SETD3-DEPENDENT GENE EXPRESSION CHANGES DURING ENDODERM DIFFERENTIATION OF MOUSE EMBRYONIC STEM CELLS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EMRE BALBAŞI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

FEBRUARY 2022

Approval of the thesis:

SETD3-DEPENDENT GENE EXPRESSION CHANGES DURING ENDODERM DIFFERENTIATION OF MOUSE EMBRYONIC STEM CELLS

submitted by **EMRE BALBAŞI** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Biology, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşe Gül Gözen
Head of the Department, **Biological Sciences** _____

Assist. Prof. Dr. Nihal Terzi Çizmecioğlu
Supervisor, **Biological Sciences, METU** _____

Examining Committee Members:

Assoc. Prof. Dr. Işık G. Yuluğ
Molecular Biology and Genetics, Bilkent University _____

Assist. Prof. Dr. Nihal Terzi Çizmecioğlu
Biological Sciences, METU _____

Prof. Dr. Mesut Muyan
Biological Sciences, METU _____

Date: 09.02.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Emre Balbaşı

Signature:

ABSTRACT

SETD3-DEPENDENT GENE EXPRESSION CHANGES DURING ENDODERM DIFFERENTIATION OF MOUSE EMBRYONIC STEM CELLS

Balbaşı, Emre
Master of Science, Biology
Supervisor: Assist. Prof. Dr. Nihal Terzi Çizmecioğlu

February 2022, 86 pages

Mouse embryonic stem cells (mESCs) are pluripotent cells that have self renewal capability. They can differentiate into all three primary germ layers: mesoderm, endoderm, and ectoderm during embryonic development. The embryonic development is controlled via spatiotemporal regulation of gene expression changes. The collaborative effects of Wnt, Nodal, BMP signaling pathways help form the primitive streak, and the subsequent definitive endoderm layer in the gastrulating embryo. Deactivation of core pluripotency network, and activation of germ layer specific transcription netwoks are required for this process. This is precisely achieved by chromatin-based regulation. SETD3 is a SET-domain containing methyltransferase that targets both histone and nonhistone proteins. An shRNA screen identified SETD3 as a key factor for mesendoderm commitment of mESCs. *Setd3* knock-out mESC cannot upregulate pioneer transcription factors that initiate mesendoderm differentiation *in vitro*. In this project, we aimed to determine SETD3-dependent gene expression changes during the endoderm differentiation of mouse embyronic stem cells (mESCs) and we performed time-course endoderm differentiation experiments and employed RNA-sequencing to identify differentially expressed genes (DEGs) in the absence of SETD3. Our results indicate a role for

SETD3 in timely downregulation of the pluripotent state, and response to key signaling pathways which leads to the delayed and defective differentiation in its absence.

Keywords: Mouse Embryonic Stem Cells, SETD3, Endoderm Differentiation, Pluripotency.

ÖZ

FARE EMBRİYONİK KÖK HÜCRELERİNİN ENDODERME FARKLILAŞMASI SIRASINDA SETD3 PROTEİNİNE BAĞIMLI GEN İFADESİ DEĞİŞİMLERİ

Balbaşı, Emre
Yüksek Lisans, Biyoloji
Tez Yöneticisi: Dr. Öğr. Üyesi Nihal Terzi Çizmecioğlu

Şubat 2022, 86 sayfa

Fare embriyonik kök hücreleri (EKH) kendi kendilerini yenileyebilen, pluripotent hücrelerdir. EKH'ler embriyonik gelişim sırasında üç ilkel tabaka olan mezoderm, endoderm, ve ektoderme farklılaşabilirler. Embriyonik gelişim gen ifadesi değişimlerinin zaman ve hücrelerin bulundukları yere bağlı olarak çok sıkı bir şekilde kontrol edildiği bir süreçtir. Embriyonik gelişimin gastrulasyon aşamasında, Wnt, Nodal, ve BMP sinyal yolaklarının ortak çalışması ile ilkel çizgi oluşumu, ve sonrasında da endoderm tabakasının oluşması sağlanır. Bu süreçte ana pluripotentlik ağının kapatılarak, ilkel tabakalara özgü ifade ağlarının açılması gereklidir; bu da kromatin temelli regülasyon ile sağlanır. SETD3, SET bölgesi içeren, hem histonları hem de histon olmayan proteinleri hedef alabilen bir metiltransferazdır. Daha önceki bir shRNA taramamızda SETD3'ün mesendoderm farklılaşması için gerekli faktörlerden biri olduğunu belirledik. *Setd3* silinmiş EKH'ler *in vitro* koşullarda mesendoderm farklılaşmasını sağlayacak öncül transkripsiyon faktörlerinin ifadesini arttıramamaktadırlar. Bu projede, endoderm farklılaşması sırasında SETD3'e bağımlı gen ifadesi değişimlerini tespit etmeyi amaçladık. Bu amaçla endoderm farklılaşması deneyleri yaptık ve RNA-sekanslama yöntemini kullanarak SETD3

yokluğunda endoderm farklılaşması sırasında farklı ifade edilen genleri belirledik. Sonuçlarımız, SETD3'ün pluripotentliğin zamanla kapatılmasında ve yokluğunda gecikmiş ve kusurlu farklılaşmaya yol açan anahtar sinyal yollarına yanıt vermedeki rolünü göstermektedir.

Anahtar Kelimeler: Fare Embriyonik Kök Hücreleri, SETD3, Endoderm Farklılaşması, Pluripotentlik.

To the Giants whose shoulders we stand on

ACKNOWLEDGMENTS

The author wishes to express his deepest gratitude to his supervisor Assist. Prof. Dr. Nihal Terzi Çizmecioğlu and for her guidance, advice, criticism, encouragements, and insight throughout the research.

The author would also like to thank Prof. Dr. Mesut Muyan for his suggestions and comments, and for the great talks during the coffee breaks.

The technical assistance of Dr. İhsan Cihan Ayanoğlu and İsmail Güderer are gratefully appreciated.

The friendship and the support of the lab mates Gözde Güven, Ceren Alganatay, and Dersu Sezginmert were always most appreciated.

The author would like to thank Ezgi Gül Keskin for her patience while teaching the techniques, methods, and protocols during the first months of author's training process in the lab.

Finally, the author would like to thank Dersu Sezginmert for her constant mental support on the bad days, and for being the only person who listens without any judgment.

This work is partially funded by Scientific and Technological Research Council of Turkey under grant number TUBİTAK 119Z405

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1 Early Mouse Embryonic Development

The mice, as all mammalians, start their journey of life as a single cell called the zygote. Through serial and controlled cell divisions without increasing the overall cytoplasmic volume, the zygote becomes an unorganized cell mass with eight small identical cells called blastomeres (Aiken et al., 2004). These 8 blastomeres are totipotent stem cells, which can form all embryonic and extraembryonic structures in the developing embryo (Lu & Zhang, 2015). Through the first morphogenetic event called the compaction, the blastomeres become flattened, increasing the overall cell-to-cell contact through adherens junctions between the neighboring blastomeres (White et al., 2016). Due to their position, the blastomeres remaining at the border show less cell-to-cell contact, while the cells in the middle are in contact all around. This polarized state of the blastomeres leads to the formation of the first extraembryonic structure called trophectoderm as an outside layer of the morula, and the apolar cells in the middle give rise to the inner cell mass (ICM), marking the first cell-fate decision (Mihajlović & Bruce, 2017).

The formation of a liquid filled cavity on the embryonic day 3.5 (E3.5) marks the blastula stage of the developing embryo (Manejwala et al., 1989; Wiley, 1984). The early blastocyst contains two types of cells: the pluripotent ICM, and the trophectoderm layer surrounding both the ICM and the cavity. A second cell-fate decision is made at this point of development, and the ICM gives rise to two cell types called the epiblast and the primitive endoderm that separates the epiblast and the cavity on E4.5 (Chazaud et al., 2006). The preimplantation stage of the development ends as the embryo in the late blastula stage is capable of implanting into the uterus (Aiken et al., 2004).

Following the implantation, the morphology of the blastocyst undergoes dramatic changes. The trophectoderm just above the epiblast and the epiblast itself transform into an elongated structure consisting of ectoplacental cone, extraembryonic ectoderm (ExE), a layer of visceral endoderm, and the pluripotent epiblast (Tam & Loebel, 2007). The gastrulation, a key event that starts with the formation of the primitive streak begins to proceed on E6.5. The visceral endoderm layer has critical roles in primitive streak formation and anterior-posterior patterning (Thomas & Beddington, 1996). It resides between the epiblast and the trophectoderm layer, virtually covering the embryo. Some of the cells of the distal visceral endoderm (DVE) starts to migrate towards the anterior side of the embryo, forming the anterior visceral endoderm (AVE). This is the site of the expression of Nodal antagonists *Cer1* and *Lefty1* (Yamamoto et al., 2004), while the posterior side of the DVE expresses *Wnt3* (Rivera-Pérez & Magnuson, 2005). *Wnt3* expression is controlled by Nodal and BMP signaling (Ben-Haim et al., 2006), and the inhibition of Nodal signaling through its antagonists results in decreased *Wnt3* expression at the anterior epiblast, providing an asymmetrical Wnt and Nodal signaling gradient between the anterior and the posterior sides. This asymmetry is one of the key factors that drive the formation of primitive streak strictly at the posterior side of the developing mouse embryo, as the high expression of the mesendoderm marker *Brachyury (T)* remains exclusive to the posterior epiblast (Rivera-Pérez & Magnuson, 2005; Yamaguchi et al., 1999).

The epiblast consists of pluripotent stem cells, which can give rise to the primary germ layers during gastrulation depending on their initial positions and extracellular signals from the neighboring cells. Through migration, epiblast cells on the posterior side ingress through the primitive streak to form the mesendodermal lineages, with the dual potential to form both the mesoderm and the definitive endoderm. The remaining epiblast cells differentiate into ectodermal lineages to form the surface ectoderm and the neural progenitors.

The migration of the epiblast is dependent on the epithelial-to-mesenchymal transition (EMT) to ingress into primitive streak to form the mesoderm, initiated by

high Wnt/β-catenin, FGF, and TGF-β signaling, and the consequent activation of the EMT transcription factors such as *Snai1-2* (Cano et al., 2000; Carver et al., 2001), *Mesp1-2* (Kitajima et al., 2000; Lindsley et al., 2008), and *Zeb1-2* (Lamouille et al., 2014; Peinado et al., 2007). Though it was widely believed that the definitive endoderm progenitors within the epiblast also go through the same EMT process and a following mesenchymal-to-epithelial transition (MET) to form the definitive endoderm layer, recent data show that the definitive endoderm forms independently of the EMT-MET cycle (Scheibner et al., 2021).

The formation of the primary germ layers ectoderm, mesoderm and definitive endoderm marks the end of the gastrulation and the early developmental stages of the mouse embryo.

For the detailed view of the pathways mentioned above, *see* Appendices G and H.

1.2 Embryonic Stem Cells

Embryonic stem cells (ESC), with their unique capabilities of self-renewal and pluripotency, are one of the versatile tools the science have for modeling and studying a variety of subjects covering a range from as early as the initial steps of the embryonic development to the Parkinson's disease seen in the elderly years of the human life (H. Kim et al., 2019). Pluripotency is defined as the capability of forming all three germ layers called ectoderm, mesoderm, and definitive endoderm. These germ layers are the progenitor cells of all the tissue and cell types in the adult body. They are derived from the inner cell mass (ICM) of the developing blastocyst and are grown *in vitro* conditions that support and maintain their pluripotent state (Martin, 1981). There are two widely used methods to grow and maintain ESCs *in vitro*. The conventional method requires the co-culturing of ESCs with a supportive feeder layer of mitotically inactivated cells in a high-serum and leukemia inhibitory factor (LIF) containing medium (Lin & Talbot, 2011). A most recent method uses

two inhibitors for GSK3β and MEK1-2, and LIF in low or no-serum containing media without the need of the feeder cells (Wray et al., 2010).

ESCs, when grown *in vitro*, grow in colonies unlike many other mammalian cells, which are observed as dome shaped and bright colonies under light microscope. This morphology is one of the important features of healthy, undifferentiated ESCs. Suboptimal media conditions may lead to uncontrolled differentiation due to loss of pluripotency of the ESCs.

The pluripotency of the ESCs is controlled by a core transcriptional network involving OCT4 (POU5F1), SOX2, and NANOG (Boyer et al., 2005; X. Chen et al., 2008; Orkin & Hochedlinger, 2011) (*see* Appendix I for detailed pathway view). These three transcription factors (TFs) control each other's and as well as the secondary transcription factors' expression to maintain the pluripotency, while repressing differentiation related genes.

Their potential to form the primary germ layers make them a great tool for research. By mimicking the microenvironment ESCs reside in during embryonic development, the ESCs can be efficiently directed towards any primary germ layer cells and further into organ progenitors *in vitro* (Gadue et al., 2006; Irion et al., 2008; Kubo et al., 2004; Martello et al., 2013; Ying et al., 2003; Ying & Smith, 2003), opening great possibilities for regenerative or personalized medicine.

1.3 Epigenetic Regulation

In eukaryotes, the DNA is found within the nucleus, in a compacted state to compensate for its large structure, to provide further control over how the genetic code is read and to ensure that cellular functions are governed from one center. One level of such compaction is carried out by nucleosomes. The nucleosomes are DNA-histone protein structures, similar to beads on a string, considered as the building blocks of the chromatin. Histones are small and basic proteins; they are highly conserved both in length and structure among the eukaryotes (Eirín-López et al.,

2009). The core histone proteins form heterodimers, H2A-H2B and H3-H4. Two of each heterodimer combine to form the octameric histone complex. Due to their basic nature, they can easily bind to acidic DNA molecules. Almost two loops (147 bp) of DNA are wrapped around the histone octamer in a left-handed helix to form the nucleosomes (Luger et al., 1997). In between two nucleosomes, there sometimes resides H1 as the linker histone, providing a further level of compaction to the long strings DNA (P. J. J. Robinson et al., 2006).

The chromatin exists as a continuum of differentially condensed regions. It contains actively transcribed regions and regions with very little or no gene expression. Heterochromatin regions are the densely packed, mostly inaccessible parts of the chromatin. Due to their tightly packed structure, gene expression is low in these regions (Chiarella et al., 2020). Even though these regions do not have actively transcribed genes, they play important roles in preserving the integrity of the genome as they are found on the telomeric regions of the DNA, and in centromeres (Bühler & Gasser, 2009). On the other hand, euchromatin is the name for the actively transcribed regions of the chromatin. These regions are loosely packed, providing multiple docking sites for DNA-binding proteins, and hence, have the most active gene loci (Chiarella et al., 2020). Euchromatin and heterochromatin regions dynamically change during development and can have variations among different cell and tissue types. Post-translational modifications of histones, DNA methylation, and chromatin remodelers play important roles in dynamic regulation of the chromatin structure (J. A. Kim et al., 2019; G. Li & Reinberg, 2011), providing a variety of functions to genetically identical cells in different tissues and organs.

Core histones have an N-terminus tail, poking out the core, containing several serine, arginine, and lysine residues available for post-translational modifications that can determine and dynamically change the function of the nucleosome as well as the accessibility of the surrounding DNA, altering the expression patterns of the genes in that region (Luger & Richmond, 1998). These modifications are carried out by protein complexes containing “writer”, “eraser”, and “reader” proteins that can determine the specificity, location, and level of the modification (Biswas & Rao,

2018). Writer proteins, as the name implies, are responsible for depositing the modifications on the designated residues on the histones. Some of the writer proteins are categorized as histone acetyltransferases (HATs), histone methyltransferases (HMTs), kinases, and ubiquitin ligases, responsible for the addition of acetyl, methyl, and phosphate groups, and ubiquitin molecules, respectively. The eraser enzymes are responsible for the removal of the modifications from the histone tails. Histone deacetylases (HDACs), histone demethylases (HDMs), and phosphatases remove the acetyl, methyl, and phosphate groups, respectively. Both writer and eraser proteins work with reader proteins in a complex that can recognize a residue to determine the specificity and the location of the modification. Additionally, the recognition of specific modifications on histones by the readers ensure proper recruitment of different members of nuclear signaling network to the chromatin, mediating crucial functions such as transcription, replication, and DNA damage response (Musselman et al., 2012). The reader proteins recognize specific epigenetic marks: for example, while the bromodomain containing readers recognize previously acetylated lysine residues of the histones, the chromodomain containing ones show higher affinity towards methylated histone residues (Gardner et al., 2011).

Similar to histones, the DNA itself is known to undergo one type of modification: the DNA methylation. CpG islands on the DNA are subjected to DNA methylation by DNA methyltransferases (DNMTs) (Bird, 1992). Cytosine residues on the CpG islands can be methylated by DNMTs, generally resulting in an inactive gene expression (Jones, 1996). DNA methylation is also a reversible modification. The methyl group from methylated cytosine can be sequentially removed by the Tet enzymes, the equivalent of a DNA demethylase (Wu & Zhang, 2014). The combination and the level of these modifications determine the local transcriptional activity of the chromatin. In general, high levels of DNA methylation is associated with low levels of gene expression. While histone acetylation correlates with active gene expression (Hebbes et al., 1988), such generalization cannot be done for histone methylation. For example, histone 3 lysine 4 trimethylation (H3K4me3) and lysine 36 trimethylation (H3K36me3) found around the promoters and gene ends,

respectively, can be associated with a transcriptionally active state of that gene. On the other hand, histone 3 lysine 9 trimethylation (H3K9me3) and lysine 27 trimethylation (H3K27me3) generally mean a repressed gene expression (Vermeulen et al., 2010).

1.4 Epigenetic Mechanisms on mESCs

The control of the pluripotent state is a complex and highly regulated process in ESCs. Uncontrolled exit from pluripotency, and activation of differentiation specific genes in the wrong time and sequence can lead to serious developmental defects, requiring a spatiotemporal control mechanism to prevent such problems. Epigenetic regulation plays an important role in the control of pluripotency and the exit from it.

Overall, the DNA methylation levels in ESCs are low compared to terminally differentiated cells. This is expected, as the ESCs have the potential to form any cell types, and the identity of the ESC will be defined through the signals from the environment, almost all genes should be kept ready to be expressed. The core pluripotency genes, *Oct4* and *Nanog* have unmethylated promoters, associated with the high expression levels in ESCs (Mitsui et al., 2003). A similar trend is also seen in the histone modification level. The lineage specific genes are kept in a bivalent state, simultaneously containing both activatory and inhibitory modifications (Azuara et al., 2006; Bernstein et al., 2006). That is, such genes are not expressed or totally repressed in the pluripotent state, but they are poised to be expressed in the case of differentiation (Boyer et al., 2006). On the other hand, the genes in the core pluripotency network are marked with H3K4 and H3K36 methylations, indicating high levels of expression. Antagonistic effects of Polycomb group complex (PcG) and Trithorax group complex (TrxG) provide the fine tuning of the decision between self-renewal and differentiation that ESC face all the time. While PcG represses the lineage specific genes by providing H3K27 methylations, TrxG complex methylates H3K4 of the pluripotency related genes (Brand et al., 2019). Both of these complexes

contain a SET domain containing methyltransferases, that have functions in both maintaining the pluripotency and differentiation (Paranjpe & Veenstra, 2015).

1.5 SET Domain Containing Proteins

Histone methylation has two variations with three different levels. It can be either lysine specific, or arginine specific. Both residues can be found in mono-, di-, trimethylated states. The lysine specific methyltransferases (KMTs) act through SET domain (Su(var)3-9, Enhancer-of-zeste (E(z)), Trithorax (Trx)), or it can be SET domain independent as in the case of Dot1 proteins (Gao & Liu, 2007). SET domain is an evolutionarily conserved, 130-140 amino acid long protein domain that functions as the catalytic domain for the transfer of one or more methyl groups from SAM (S-adenosylmethionine) to lysine residues on histones, or non-histone proteins (Jenuwein et al., 1998; Rea et al., 2000).

The SET domain containing methyltransferases are divided into three subgroups: SET1, SET2, and SUV39 families. Well known SET1 family methyltransferases are EZH1 and EZH2, with a SET domain following a pre-SET domain. SET2 family lysine methyltransferases contain a SET domain followed by post-SET and AWS domains required for nuclear receptor binding to the SET domain. This family includes NSD1-3, SETD2 and SMYD family methyltransferases. The last family of SET domain containing KMTs is the SUV39 family, which contains SUV39H1, SUV39H2, G9a, GLP, ESET, and CLLL8 enzymes with a pre-SET domain (Rea et al., 2000).

Just like most epigenetic regulators, SET domain containing KMTs work in a complex that determine the specificity, location, and the level of the modification. Polycomb (PcG) repressive complexes, Trithorax group complexes, and COMPASS (complex of proteins associated with Set1) complexes are some of the examples that contain an active methyltransferase with a SET domain (Schuettengruber et al., 2017). In mammals, one of the PcG complexes called PRC2 complex consists of

EZH1/2, SUZ12, EED164 and RBBP4/7 core proteins. While EZH1/2 are the H3K27 KMTs with a SET domain, other members allow the complex to recognize H3K27me1 and H3K36me3 residues and to bind the DNA (Blackledge et al., 2014).

The Set1/COMPASS is the first H3K4 methyltransferase complex identified in yeast as the homolog of Trithorax in Drosophila (Miller et al., 2001; Shilatifard, 2012). Since then, many complexes resembling the COMPASS (COMPASS-like) have been identified in mammals including the SET1/COMPASS, containing KMTs SET1A/B, HCF1, WDR82 and CFP1 in its core (Schuettengruber et al., 2017). SET1A has been shown to mediate H3K4 methylation in the inner cell mass (ICM) of the developing mouse embryo, and it is required for the proliferation, and pluripotency related gene expression in epiblast (Bledau et al., 2014). In another study, wild-type and catalytically dead SET1A expressing mESCs showed minor differences in ESC state. However, when differentiated, these cells showed lower H3K4me3 on the promoters of differentiation related genes, and impaired proliferation rates (Sze et al., 2017). Overall, these results indicate that methyltransferase-independent activity of SET1A is dispensable for the self-renewal, while methyltransferase-dependent activity is crucial for pluripotency of mESCs.

1.6 SETD3 Histone Methyltransferase

SETD3 is a histone and non-histone protein methyltransferase. It contains two domains: a catalytic SET domain, and a Rubis-subs-bind (Rubisco LSMT) domain for substrate binding (Eom et al., 2011). It was shown to be responsible for H3K4 and H3K36 methylations both *in vitro* and *in vivo* (Z. Chen et al., 2013; Eom et al., 2011). It is highly expressed in muscle tissues and some of the internal organs including stomach, colon, and small intestine. SETD3 was shown to play a critical role in muscle differentiation by directly interacting with MyoD and activating muscle differentiation specific genes in myoblasts (Eom et al., 2011). More recent

studies identified an actin histidine methyltransferase activity of SETD3 (Guo et al., 2019; Kwiatkowski et al., 2018; Wilkinson et al., 2019).

SETD3 has been proposed as an epigenetic marker for renal cell carcinoma (Z. Chen et al., 2013). Additionally, SETD3 was shown to methylate FOXM1, which is associated with *Vegf* expression. SETD3 negatively regulates *Vegf* expression by physically interacting with FoxM1 on *Vegf* promoter (Cohn et al., 2016; Jiang et al., 2018), which is the key factor in angiogenesis. In cervical cancer, SETD3 overexpression leads to decreased KLC4 expression, which in turn allows the sensitization of cancer cells to radiotherapy (Q. Li et al., 2019). In colon cancer, SETD3 is shown to positively regulate apoptosis by physically interacting with p53 and controlling its recruitment to its target genes (Abaev-Schneiderman et al., 2019). Apart from SETD3's role in cancer, it was shown that host SETD3 is required for the RNA replication in the viral life cycle of enteroviruses, independent of its methyltransferase activity (Diep et al., 2019). Moreover, the overexpression of CXCR5 in $CD4^+$ T cells of systemic lupus erythematosus (SLE) patients has been linked with increased H3K4me3 and H3K36me3 modifications in the protomer of CXCR5 due to increased SETD3 expression (Liao et al., 2020).

1.7 Preliminary Data

In our previous study, a pooled shRNA screening for the epigenetic factors that affect mesendoderm differentiation was conducted (Terzi Cizmecioglu et al., 2020). SETD3 was among the important epigenetic factors, leading to decreased endoderm differentiation efficiency upon knockdown. CRISPR/Cas9 mediated deletion of *Setd3* in mouse ESCs and endoderm differentiation efficiency of setd3Δ mESCs into mesendoderm was investigated (Figure 1.1).

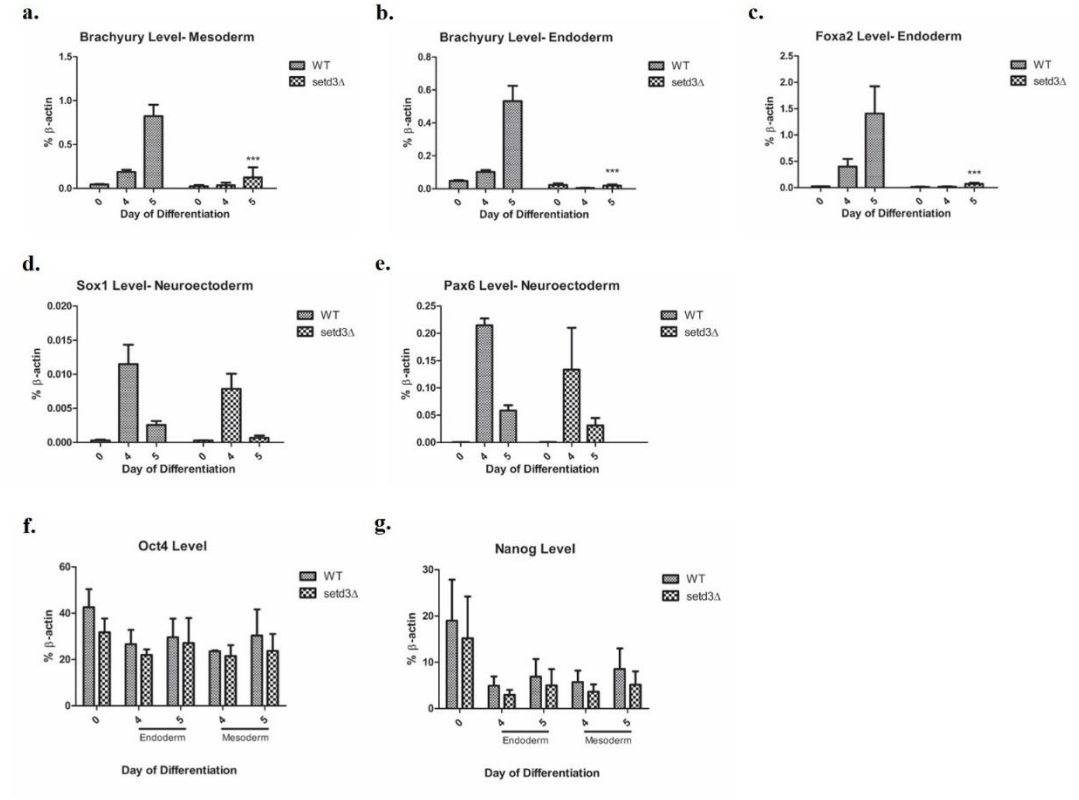

Figure 1.1 qRT-PCR analysis of wild-type (WT) and setd3Δ cells in mESC state and differentiation towards endoderm, mesoderm, and ectoderm. SETD3 loss leads to defective meso/endoderm differentiation. **a**. *Brachyury* (*Bry*) expression levels in mesoderm differentiation. **b-c**. *Bry* and *Foxa2* expression levels in endoderm differentiation. **d-e**. *Sox1* and *Pax6* expression levels in ectoderm differentiation. **f-g**. *Oct4* and *Nanog* expression levels in meso/endoderm differentiation. Time-course differentiation experiments were done for 5 days. Error bars indicate ±SEM of three biological replicates. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. **: p-value < 0.01, ***: p-value < 0.001.

The results showed that, when setd3Δ mESCs were differentiated towards mesoderm and endoderm, they failed to express mesendoderm specific marker *Brachyury (Bry, T)* and endoderm specific marker *Foxa2*, consistent with the pooled shRNA screening results. Neuroectodermal differentiation normally proceeded in the absence of SETD3, as *Sox1* and *Pax6* expression levels were similar to the wild type cells. Similarly, expression of the core pluripotency transcription factors *Oct4* and *Nanog* were similar among wild type and setd3Δ cells both in mESC state and during

meso/endoderm differentiation, suggesting that the deficiency in the absence of SETD3 is confined only to mesendodermal lineages.

1.8 Aim of the Study

mESCs are pluripotent cells, that is they can form all primary germ layers. However, in the absence of SETD3, they fail to differentiate towards mesoderm and endoderm. We hypothesized that this defect stems from SETD3's nuclear function. SETD3 is a histone methyltransferase that provides H3K4me2-3 and H3K36me2-3 modifications; thus, it may directly or indirectly control the expression of genes during the exit from pluripotency and differentiation towards endoderm in mESCs. Our aim was to identify these genes, whose expression were directly or indirectly controlled by SETD3. Upon RNA-sequencing of wild type and setd3Δ cells both in pluripotent state and during endoderm differentiation, we identified thousands of differentially expressed genes (DEGs) among wild type and setd3Δ cells. We focused on key pathways in pluripotency maintenance and endoderm differentiation; and selected the DEGs with a significant expression difference for further analysis. Furthermore, we rescued the defective phenotype by ectopic expression of SETD3.

CHAPTER 2

MATERIALS AND METHODS

2.1 mESC Culture and Endoderm Differentiation

Primary BALB-c MEFs obtained from Koç University were grown and expanded in MEF medium in gelatin-covered 15-cm tissue culture plates at 37°C, 5% CO_2. When 80-90% confluency was achieved, MEFs were passaged using Trypsin EDTA Solution B (0.25%), EDTA (0.05%), with Phenol Red (Cat. No.: BI03-052-1B, Biological Industries). After 4-5 passages, MEFs were treated with 10 µg/mL Mitomycin C lyophil. research grade (Cat. No.: SE2980501, Serva) in MEF medium (*see* Appendix A for recipe) to block cell division. These MEFs were referred to as mitoMEFs.

CJ9 (normal, wild-type, WT) mESCs and *Setd3* gene lacking mESCs (setd3Δ, knock-out, KO) were obtained from Prof. Stuart Orkin's laboratory in Boston Children's Hospital. Wild type, setd3Δ, and setd3Δ+pEF1α-Setd3 mESCs were grown and maintained in high-serum containing ESC medium (*see* Appendix A for recipe) on mitoMEFs in gelatin covered 6-well cell culture plates at 37°C, 5% CO_2. mESCs were passaged every 2-3 days when healthy, bright, and dome-shaped mESC colony morphology was achieved. mESC colonies were detached from the plate bottom using Trypsin EDTA Solution B.

For endoderm differentiation, mESC colonies were detached from the well bottom using Trypsin EDTA Solution B and single cell suspension was obtained. mitoMEFs were removed by 45 minutes incubation of the cell suspension in a gelatin coated one well of a 6-well plate and collecting the mESCs yet to be attached. mESCs were counted via Trypan Blue Stain (0.4%) for use with the Countess™ Automated Cell Counter (Cat. No.: T10282, Invitrogen). 7.5 x 10^5 live mESCs were seeded in a 10-

cm Petri dish in Serum-free Base Differentiation Medium (*see* Appendix A for recipe), and incubated at 37°C, 5% CO_2 for 48 hours. During this period, mESCs formed embryoid bodies (EBs) in suspension. After 48 hours, the EBs were collected in IMDM (Cat. No.: 21980032, Gibco) and dissociated into single cells using Accutase cell detachment solution (Cat. No. SCR005, Millipore). Cells were counted via Trypan Blue staining. 5.0 x 10^5 cells were seeded in Serum-free Endoderm Differentiation Medium (*see* Appendix A for recipe) in 6-cm petri dishes, and incubated at 37°C, 5% CO_2.

2.2 Lipofection (Lipid Transfection) and Re-expression of SETD3 in setd3Δ mESCs

setd3Δ mESCs were grown in low-serum containing 2i4 medium (*see* Appendix A for recipe) in gelatin coated wells to prevent MEF contamination during the transfection process. Once the healthy mESC morphology was achieved, the colonies were detached using TrypLe Express Enzyme (1X), phenol red (Cat. No.: 12605-010, Gibco) and single cells were obtained. mESCs were counted using Trypan Blue staining. 7.5 x 10^5 mESCs were seeded into a gelatin coated one well of a 6-well plate in 2i4 medium. In this stage, just before attaching to the bottom of the well, mESCs were transfected using Lipofectamine 3000 (Cat. No.: L3000008, Thermo) with 2500 ng pEF1αFLBIOsetd3-puro plasmid (obtained from Prof. Stuart Orkin's laboratory in Boston Children's Hospital) and incubated at 37°C, 5% CO_2. Two days later, the colonies were detached using TrypLe Express and single cell suspension was obtained. mESCs were seeded into a gelatin coated 10-cm cell culture plate in 2i4 medium supplemented with 1 µg/mL puromycin for selection. The cells were incubated at 37°C, 5% CO_2 for a week, and the medium was refreshed every other day. At the end of one week, the remaining mESC colonies were passaged in 2i4 medium with 1/6 of the initial puromycin concentration and the selection process continued for 3-4 passages. After the selection process, these mESCs were referred to as setd3Δ+pEF1α-Setd3 mESCs.

For further experiments, setd3Δ+pEF1α-Setd3 mESCs needed to be adapted to high-serum containing ESC medium. 2.5-3.0 x 10^5 setd3Δ+pEF1α-Setd3 mESCs were seeded into gelatin coated 6-well plates on mitoMEFs in ESC medium. The cells maintained in ESC medium on mitoMEFs until healthy mESC colony morphology was achieved.

2.3 RNA Sample Preparation, Isolation, and cDNA Synthesis

After the mitoMEFs were removed, mESCs were counted using Trypan Blue staining. 1.0 x 10^6 live mESCs were lysed in 1 mL TRIzol™ Reagent (Cat. No.: A33251, Invitrogen) and stored at -20°C.

On days 2,3 and 4 of endoderm differentiation, EBs were dissociated in Accutase solution, and the single cells were counted using Trypan Blue stain. 1.0 x 10^6 live cells were lysed in 1 mL TRIzol solution and stored at -20°C.

RNA isolation was done using RNeasy Plus Mini Kit (Cat. No.: 74134, Qiagen) from the lysates in TRIzol reagent following the protocol of the manufacturer. For RNA samples sent to RNA-sequencing, DNA was digested on column, using RNase-free DNase set (Cat. No.: 79254, Qiagen) instead of gDNA Eliminator columns provided with the kit. RNA concentrations were measured using NanoDrop (Cat. No.: MN-913, MaestroGen). For RNA-sequencing samples, RNA concentrations, RIN values, and rRNA ratios were measured by the 2100 Bioanalyzer Instrument (Cat. No.: G2939BA, Agilent). Using the concentrations measured by Bioanalyzer, RNA samples with 100 ng/µL concentration in 60 µL total volume were sent to Macrogen Europe in Holland for RNA-sequencing and TruSeq Library creation.

1000 ng RNA was converted to cDNA using iScript cDNA Synthesis Kit (Cat. No.: 1708891, Bio-Rad) by following the instructions provided by the manufacturer.

2.4 Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR)

qRT-PCR analyses were done using SsoAdvancedTM Universal Inhibitor-Tolerant SYBR® Green Supermix (Cat. No.: 1725018, Bio-Rad) in 10 µL reaction volume in LightCycler 96 (Cat. No.: 05815916001, Roche) thermal cycler. Expression levels were normalized to *β-actin* expression levels in each sample using the formula $2^{(-\Delta Ct)}$ x 100. SETD3 deletion does not affect *β-actin* expression levels, which are validated by differential expression analysis of WT/KO samples (FDR values are greater than 0.05 on all days). The list of primers used for qRT-PCR analyses are shown in the appendices (*see* Appendix C).

2.5 Protein Sample Preparation and Isolation

Wild type and setd3Δ+pEF1α-Setd3 mESCs grown in 2i4 medium were detached from culture plate bottom using TrypLe Express and single cell suspensions were obtained. mESCs were counted using Trypan Blue staining. 5.0 x 10^6 live mESCs were transferred to 1.5 mL reaction tubes, and washed 3 times in 1X cold PBS, and snap frozen in liquid nitrogen. Frozen samples were stored at -80°C.

Wild type, setd3Δ, and setd3Δ+pEF1α-Setd3 mESCs grown in ESC medium were detached from culture plate bottom using Trypsin and single cell suspensions were obtained. mitoMEFs were removed and the mESCs were counted using Trypan Blue staining. 2.0 x 10^6 live mESCs were transferred to 1.5 mL reaction tubes, and washed 3 times in cold PBS, and snap-frozen in liquid nitrogen. Frozen samples were stored at -80°C.

Frozen 5.0 x 10^6 cells were separated into nuclear and cytoplasmic fractions using Universal Magnetic Co-IP Kit (Cat. No.: 54002, Active Motif) following the instructions of the manufacturer. Protein concentrations were measured using Pierce™ BCA Protein Assay Kit (Kat. No.: 23227, Thermo) following the instructions of the manufacturer. The isolation and concentration measurements

were conducted by Dersu Sezginmert and Ceren Alganatay due to their better knowledge and experience on the protocols.

2.6 SDS-PAGE and Western Blotting

Snap-frozen 2.0 x 10^6 cells were thawed on ice and resuspended in 80 µL 2X Laemmli solution (950 µL 2X Laemmli Buffer (Cat. No.: 1610737, Bio-Rad) + 50 µL β-Mercaptoethanol (Cat. No.: M-6250, Sigma)). The suspension was boiled at 95°C for 5 minutes and centrifuged at 10000 rcf for 30 seconds. Incubated on ice for 1 minute, and 20 µL of the supernatant (4.0 x 10^5 cells) was loaded onto %12 SDS-PA gel and ran at constant 100V.

10 µg protein from nuclear and cytoplasmic fractions of wild type and setd3Δ+pEF1α-Setd3 mESCs was loaded onto %12 SDS-PA gel and run at constant 100V.

Protein bands on the SDS-PA gel were transferred onto nitrocellulose membrane using semi-dry transfer method with Bio-Rad TransBlot Turbo Transfer System (Cat. No.: 1704150, Bio-Rad). The blocking was done in 5% skimmed milk in 1X TBS. The membrane was cut into smaller pieces according to protein band sizes, and the pieces were incubated overnight with anti-SETD3, anti-GAPDH and anti-H3 antibodies at 4°C. The membranes were washed 3 times in 1X TBS-T and then incubated in HRP conjugated secondary antibodies for 1 hour at room temperature. The membranes were washed 3 times in TBS-T, and membrane images were taken using Clarity ECL Substrate (Cat. No.: 1705060, Bio-Rad). The list of antibodies used for western blotting were shown in the appendices (*see* Appendix D).

2.7 RNA sequencing and Library Preparation

RNA isolated from the three replicates of wild type and setd3Δ mESCs and from the days 2,3, and 4 of endoderm differentiation were sent to Macrogen Europe for RNA

sequencing and library formation. A total of 24 RNA samples were delivered to the company on dry ice. Each sample contained more than 1 µg of RNA, with RIN values greater than 9, and rRNA ratios greater than 1.1 as measured by Bioanalyzer. All samples passed the initial quality control.

4.0 x 10^7 reads per sample were obtained, and Illumina TruSeq stranded mRNA libraries were created. All libraries passed the quality control.

2.8 Bioinformatic Analyses

The obtained Truseq stranded mRNA libraries were delivered to Gen-era, and bioinformatic analyses were conducted by the company using tools and methods below:

After the reading process, the FASTQC tool (Babraham Bioinformatics, USA, http://www.bioinformatics.babraham.ac.uk) was used for the quality control of the acquired data. During the sequencing process, poor quality base reads and possible adaptor-index contaminations in raw data reads were cropped to prevent bias in the following analyses, using the Trimmomatic tool (Bolger et al., 2014). The HISAT2 tool (Trapnell et al., 2009) was used for post-crop alignment. For this purpose, Mus musculus reference genome M25 (GRCm38.p6) was used as the standard reference genome. Ensembl dataset was used for gene, exon, and transcript information in post-alignment annotation. After alignment, the number of reads on each transcript was calculated and then normalized to the total number of reads. The Subread tool (Y. Liao et al., 2013) was used to determine the number of reads for the transcriptome elements. R:edgeR tool (M. D. Robinson et al., 2009), and the R::limma tool (Ritchie et al., 2015) were used for normalization and filtering of reads per gene, and to identify genes with varying expression between groups (Differentially Expressed Genes), respectively. R scripts were used in statistical comparison studies within and between groups and data visualization applications.

2.8.1 Integrative Genomics Viewer (IGV) Analysis

The BAM files obtained from the RNA-seq were submitted to Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011) software to visualize transcript alignments on the desired gene loci. IGV analysis was used as a checkpoint before the qRT-PCR validations to confirm whether the differential expression of a few selected genes can be validated. For this purpose, Mus musculus genome 10 was used. The BAM files obtained from RNA-sequencing of the first replicate of each experimental group were loaded to the IGV software and the gene loci for *Bry*, *Foxa2*, *Cer1*, *Dkk1*, *Smad7*, and *Bmp7* were visualized.

2.8.2 Short Time-series Expression Miner (STEM) Analysis

Short Time-series Expression Miner (STEM) analysis is used for the expression change analysis in time-course experiments, usually to analyze microarray data (Ernst & Bar-Joseph, 2006). However, any data from an experimental setup with more than three time points, different doses of a drug, etc. can be analyzed using STEM. Basically, STEM clusters the genes into expression profiles with similar trends over the course of the experiment.

Obtained transcript counts from RNA-sequencing were converted to RPKM (Reads Per Kilobase of transcript per Million mapped reads) values using the formula "(109 * C) / (N * L)", where the "C" is the number of reads mapped to a gene, "N" is the total mapped reads in the replicate, and "L" is the gene length in base-pairs for a gene. RPKM values of 3 replicates were averaged. The list of genes with the RPKM values were submitted to the STEM java software. RPKM values were log-normalized using the STEM software. Genes with more than one missing value for four time-points were filtered out from the analysis. If the absolute expression change of a gene between any time point is lower than 0.585 ($=\log_2(1.5)$), then it is filtered out as well.

For pathway enrichment analysis (Gene Ontology (GO)), only the Biological Processes were included. To prevent too general terms to be enriched in the analysis, minimum GO hierarchy level was set to 6.

The results were visualized using REVIGO (http://revigo.irb.hr) (Supek et al., 2011).

2.8.3 Transcription Factor Enrichment Analysis via ChEA3

ChEA3 (https://maayanlab.cloud/chea3/#top) is a web-based transcription factor (TFs) enrichment analysis tool that can identify the TF network controlling the expression of a used-submitted list of genes by integrating the data from multiple sources including ENCODE, ReMap, GTEx, ARCHS4, Enrichr, and ChIP data from literature (Keenan et al., 2019).

The list of significantly differentially expressed genes (DEGs) among wild type and setd3Δ cells on each day of differentiation was submitted to the web-based ChEA3 tool to determine the TFs controlling the expression of genes in the absence or presence of SETD3 during endoderm differentiation. DEGs were submitted separately for upregulated and downregulated genes. TF networks, constructed by the upregulated genes in wild type cells, containing top 15 or 20 TFs were visualized depending on the compactness of the generated network.

CHAPTER 3

RESULTS

3.1 Detection of Transcriptomic Effects of SETD3 Protein via Total RNA-Sequencing

3.1.1 Growth and Differentiation of Wild Type and setd3Δ mESCs

Our preliminary data suggested that SETD3 methyltransferase enzyme is required for endoderm differentiation of mESCs. To understand the effects of SETD3 deficiency, wild-type (WT, normal) and setd3Δ mESCs were grown on mitotically inactivated MEFs (mitoMEFs). When the healthy mESC morphology was achieved, the mESCs were detached from the plate bottom and the mitoMEFs were removed to obtain a pure population of mESCs. Using the previously adapted endoderm differentiation method, wild type and setd3Δ mESCs were differentiated to the definitive endoderm for four days. Embryoid body (EB) formation was observed starting from the 2nd day of differentiation. Three independent biological replicates were obtained using the same differentiation method.

3.1.2 Validation of Endoderm Differentiation via qRT-PCR

On the mESC state (day 0) and from the 3rd and the 4th days of endoderm differentiation, 1.0 x 10^6 cells were lysed in TRIzol solution. Using these lysates, total RNA from days 0, 3 and 4 of all three replicates were isolated. 1000 ng RNA was converted to cDNA and qRT-PCR analysis for the mesendoderm marker *Brachyury* (*Bry*, *T*) and the definitive endoderm marker *Foxa2* was done on each day to validate the endoderm differentiation efficiency (Figure 3.1).

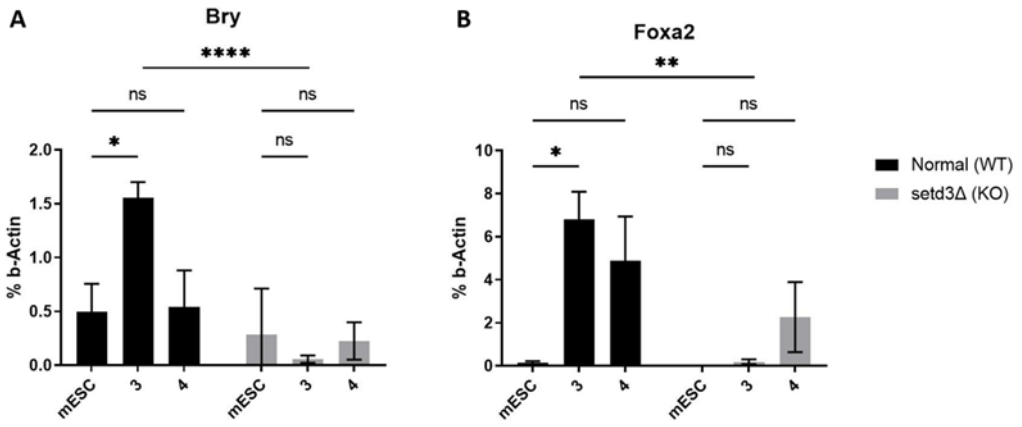

Figure 3.1 qRT-PCR analyses of wild type and setd3Δ cells at both mESC state and throughout the differentiation. Relative expression levels of the endoderm markers **A**. *Bry* and **B**. *Foxa2*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of three independent biological replicates. mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ****: p-value<0.0001, ns: not significant.)

The qRT-PCR results showed that when wild type mESCs grown on mitoMEFs were differentiated to definitive endoderm, *Bry* and *Foxa2* expression levels simultaneously peaked on the third day. While the wild type cells successfully differentiated in four days, setd3Δ mESCs failed to do so. As a result, the previously observed defective phenotype was observed again.

3.1.3 TruSeq Library Formation and RNA-sequencing

setd3Δ mESCs failed to differentiate to definitive endoderm in four days. To understand the underlying mechanism of this defect, the gene expression changes occurring in the time frame between the mESC state, and the peak of differentiation were investigated using RNA-sequencing.

The isolated total RNA from mESCs and the days 2, 3 and 4 of the endoderm differentiation were analyzed using Bioanalyzer to determine the RNA quality and

concentration for RNA sequencing. The results of the analysis are shown in appendices (*see* Appendix E).

The expected parameters for RNA sequencing were reached. Using the concentrations determined by Bioanalyzer, more than 1 µg RNA sample for each day was sent to Macrogen Europe on dry ice. All samples passed the quality control and the TruSeq stranded mRNA libraries were created by Macrogen Europe. Following the successful quality control phase, the RNA sequencing of the libraries was conducted by Macrogen Europe. The quality control results of the libraries were shown in appendices (*see* Appendix F).

3.2 Comparative Analyses of RNA-seq Libraries of Wild Type and setd3Δ Cells

3.2.1 Hierarchical Clustering Analysis (HCA)

Hierarchical clustering analysis was done on 24 groups including day 0, 2, 3 and 4 samples from three biological replicates. The expression levels of all transcripts were normalized, and each sample was compared with 23 other samples to determine the holistic changes. The results were shown on a dendrogram (Figure 3.2).

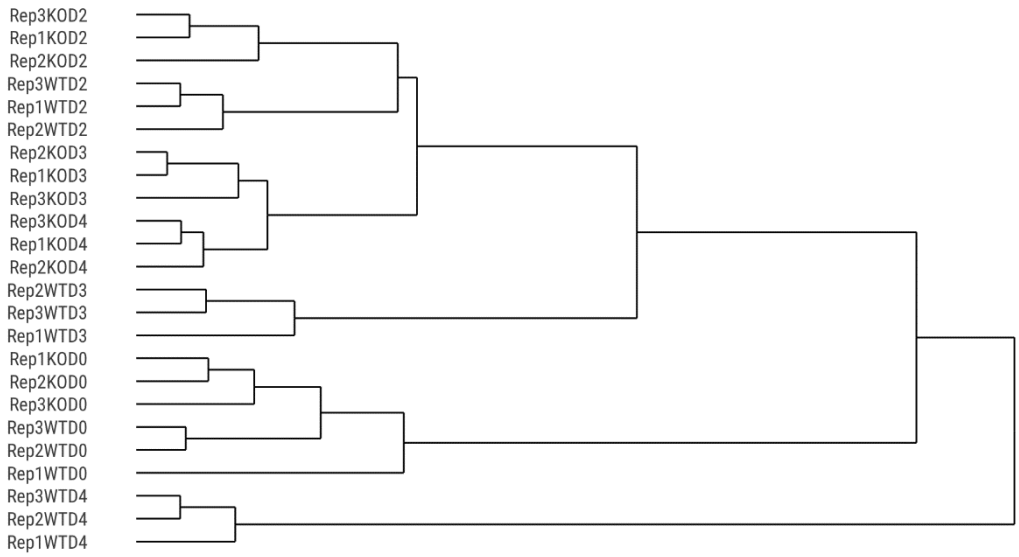

Figure 3.2 Hierarchical clustering analysis (HCA) results. For 24 samples containing 8 groups and 3 replicates, Euclidian distance was measured, and clustering was done via ward.D2 method using normalized reads. Three biological replicates for each group clustered on the same arm. (Rep: Replicate, D0: mESC. D2, D3, D4: Endoderm differentiation days 2, 3, and 4.)

All three replicates of the same day's sample grouped together, meaning that the experimental variation amongst the replicates was minimal and the observed differences were not caused by the heterogeneity or the variation in the experimental protocols.

Wild type and setd3Δ cells on day 0 (mESC state) grouped together on a split arm on the dendrogram. It means that the mESCs undergo critical gene expression changes during the differentiation as expected.

Similarly, wild type and setd3Δ cells grouped together on the 2nd day. However, this trend changed starting from the 3rd day. On the 4th day, wild type samples grouped far from other samples. Day 3 and day 4 samples of setd3Δ cells grouped together with the wild type day 3 samples. This may indicate that setd3Δ cells may be retarding on the 3rd day of normal differentiation.

3.2.2 Principal Component Analysis (PCA)

Similar to the HCA, principal component analysis is a method that examines the similarities and the differences amongst all groups in a holistic manner. PCA uses multi-dimensional scaling (MDS) to determine the differences among groups with high variability. All 24 samples were placed on a three-dimensional grid depending on different features, and the two features that explain the variation the best were chosen as x and y axes on the graph (Figure 3.3).

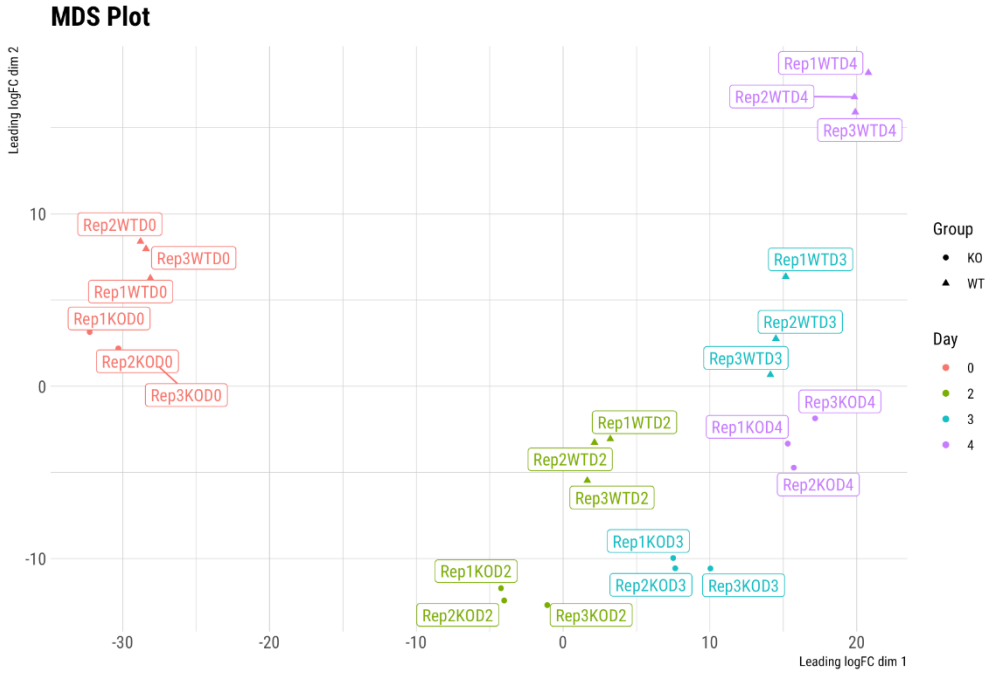

Figure 3.3 Principal component analysis (PCA) results. 24 samples were grouped according to the best 2 features explaining the variation. Similar samples were grouped together, while different ones grouped apart. (Rep: Biological replicate, WT: wild-type, normal. KO: Knock-out, setd3Δ. 0: mESC. 2,3,4: days of endoderm differentiation.)

Wild type and setd3Δ mESCs were clustered together on the left side of the MSD plot, verifying the results from the HCA (Figure 3.2). As the differentiation proceeded, wild type and setd3Δ samples of the same day started to cluster far from each other starting from the 2nd day. Similar to the HCA results, day 4 samples of wild type cells grouped together on the right-top corner, further away from all other groups (Figure 3.3). This indicates that the differentiation defect started on the second day, and the difference between wild type and setd3Δ cells became more evident as the normal differentiation proceeded.

Day 3 samples of setd3Δ cells grouped with day 2 samples of wild type cells. Similarly, day 4 samples of setd3Δ cells grouped with day 3 samples of wild type cells, suggesting that setd3Δ cells followed the normal differentiation with one day delay, consistent with the HCA results (Figure 3.2).

3.2.3 Differential Expression Analysis

Following the unsupervised analyses, wild type and setd3Δ groups were compared with each other on each day of differentiation and on mESC state. For this purpose, differential expression of genes was analyzed using R::limma tool (Ritchie et al., 2015). All biological replicates were averaged, and seven different groups of comparison were created (Table 1 and Figure 3.4).

Table 1 The comparison groups for differential expression analysis.

Wild Type vs. setd3Δ	Wild Type vs. Wild Type
WT/setd3Δ mESC (WT0 – KO0)	WT mESC / WT Day 2 (WT0 – WT2)
WT/setd3Δ Day 2 (WT2 – KO2)	WT mESC / WT Day 3 (WT0 – WT3)
WT/setd3Δ Day 3 (WT3 – KO3)	WT mESC / WT Day 4 (WT0 – WT4)
WT/setd3Δ Day 4 (WT4 – KO4)	

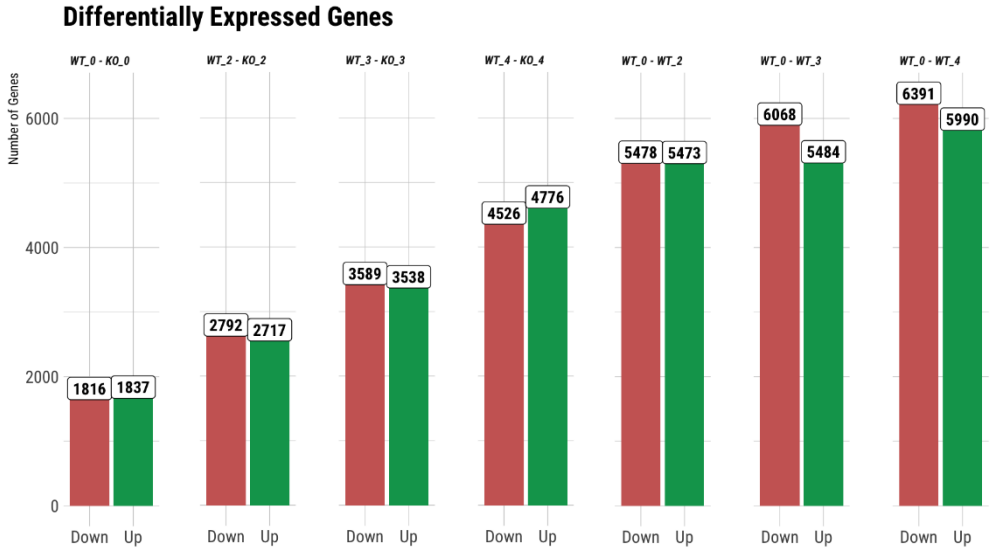

Figure 3.4 Differential expression analysis results. The numbers of differentially expressed genes (DEGs) were shown for each comparison group. (Down: Numbers of DEGs downregulated in the first sample in each group. Up: Numbers of DEGs upregulated in the first sample in each group. WT: Wild-type, normal. KO: Knock-out, setd3Δ. 0: mESC. 2,3,4: days of endoderm differentiation.)

After the differential expression analysis, lists of differentially expressed genes (DEGs) on each day of differentiation and on mESC state were obtained. Starting from the mESC state, there was a linear increase in the numbers of DEGs throughout the differentiation. These data confirmed that both wild type and setd3Δ cells were similar on the mESC state, and the difference gap between these two groups expanded as the differentiation proceeded.

3.2.4 Pathway Enrichment Analysis using KEGG Database

On mESC state and throughout the differentiation, the biological processes and the pathways including the genes that showed strong and significant expression change ($|(\text{fold change})| \geq 1.5$, and q-value < 0.05) were analyzed on KEGG pathway database. Statistically enriched pathways may be directly or indirectly controlled by

SETD3. While gene cluster enrichment analyses allow statistically significant ontologies to be discovered, it also provides insightful information on up and downregulation of these ontologies, allowing the examination of pathways activated or inhibited primarily by SETD3.

Following the differential expression analysis, DEGs were matched with the pathways on the KEGG database to determine which pathways were enriched. Wild type mESCs were compared to wild type cells on each day of differentiation. Results showed that the Wnt pathway was significantly enriched in wild type mESCs, but not in differentiated wild type cells (Figure 3.5). Even though it was not shown in the figure, Wnt signaling was the 12^{th} term enriched by the downregulated genes in wild type mESCs on the 2^{nd} day of differentiation. Considering the fact that the Wnt pathway is downregulated in the definitive endoderm layer following the formation of the primitive streak (Arnold & Robertson, 2009; Tam & Loebel, 2007), our endoderm differentiation process successfully completed in wild type cells.

Especially, on the 2^{nd} and the 3^{rd} days of endoderm differentiation, pathways related to the neuroectoderm differentiation were enriched in the mESC state. This may indicate that these pathways were successfully repressed during the endoderm differentiation as expected.

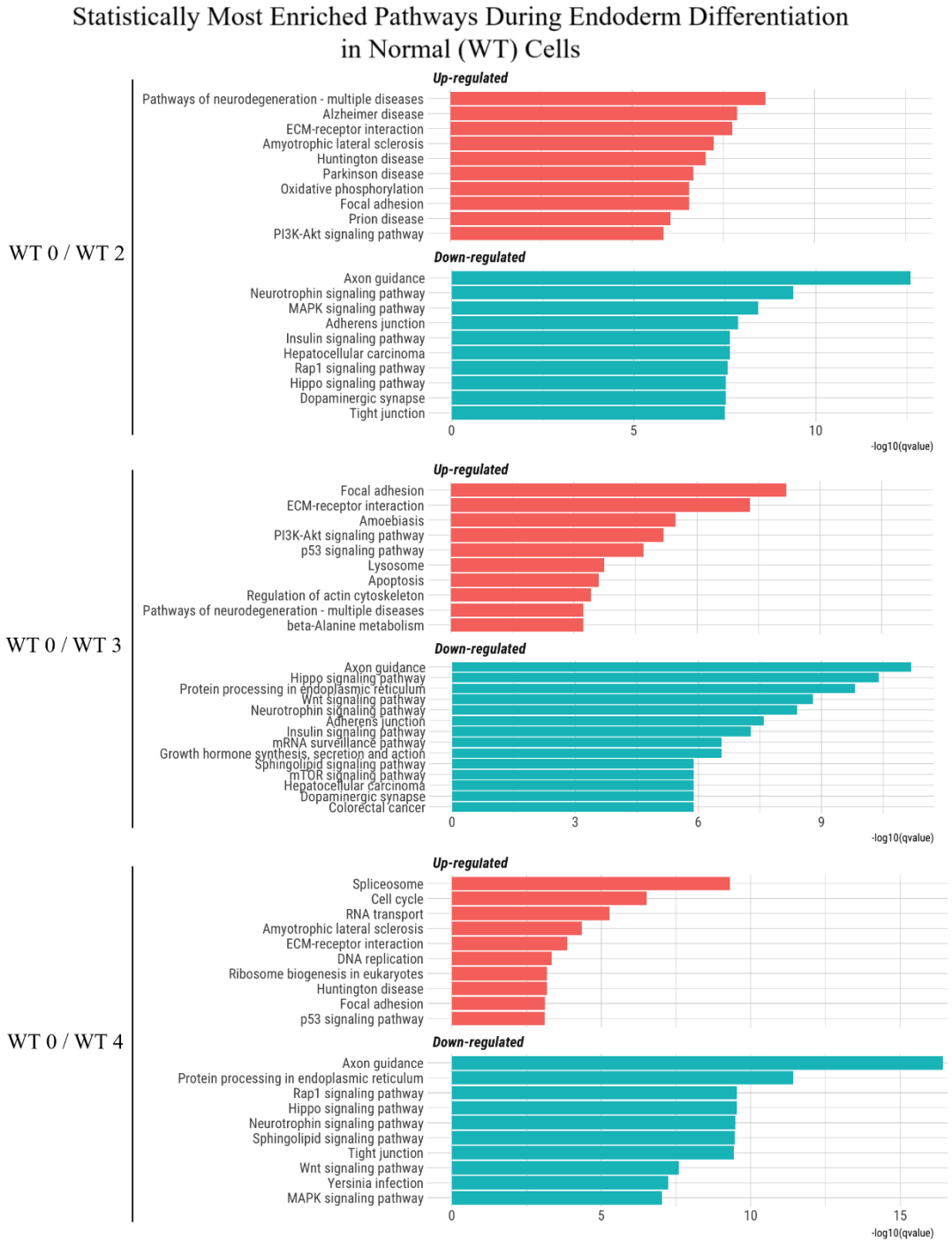

Figure 3.5 Pathway enrichment analysis results for wild type (normal, WT) cells. Statistically the most significant and most enriched 10 pathways in the KEGG pathway were shown. The most enriched pathways by the upregulated genes in wild type mESCs (WT 0) were shown in blue, while the pathways enriched by the downregulated genes were shown in red.

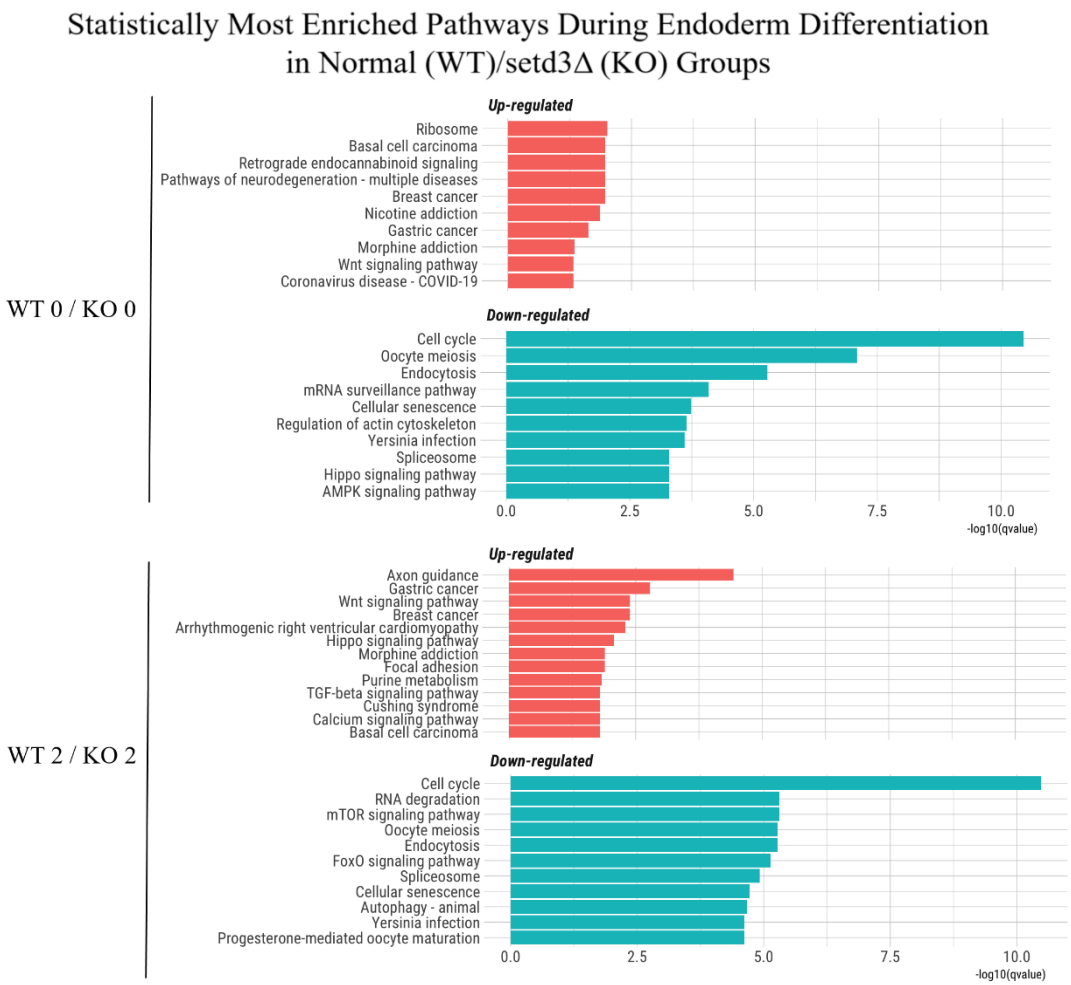

Figure 3.6 Pathway enrichment analysis results for wild type (normal, WT) vs. setd3Δ cells. Statistically the most significant and most enriched 10 pathways in the KEGG pathway database were shown. The most enriched pathways by the upregulated genes in wild type cells were shown in blue, while the pathways enriched by the downregulated genes were shown in red.

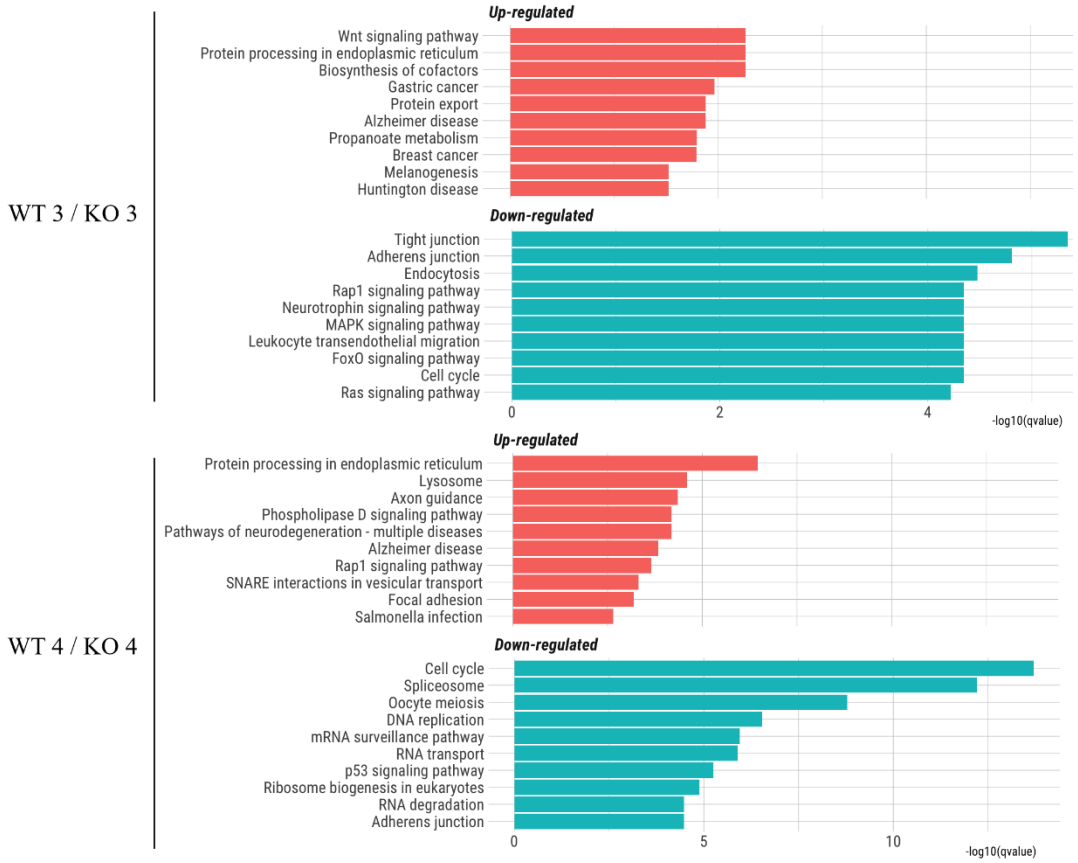

Figure 3.6 (continued)

When setd3Δ cells were compared to wild type cells on mESC state and throughout the differentiation, the Wnt pathway was highly enriched on wild type cells on days 0, 2 and 3. This suggested that the genes in the Wnt pathway were upregulated in wild type cells on these days (Figure 3.6). When the pathway was closely examined, the genes showed a strong and significant change among wild type and setd3Δ cells were selected to be validated via qRT-PCR. Similarly, the cell cycle pathway was enriched on all days in setd3Δ cells (Figure 3.6), suggesting that these cells may have a higher division rate than the wild type ones. If setd3Δ cells could not adapt to the differentiation process, it might have been caused by their inability to switch off the self-renewal properties, thus showing high cell division rates.

3.2.5 Integrative Genomics Viewer (IGV) Analysis

The RNA-seq dataset was analyzed on Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011) to determine where RNA fragments of some of the genes of interest were aligned in the mouse genome (Figure 3.7).

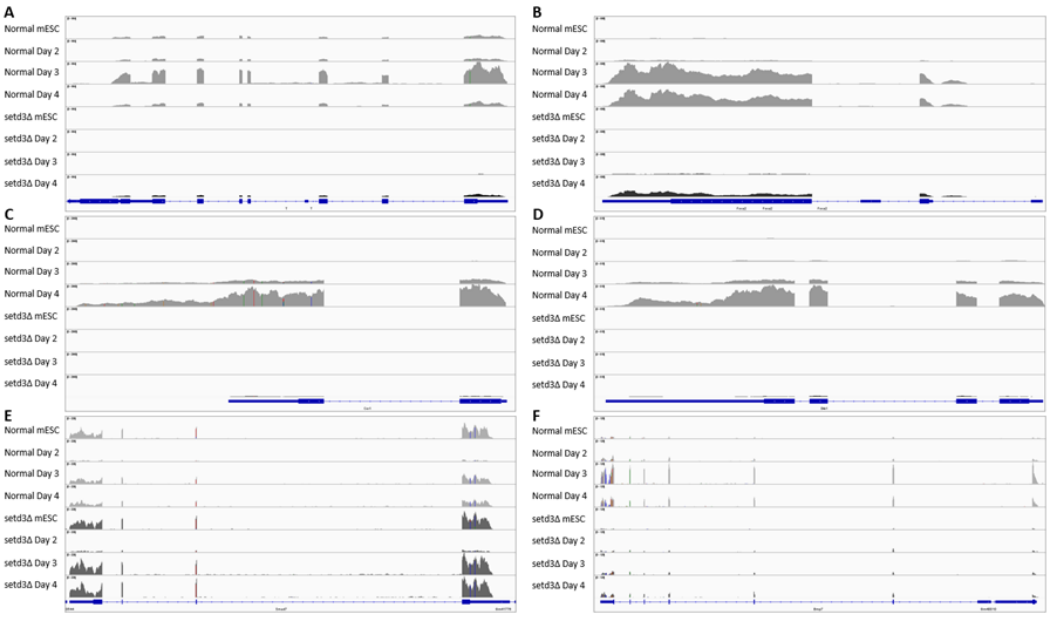

Figure 3.7 IGV analysis results. RNA-seq results were analyzed on IGV using the reference genome *Mus musculus* 10. **A.** *Bry (T)*, **B.** *Foxa2*, **C.** *Cer1*, **D.** *Dkk1*, **E.** *Smad7*, **F.** *Bmp7* gene expression level differences between wild type and setd3Δ cells at mESC state and throughout the differentiation were shown.

The number of RNA fragments aligned to the endoderm marker *Foxa2* gene locus was much higher on the 3rd and the 4th days of differentiation in wild type cells, while *Bry (T)* alignment seemed much higher only on the 3rd day in wild type cells (Figure 3.7 A-B).

The RNA alignment on the gene loci of extracellular Wnt antagonists *Cer1* and *Dkk1* started to become visible on the 3rd day and became evident on the 4th day of differentiation in wild type cells (Figure 3.7 C-D).

Smad7, an internal inhibitor of the Activin/Nodal pathway, showed slightly higher RNA alignment on the gene locus on the 3rd day in setd3Δ cells (Figure 3.7 E).

One of the ligands of the BMP signaling pathway, *Bmp7*, had increased RNA alignment in wild type cells on the 3rd and the 4th days, while in setd3Δ cells it remained low (Figure 3.7 F). The genes examined here will be justified in the following sections.

3.2.6 Short Time-series Expression Miner (STEM) Analysis

STEM is an analysis software that can track gene expression changes through a time-course and group the genes that show similar gene expression change patterns under pre-determined profiles (Ernst & Bar-Joseph, 2006). Using STEM, the genes showing similar expression change patterns throughout the endoderm differentiation process were determined and the enriched terms were obtained using the integrated pathway enrichment analysis (GO) feature of the STEM software. This allows us to determine co-expressed genes following a specific trend throughout the endoderm differentiation process, and the GO terms enriched by these genes.

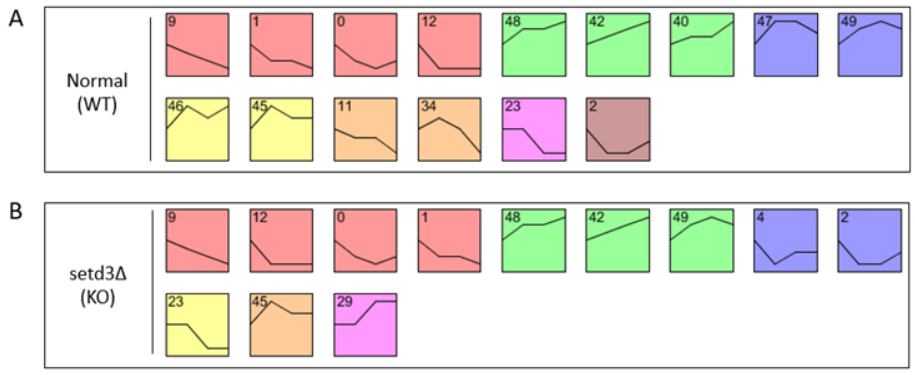

Figure 3.8 STEM analysis results. Only the significantly enriched profiles were shown. **A.** Significantly enriched profiles in WT cells. **B**. Significantly enriched profiles in setd3Δ cells. Numbers on the top left corner shows the number of the profile. Each profile has a unique gene expression change pattern. Similar profiles were clustered together and shown in the same color.

Gene expression changes in wild type and setd3Δ cells were tracked at four different time points (mESCs (day 0), and days 2, 3, and 4 of endoderm differentiation), and grouped under pre-determined profiles. Only the significantly enriched profiles were shown (Figure 3.8). The profiles with similar trends were collected under two major clusters showing either downregulated genes (red, cluster 1) or upregulated genes (green, cluster 2) (Figure 3.8). Also, the profile 23 was significantly enriched in both wild type and setd3Δ groups (Figure 3.8). The genes enriched in cluster 1, cluster 2, and profile 23 were analyzed using built-in GO analysis feature of STEM software, and the enriched terms were visualized using REVIGO (Supek et al., 2011) (Figure 3.9, Figure 3.10, and Figure 3.11). In the graphs generated via REVIGO, the size of the dots on the graph increases as the fold enrichment of the terms increases. Similarly, their color turns from red to yellow as the fold enrichment value increases.

3.2.6.1 GO Enrichment Analysis of Cluster 1 (Downregulated Genes)

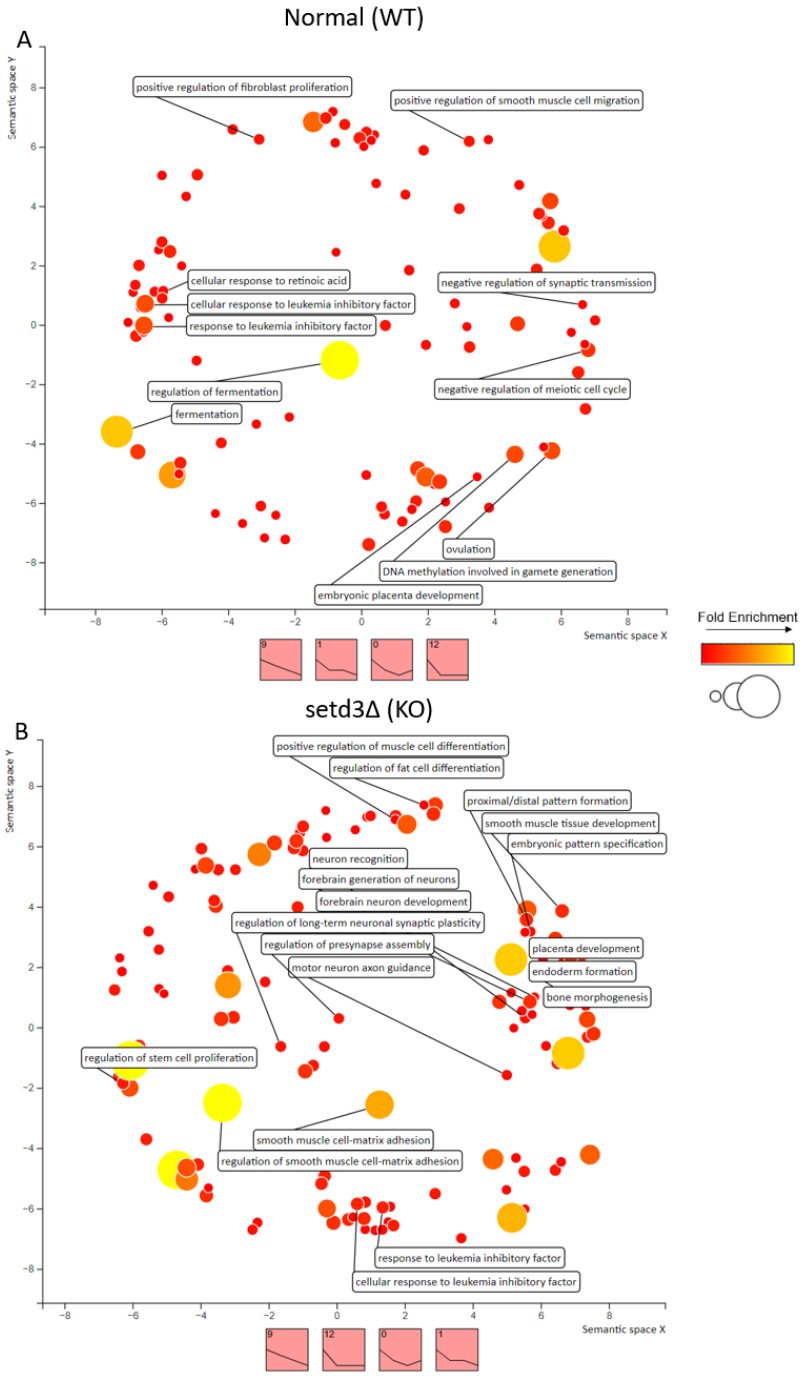

Figure 3.9 GO analysis results of the cluster 1 (downregulated profiles) in STEM analysis. Enriched GO terms were visualized via REVIGO. **A**. Enriched GO terms of WT cells, **B**. Enriched GO terms of setd3Δ cells. The dot size and the color show the fold enrichment value of the term.

Cluster 1 shows the profiles enriched by downregulated genes during endoderm differentiation. In the GO analysis of cluster 1 of wild type cells, fermentation related pathways were enriched (Figure 3.9 A). The endoderm differentiation method we used includes embryoid body (EB) formation. Cells in the center of the large EBs might have reduced access to nutrients, vitamins, cytokines, and the soluble gasses like O_2 and CO_2 in the differentiation media. Although not quantified, wild type EBs were visibly bigger than the setd3Δ EBs. This size difference may lead to altered glucose metabolism in the cells in the center of the wild type EBs, resulting in fermentation related pathways to be enriched in GO analysis.

Unsurprisingly, ovum and sperm related pathways involving meiosis, ovulation, and DNA methylation in gamete formation were downregulated in wild type cells (Figure 3.9 A). These pathways show no significant enrichment in setd3Δ cells.

Leukemia inhibitory factor (LIF) is an essential cytokine for maintaining the pluripotency in mESCs. During the differentiation, mESCs were transferred to a differentiation medium without LIF. Therefore, LIF response related pathways were shown to be downregulated in both wild type and setd3Δ cells (Figure 3.9 A-B).

Most of the pathways downregulated in setd3Δ cells were neuroectoderm specific. These include neuron recognition, motor neuron axon guidance, and forebrain neuron development (Figure 3.9 B). Interestingly, pathways related to mesendoderm differentiation and gastrulation were also downregulated (Figure 3.9 B). This suggests that SETD3 may have a role in activation of cytokine-dependent differentiation programs after the loss of pluripotency.

3.2.6.2 GO Enrichment Analysis of Cluster 2 (Upregulated Genes)

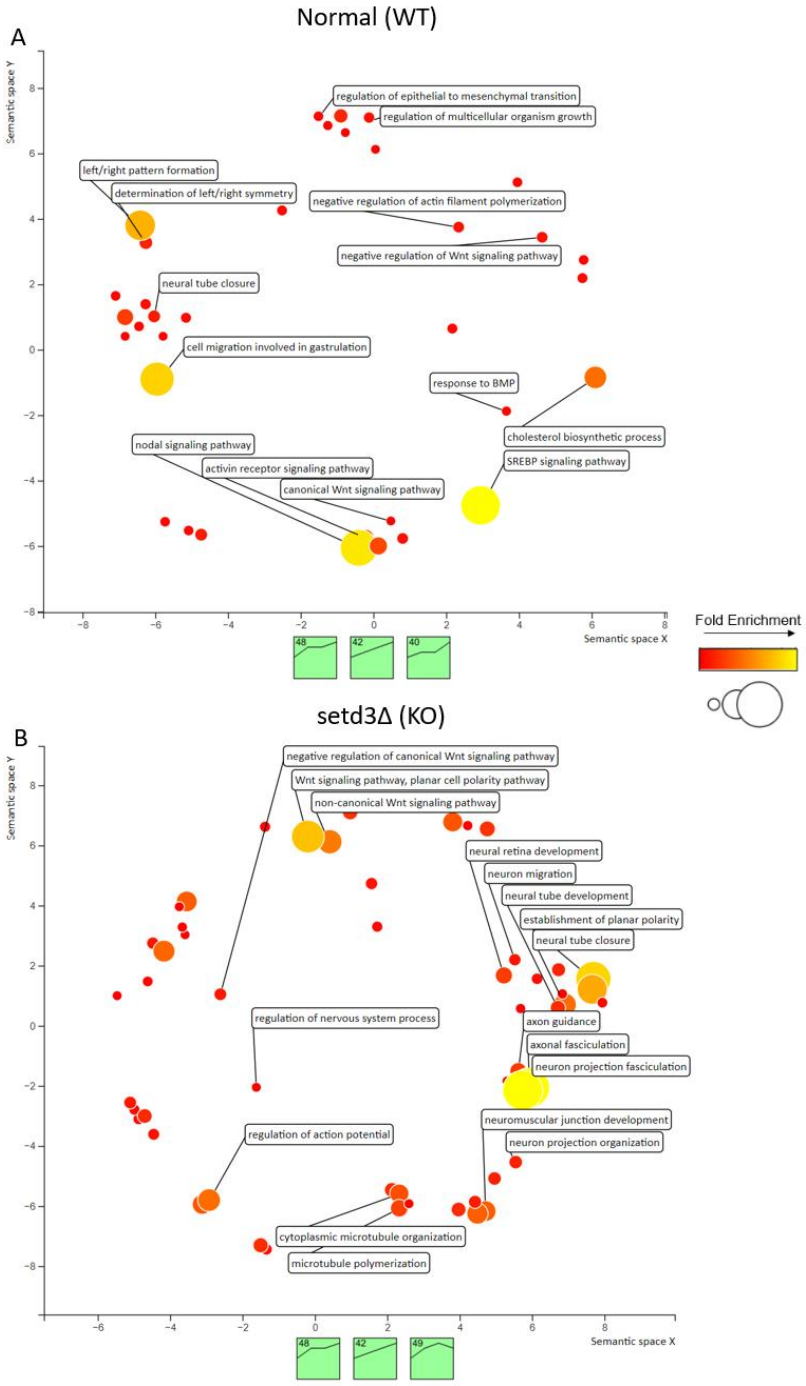

Figure 3.10 GO analysis results of the cluster 2 (upregulated profiles) in STEM analysis. Enriched GO terms were visualized via REVIGO. **A**. Enriched GO terms of WT cells, **B**. Enriched GO terms of setd3Δ cells. The dot size and the color show the fold enrichment value of the term.

Cluster 2 includes the profiles enriched by the upregulated genes. Enrichment of gastrulation related terms like left/right pattern formation, and cell migration involved in gastrulation suggests that endoderm differentiation properly proceeds in wild type cells (Figure 3.10 A). Most of the enriched terms were gastrulation-related, Nodal, Activin, canonical Wnt, and BMP pathways, which are essential for the primitive streak formation and subsequent mesendoderm differentiation.

Additionally, cholesterol synthesis related pathways were enriched in wild type cells (Figure 3.10 A). Cholesterol is found in the plasma membrane, providing rigidity, and allowing lipid raft formation. Lipid rafts play roles in the formation of receptor dimers, providing an important function in the cell signal initiation (Simons & Toomre, 2000). Increased cholesterol biosynthesis may have resulted from increased need for lipid raft formation to provide proper signaling for the differentiation cytokine Activin A. A similar trend for receptor clustering in lipid rafts was previously observed for the type I TGF-β receptors upon TGF-β stimulation (Ma et al., 2007).

In setd3Δ cells, most of the terms enriched by upregulated genes are related to neural development (Figure 3.10 B). These include neural retina development, neuron migration, and neural tube development. Enrichment of neural pathways rather than mesendoderm related ones suggests that the mesendoderm differentiation process is ablated in the absence of SETD3.

Additionally, while canonical Wnt pathway related terms were enriched in wild type cells, non-canonical Wnt pathway, and negative regulation of canonical Wnt pathway terms were enriched in setd3Δ cells (Figure 3.10 B).

3.2.6.3 GO Enrichment Analysis of Profile 23

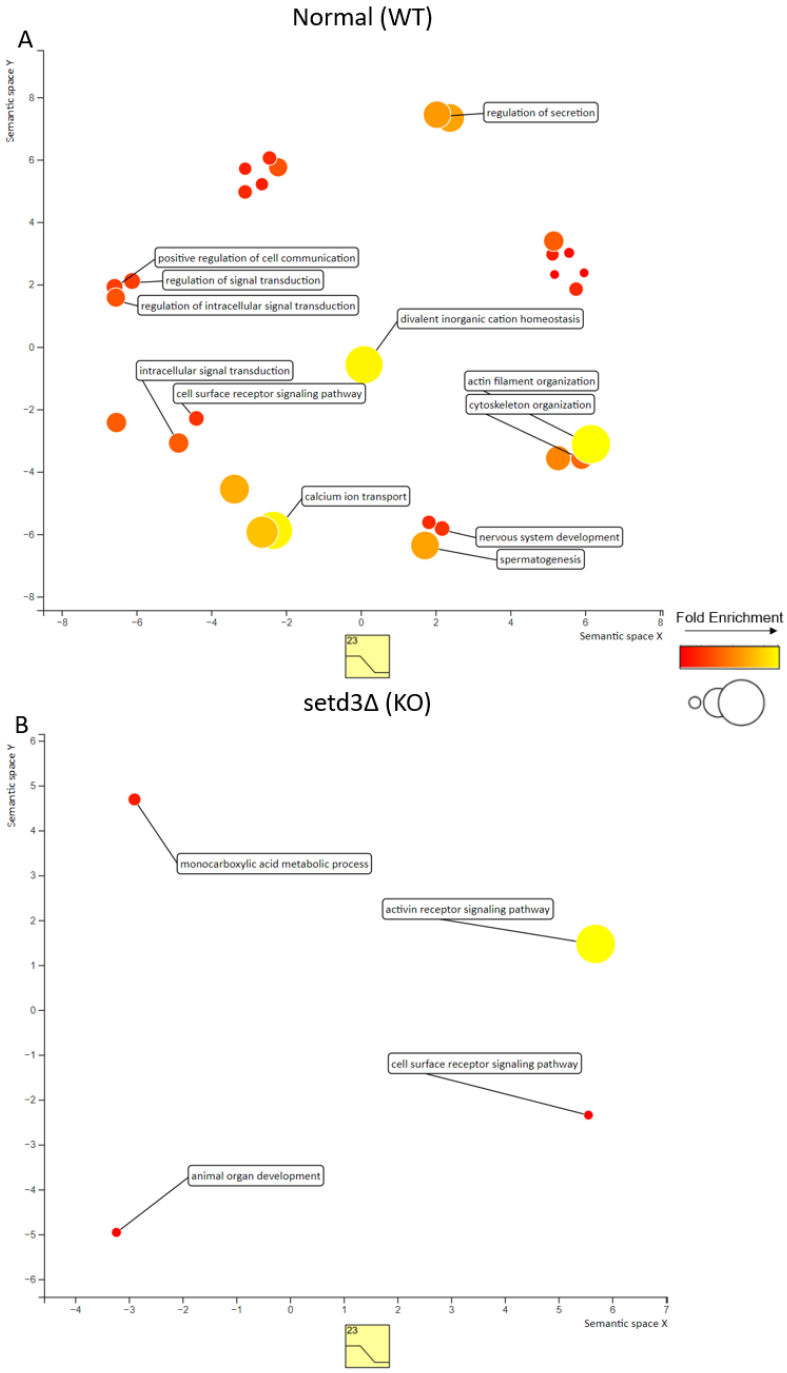

Figure 3.11 GO analysis results of the profile 23 in STEM analysis. Enriched GO terms were visualized via REVIGO. **A**. Enriched GO terms of WT cells, **B**. Enriched GO terms of setd3Δ cells. The dot size and the color show the fold enrichment value of the term.

Profile 23 contains the genes that remained steady first, then downregulated between days 2 and 3, and finally remained steady again. The GO analysis of these genes yielded 4 significant terms in setd3Δ cells (Figure 3.11 B). The most significant among the four is the Activin receptor signaling pathway (Figure 3.11 B). When the GO results of Cluster 2 are considered (Figure 3.10), setd3Δ cells may not be properly responding to Activin A signaling when compared to wild type cells.

Consistent with the previous results, the enriched terms in profile 23 in wild type cells are nervous system development, spermatogenesis, etc. (Figure 3.11 A). Upon endoderm differentiation, downregulation of pathways related to ectoderm differentiation, and the pathways related to gamete formation that may still be active during the earlier stages of the embryonic development were expected.

3.2.7 Transcription Factor Enrichment Analysis (ChEA3)

The ChEA3 tool is used to determine the TFs controlling the expression of a list of genes by mining the literature from a variety of sources (*see* Section 2.8.3). Using ChEA3, we identified the TFs controlling the expression of DEGs among wild type and setd3Δ cells on each day of differentiation. These TFs may be interacting with SETD3 to control the expression of a variety of genes involved in the pluripotency exit and endoderm differentiation.

Up and downregulated DEGs of each time point were separately analyzed using the ChEA3 tool. Unfortunately, the analysis of downregulated DEGs did not yield any meaningful data (data not shown). Only the TFs enriched by the upregulated DEGs were shown (Figure 3.12).

Especially in the mESC state and on the 2nd day, ectoderm specific TFs such as NEUROD1, SOX1, POU3F2, FOXG1, along with an endoderm specific one, LHX1, were enriched (Figure 3.12 A-B). On these days, the cells were not supplemented with Activin A. That is, the differentiation was not directed towards endoderm, yet. Therefore, a combination of TFs with functions in any germ layer might be enriched.

However, on 3rd and 4th days, most of the TFs were mesendoderm specific. LHX1 and GSC were enriched on the 3rd day (Figure 3.12 C), while LHX1, GATA4, GATA6, FOXA2, and SOX17 were enriched on the 4th day (Figure 3.12 D).

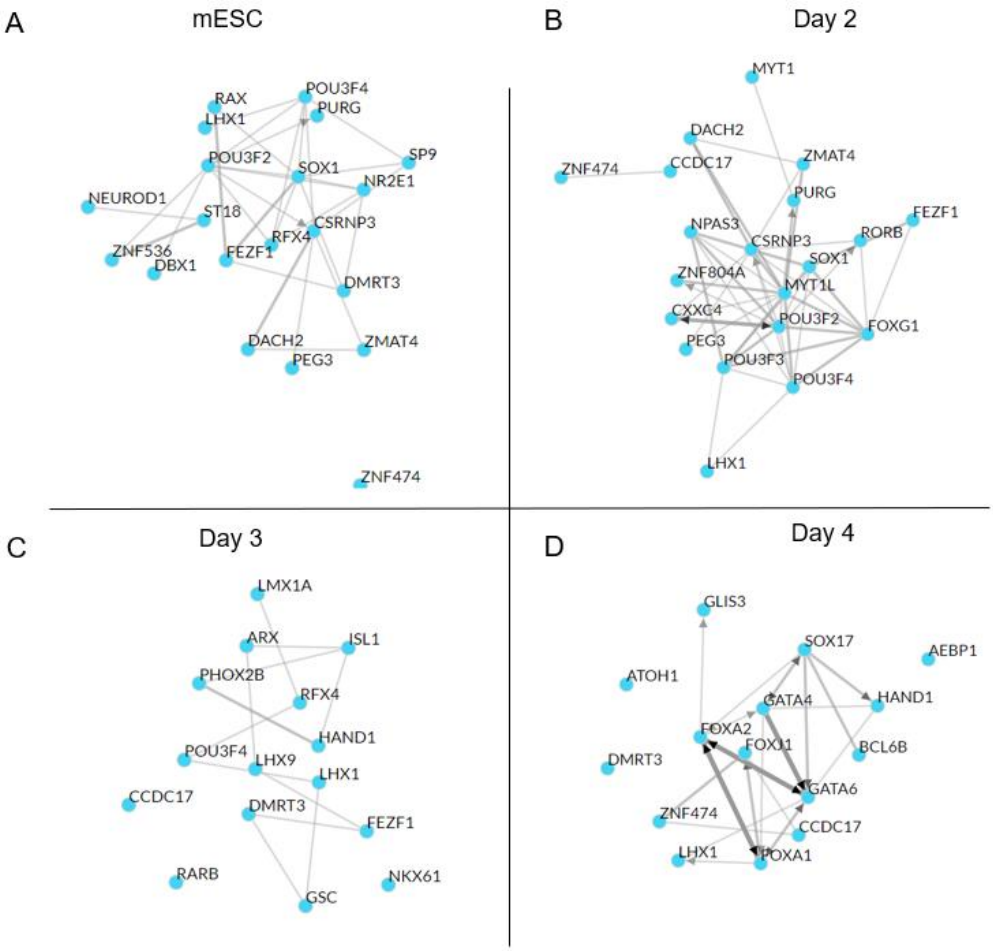

Figure 3.12 ChEA3 TF network that controls the expression of upregulated genes in wild type cells compared to setd3Δ cells on **A**. mESC state and on the **B**. 2nd, **C**. 3rd, and **D**. 4th days of endoderm differentiation. 15 TFs for mESCs and day 2, 20 TFs for days 3 and 4 were shown.

3.3 Validation of Differentially Expressed Genes (DEGs) in Wild type and setd3Δ Cells via qRT-PCR

Upon mining the data obtained from the bioinformatic analyses following RNA-sequencing, several genes from Wnt/β-catenin, Nodal/Activin, and BMP signaling pathways, and genes related to pluripotency, and endoderm differentiation were selected to be validated by qRT-PCR. Only the statistically significant genes (|fold change (FC)| ≥ 1.5, and FDR < 0.05) were selected. The list of selected genes was shown in Table 2 below.

Table 2 The list of selected DEGs. DEGs with the |(fold change)| ≥ 1.5 and FDR < 0.05 were selected as significant. Fold changes were calculated as WT/setd3Δ. Blue cells show on which day the qRT-PCR validation was done for that gene. *: *Gsc* was not validated initially but added to the list for rescue experiments.)

		mESC (Day 0)		Day 2		Day 3		Day 4	
	Transcript	**Log_2(FC)**	**FDR**	**Log_2(FC)**	**FDR**	**Log_2(FC)**	**FDR**	**Log_2(FC)**	**FDR**
Pluripotency Network	*Pou5f1 (Oct4)*	0.06440	0.74729	-0.25158	0.10840	-0.12734	0.41133	-1.75665	0.00000
	Nanog	0.12737	0.33051	-0.73084	0.00000	0.94026	0.00000	-1.15830	0.00000
Activin/Nodal Signaling Pathway	*Nodal*	0.37842	0.04687	1.02241	0.00018	0.82756	0.00002	-0.11025	0.49771
	Smad7	-0.33064	0.03936	-0.87617	0.00039	-0.93137	0.00000	-1.10754	0.00000
Endoderm Markers	*Foxa2*	3.22122	0.00020	4.74250	0.00000	4.74277	0.00000	1.53421	0.00000
	T (Bry)	7.06122	0.00000	7.22079	0.00000	6.32828	0.00000	1.64659	0.00009
	Gata6	2.47581	0.03393	4.00523	0.03195	7.41131	0.00000	5.10765	0.00000
	Lhx1	8.40732	0.00001	5.72301	0.00008	8.90062	0.00000	4.88109	0.00000
	*Gsc**	1.58459	0.04465	2.81593	0.02441	4.89443	0.00000	2.37070	0.00000
Wnt/β-catenin Signaling Pathway	*Cer1*	8.33001	0.08216	10.53783	0.00114	10.20576	0.00000	4.41177	0.00000
	Dkk1	2.37558	0.14871	5.52329	0.00049	6.27178	0.00000	4.30320	0.00000
	Axin2	1.39013	0.00000	-0.01397	0.96536	2.23055	0.00000	0.44044	0.02460
	Ror2	0.27408	0.21878	0.95319	0.00016	2.63754	0.00000	2.14252	0.00000
BMP Signaling Pathway	*Bmp7*	0.60469	0.17417	0.84841	0.02592	2.86662	0.00000	1.62551	0.00001
	Smad1	-0.27489	0.07545	-0.12543	0.39496	0.56583	0.00008	1.63745	0.00000
	Bambi	0.23307	0.25603	0.18117	0.36360	0.59960	0.00102	1.94272	0.00000

One reason why setd3Δ cells are unable to differentiate towards the endoderm layer could be that they were not able to exit pluripotency in the absence of SETD3. To differentiate into the germ layers, mESCs should repress the pluripotency related genes and activate lineage specific ones. The differential expression analysis results showed that the pluripotency markers *Pou5f1 (Oct4)* and *Nanog* expressions were significantly higher in setd3Δ cells on the 4th day of differentiation. qRT-PCR analyses revealed that *Oct4* levels remained significantly higher in setd3Δ cells while its expression in wild type cells drops on the fourth day (Figure 3.13 A). *Nanog* levels showed a decrease in wild type cells as differentiation progressed, while its expression remained high in setd3Δ cells (Figure 3.13 B). The data suggest that mESCs may not be able to efficiently exit pluripotency in the absence of SETD3.

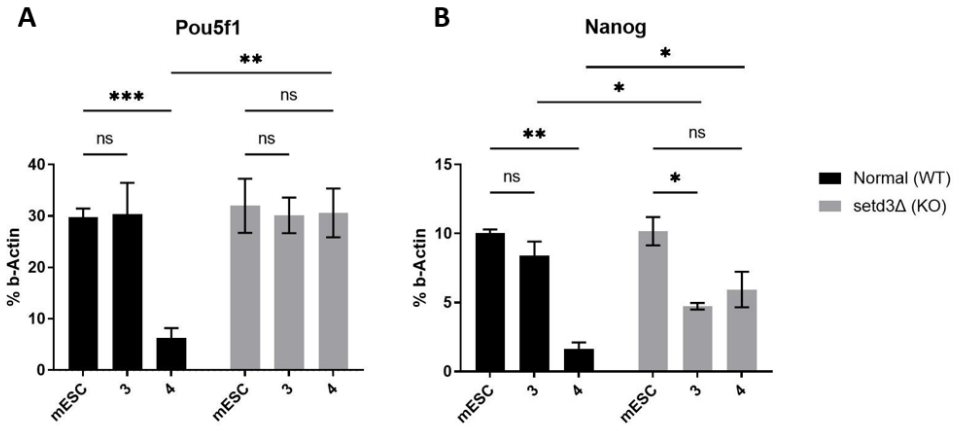

Figure 3.13 qRT-PCR analysis for pluripotency markers. **A.** *Pou5f1* (*Oct4*), **B.** *Nanog*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of three independent biological replicates. (mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ***: p-value<0.001, ns: not significant.)

The differentiation defect may be a result of a lack of cellular response to Activin A cytokine. Activin A is widely used for endoderm differentiation protocols, as it activates the Nodal/Activin pathway, a crucial pathway for endoderm lineage

specification. RNA-seq data indicate that *Smad7*, an intracellular inhibitor of the Nodal/Activin pathway, was significantly upregulated in setd3Δ cells on the 3rd and 4th days of differentiation. *Nodal* expression was higher in wild type cells on the 3rd day. When the expression levels of *Smad7* and *Nodal* were analyzed by qRT-PCR, *Smad7* levels were indeed higher in setd3Δ cells on the 3rd day (Figure 3.14 A). *Nodal* expression seemed higher in wild type cells on the 3rd day, yet the difference was not significant (Figure 3.14 B).

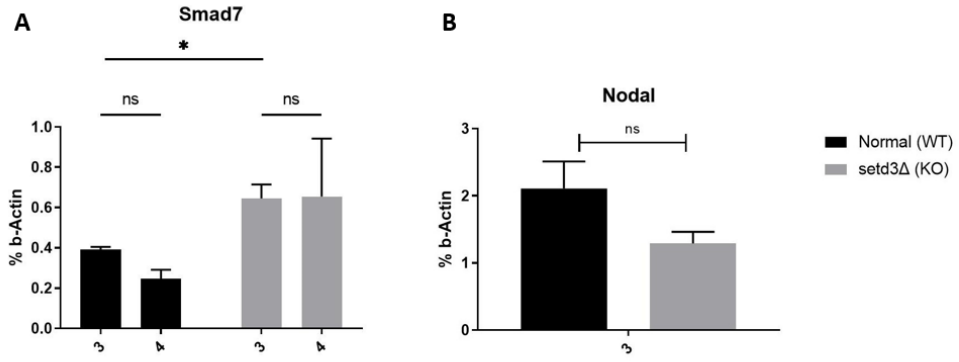

Figure 3.14 qRT-PCR analysis for Nodal/Activin A pathway. **A.** *Smad7*, **B.** *Nodal.* Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of two independent biological replicates. (3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, ns: not significant.)

Next, the endoderm differentiation markers *Brachyury* (*T*, *Bry*), *Foxa2*, *Gata6* and *Lhx1* were analyzed via qRT-PCR. As expected, all four markers showed higher expression levels in wild type cells throughout the differentiation (Figure 3.15).

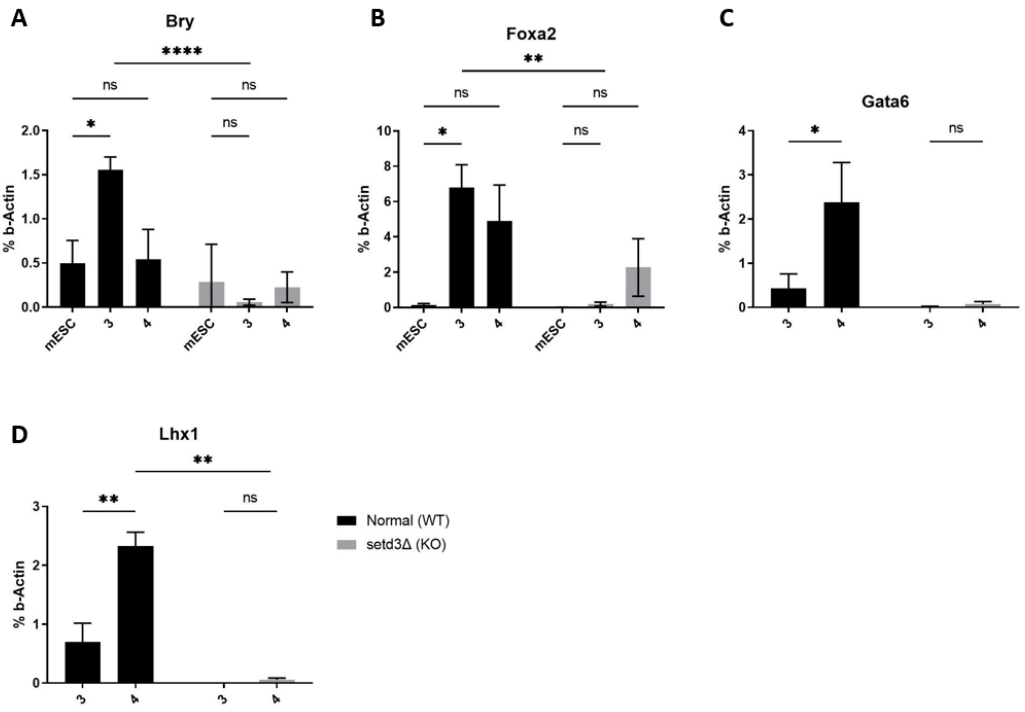

Figure 3.15 qRT-PCR analysis for endoderm markers. **A.** *Bry (T),* **B.** *Foxa2,* **C.** *Gata6,* **D.** *Lhx1*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of at least two independent biological replicates. (mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ****: p-value<0.0001, ns: not significant.)

The pathway enrichment analyses of the RNA-seq data indicated that Wnt signaling pathway was significantly enriched in wild type cells on both ESC state and during the differentiation (Figure 3.5). When the Wnt pathway was closely examined in wild type cells, *Cer1*, *Dkk1* and *Ror2* were expressed at higher levels on the 3rd and 4th days, and *Axin2* expression was higher on the 3rd day in wild type cells. The expression levels were validated via qRT-PCR (Figure 3.16).

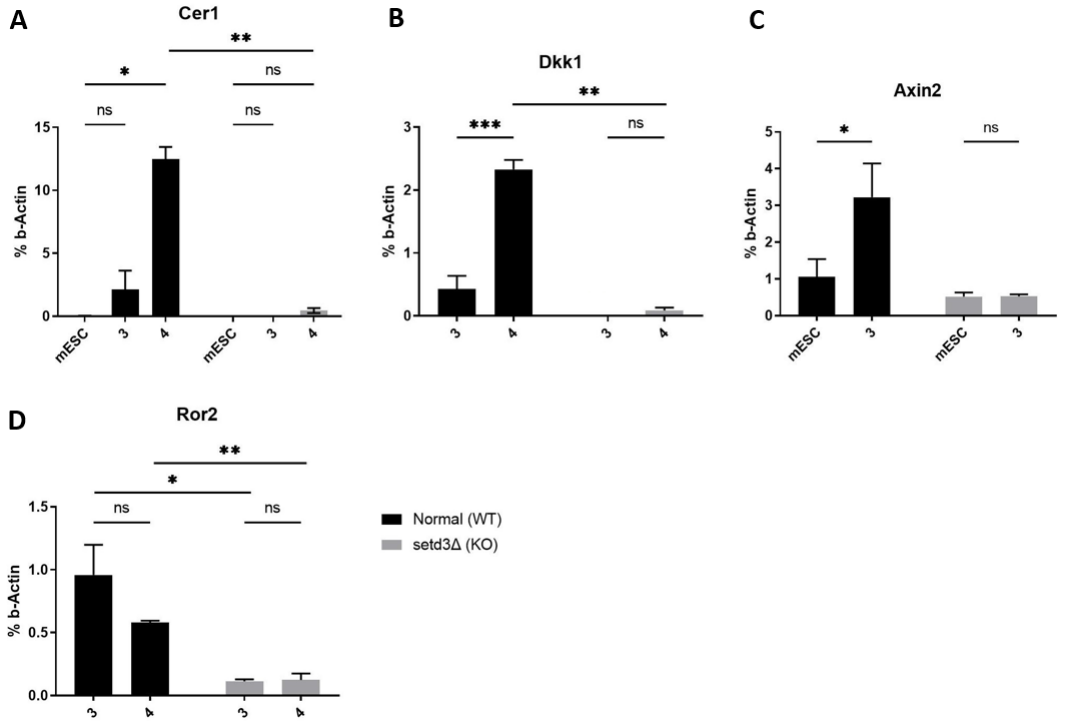

Figure 3.16 qRT-PCR analysis for Wnt pathway. **A.** *Cer1*, **B.** *Dkk1*, **C.** *Axin2*, **D.** *Ror2*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of two independent biological replicates. mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ***: p-value<0.001, ns: not significant.

Cer1, *Dkk1*, and *Axin2*, the repressors of the Wnt pathway, were upregulated throughout the differentiation in wild type cells (Figure 3.16 A-B-C). While *Ror2* levels appeared to be decreased, it was not significant (Figure 3.16 D). Their expression levels remained low in setd3Δ cells (Figure 3.16).

During mouse embryonic development, Wnt and Nodal ligands expressed from the epiblast lead to the expression of *Foxa2* and *Lhx1* in anterior visceral endoderm (AVE). This creates a negative feedback loop in AVE and upregulates the expression of Wnt and Nodal antagonists, namely DKK1 and LEFTY1 (Arnold & Robertson, 2009). Inhibition of Wnt and Nodal signaling pathways in the anterior epiblast prevents primitive streak formation and triggers the differentiation towards the

ectoderm layer. In wild type cells, these antagonists (*Dkk1, Lefty1* (data not shown), and additionally, *Cer1*) were expressed in high levels, while this was not the case in setd3Δ cells. Therefore, SETD3 may have a role in the regulation of the Wnt signaling pathway. The differentiation defect in setd3Δ cells might originate from the mis-regulation of the Wnt pathway in the ESC state.

Similar to the Wnt pathway, BMP signaling pathway plays key roles in keeping the mESCs in pluripotent state, as well as in the differentiation process. When suppressed, primitive endoderm layer could not form in the developing mouse embryo (Graham et al., 2014). Similarly, the defect in setd3Δ cells may stem also from the mis-regulation of the BMP signaling pathway. When the BMP signaling pathway was closely examined using the RNA-seq data, a BMP ligand *Bmp7* and one of the inhibitors of the BMP pathway called *Bambi* were upregulated on the 3rd and 4th days of differentiation in wild type cells. Likewise, *Smad1*, a member of the activator complex of the BMP signaling, was upregulated on the fourth day in wild type cells. When the differences were validated by qRT-PCR, *Bmp7* expression levels increased during differentiation in wild type cells (Figure 3.17 A). While *Smad1* and *Bambi* levels seemed relatively higher in wild type cells, the differences were not statistically significant (Figure 3.17 B-C).

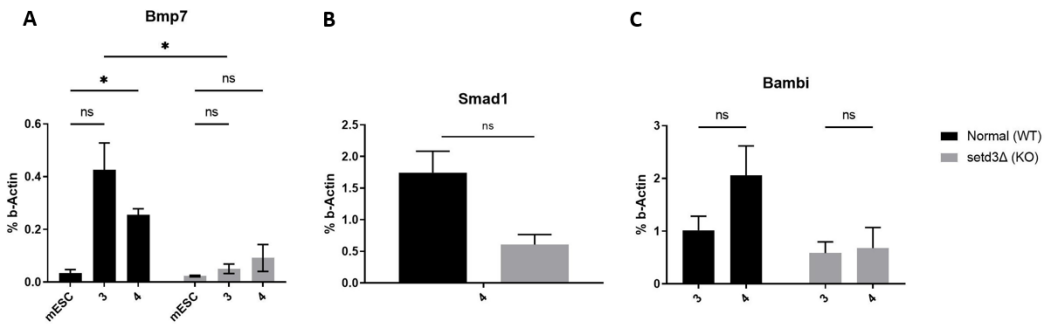

Figure 3.17 qRT-PCR analysis for BMP pathway. **A.** *Bmp7*, **B.** *Smad1*, **C.** *Bambi.* Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of two independent biological replicates. (mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, ns: not significant.)

The exit from the pluripotency and the activation of differentiation related pathways are quite complex and highly regulated processes. Specifically, the endoderm differentiation process is regulated by Wnt, Nodal and BMP pathways through controlled up and down regulation. The data so far indicated that these pathways were not properly regulated in the absence of SETD3, and they directly or indirectly contribute to the differentiation defect observed in setd3Δ cells.

3.4 Rescue of the Defective Differentiation Phenotype in setd3Δ Cells via Re-expression of SETD3

3.4.1 Re-expression of SETD3 in setd3Δ mESCs

The changes observed after knocking-down or knocking-out a gene could result from off-target or secondary effects. To determine whether the changes are the direct effects of the gene manipulation or not, rescue experiments are required.

For this purpose, a SETD3 rescue cell line was created using pEF1αFLBIOSetd3-puro (pEF1α-Setd3) plasmid. In this construct, *Setd3* is expressed with a FLAG tag

under EF1α promoter, which provides strong and stable gene expression in mammalian cells. To determine the stability, and the sub-localization of the exogenous SETD3 protein, western blot analysis was conducted for the protein isolated from the whole cell extract as well as cytoplasmic and nuclear fractions of wild type, setd3Δ, and setd3Δ+pEF1α-Setd3 mESC lines. GAPDH was used as a whole cell and cytoplasmic fraction loading control, while Histone 3 (H3) was used as a nuclear loading control. WB results showed that the fractionation was clear as H3 was absent in the cytoplasmic fraction, and GAPDH did not show up in the nuclear fraction (Figure 3.18).

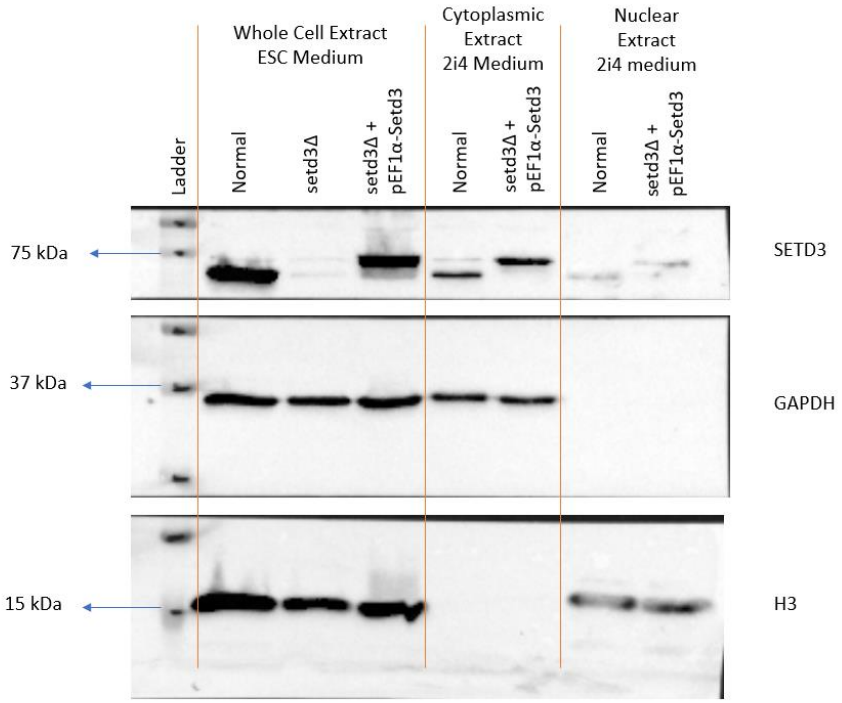

Figure 3.18 Stable re-expression of SETD3 in setd3Δ mESCs. Western blotting image of wild type, setd3Δ and setd3Δ+pEF1α-Setd3 mESCs. Overall SETD3 protein levels from the whole cell extracts were similar in wild type and setd3Δ+pEF1α-Setd3 mESCs. SETD3 levels in cytoplasmic and nuclear fractions were also similar in wild type and setd3Δ+pEF1α-Setd3 mESCs. GAPDH was used for whole cell lysates and cytoplasmic extracts as loading control. Histone 3 (H3) was used for nuclear loading control.

SETD3 protein levels showed close similarity between wild type and setd3Δ+pEF1α-Setd3 mESCs as seen in whole cell lysates. Exogenous SETD3 was tagged with FLAG, therefore it is located slightly above the endogenous SETD3. In both cytoplasmic and nuclear fractions, SETD3 levels in the setd3Δ+pEF1α-Setd3 mESCs were similar to the wild type mESCs. Additionally, *Setd3* expression levels did not change during the differentiation of WT cells when compared to the levels in WT mESCs as suggested by the differential expression analysis (WT0/WT2 LogFC: 0.32 FDR: 0.0007, WT0/WT3 LogFC: 0.02 FDR: 0.78, WT0/WT4 LogFC: -0.023 FDR: 0.74). Therefore, any possible negative effect of stable and steady expression of *Setd3* from the EF1α promoter in the rescue line is negligible.

3.4.2 Validation of Differentially Expressed Genes (DEGs) in setd3Δ+ pEF1α-Setd3 Cells via qRT-PCR

Wild type, setd3Δ, and setd3Δ+ pEF1α-Setd3 cells were differentiated towards definitive endoderm as previously explained, and qRT-PCR analysis was done on days 0, 3, and 4 for several genes which showed prominent difference in expression levels in the absence of SETD3.

It was shown that the endoderm differentiation process is defective in the absence of SETD3. The expression levels of the endoderm markers *Brachyury (T, Bry)*, *Foxa2*, *Gata6*, *Lhx1* and *Gsc* returned to the levels in the wild type cells upon re-expression of SETD3 (Figure 3.19). Normally, *Bry* expression starts to increase on the 3rd day. In setd3Δ+pEF1α-Setd3 cells, it did not increase as much on the 3rd day, yet it peaked on the 4th day (Figure 3.19 A). *Foxa2* expression in setd3Δ+pEF1α-Setd3 cells followed the trend in wild type cells (Figure 3.19 B). Similarly, *Gata6, Lhx1,* and *Gsc* expression levels in setd3Δ+pEF1α-Setd3 reached the levels in the wild type cells (Figure 3.19 C-D-E). The fact that the levels of endoderm differentiation markers reach to the levels in wild type cells upon SETD3 re-expression indicate that the differentiation defect is indeed caused by the loss of SETD3, and it can be rescued by ectopic SETD3 expression.

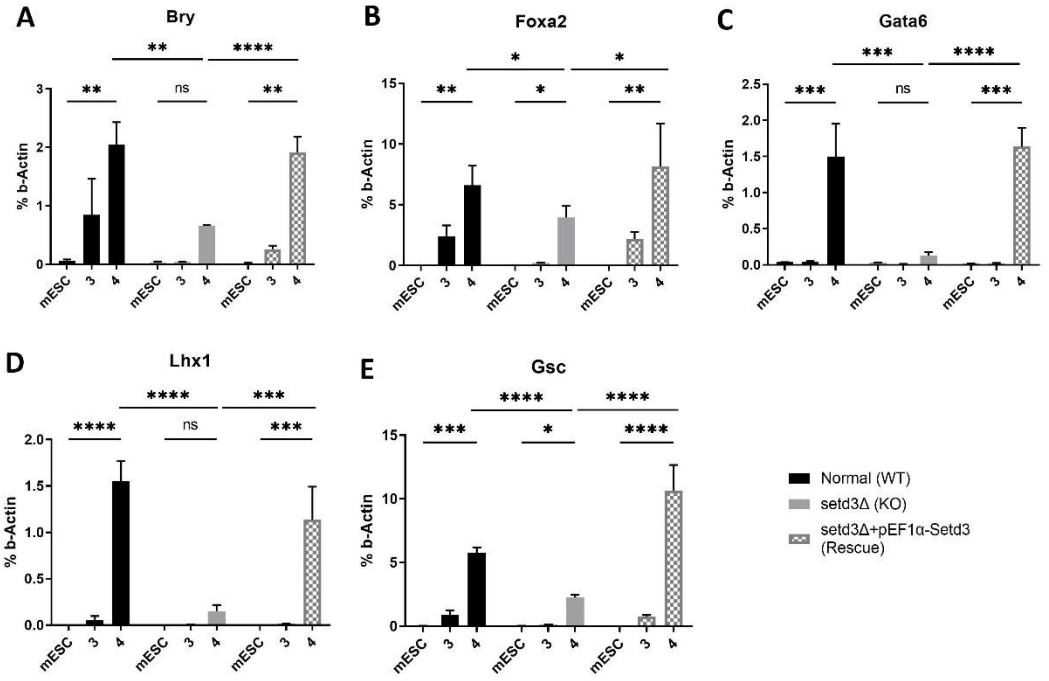

Figure 3.19 qRT-PCR analysis for endoderm markers. **A.** *Bry (T)*, **B.** *Foxa2*, **C.** *Gata6*, **D.** *Lhx1*, **E.** *Gsc.* Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of at least two independent biological replicates. (mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ***: p-value<0.001, ****: p-value<0.0001, ns: not significant.)

Similar results were obtained for the Wnt antagonists *Cer1* and *Dkk1*. While their expression peaked on the 4th day in wild type cells, they remained low in setd3Δ cells. Ectopic SETD3 expression was efficient to bring the expression levels back to the levels in wild type cells (Figure 3.20).

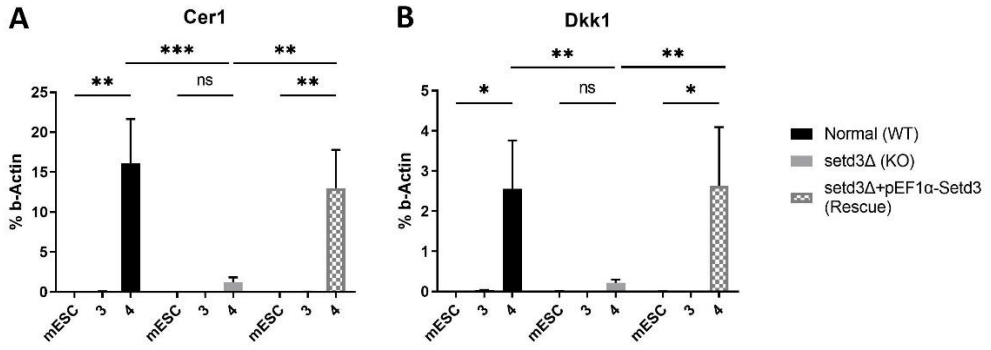

Figure 3.20 qRT-PCR analysis for Wnt pathway. **A.** *Cer1*, **B.** *Dkk1*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of two independent biological replicates. (mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ns: not significant.)

In the absence of SETD3, the pluripotency network is also affected. *Pou5f1* (*Oct4*) and *Nanog* expression levels were higher in setd3Δ cells especially on the 4th day of differentiation (Figure 3.13). When they were examined in setd3Δ+pEF1α-Setd3 cells, *Pou5f1* expression dropped on the 4th day similar to wild type cells (Figure 3.21 A). Although the drop in the expression of *Nanog* on the 4th day was clearly visible in wild type and rescue lines, we were not able to validate it statistically (Figure 3.21 B). *Nanog* expression levels remained high on the 4th day in setd3Δ cells (Figure 3.21 B) as previously observed (Figure 3.13 B). Re-expression of SETD3 was sufficient to rescue the *Pou5f1* and *Nanog* expression.

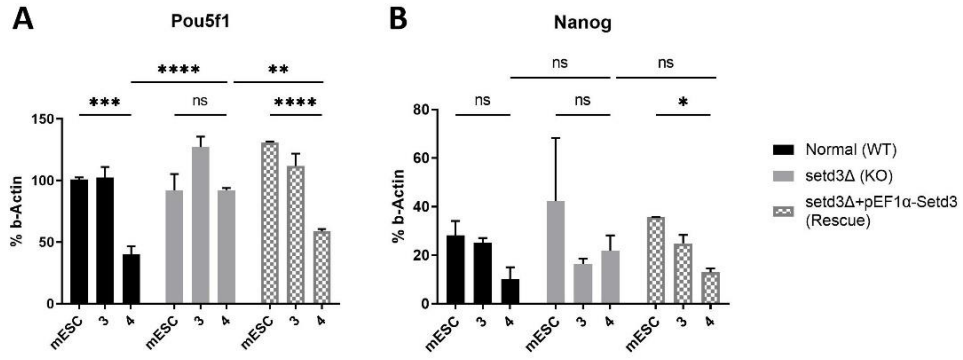

Figure 3.21 qRT-PCR analysis for pluripotency network. **A.** *Pou5f1 (Oct4)*, **B.** *Nanog*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of two independent biological replicates. mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ***: p-value<0.001, ****: p-value<0.0001, ns: not significant.)

Finally, the expression level of *Smad7*, an inhibitory mediator of the Nodal/Activin signaling pathway was examined in rescue cells upon endoderm differentiation. Its expression drops significantly on the 4th day in wild type cells, while it remains high when SETD3 is absent. The ectopic expression of SETD3 in setd3Δ cells allows the expression level to drop on the 4th day, similar to wild type cells (Figure 3.22)

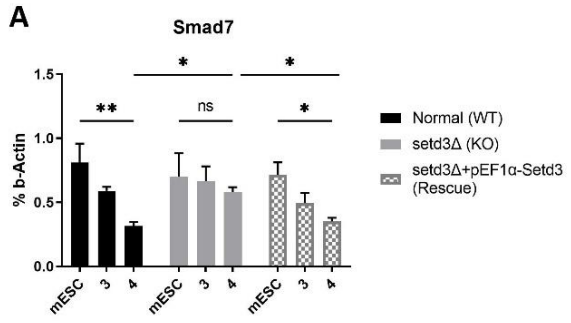

Figure 3.22 qRT-PCR analysis for Nodal/Activin inhibitor *Smad7*. Expression levels were normalized to *β-actin* levels. Error bars were shown as ±SEM of two independent biological replicates. (mESC: Day 0. 3,4: Days of endoderm differentiation. Statistical analysis (two-way ANOVA or t-test) was done on GraphPad Prism software. *: p-value<0.05, **: p-value<0.01, ns: not significant.)

CHAPTER 4

DISCUSSION

Embryonic development is a quite complex process requiring spatiotemporal regulation of gene expression to direct the pluripotent stem cells towards a collection of highly specified cell types that form a functional, healthy organism. That regulation is mainly provided by the epigenetic factors that can selectively activate and deactivate gene expression in specific parts of the embryo throughout the developmental stages. We identified one of those epigenetic factors, SETD3, as an important element for mouse embryonic stem cell (mESC) differentiation towards mesendodermal lineages (Figure 1.1).

The earlier research on the SETD3 deemed it as a H3K4 and H3K36 methyltransferase. It has been shown to catalyze H3K4me2-3 and H3K36me2-3 modifications both *in vitro* and *in vivo,* and it physically interacts with MyoD to activate myogenin (*Myog*) expression to drive muscle differentiation (Eom et al., 2011), which originates from the mesoderm lineage. The same principle may apply to our case: SETD3 may be interacting with other TFs to activate the endoderm differentiation network via its transcriptional activation activity. However, the literature is limited, and there is not much known about the nuclear function of SETD3. The most recent literature considers SETD3 only as an actin His73 methyltransferase, and not as a H3 lysine methyltransferase (Kwiatkowski et al., 2018; Wilkinson et al., 2019). However, we believe that the differentiation defect may be caused by SETD3's nuclear function, which is yet to be investigated.

We started by identifying the expression changes in the absence of SETD3 during early endoderm differentiation. Our time frame starts at the mESC state, and ends when the expression of the key definitive endoderm marker *Foxa2* reaches its peak at around the 4^{th} day of endoderm differentiation. We identified the differentially

expressed genes (DEGs) in the differentiation process by comparing wild type endoderm cells with wild type mESCs, and also by comparing wild type and setd3Δ cells on each day by using RNA-seq. The initial unsupervised analyses (PCA and HCA) showed that both wild type and setd3Δ cells were similar in the mESC state as they clustered together (Figure 3.2 and Figure 3.3). Surprisingly, both analysis results suggested that the differentiation does occur in the absence of SETD3 with one day delay (Figure 3.2 and Figure 3.3), which was confirmed by low expression levels of mesendoderm markers *Bry* (*T*) and *Foxa2* via qRT-PCR analysis (Figure 3.1).

A similar trend was also observed after the differential expression analysis. The number of DEGs between the wild type mESCs and the wild type endoderm cells on each day of differentiation was quite high, and it increased steadily throughout the differentiation (Figure 3.4). This is expected as the differentiation process involves massive gene expression changes to ensure the exit from pluripotency, and the activation of lineage specific gene networks. The comparison of wild type and setd3Δ cells suggested that both lines are relatively similar in the mESC state as the number of DEGs were limited (Figure 3.4). However, the number of DEGs between wild type and setd3Δ cells increased as the differentiation processes, suggesting that the differentiation defect starts after the 2nd day (Figure 3.4).

One reason for the observed differentiation defect may stem from the mESCs' inability to exit from pluripotency in the absence of SETD3. An over-active pluripotency network may prevent the activation of differentiation specific networks. TF enrichment analysis results of the downregulated DEGs in setd3Δ cells backed up our hypothesis, but the data was not very clear (data not shown). There was one pluripotency specific TF, KLF4, enriched on the 2nd day by the upregulated DEGs in setd3Δ cells (data not shown). Therefore, we checked whether the core pluripotency markers were differentially expressed among WT and setd3Δ cells.

Pou5f1 (*Oct4*) and *Nanog* were among the most prominent markers showing differential expression in the absence of SETD3 (Table 2). When we checked their expression levels via qRT-PCR, *Pou5f1* levels remained high on the 4th day in setd3Δ cells while it significantly dropped in wild type cells (Figure 3.13 A). *Nanog* levels also showed a similar trend. In wild type cells, it dropped significantly on the 4th day (Figure 3.13 B). Although it dropped on the 3rd day in setd3Δ cell, its level was not significantly low on the 4th day (Figure 3.13 B). The re-expression of SETD3 in setd3Δ cells was sufficient to rescue the expression of *Pou5f1* and *Nanog*. (Figure 3.21). These data suggest that mESCs fail to exit from pluripotency in the absence of SETD3. Mechanistically, due to its suggested activatory function, SETD3 may be activating one or more TFs that control the repression of the pluripotency network.

Another reason for the defect is that the cells may not respond to the differentiation cytokine, Activin A, during the endoderm differentiation when SETD3 is absent. The STEM analysis of the profile 23 supports this hypothesis, as the "activin receptor signaling pathway" was among the GO terms enriched by the downregulated genes in setd3Δ cells (Figure 3.11). We thought that this may result from the decreased levels of Activin receptors, or the mediators of the signaling pathway. Once we checked the DEGs, we could not find any significant expression difference in the Nodal/Activin receptors, or the positive mediators of the pathway, SMAD2-3-4 (data not shown). However, a negative mediator of the pathway, *Smad7*, was slightly upregulated in setd3Δ cells on the 3rd day (Table 2). It was validated by qRT-PCR analysis, showing significantly higher *Smad7* expression on the 3rd day in setd3Δ cells (Figure 3.14 A). Additionally, SETD3 re-expression brought the expression levels back to the levels in wild type cells (Figure 3.22). Increased SMAD7 levels may inhibit the effect of Activin A, resulting in decreased response to Activin A signal, and hence, the decreased differentiation efficiency in the absence of SETD3.

Our STEM analysis results suggested that the gastrulation, and subsequently left/right pattern formation related GO terms were enriched by the upregulated genes in wild type cells (Figure 3.10), while similar terms were among the terms enriched by the downregulated genes in setd3Δ (Figure 3.9). Gastrulation stage is marked by the formation of primitive streak, and the subsequent mesendoderm development. Similarly, TF enrichment analysis of the DEGs upregulated in wild type cells showed the enrichment of mesendoderm related TFs, such as LHX1, FOXA2, SOX17, GSC, GATA4, and GATA6, especially on the 3rd and 4th days of differentiation (Figure 3.12). Therefore, well known mesendoderm markers were also manually checked to see if they show significant differential expression among wild type and setd3Δ cells. *Bry* (*T*), *Foxa2*, *Gata6*, *Lhx1*, and *Gsc* were among the most significant DEGs (Table 2). Their expression levels show similar trends, as they were low at mESC state, and peaked on either 3rd or 4th day of differentiation in wild type cells. Although *Bry* and *Foxa2* expression levels slightly increased on the 4th day in setd3Δ cells, it is not statistically significant (Figure 3.15 A-B). *Gata6* and *Lhx1* levels peaked on the 4th day in wild type cells, while they remained low in setd3Δ cells (Figure 3.15 C-D). With the re-expression of SETD3 in setd3Δ cells, their expressions increased back to normal levels, rescuing the defective differentiation phenotype (Figure 3.19).

Wnt signaling plays a crucial role in primitive streak formation, and the subsequent mesendoderm differentiation. The enrichment analysis done on the DEGs, and the GO enrichment analysis following the STEM analysis revealed that the Wnt signaling was enriched in wild type cells. Close examination of the DEGs and the STEM analysis results yielded that the negative regulators of the pathway, *Cer1*, *Dkk1*, and *Axin2*, were downregulated in the absence of SETD3 (Table 2). Our qRT-PCR results validate the differential expression of these genes in the absence of SETD3. *Cer1* and *Dkk1* expression levels were significantly high on the 4th day of differentiation in wild type cells, while they remained low throughout the differentiation in setd3Δ cells (Figure 3.16). Similarly, *Axin2* levels peaked on the 3rd day in wild type cells, but it remained low in the absence of SETD3 (Figure 3.16

C). Additionally, the re-expression of SETD3 was efficient to bring *Cer1* and *Dkk1* expression to the levels in wild type cells (Figure 3.20). Proper downregulation of the Wnt signaling after mesendoderm formation, and activation of Nodal signaling drive the endoderm differentiation. Though other components of the Wnt signaling pathway were not differentially expressed among wild type and setd3Δ cells, further validation of the downregulation of the Wnt signaling in wild type and setd3Δ cells during endoderm differentiation may yield more valuable data.

BMP signaling is another important pathway in the regulation of pluripotency, and during pattern determination in the gastrulating embryo. Its mis-regulation may cause the observed differentiation defect in the absence of SETD3. Response to BMP was among the STEM analysis results of upregulated DEGs in wild type cells (Figure 3.10). The initial mining of the DEG lists yielded several components of the BMP signaling pathway as differentially expressed (Table 2). However, only the differential expression of *Bmp7* was validated as significantly high on the 3rd day compared to the setd3Δ cells (Figure 3.17 A). *Smad1* and *Bambi* expression differences were found to be statistically insignificant (Figure 3.17 B-C).

Overall, the key pathways including Wnt, Nodal, and BMP signaling pathways, along with the pluripotency and the mesendoderm differentiation networks were affected by the absence of SETD3. The re-expression of SETD3 was sufficient to rescue the defective differentiation phenotype. *Bry*, *Foxa2*, *Gata6*, *Lhx1*, *Gsc*, *Cer1*, and *Dkk1* genes failed to be expressed at significant levels during endoderm differentiation in the absence of SETD3. When the activatory function of SETD3 dependent histone methylation is considered, these genes' expression may be directly controlled by SETD3 during endoderm differentiation, thus, they are selected as candidate genes for further investigation via ChIP experiments. Additionally, the expression levels of *Pou5f1*, *Nanog*, and *Smad7* failed to drop in the absence of SETD3 indicating an indirect control of SETD3 over their expression.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

SETD3 was first coined as a H3K4me2-3 and H3K36me2-3 methyltransferase (Eom et al., 2011). As these modifications mark the actively transcribed genes, we believe that SETD3 has an active role in the activation of some genes, especially during mesendoderm differentiation as suggested by our preliminary data. In this research, we aimed to identify the SETD3 dependent gene expression changes throughout the endoderm differentiation process. For this purpose, we differentiated WT and setd3Δ mouse embryonic stem cells towards endoderm for 4 days and collected total RNA samples from each day. The total RNAs were sequenced, and unsupervised bioinformatic analyses were conducted on the results. In this context, the hierarchical clustering analysis (HCA), and the principal component analysis (PCA) results showed that the differentiation defects started after the 2nd day, and the endoderm differentiation was delayed by one day in the absence of SETD3. Lists of differentially expressed genes (DEGs) for throughout the differentiation of wild type cells, and among WT and setd3Δ cells on each day of differentiation were obtained via differential expression analysis. The detailed bioinformatic analyses including pathway enrichment (using KEGG pathway), STEM, and TF enrichment (ChEA3) on the DEGs revealed that the key pathways involved in mesendoderm differentiation including Wnt, Activin/Nodal, and BMP signaling pathways, as well as the pluripotency network were affected in the absence of SETD3.

A selection of DEGs from the pluripotency network, endoderm differentiation network, and Wnt/β-catenin, Nodal/Activin, BMP signaling pathways were validated to determine if the changes observed via RNA-seq can be confirmed upon close examination via qRT-PCR. The endoderm markers *Bry (T)*, *Foxa2*, *Gata6*, *Lhx1*, and *Gsc*, along with Wnt pathway antagonists *Cer1* and *Dkk1* showed significantly decreased expression levels in the absence of SETD3. Additionally, the

inhibitory mediator of the Nodal/Activin signaling pathways, *Smad7* was expressed at higher levels in the absence of SETD3. Upon re-expression of SETD3, these genes showed normal expression levels when compared to the wild type cells. This suggests that these genes' expression may be directly or indirectly controlled by SETD3.

Surprisingly, pluripotency TFs, *Pou5f1 (Oct4)* and *Nanog* expression levels remained high throughout the differentiation when SETD3 was knocked out, indicating that the exit from the pluripotency is defective in the absence of SETD3. With the re-expression of SETD3, *Pou5f1* expression levels showed a similar pattern as seen in wild type cells. Though not examined, the TF in the network governing the pluripotency exit might be controlled by SETD3 as well.

In conclusion, Wnt, Nodal/Activin, and BMP signaling pathways, along with mesendoderm differentiation and pluripotency networks are affected in the absence of SETD3.

Soon, we plan on conducting ChIP experiments to determine if the SETD3's control on the identified genes is direct or not. As SETD3 catalyzes H3K4 di/trimethylation and H3K36 di/trimethylation, we expect to see enriched immunoprecipitation of gene promoters and/or ends, respectively. For the endoderm markers *Foxa2*, *Bry*, *Gata6*, *Lhx1*, *Gsc*, and for Wnt antagonists *Cer1*, and *Dkk1* ChIP primers targeting both transcription start sites and transcription end sites will be designed.

A more recently identified function of SETD3 is that it methylates His73 residues on actin to provide stabilization (Dai et al., 2019; Guo et al., 2019; Kwiatkowski et al., 2018). Therefore, it also has functions in the cytoplasm. To elucidate whether the mesendoderm differentiation defect we observed in the absence of SETD3 is caused by the nuclear or cytoplasmic function of the SETD3, we plan to repeat the rescue experiments with a nuclear localization signal (NLS) deleted SETD3. We hypothesize that the nuclear function of SETD3 is playing a role in both the exit from pluripotency and activation of differentiation specific genes after the exit from it.

Therefore, SETD3 incapable of entering the nucleus should not be sufficient to rescue the defective phenotype observed in the absence of SETD3.

Lately, peptide-based inhibitors of SETD3 that prevent His73 methylation of actin have been produced (Hintzen et al., 2021). These inhibitors are 16mer peptides of actin that contain a modified histidine residue that is normally methylated by SETD3 (Hintzen et al., 2021), therefore they should selectively inhibit His73 methylation. If the endoderm differentiation defect is caused by the His73 methyltransferase function of SETD3, the effect of these inhibitors can be further investigated during the endoderm differentiation of mESCs. In this case, we would expect to see the same defective phenotype when treated with the inhibitors. However, if the differentiation defect stems from its nuclear function as we hypothesized, we expect to see normal endoderm differentiation of mESCs in the presence of these inhibitors.

After confirming that the nuclear function of SETD3 is necessary for proper mesendoderm differentiation, SET or RSB domain deleted/mutated SETD3 can be stably re-expressed in the setd3Δ cells, and the endoderm differentiation can be repeated. SET domain is the catalytic domain of the SETD3, while RSB domain is required for substrate recognition. In either case, we expect to see the same differentiation defect, as SETD3 cannot function without either domain.

As a future direction, we plan on confirming the gene expression changes seen in the absence of SETD3 are also seen on the protein level as well. The decreased expression levels of Wnt antagonists (*Cer1* and *Dkk1*) and endoderm markers (*Bry*, *Foxa2*, *Gata6*, *Lhx1*, and *Gsc*), and the increased expression levels of pluripotency marker *Pou5f1*, and the inhibitory SMAD of the Activin/Nodal signaling pathways, *Smad7*, will be validated via western blot analysis.

REFERENCES

Abaev-Schneiderman, E., Admoni-Elisha, L., & Levy, D. (2019). SETD3 is a positive regulator of DNA-damage-induced apoptosis. *Cell Death and Disease*, *10*(2). https://doi.org/10.1038/s41419-019-1328-4

Aiken, C. E. M., Swoboda, P. P. L., Skepper, J. N., & Johnson, M. H. (2004). The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. *Reproduction*, *128*(5), 527–535. https://doi.org/10.1530/rep.1.00281

Arnold, S. J., & Robertson, E. J. (2009). Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, *10*(2), 91–103. https://doi.org/10.1038/nrm2618

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., & Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines. *Nature Cell Biology*, *8*(5), 532–538. https://doi.org/10.1038/ncb1403

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E. J. J., & Constam, D. B. (2006). The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. *Developmental Cell*, *11*(3), 313–323. https://doi.org/10.1016/j.devcel.2006.07.005

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, *125*(2), 315–326. https://doi.org/10.1016/j.cell.2006.02.041

Bird, A. (1992). The essentials of DNA methylation. *Cell*, *70*(1), 5–8.

https://doi.org/https://doi.org/10.1016/0092-8674(92)90526-I

Biswas, S., & Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology*, *837*(June), 8–24. https://doi.org/10.1016/j.ejphar.2018.08.021

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H., & Klose, R. J. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, *157*(6), 1445–1459. https://doi.org/10.1016/j.cell.2014.05.004

Bledau, A. S., Schmidt, K., Neumann, K., Hill, U., Ciotta, G., Gupta, A., Torres, D. C., Fu, J., Kranz, A., Stewart, A. F., & Anastassiadis, K. (2014). The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development (Cambridge)*, *141*(5), 1022–1035. https://doi.org/10.1242/dev.098152

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., & Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, *441*(7091), 349–353. https://doi.org/10.1038/nature04733

Boyer, L. A., Tong, I. L., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., & Young, R. A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, *122*(6), 947–956. https://doi.org/10.1016/j.cell.2005.08.020

Brand, M., Nakka, K., Zhu, J., & Dilworth, F. J. (2019). Polycomb/Trithorax Antagonism: Cellular Memory in Stem Cell Fate and Function. *Cell Stem Cell*, *24*(4), 518–533. https://doi.org/10.1016/j.stem.2019.03.005

Bühler, M., & Gasser, S. M. (2009). Silent chromatin at the middle and ends: Lessons from yeasts. *EMBO Journal*, *28*(15), 2149–2161. https://doi.org/10.1038/emboj.2009.185

Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., Del Barrio, M. G., Portillo, F., & Nieto, M. A. (2000). The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biology*, *2*(2), 76–83. https://doi.org/10.1038/35000025

Carver, E. A., Jiang, R., Lan, Y., Oram, K. F., & Gridley, T. (2001). The Mouse Snail Gene Encodes a Key Regulator of the Epithelial-Mesenchymal Transition. *Molecular and Cellular Biology*, *21*(23), 8184–8188. https://doi.org/10.1128/mcb.21.23.8184-8188.2001

Chazaud, C., Yamanaka, Y., Pawson, T., & Rossant, J. (2006). Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Developmental Cell*, *10*(5), 615–624. https://doi.org/10.1016/j.devcel.2006.02.020

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., … Ng, H. H. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*, *133*(6), 1106–1117. https://doi.org/10.1016/j.cell.2008.04.043

Chen, Z., Yan, C. T., Dou, Y., Viboolsittiseri, S. S., & Wang, J. H. (2013). The role of a newly identified SET domain-containing protein, SETD3, in oncogenesis. *Haematologica*, *98*(5), 739–743. https://doi.org/10.3324/haematol.2012.066977

Chiarella, A. M., Lu, D., & Hathaway, N. A. (2020). Epigenetic control of a local chromatin landscape. *International Journal of Molecular Sciences*, *21*(3), 6–8.

https://doi.org/10.3390/ijms21030943

Cohn, O., Feldman, M., Weil, L., Kublanovsky, M., & Levy, D. (2016). Chromatin associated SETD3 negatively regulates VEGF expression. *Scientific Reports*, *6*(October), 1–10. https://doi.org/10.1038/srep37115

Dai, S., Horton, J. R., Woodcock, C. B., Wilkinson, A. W., Zhang, X., Gozani, O., & Cheng, X. (2019). Structural basis for the target specificity of actin histidine methyltransferase SETD3. *Nature Communications*, *10*(1), 1–8. https://doi.org/10.1038/s41467-019-11554-6

Diep, J., Ooi, Y. S., Wilkinson, A. W., Peters, C. E., Foy, E., Johnson, J. R., Zengel, J., Ding, S., Weng, K. F., Laufman, O., Jang, G., Xu, J., Young, T., Verschueren, E., Kobluk, K. J., Elias, J. E., Sarnow, P., Greenberg, H. B., Hüttenhain, R., … Carette, J. E. (2019). Enterovirus pathogenesis requires the host methyltransferase SETD3. *Nature Microbiology*, *4*(12), 2523–2537. https://doi.org/10.1038/s41564-019-0551-1

Eirín-López, J. M., González-Romero, R., Dryhurst, D., Méndez, J., & Ausió, J. (2009). Long-Term Evolution of Histone Families: Old Notions and New Insights into Their Mechanisms of Diversification Across Eukaryotes. In *Evolutionary Biology* (pp. 139–162). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-00952-5_8

Eom, G. H., Kim, K. B., Kim, J. H., Kim, J. Y., Kim, J. R., Kee, H. J., Kim, D. W., Choe, N., Park, H. J., Son, H. J., Choi, S. Y., Kook, H., & Seo, S. B. (2011). Histone methyltransferase SETD3 regulates muscle differentiation. *Journal of Biological Chemistry*, *286*(40), 34733–34742. https://doi.org/10.1074/jbc.M110.203307

Ernst, J., & Bar-Joseph, Z. (2006). STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, *7*, 1–11. https://doi.org/10.1186/1471-2105-7-191

Gadue, P., Huber, T. L., Paddison, P. J., & Keller, G. M. (2006). Wnt and TGF-β

signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(45), 16806–16811. https://doi.org/10.1073/pnas.0603916103

Gao, W.-L., & Liu, H.-L. (2007). DOT1: a distinct class of histone lysine methyltransferase. *Yi chuan = Hereditas*, *29*(12), 1449–1454.

Gardner, K. E., Allis, C. D., & Strahl, B. D. (2011). Operating on chromatin, a colorful language where context matters. *Journal of Molecular Biology*, *409*(1), 36–46. https://doi.org/10.1016/j.jmb.2011.01.040

Graham, S. J. L., Wicher, K. B., Jedrusik, A., Guo, G., Herath, W., Robson, P., & Zernicka-Goetz, M. (2014). BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. *Nature Communications*, *5*(May). https://doi.org/10.1038/ncomms6667

Guo, Q., Liao, S., Kwiatkowski, S., Tomaka, W., Yu, H., Wu, G., Tu, X., Min, J., Drozak, J., & Xu, C. (2019). Structural insights into SETD3-mediated histidine methylation on β-actin. *BioRxiv*, *73*, 1–20. https://doi.org/10.1101/476705

Hebbes, T. R., Thorne, A. W., & Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO Journal*, *7*(5), 1395–1402. https://doi.org/10.1002/j.1460-2075.1988.tb02956.x

Hintzen, J. C. J., Moesgaard, L., Kwiatkowski, S., Drozak, J., Kongsted, J., & Mecinović, J. (2021). β-Actin Peptide-Based Inhibitors of Histidine Methyltransferase SETD3. *ChemMedChem*, *16*(17), 2695–2702. https://doi.org/10.1002/cmdc.202100296

Irion, S., Nostro, M. C., Kattman, S. J., & Keller, G. M. (2008). Directed differentiation of pluripotent stem cells: From developmental biology to therapeutic applications. *Cold Spring Harbor Symposia on Quantitative*

Biology. https://doi.org/10.1101/sqb.2008.73.065

Jenuwein, T., Laible, G., Dorn, R., & Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cellular and Molecular Life Sciences*, *54*(1), 80–93. https://doi.org/10.1007/s000180050127

Jiang, X., Li, T., Sun, J., Liu, J., & Wu, H. (2018). SETD3 negatively regulates VEGF expression during hypoxic pulmonary hypertension in rats. *Hypertension Research*, *41*(9), 691–698. https://doi.org/10.1038/s41440-018-0068-7

Jones, P. A. (1996). DNA methylation errors and cancer. *Cancer Research*, *56*(11), 2463–2467.

Kim, H., Park, H. J., Choi, H., Chang, Y., Park, H., Shin, J., Kim, J., Lengner, C. J., Lee, Y. K., & Kim, J. (2019). Modeling G2019S-LRRK2 Sporadic Parkinson's Disease in 3D Midbrain Organoids. *Stem Cell Reports*, *12*(3), 518–531. https://doi.org/10.1016/j.stemcr.2019.01.020

Kim, J. A., Kwon, M., & Kim, J. (2019). Allosteric Regulation of Chromatin-Modifying Enzymes. *Biochemistry*, *58*(1), 15–23. https://doi.org/10.1021/acs.biochem.8b00894

Kitajima, S., Takagi, A., Inoue, T., & Saga, Y. (2000). MesP1 and MesP2 are essential for the development of cardiac mesoderm. *Development (Cambridge, England)*, *127*(15), 3215–3226.

Kubo, A., Shinozaki, K., Shannon, J. M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H. J., & Keller, G. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development*, *131*(7), 1651–1662. https://doi.org/10.1242/dev.01044

Kwiatkowski, S., Seliga, A. K., Veiga-da-Cunha, M., Vertommen, D., Terreri, M., Ishikawa, T., Grabowska, I., Jagielski, A. K., & Drozak, J. (2018). SETD3

protein is the actin-specific histidine N-methyltransferase. *BioRxiv*, *3*(2016), 1–42. https://doi.org/10.1101/266882

Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nature Reviews Molecular Cell Biology*, *15*(3), 178–196. https://doi.org/10.1038/nrm3758

Li, G., & Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics and Development*, *21*(2), 175–186. https://doi.org/10.1016/j.gde.2011.01.022

Li, Q., Zhang, Y., & Jiang, Q. (2019). SETD3 reduces KLC4 expression to improve the sensitization of cervical cancer cell to radiotherapy. *Biochemical and Biophysical Research Communications*, *516*(3), 619–625. https://doi.org/10.1016/j.bbrc.2019.06.058

Liao, J., Luo, S., Yang, M., & Lu, Q. (2020). Overexpression of CXCR5 in CD4+ T cells of SLE patients caused by excessive SETD3. *Clinical Immunology*, *214*(January), 108406. https://doi.org/10.1016/j.clim.2020.108406

Lin, S., & Talbot, P. (2011). Methods for Culturing Mouse and Human Embryonic Stem Cells. In *Nature cell biology* (Vol. 7, Issue 3, pp. 31–56). https://doi.org/10.1007/978-1-60761-962-8_2

Lindsley, R. C., Gill, J. G., Murphy, T. L., Langer, E. M., Cai, M., Mashayekhi, M., Wang, W., Niwa, N., Nerbonne, J. M., Kyba, M., & Murphy, K. M. (2008). Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. *Cell Stem Cell*, *3*(1), 55–68. https://doi.org/10.1016/j.stem.2008.04.004

Lu, F., & Zhang, Y. (2015). Cell totipotency: Molecular features, induction, and maintenance. *National Science Review*, *2*(2), 217–225. https://doi.org/10.1093/nsr/nwv009

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J.

(1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luger, K., & Richmond, T. J. (1998). The histone tails of the nucleosome. *Current Opinion in Genetics and Development*, *8*(2), 140–146. https://doi.org/10.1016/S0959-437X(98)80134-2

Ma, X., Wang, Q., Jiang, Y., Xiao, Z., Fang, X., & Chen, Y. G. (2007). Lateral diffusion of TGF-β type I receptor studied by single-molecule imaging. *Biochemical and Biophysical Research Communications*, *356*(1), 67–71. https://doi.org/10.1016/j.bbrc.2007.02.080

Manejwala, F. M., Cragoe, E. J., & Schultz, R. M. (1989). Blastocoel expansion in the preimplantation mouse embryo: Role of extracellular sodium and chloride and possible apical routes of their entry. *Developmental Biology*, *133*(1), 210–220. https://doi.org/10.1016/0012-1606(89)90312-6

Martello, G., Bertone, P., & Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO Journal*, *32*(19), 2561–2574. https://doi.org/10.1038/emboj.2013.177

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(12 II), 7634–7638. https://doi.org/10.1073/pnas.78.12.7634

Mihajlović, A. I., & Bruce, A. W. (2017). The first cell-fate decision of mouse preimplantation embryo development: Integrating cell position and polarity. *Open Biology*, *7*(11). https://doi.org/10.1098/rsob.170210

Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., & Shilatifard, A. (2001). COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 12902–12907. https://doi.org/10.1073/pnas.231473398

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., & Yamanaka, S. (2003). The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, *113*(5), 631–642. https://doi.org/10.1016/S0092-8674(03)00393-3

Musselman, C. A., Lalonde, M. E., Côté, J., & Kutateladze, T. G. (2012). Perceiving the epigenetic landscape through histone readers. *Nature Structural and Molecular Biology*, *19*(12), 1218–1227. https://doi.org/10.1038/nsmb.2436

Orkin, S. H., & Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. *Cell*, *145*(6), 835–850. https://doi.org/10.1016/j.cell.2011.05.019

Paranjpe, S. S., & Veenstra, G. J. C. (2015). Establishing pluripotency in early development. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1849*(6), 626–636. https://doi.org/10.1016/j.bbagrm.2015.03.006

Peinado, H., Olmeda, D., & Cano, A. (2007). Snail, ZEB and bHLH factors in tumour progression: An alliance against the epithelial phenotype? *Nature Reviews Cancer*, *7*(6), 415–428. https://doi.org/10.1038/nrc2131

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtier, K., Ponting, C. P., Allis, C. D., & Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, *406*(6796), 593–599. https://doi.org/10.1038/35020506

Rivera-Pérez, J. A., & Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. *Developmental Biology*, *288*(2), 363–371. https://doi.org/10.1016/j.ydbio.2005.09.012

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Robinson, P. J. J., Fairall, L., Huynh, V. A. T., & Rhodes, D. (2006). EM measurements define the dimensions of the “30-nm” chromatin fiber: Evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(17), 6506–6511. https://doi.org/10.1073/pnas.0601212103

Scheibner, K., Schirge, S., Burtscher, I., Büttner, M., Sterr, M., Yang, D., Böttcher, A., Ansarullah, Irmler, M., Beckers, J., Cernilogar, F. M., Schotta, G., Theis, F. J., & Lickert, H. (2021). Epithelial cell plasticity drives endoderm formation during gastrulation. *Nature Cell Biology*. https://doi.org/10.1038/s41556-021-00694-x

Schuettengruber, B., Bourbon, H. M., Di Croce, L., & Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*, *171*(1), 34–57. https://doi.org/10.1016/j.cell.2017.08.002

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: Mechanisms of regulation in development and disease pathogenesis. *Annual Review of Biochemistry*, *81*, 65–95. https://doi.org/10.1146/annurev-biochem-051710-134100

Simons, K., & Toomre, D. (2000). Lipid rafts and signal transduction. *Nature Reviews Molecular Cell Biology*, *1*(1), 31–39. https://doi.org/10.1038/35036052

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0021800

Sze, C. C., Cao, K., Collings, C. K., Marshall, S. A., Rendleman, E. J., Ozark, P. A., Chen, F. X., Morgan, M. A., Wang, L., & Shilatifard, A. (2017). Histone H3K4 methylation-dependent and -independent functions of set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes and Development*, *31*(17), 1732–1737.

https://doi.org/10.1101/gad.303768.117

Tam, P. P. L., & Loebel, D. A. F. (2007). Gene function in mouse embryogenesis: Get set for gastrulation. *Nature Reviews Genetics*, *8*(5), 368–381. https://doi.org/10.1038/nrg2084

Terzi Cizmecioglu, N., Huang, J., Keskin, E. G., Wang, X., Esen, I., Chen, F., & Orkin, S. H. (2020). ARID4B is critical for mouse embryonic stem cell differentiation towards mesoderm and endoderm, linking epigenetics to pluripotency exit. *Journal of Biological Chemistry*, *295*(51), 17738–17751. https://doi.org/10.1074/jbc.RA120.015534

Thomas, P., & Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Current Biology*, *6*(11), 1487–1496. https://doi.org/10.1016/S0960-9822(96)00753-1

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., & Mann, M. (2010). Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell*, *142*(6), 967–980. https://doi.org/10.1016/j.cell.2010.08.020

White, M. D., Bissiere, S., Alvarez, Y. D., & Plachta, N. (2016). Mouse Embryo Compaction. In *Current Topics in Developmental Biology* (1st ed., Vol. 120). Elsevier Inc. https://doi.org/10.1016/bs.ctdb.2016.04.005

Wiley, L. M. (1984). Cavitation in the mouse preimplantation embryo: Na K-ATPase and the origin of nascent blastocoele fluid. *Developmental Biology*, *105*(2), 330–342. https://doi.org/10.1016/0012-1606(84)90290-2

Wilkinson, A. W., Diep, J., Dai, S., Liu, S., Ooi, Y. S., Song, D., Li, T. M., Horton, J. R., Zhang, X., Liu, C., Trivedi, D. V., Ruppel, K. M., Vilches-Moure, J. G., Casey, K. M., Mak, J., Cowan, T., Elias, J. E., Nagamine, C. M., Spudich, J. A., … Gozani, O. (2019). SETD3 is an actin histidine methyltransferase that prevents primary dystocia. *Nature*, *565*(7739), 372–376. https://doi.org/10.1038/s41586-018-0821-8

Wray, J., Kalkan, T., & Smith, A. G. (2010). The ground state of pluripotency. In *Biochemical Society Transactions* (Vol. 38, Issue 4, pp. 1027–1032). https://doi.org/10.1042/BST0381027

Wu, H., & Zhang, Y. (2014). Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell*, *156*(1–2), 45–68. https://doi.org/10.1016/j.cell.2013.12.019

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N., & McMahon, A. P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes & Development*, *13*(24), 3185–3190. https://doi.org/10.1101/gad.13.24.3185

Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawiot, W., Behringer, R. R., Ang, S. L., Hamada, H., & Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature*, *428*(6981), 387–392. https://doi.org/10.1038/nature02418

Ying, Q. L., & Smith, A. G. (2003). Defined Conditions for Neural Commitment and Differentiation. *Methods in Enzymology*, *365*, 327–341. https://doi.org/10.1016/S0076-6879(03)65023-8

Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., & Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnology*, *21*(2), 183–186. https://doi.org/10.1038/nbt780

APPENDICES

A. Media Recipes for Cell Culture

2i4 Medium (Low-serum medium, 100 mL): 50 mL Neurobasal Medium (Cat no.: 21103049, Gibco), 50 mL DMEM/F-12 (Cat. No.: 11320074, Gibco), 500 µL N-2 Supplement (100X) (Cat. No.: 17502048, Gibco), B-27™ Supplement (50X), serum free (Cat. No.: 17504044, Gibco), 500 µL 10% BSA, 1 mL GlutaMAX I (Cat. No.: 35050061, Gibco), 1 mL Pen/Strep (Cat. No.: 15140122, Gibco), 1.3 µL MTG (1-Thioglycerol) (Cat. No.: M6145-25ML, Sigma), 4 mL Fetal Bovine Serum (FBS) (Cat. No.:10270106, Gibco). Supplement with final concentrations of 1000 units/mL Leukemia Inhibitory Factor (LIF) (Cat. No.: ESG1107, Millipore), 3 µM CHIR-99021 (Cat. No.: S2924, Selleckchem) and 1 µM PD0325901 (Cat. No.: S1036, Selleckchem).

MEF Medium (100 mL): 88 mL DMEM (Cat. No.: 41966029, Gibco), 10 mL Fetal Bovine Serum (FBS) (Cat. No.:10270106, Gibco), 1 mL GlutaMAX I (Cat. No.: 35050061, Gibco), 1 mL Pen/Strep (Cat. No.: 15140122, Gibco).

ESC Medium (High-serum medium, 100 mL): 80 mL DMEM (Cat. No.: 41966029, Gibco), 15 mL Fetal Bovine Serum (FBS) (Cat. No.:10270106, Gibco), 2 mL Pen/Strep (Cat. No.: 15140122, Gibco), 1 mL Nucleoside Mix, 1 mL GlutaMAX I (Cat. No.: 35050061, Gibco), 1 mL MEM NEAA (Cat. No.: 11140-035, Gibco), 0.704 µL β-Mercaptoethanol (Cat. No.: M-6250, Sigma), 1000 units/mL Leukemia Inhibitory Factor (LIF) (Cat. No.: ESG1107, Millipore).

Serum-free Differentiation Base Medium (100 mL): 75 mL IMDM (Cat. No.: 21980032, Gibco), 25 mL Ham's F-12 Nutrient Mix with GlutaMAX™ Supplement (Cat. No.: 31765027, Gibco), 5 mL 10% BSA, 1 mL B-27 Supplement without Vitamin A (50X) (Cat. No.: 12587010, Gibco), 500 µL N-2 Supplement (100X) (Cat. No.: 17502048, Gibco), 1 mL GlutaMAX I (Cat. No.: 35050061, Gibco), 1 mL L-Ascorbic Acid Solution, 300 µL MTG Solution. **For endoderm commitment**,

Serum-free Differentiation Base Medium was supplemented with 75 ng/mL Activin A (Cat. No.: 120-14P, Peprotech) after the second day of differentiation (referred to as "**Serum-free Endoderm Differentiation Medium**").

B. Solution Recipes

10% BSA (w/v): 5 g Bovine Serum Albumin (BSA) (Cat. No.: A3311-50G, Sigma) in 50 mL Dulbecco's Phosphate Buffered Saline (DPBS) (Cat. No.: 02-023-1A, Biological Industries). Filter-sterilized and stored at 4°C.

Nucleoside Mix: 80 mg Adenosine (Cat. No.: A4036-5G, Sigma), 85 mg Guanosine (Cat. No.: G5264-1G, Sigma), 73 mg Uridine (Cat. No.: U3003-5G, Sigma), 73 mg Cytidine (Cat. No.: C4654-1G, Sigma), and 24 mg Thymidine (Cat. No.: T1895-1G, Sigma) in 100 mL distilled water. Dissolved by warming to 45°C. Filter-sterilized and aliquoted while warm. Stored at -20°C.

L-Ascorbic Acid Solution: 5 mg L-Ascorbic acid (Cat. No.: A4544-25G, Sigma) in 10 mL cell culture grade water (Cat. No.: BI03-055-1A, Biological Industries). Dissolved by heating the solution to 37°C.

MTG (1-Thioglycerol) Solution: 13 µL MTG (Cat. No.: M6145-25ML, Sigma) diluted in 1 mL IMDM.

TBS: 48.4 g Tris-base (Cat. No.: 37190.02, Serva) and 160 g NaCl (Cat. No.: 1.06404.1000, Merck) in 1 L distilled water to obtain 20X TBS. Diluted to 1X with distilled water before use. Stored at room temperature.

TBS-T: 10 mL %10 Tween-20 (Cat. No.:0777-1L, VWR) stock and 50 mL 20X TBS in 1 L distilled water.

C. Primers Used in qRT-PCR Analysis

Table 3 The list of primers used in qRT-PCR analyses.

Primer Name	*Primer Sequence (5'- 3')*
β-actin - Forward	ATG AAG ATC CTG ACC GAG CG
β-actin - Reverse	TAC TTG CGC TGA GGA GGA GC
Brachyury (Bry, T) – Forward	CAT GTA CTC TTT CTT GCT GG
Brachyury (Bry, T) – Reverse	GGT CTC GGG AAA GCA GTG GC
Foxa2 - Forward	TGG TCA CTG GGG ACA AGG GAA
Foxa2 - Reverse	GCA ACA ACA GCA ATA GAG AAC
Oct4 (Pou5f1) - Forward	CTG AGG GCC AGG CAG GAG GAC GAG
Oct4 (Pou5f1) - Reverse	CTG TAG GGA GGG CTT CGG GCA CTT
Nanog - Forward	ATG AAG TGC AAG CGG TGG CAG AAA
Nanog - Reverse	CCT GGT GGA GTC ACA GAG TAG TTC
Smad7 - Forward	GTC CAG ATG CTG TAC CTT CCT C
Smad7 - Reverse	GCG AGT CTT CTC CTC CCA GTA T
Gata6 - Forward	GCC GCA GGC CTG ACT CCT G
Gata6 - Reverse	ACG CGC TTC TGT GGC TTG ATG A
Cer1 - Forward	ATC CTG CCC ATC AAA AGC CAC G
Cer1 - Reverse	CGA ATG GAA CTG CAT TTG CCA AAG
Dkk1 - Forward	GGA AAT TGA GGA AAG CAT CAT TGA A
Dkk1 - Reverse	CAG ATC TTG GAC CAG AAG TGT CTT G
Ror2 - Forward	ATC GAC ACC TTG GGA CAA CC
Ror2 - Reverse	AGT GCA GGA TTG CCG TCT G
Axin2 - Forward	TGA CTC TCC TTC CAG ATC CCA
Axin2 - Reverse	TGC CCA CAC TAG GCT GAC A
Bmp7 - Forward	GTG GTC AAC CCT CGG CAC A
Bmp7 - Reverse	GGC GTC TTG GAG CGA TTC TG
Smad1 - Forward	ATG GTT TCA CAG ATC CGT CCA
Smad1 - Reverse	TCC CAA TAT GTC GCC TGG TGT
Bambi - Forward	GAT CGC CAC TCC AGC TAC TTC
Bambi - Reverse	GCA GGC ACT AAG CTC AGA CTT
Gsc - Forward	ACC ATC TTC ACC GAT GAG CAG C
Gsc - Reverse	CTT GGC TCG GCG GTT CTT AAA C
Lhx1 - Forward	CCC AGC TTT CCC GAA TCC T
Lhx1 – Reverse	GCG GGA CGT AAA TAA ATA AAA TGG

D. Antibodies Used in Western Blotting

Table 4 The list of antibodies used in western blotting.

Antibody	*Host*	*Catalog Number, Vendor*
Anti-SETD3	Rabbit	a304-071a, Bethyl
Anti-GAPDH	Rabbit	2118S, CST
Anti-H3	Mouse	sc-517576, Santa Cruz
Anti-Rabbit	Goat	ab97051, Abcam
Anti-Mouse	Goat	ab97023, Abcam

E. The Quality of RNA Samples Sent for RNA-sequencing

Table 5 The quality and the concentrations of the total RNA samples sent for RNA-seq measured via Bioanalyzer 2100 (Rep: Replicate. WT: Wild type, KO: Knock-out. mESC: Mouse Embryonic Stem Cell).

Sample Name	*Concentration (ng/uL)*	*RIN*	*rRNA Ratio*
Rep 1 - WT mESC	315	9.7	1.3
Rep 2 - WT mESC	427	9.8	1.7
Rep 3 - WT mESC	355	10	1.6
Rep 1 - WT Day 2	328	10	1.8
Rep 2 - WT Day 2	337	9.7	1.4
Rep 3 - WT Day 2	325	10	1.9
Rep 1 - WT Day 3	323	9.8	1.3
Rep 2 - WT Day 3	496	9.9	1.7
Rep 3 - WT Day 3	274	10	2.0
Rep 1 - WT Day 4	234	9.8	1.3
Rep 2 - WT Day 4	387	9.8	1.4
Rep 3 - WT Day 4	302	10	1.6
Rep 1 - KO mESC	377	10	1.7
Rep 2 - KO mESC	401	10	1.8
Rep 3 - KO mESC	275	10	2.1
Rep 1 - KO Day 2	331	9.9	1.8
Rep 2 - KO Day 2	372	10	1.9
Rep 3 - KO Day 2	298	10	2.1
Rep 1 - KO Day 3	340	10	1.9
Rep 2 - KO Day 3	350	10	1.9
Rep 3 - KO Day 3	282	10	1.9
Rep 1 - KO Day 4	387	9.9	1.8
Rep 2 - KO Day 4	227	9.9	1.5

F. Quality Control Results of RNA-seq Libraries

Table 6 Quality control results of TruSeq Stranded mRNA libraries (Rep: Replicate. WT: Wild type, KO: Knock-out. mESC: Mouse Embryonic Stem Cell).

Library Name	*Conc. (ng/uL)*	*Conc. (nM)*	*Base Pair*
Rep 1 - WT mESC	37.69	185.24	313
Rep 2 - WT mESC	81.71	404.22	311
Rep 3 - WT mESC	57.09	280.59	313
Rep 1 - WT Day 2	50.64	251.3	310
Rep 2 - WT Day 2	32.95	164.57	308
Rep 3 - WT Day 2	64.85	322.88	309
Rep 1 - WT Day 3	30.93	158.07	301
Rep 2 - WT Day 3	55.76	280.34	306
Rep 3 - WT Day 3	39.43	190.76	318
Rep 1 - WT Day 4	32.34	152.6	326
Rep 2 - WT Day 4	44.00	209.59	323
Rep 3 - WT Day 4	27.78	130.69	327
Rep 1 - KO mESC	35.19	164.55	329
Rep 2 - KO mESC	29.63	139.81	326
Rep 3 - KO mESC	56.72	261.26	334
Rep 1 - KO Day 2	35.12	167.26	323
Rep 2 - KO Day 2	68.39	331.89	317
Rep 3 - KO Day 2	59.44	287.59	318
Rep 1 - KO Day 3	40.42	193.1	322
Rep 2 - KO Day 3	80.67	397.76	312
Rep 3 - KO Day 3	68.48	336.59	313
Rep 1 - KO Day 4	52.23	254.29	316
Rep 2 - KO Day 4	49.85	239.65	320
Rep 3 - KO Day 4	69.88	346.82	310

G. Wnt Signaling Pathway

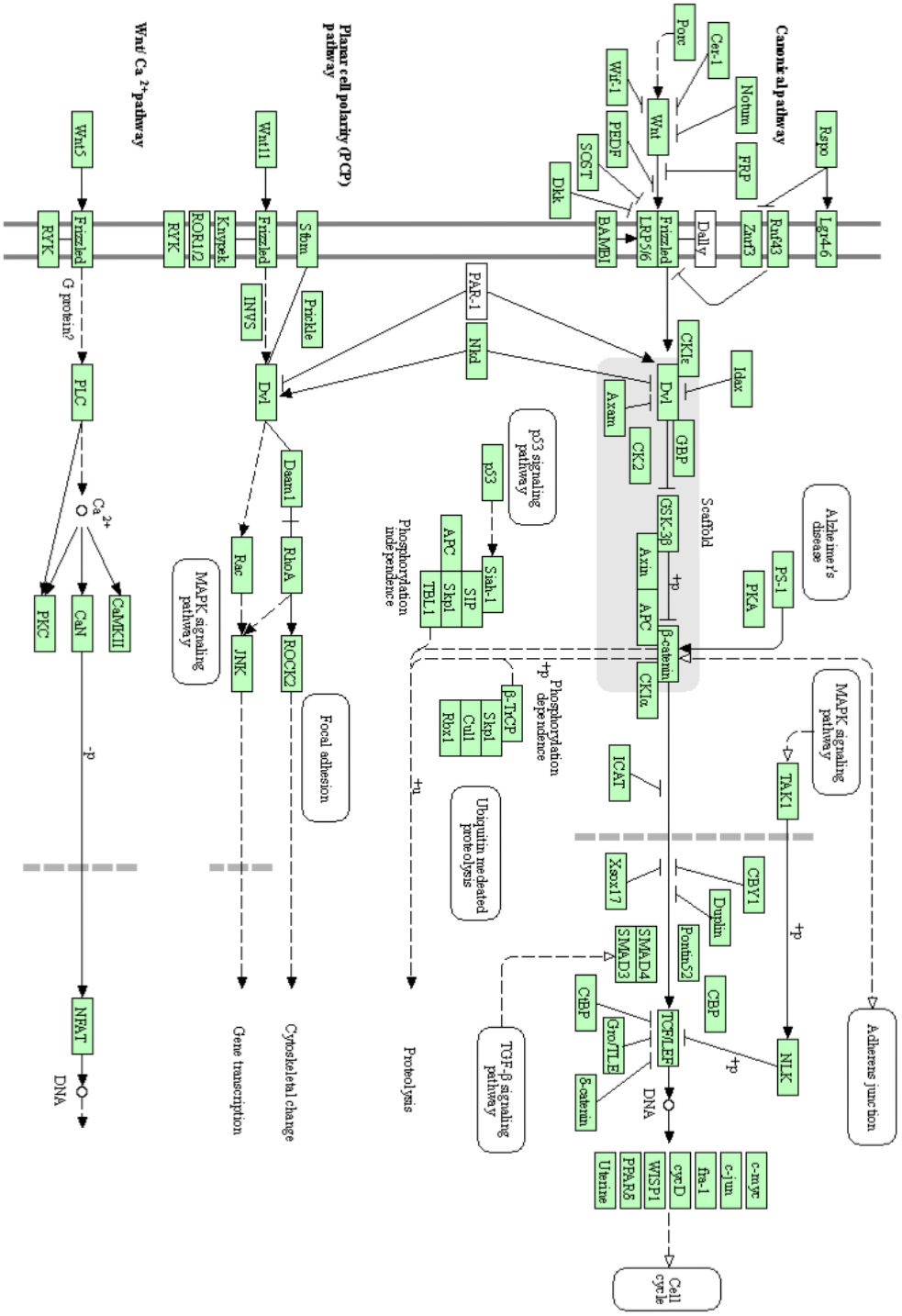

Figure S.1 Wnt signaling pathway derived from the KEGG database.

H. BMP, Nodal, and Activin Signaling Pathways

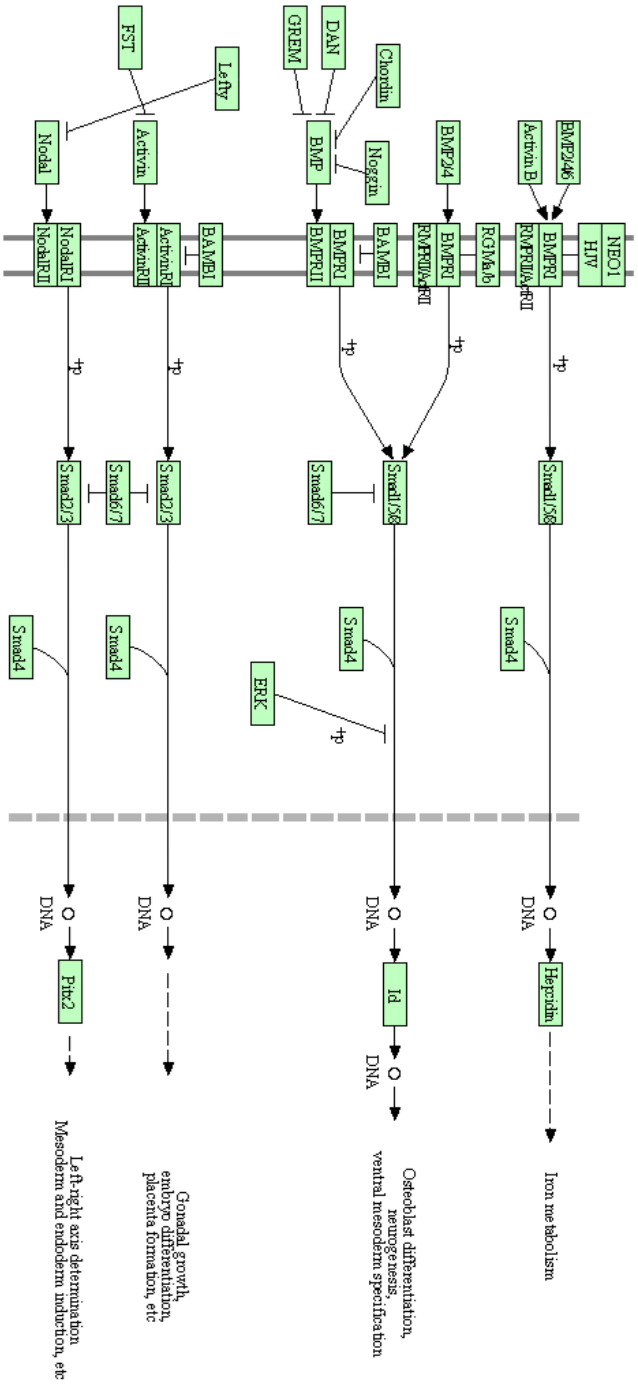

Figure S.2 BMP, Nodal, Activin signaling pathways derived from the KEGG database.

I. Pathways Related to the Pluripotency Network

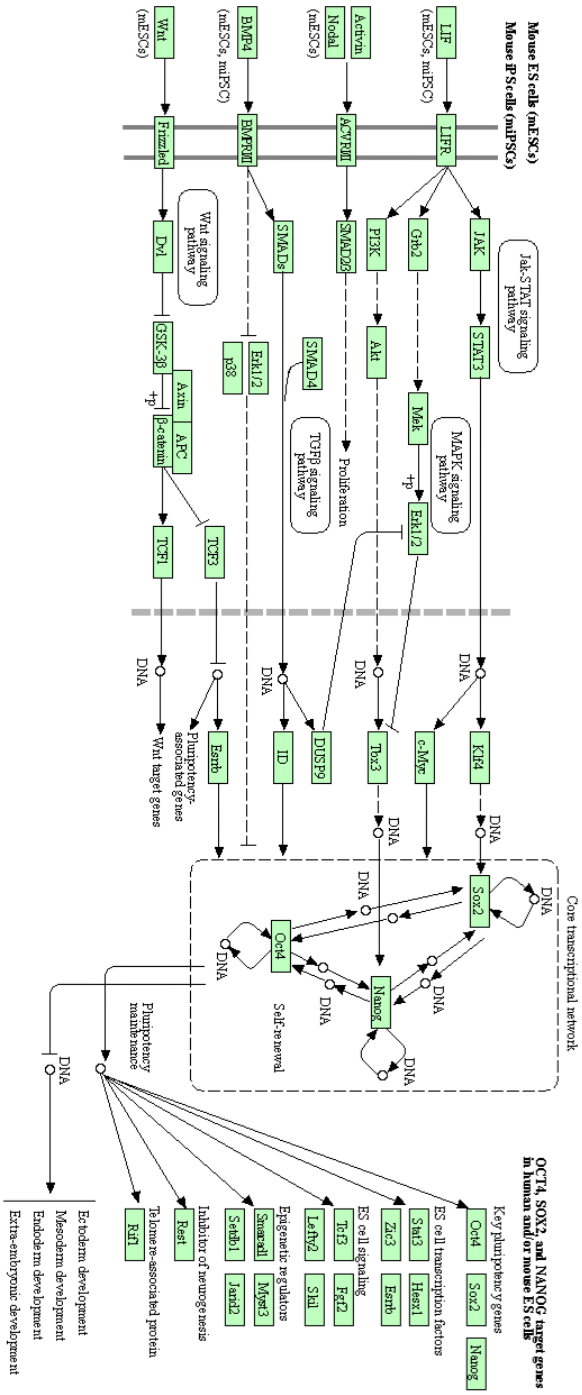

Figure S.3 Pathways regulating the pluripotency of mESCs. The pathway view was derived from the KEGG database.