer name	Sequence
SMAD1, forward	TGCGCTCGAGCCACCATGAATGTGACAAGTTTATTTTCCTTTAC
Wild-type SMAD1, reverse	TATAGCGGCCGCTTAAGATACAGATGAAATAGGATTATGAGGTGAAC
Constitutively active SMAD1, reverse	CGAATGCGGCCGCTTAGTCTACATCGTCAATAGGATTATGAGGTGAAC
SMAD2, forward	TACGCTCGAGCCACCATGTCGTCCATCTTGCCA
Wild-type SMAD2, reverse	TATAGCGGCCGCTTATGACATGCTTGAGCAACGCAC
Constitutively active SMAD2, reverse	TATAGCGGCCGCTTACTCCATCTCTGAGCAACGCACTGAAGGGGATC
SMAD3, forward	TACGCTCGAGCCACCATGTCGTCCATCCTGCCT
Wild-type SMAD3, reverse	TATAGCGGCCGCTTAAGACACACTGGAACAGCGGATGCTTGGGGAG
Constitutively active SMAD3, reverse	TATAGCGGCCGCTTAGTCTACATCGTCACAGCGGATGCTTGGGGAG

SOX2 AID-knockin hESC line setup The generation of minimal AID-knockin mutant for conditional Sox2 knockout followed the general procedures described in (Natsume et al. 2016). To construct donor vector for generation of homozygous SOX2 AID-knockin mutant for conditional Sox2 knockout, short homology arms for SOX2 were synthesized and cloned into the pUC19 vector between the HindIII and EcoRI restriction sites, with a BamHI restriction site between the left and the right arm. The minimal AID (mAID)- mNectarine-IRES-Blasticidin (Bsr) cassette was amplified by NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs) and cloned into the donor vector at the BamHI site between the SOX2 donor arms. CRISPR/Cas9 vectors were constructed by cloning SOX2 guide RNAs into the pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid from Feng Zhang (Addgene #42230) at the BbsI site. Cells transfected with the SOX2 donor plasmid and the CRISPR/Cas9 plasmid was maintained in hESC media supplemented with blasticidine at the concentration of 1 µg/ml, and the drug-selected cells that were positive for red fluorescence went through single-cell fluorescence activated cell sorting (FACS) on a Beckman Coulter MoFlo XDP. Colonies growing in 96-well plates from single cells were collected for genomic PCR to detect the genotypes. The cell line with homozygous AID-knockin mutant genotype at SOX2 loci was used for further studies.

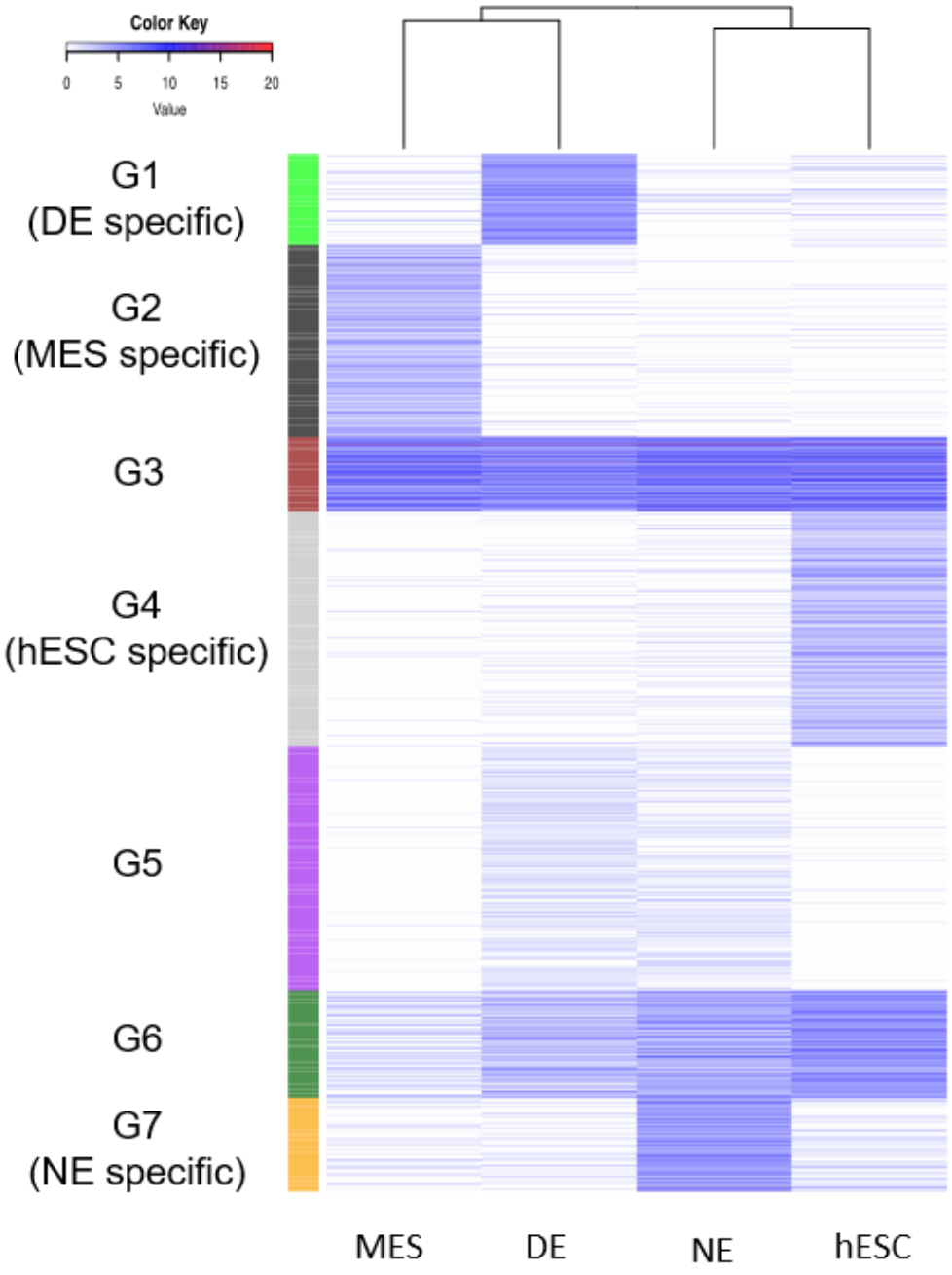


Figure 2.1.1 | K-Means Clustering of Open Chromatin Sites in hESC, Definitive Endoderm (DE), Mesoderm (MES) and Neuroectoderm (NE) Loci of open chromatin were indicated by the pooled ATAC-seq signals of hESCs, DE, ME and NE. The loci were categorized into groups specific to each cell type and common among two or three or all cell types.

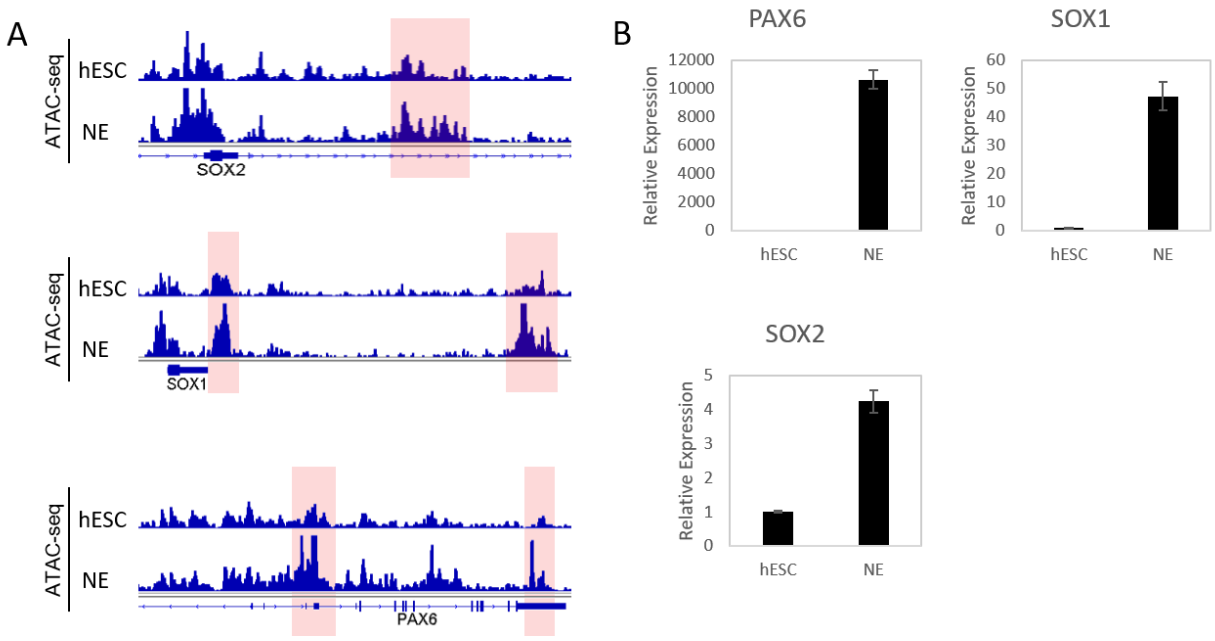
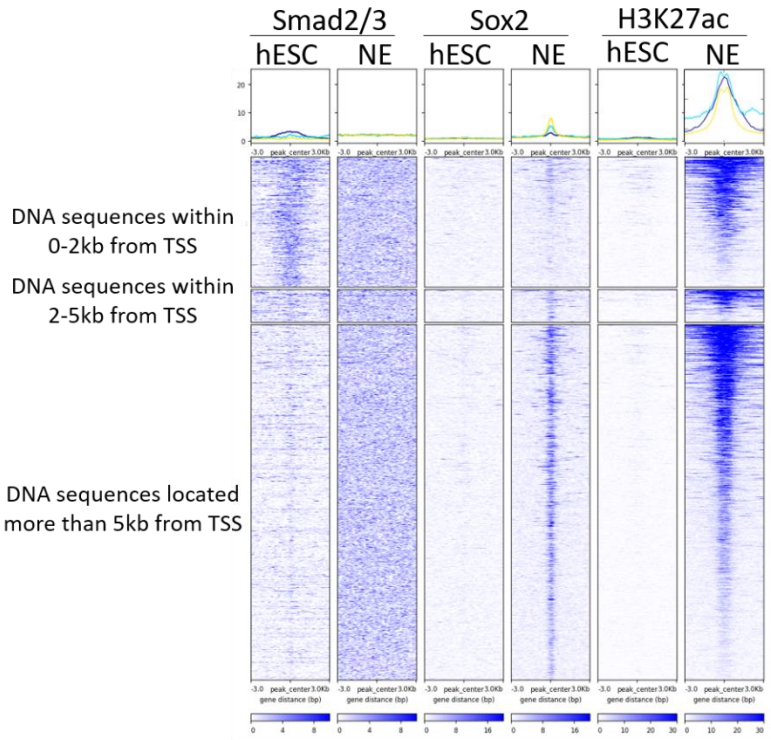


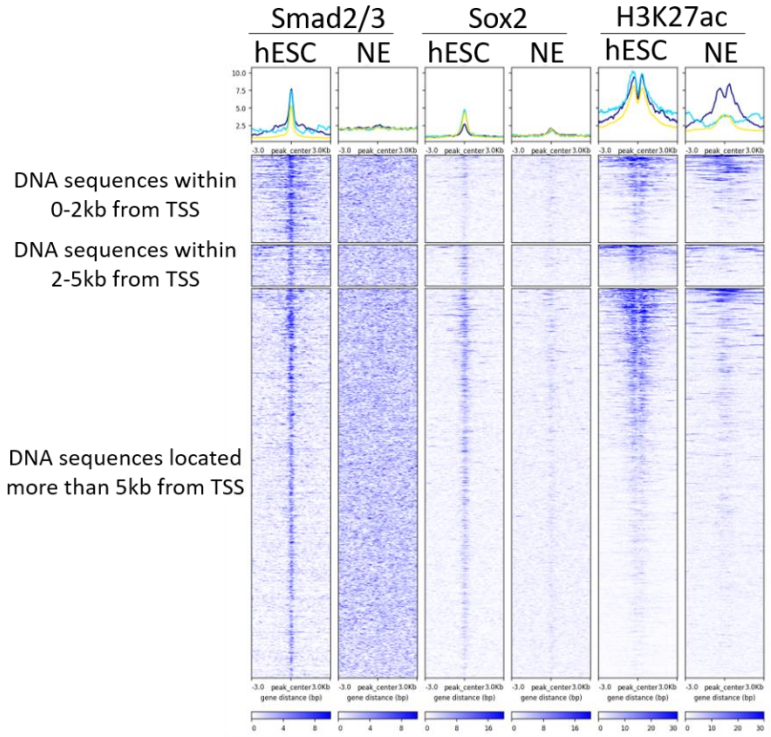
Figure 2.1.2 | Representative Genomic View of Enhanced ATAC-seq Signal at Neuroectodermal Gene Loci (A) ATAC-seq (400 million single-end reads) signals in all tracks were normalized to the same scale (0-50, normalized tag count). The sites highlighted in pink boxes represented regions where ATAC-seq signal is increased in NE compared to hESC. (B) Real-time reverse transcription polymerase chain reaction (RT-qPCR) analysis of the genes in (A). Each bar represented the average result from experiments in triplicate. Error bars reflected triplicate reactions.

A



Regions with enhanced ATAC-seq signal in NE compared to hESC in NE-specific genes

B



Regions with enhanced ATAC-seq signal in hESC compared to NE in hESC-specific genes

Figure 2.2.1 | Enrichment Profiles of Histone Modification and Transcription Factor binding Differ at Cell Type-Specific Open Chromatin Regions (A) ChIP-seq heatmaps of Smad2/3, Sox2 and H3K27ac at regions where ATAC-seq signals were upregulated in NE compared to hESC in NE-upregulated genes. The ChIP-seq peaks were clustered into three groups based on the distance from the transcription start site (TSS), namely sequences located more than 5 kb away, within 2-5 kb and within 0-2 kb, from bottom to top. Peak center was used as the reference point. (B) ChIP-seq heatmaps of Smad2/3, Sox2 and H3K27ac at regions where ATAC-seq signals were upregulated in hESC compared to NE in hESC-upregulated genes. The ChIP-seq peaks were clustered in the same way as had been described in (A).

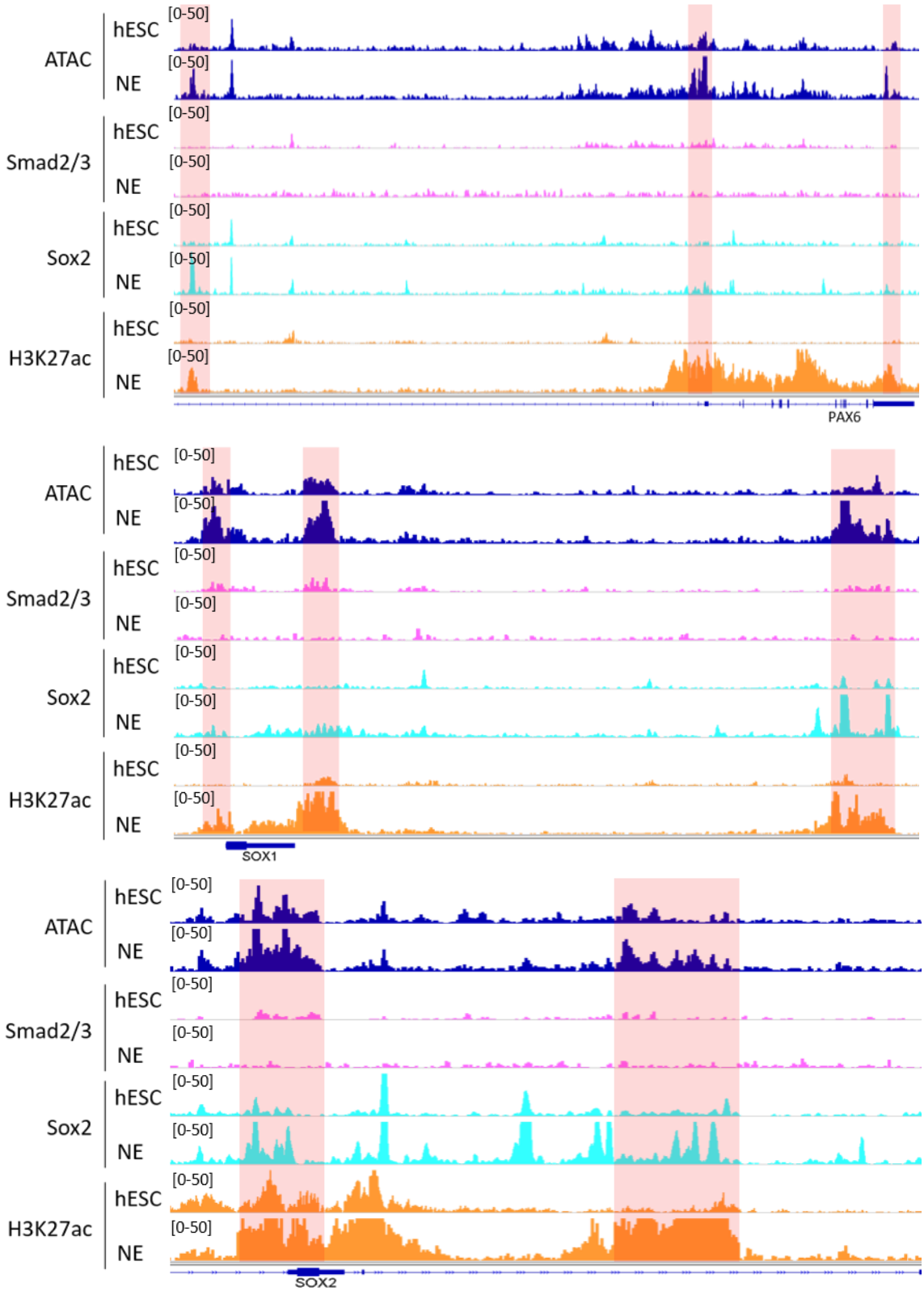

ATAC
hESC
NE
Smad2/3
hESC
NE
Sox2
hESC
NE
H3K27ac
hESC
NE
[0-50]
PAX6
SOX1
SOX2

Figure 2.2.2 | Representative Genomic View of Histone Modification and Transcription Factor Binding at Neuroectodermal Gene Loci ATAC-seq signals and ChIP-seq signals in all tracks were normalized to the same scale (0-50, normalized tag count). The sites highlighted in pink boxes represented regions where ATAC-seq signal is increased in NE compared to hESC.

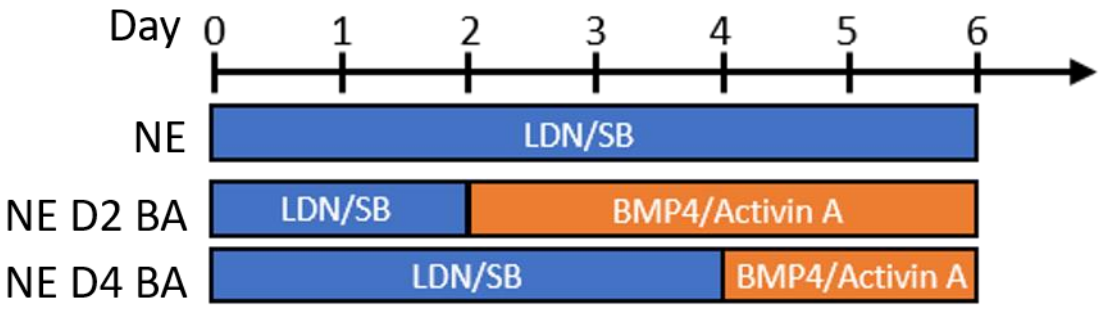


Figure 2.3.1 | Schematic Diagram of hESC-Derived Neuroectoderm Differentiation Strategy Neuroectoderm was differentiated from hESC in vitro by dual-Smad inhibition method using SB431542 and LDN193189. SB and LDN were replaced by BMP4 and Activin A on Day 2 or Day 4 of the differentiation, respectively, to restore Smad signaling.

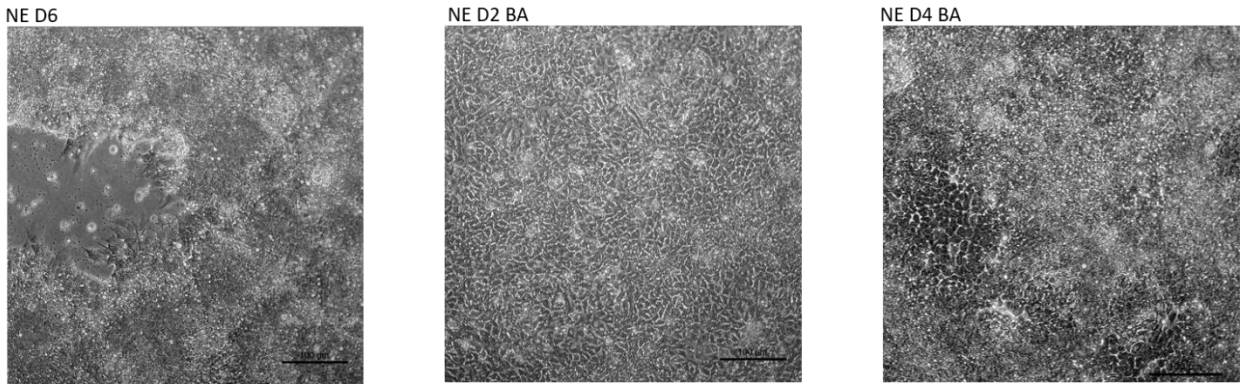


Figure 2.3.2 | Day-6 Cell Morphology Changed upon Restoration of Smad Signaling During hESC-Derived Neuroectoderm Differentiation Phase contrast images of Day-6 hESC-derived cells. NE: 6-day differentiation using SB and LDN; NE D2 BA: differentiation using SB and LDN for 2 days, followed by replacement with BMP4 and Activin A till Day 6; NE D4 BA: differentiation using SB and LDN for 4 days, followed by replacement with BMP4 and Activin A till Day 6. Scale bar 100 μm.

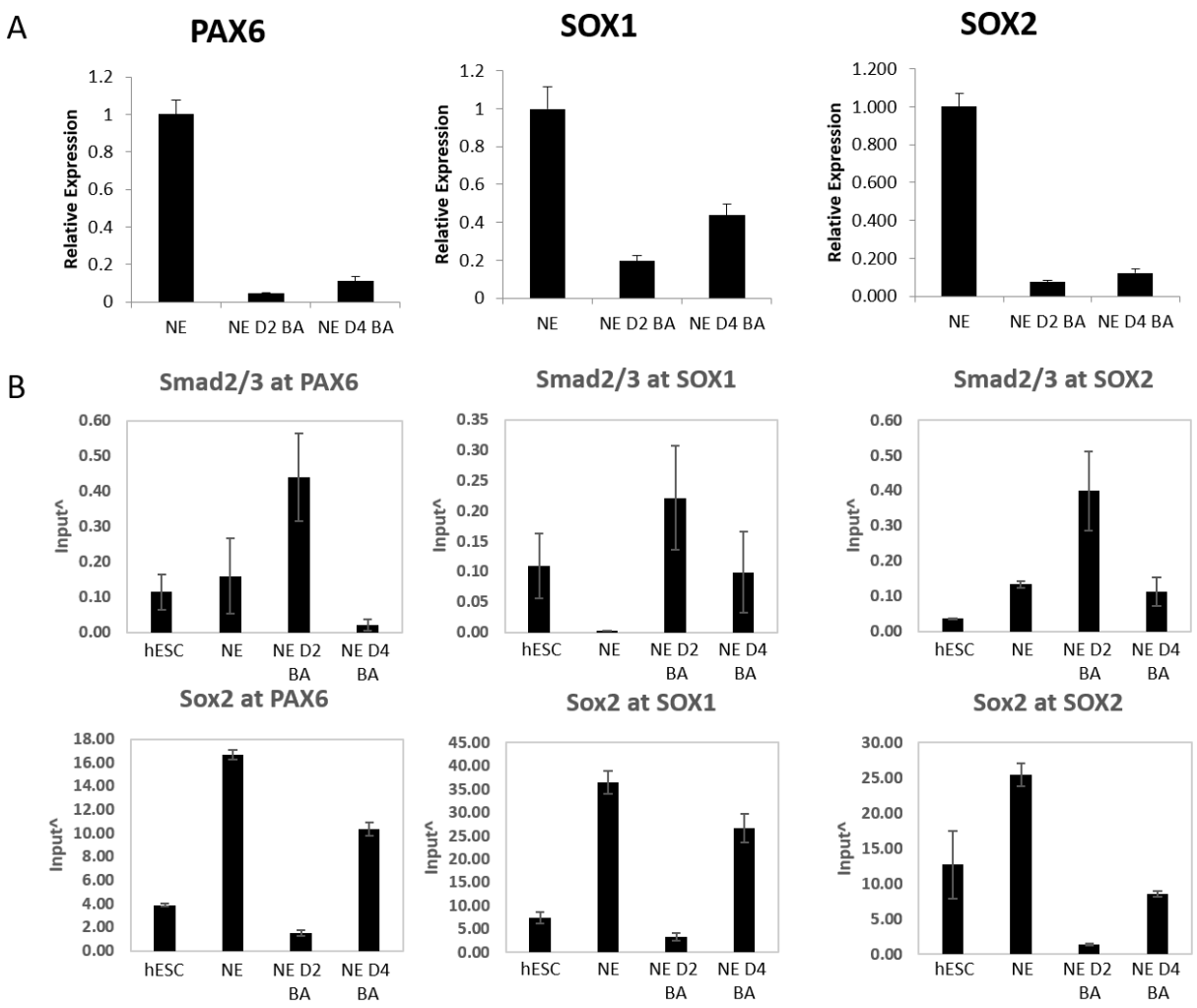


Figure 2.3.3 | Neuroectoderm Differentiation was Suppressed upon Smad Restoration, with Changes in Transcription Factor Binding Patterns (A) RT-qPCR analysis of PAX6, SOX1 and SOX2 in the hESC-derived neuroectoderm cells with/without Smad signaling restoration during differentiation. (B) ChIP-qPCR results of Smad2/3 and Sox2 binding at neuroectoderm-specific open chromatin sites in PAX6, SOX1 and SOX2 genes of hESC and hESC-derived neuroectoderm cells with/without Smad signaling restoration during differentiation. Data represent the relative level to input. NE: 6-day differentiation using SB and LDN; NE D2 BA:

differentiation using SB and LDN for 2 days, followed by replacement with BMP4 and Activin A till Day 6; NE D4 BA: differentiation using SB and LDN for 4 days, followed by replacement with BMP4 and Activin A till Day 6.

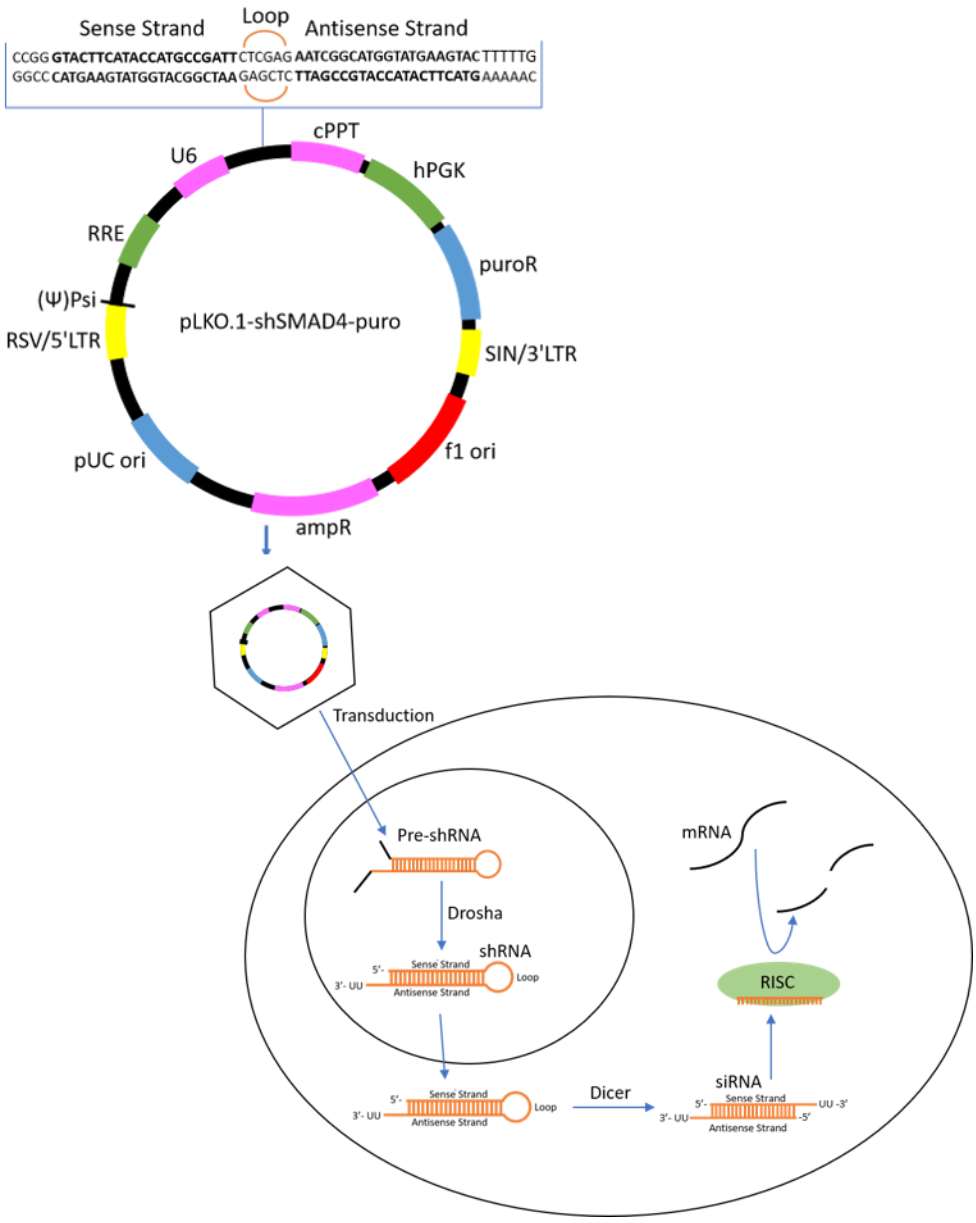


Figure 2.4.1 | Scheme of Lentivirus-Transduced Smad4 Knockdown in hESCs Smad4 shRNA plasmid was packaged into lentivirus using HEK293T cells and used to infect hESCs. The shRNA was processed inside the cell into siRNA that eventually degraded the Smad4 mRNA to reduce gene expression.

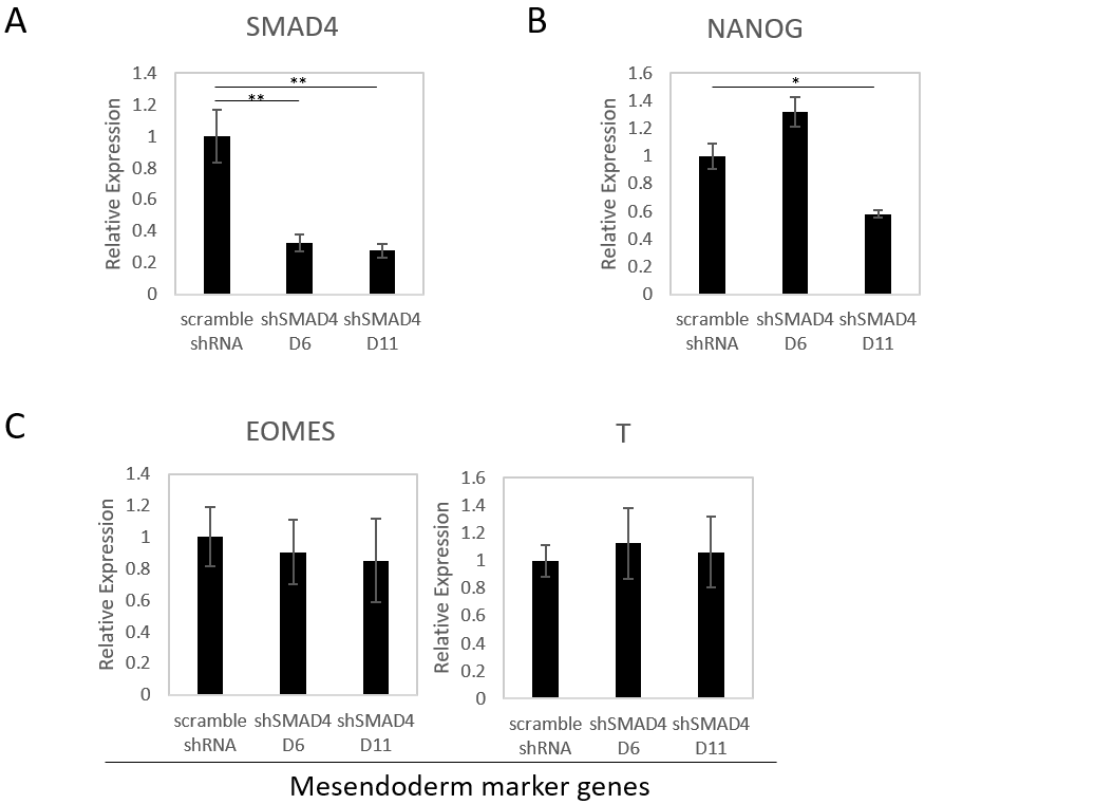

A
SMAD4
Relative Expression
scramble shRNA
shSMAD4 D6
shSMAD4 D11
B
NANOG
C
EOMES
T
Mesendoderm marker genes
D
OTX2
PAX6
SOX1
Neuroectoderm marker genes
SOX2
ZIC2
Neuroectoderm marker genes

Figure 2.4.2 | Smad4 Knockdown Led to hESC Spontaneous Differentiation Favoring Neuroectodermal Fate Smad4 was knocked down in hESCs via lentivirus-transduced shRNA. RT-qPCR was performed on the cells cultured in hESC maintenance media under puromycin selection on Day 6 and Day 11, respectively. Scramble shRNA-transduced hESCs were used as control. Gene expression levels of (A) SMAD4, (B) NANOG, (C) mesendodermal markers EOMES and T, (D) neuroectodermal markers OTX2, PAX6, SOX1, SOX2 and ZIC2 were measured. * ($P<0.05$), ** ($P<0.01$) and *** ($P<0.001$) indicate Ct values significantly different from the control samples based on two-tailed Student's t-test.

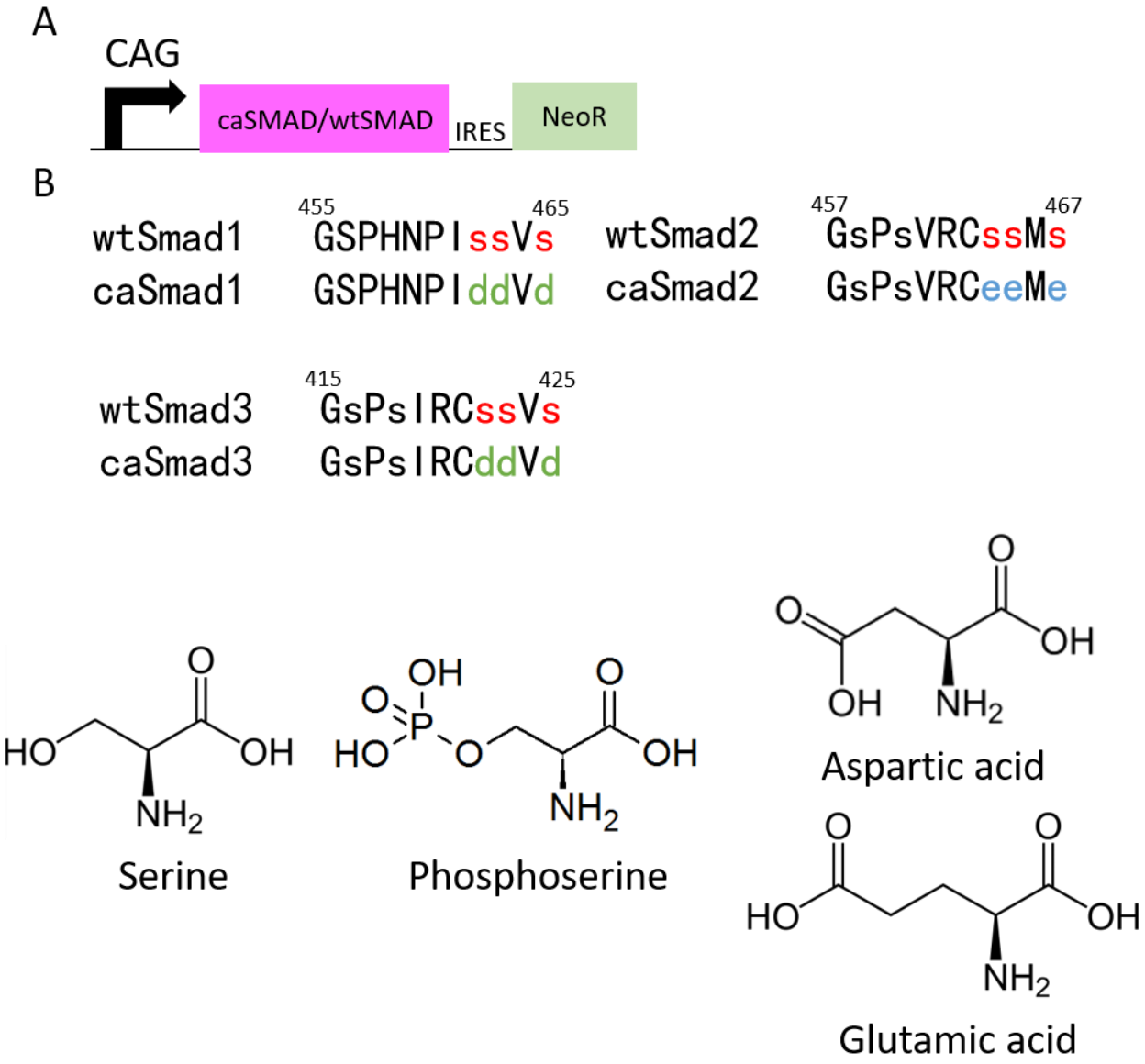
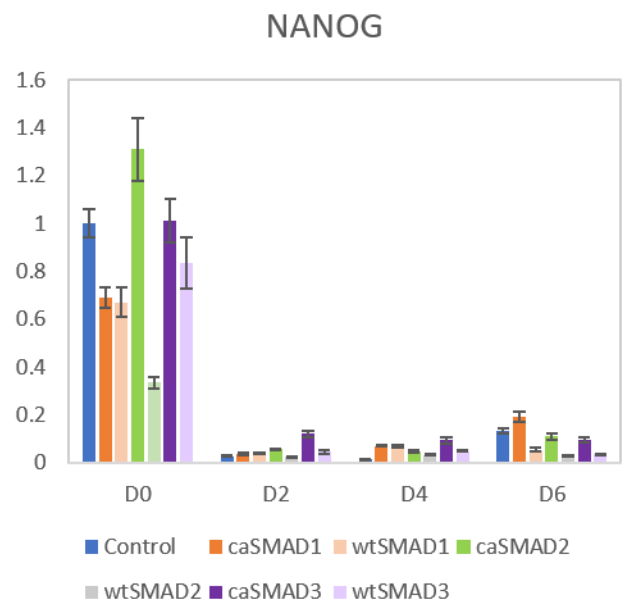


Figure 2.5.1 | Design of Constitutively Active (ca) Smad Overexpression System (A) Wild-type (wt) SMAD or caSMAD open reading frame was driven by a CAG promoter and linked to a neomycin (NeoR) selection marker via an internal ribosome entry site (IRES) sequence. (B) The caSMADs were designed by mutating the C-terminal serine (S) residues into aspartic acid (D) or glutamic acid (E) residues. Aspartic acid and Glutamic acid served as the substitute for phosphoserine because of the additional carboxyl group.

A


NANOG
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
D0
D2
D4
D6
Control
caSMAD1
wtSMAD1
caSMAD2
wtSMAD2
caSMAD3
wtSMAD3

B

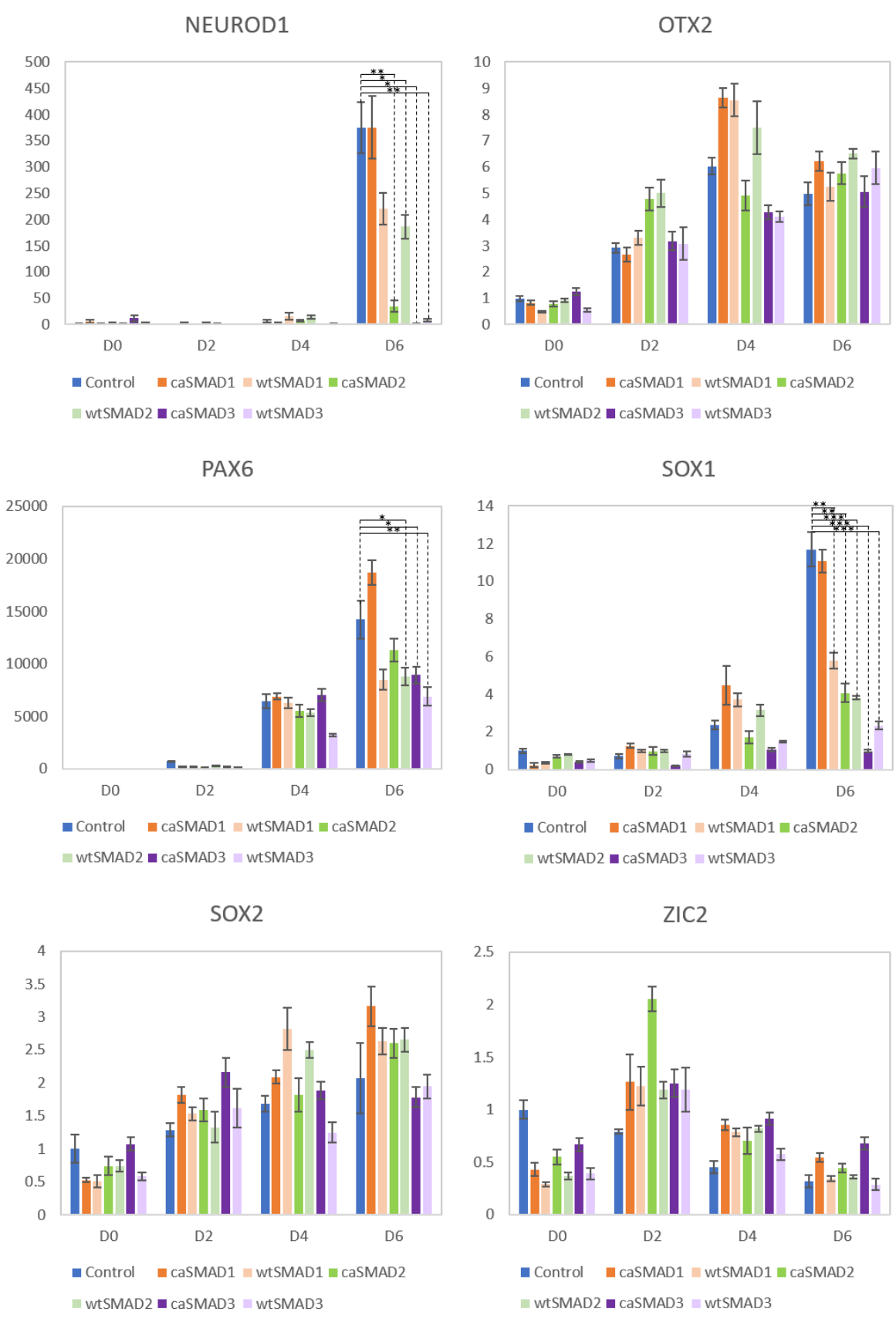

NEUROD1
OTX2
PAX6
SOX1
SOX2
ZIC2
D0
D2
D4
D6
Control
caSMAD1
wtSMAD1
caSMAD2
wtSMAD2
caSMAD3
wtSMAD3

Figure 2.5.2 | Smad Overexpression Suppressed hESC-Derived Neuroectodermal Differentiation The hESCs transfected with caSMAD or wtSMAD-overexpressing plasmids were differentiated *in vitro* towards neuroectoderm for 6 days with the non-transfected hESC serving as control. RT-qPCR was performed on the 6-day neuroectoderm to measure the expression levels of (A) NANOG and (B) neuroectodermal marker genes NEUROD1, OTX2, PAX6, SOX1, SOX2 and ZIC2. * ($P<0.05$), ** ($P<0.01$) and *** ($P<0.001$) indicate Ct values significantly different from the control samples based on two-tailed Student's t-test.

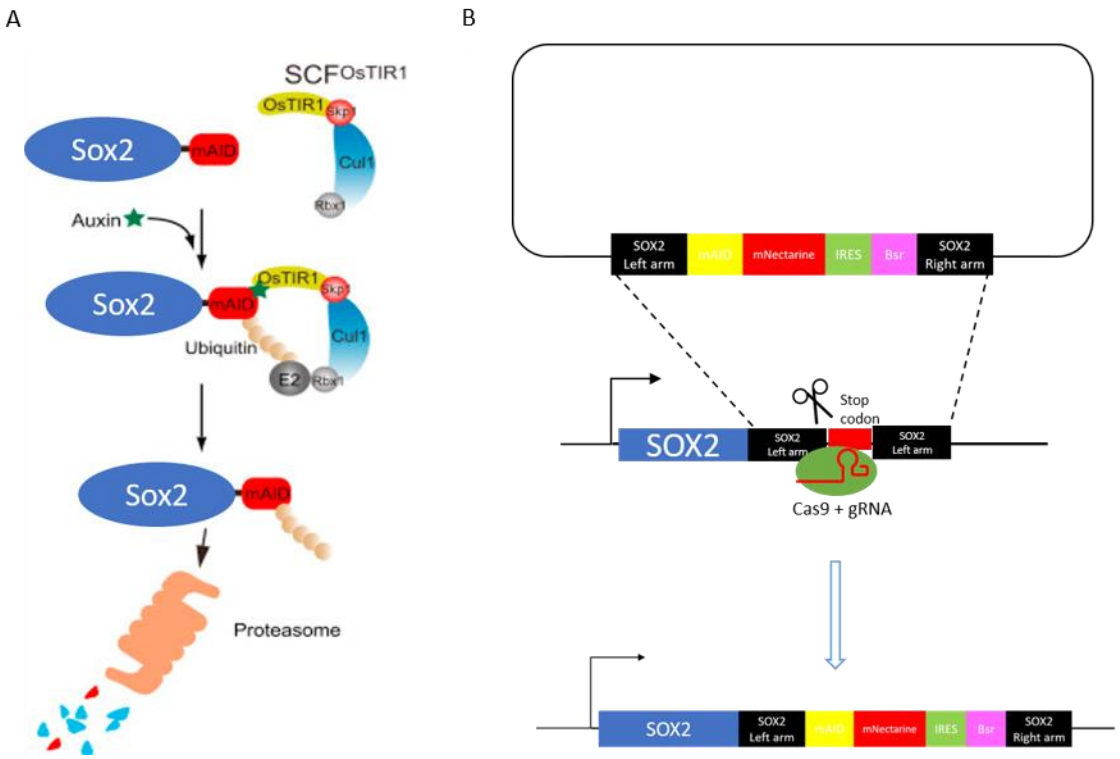


Figure 2.6.1 | Principle and Design of the Sox2 Conditional Auxin-Inducible Degron (AID) System (A) Scheme of the AID system. OsTIR1 expression generated Skp1-Cullin-F-box (SCF) ubiquitin ligase. The AID-tagged Sox2 protein was poly-ubiquitinated by SCF upon addition of auxin for degradation. (B) Schematic illustration of tagging SOX2 with mNectarine. A synthetic donor DNA harboring around 200bp homology arms was cloned into pUC19 vector before being transfected with a CRISPR/Cas9 plasmid into cells to target the SOX2 locus. Adapted from (Natsume et al. 2016)

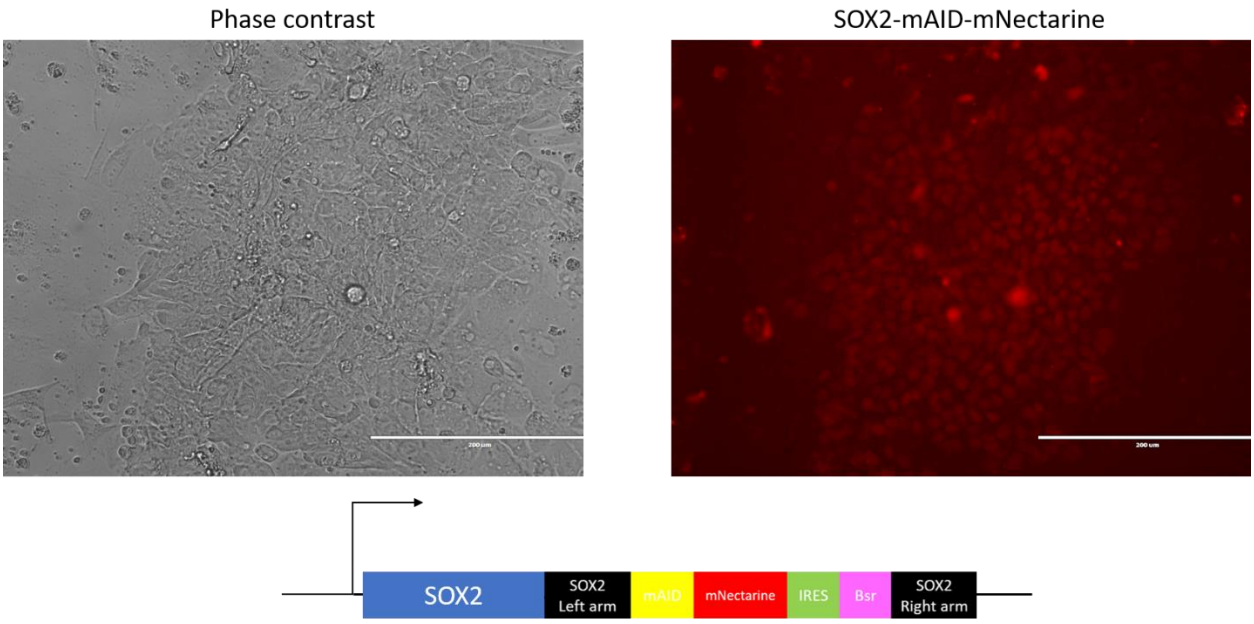


Figure 2.6.2 | Representative Colony of SOX2 AID-Knockin hESCs showing red fluorescence SOX2-mAID- mNectarine localized in the nucleus of the knock-in cells was detected under fluorescent microscopic view.

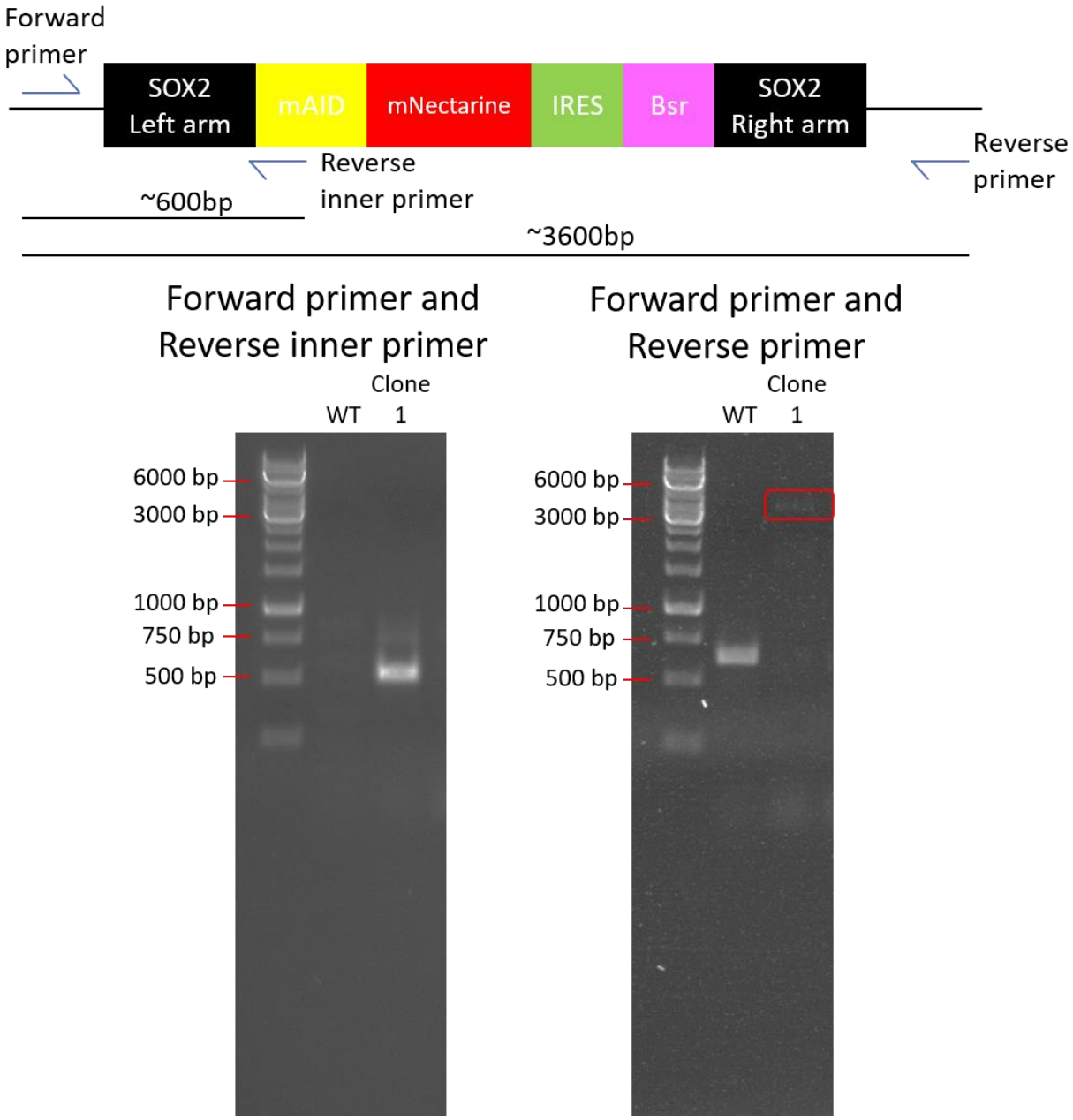


Figure 2.6.3 | Genomic PCR to Detect the Genotype of mNectarine -Positive Clones The primer sets and the expected PCR product sizes were shown in the illustration. WT: Wild-type hESCs. Clone 1: a SOX2 mAID-knockin clone that exhibited red fluorescence under the fluorescence microscope.

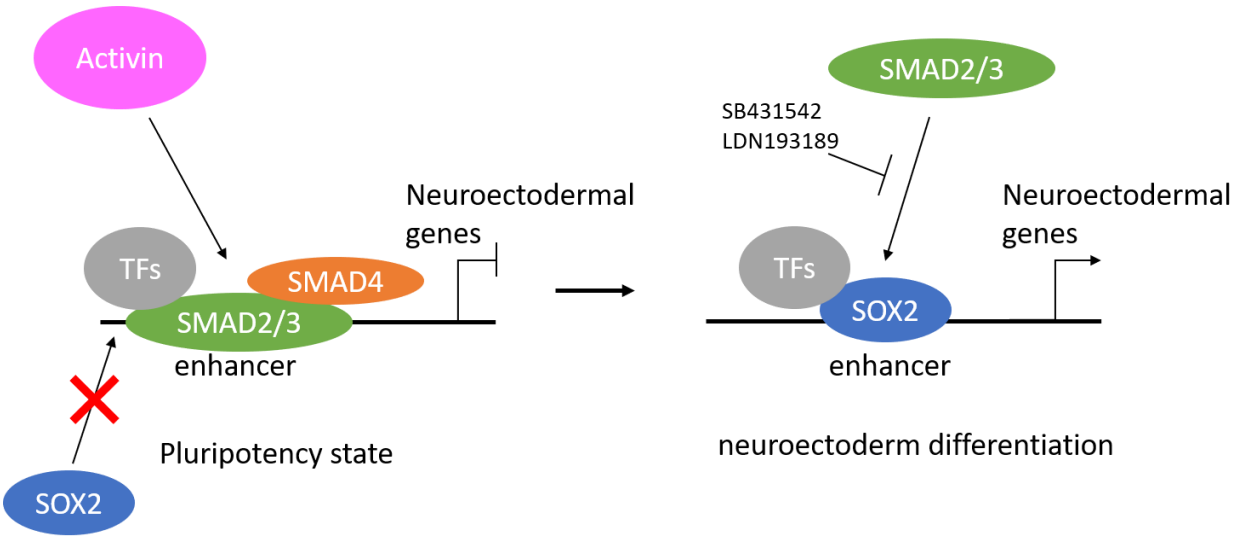


Figure 2.7 | Summary of the Relationship Between Sox2 and Smad2/3 at Neuroectoderm-Specific Enhancers in hESCs and Neuroectoderm In hESCs where pluripotency state was maintained and Smad signaling was active, Sox2 relocation to neuroectoderm enhancers was inhibited, repressing neuroectoderm gene expression. In neuroectoderm differentiation, Smad signaling was inhibited, allowing Sox2 to relocate to the neuroectoderm gene enhancers and facilitate their expressions, thus triggering the differentiation process.

REFERENCES

1. Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Current protocols in molecular biology, 109(1), 21-29.
2. Du, Y., Liu, Z., Cao, X., Chen, X., Chen, Z., Zhang, X., ... & Jiang, C. (2017). Nucleosome eviction along with H3K9ac deposition enhances Sox2 binding during human neuroectodermal commitment. Cell death and differentiation, 24(6), 1121.
3. Fathi, A., Rasouli, H., Yeganeh, M., Salekdeh, G. H., & Baharvand, H. (2015). Efficient differentiation of human embryonic stem cells toward dopaminergic neurons using recombinant LMX1A factor. Molecular biotechnology, 57(2), 184-194.
4. Huse, M., Muir, T. W., Xu, L., Chen, Y. G., Kuriyan, J., & Massagué, J. (2001). The TGFβ receptor activation process: an inhibitor-to substrate-binding switch. Molecular cell, 8(3), 671-682.
5. Lodato, M. A., Ng, C. W., Wamstad, J. A., Cheng, A. W., Thai, K. K., Fraenkel, E., ... & Boyer, L. A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS genetics, 9(2), e1003288.
6. Mullen, A. C., Orlando, D. A., Newman, J. J., Lovén, J., Kumar, R. M., Bilodeau, S., ... & Young, R. A. (2011). Master transcription factors determine cell-type-specific responses to TGF-β signaling. Cell, 147(3), 565-576.
7. Natsume, T., Kiyomitsu, T., Saga, Y., & Kanemaki, M. T. (2016). Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. Cell reports, 15(1), 210-218.

8. Shi, Y., & Massagué, J. (2003). Mechanisms of TGF-β signaling from cell membrane to the nucleus. cell, 113(6), 685-700.
9. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7(3), 562.
10. Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D. C., DiBiase, A., ... & Li, P. (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell, 147(3), 577-589.
11. Tsankov, A. M., Gu, H., Akopian, V., Ziller, M. J., Donaghey, J., Amit, I., ... & Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. Nature, 518(7539), 344.
12. Voss, T. C., & Hager, G. L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nature Reviews Genetics, 15(2), 69.