INVESTIGATING THE ROLE AND REGULATION OF H3K36 METHYLATION IN NEUROECTODERMAL LINEAGE COMMITMENT OF MOUSE EMBRYONIC STEM CELLS

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

BY

DERSU SEZGİNMERT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULER BIOLOGY AND GENETICS

AUGUST 2023

Approval of the thesis:

INVESTIGATING THE ROLE AND REGULATION OF H3K36 METHYLATION IN NEUROECTODERMAL LINEAGE COMMITMENT OF MOUSE EMBRYONIC STEM CELLS

submitted by **DERSU SEZGİNMERT** in partial fulfillment of the requirements for the degree of **Master of Science** in **Molecular Biology and Genetics, Middle East Technical University** by,

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

Prof. Dr. Mesut Muyan
Head of the Department, **Biological Sciences, METU** _____

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

Examining Committee Members:

Assoc. Prof. Dr. Erkan Kiriş
Biological Sciences, METU _____

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

Assist. Prof. Dr. Bahar Değirmenci Uzun
Molecular Biology and Genetics, Bilkent University _____

Date: 28.08.2023

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: Dersu Sezginmert

Signature:

ABSTRACT

INVESTIGATING THE ROLE AND REGULATION OF H3K36 METHYLATION IN NEUROECTODERMAL LINEAGE COMMITMENT OF MOUSE EMBRYONIC STEM CELLS

Sezginmert, Dersu
Master of Science, Molecular Biology and Genetics
Supervisor : Assist. Prof. Dr. Nihal Terzi Çizmecioğlu

August 2023, 86 pages

Embryonic development is a complex process orchestrated by spatiotemporally regulated factors. Embryonic stem cells (ESCs) serve as invaluable *in vitro* models for studying these developmental dynamics. Epigenetic mechanisms, crucial in this regulation, guide pluripotent ESCs towards specific lineages. A deeper understanding of these mechanisms is essential for addressing developmental disorders. A previously conducted mass spectrometry analysis revealed an elevation in H3K36me2 levels during neuroectoderm differentiation, while such an increase was absent in endoderm differentiation. Our preliminary data hints a distinctive role for H3K36 methylation in neuroectoderm differentiation. After these findings were validated by Western blot, expression changes in H3K36 methylation regulatory system, including methyltransferases, demethylases, and readers, were analyzed. RT-qPCR analyses of neuroectoderm, complemented with existing RNA-seq data for endoderm, indicated a central role for H3K36 methylation in pluripotency exit and/or differentiation. Notably, *Nsd2* and *Zmynd11* expression levels were upregulated during differentiation into both lineages, whereas *Kdm4c* and *Msh6* were found to be expressed higher in mESCs compared to differentiated states. Conversely, lineage-specific expression patterns were noted for *Nsd1* and *Setd5*.

Finally, through a CRISPR knock out screening done to identify the critical epigenetic factors for neuroectoderm differentiation, we revealed that *Nsd1, Setd5* and *Kdm2b* are critical for neuroectoderm differentiation. These findings highlight the importance of H3K36me2 regulation for lineage choice and exit from pluripotency, with regulatory proteins showing stage- or lineage-specific expression changes. Further work is needed to uncover the precise mechanisms by which this modification and its associated proteins influence neuroectoderm differentiation.

Keywords: Epigenetics, H3K36 Methylation, Mouse Embryonic Stem Cells, Neuroectoderm Differentiation

ÖZ

H3K36 METİLASYONUN FARE EMBRİYONİK KÖK HÜCRELERİNİN NÖROEKTODERMAL HÜCRE HATTINA YÖNELİMİNDEKİ ROLÜ VE DÜZENLENMESİNİN İNCELENMESİ

Sezginmert, Dersu
Yüksek Lisans, Moleküler Biyoloji ve Genetik
Tez Yöneticisi : Dr. Öğr. Üyesi Nihal Terzi Çizmecioğlu

Ağustos 2023, 86 sayfa

Embriyonik gelişim, zamansal ve mekansal olarak düzenlenmiş faktörler tarafından yönetilen karmaşık bir süreçtir. Embriyonik kök hücreler (EKH'ler), bu gelişim dinamiklerini incelemek için önemli *in vitro* modeller olarak hizmet eder. Bu düzenlemede kritik öneme sahip olan epigenetik mekanizmalar, pluripotent EKH'leri belirli soylara yönlendirir. Bu mekanizmaların daha derin bir şekilde incelenmesi, gelişimsel bozuklukları anlamak için gereklidir. Daha önce yürütülen bir kütle spektrometresi analizi, nöroektoderm farklılaşması sırasında H3K36me2 seviyelerinde bir yükselme olduğunu ortaya koyarken, endoderm farklılaşmasında bu artışın olmadığı gözlemlendi. Ön verilerimiz, nöroektoderm farklılaşmasında H3K36 metilasyonunun belirgin bir rolünün olduğuna işaret ediyor. Bu bulgular Western blot yöntemi ile doğrulandıktan sonra, farklılaşma sırasında metiltransferazlar, demetilazlar ve okuyucular dahil olmak üzere H3K36 metilasyon düzenleyici sistemindeki ifade değişiklikleri analiz edildi. Endoderm için mevcut RNA-sekanslama verileriyle birlikte nöroektoderm farklılaşmasına dair RT-qPCR analizleri, pluripotentlikten çıkışta ve/veya farklılaşmada H3K36 metilasyonun önemli bir rolü olabileceğini gösterdi. Özellikle, *Nsd2* ve *Zmynd11* ifade seviyeleri,

her iki soyda farklılaşma sırasında artarken, *Kdm4c* ve *Msh6'*nın, farklılaşmış durumlara kıyasla mESC'lerde daha yüksek ifade edildiği bulundu. Tersine, *Nsd1* ve *Setd5* için soya özgü ifade değişimleri gözlemlendi. Son olarak, nöroektoderm farklılaşması için kritik epigenetik faktörleri belirlemek üzere yapılan bir CRISPR gen silme taraması yoluyla, *Nsd1*, *Setd5* ve *Kdm2b*'nin nöroektoderm farklılaşması için kritik olduğunu ortaya çıkardık. Bu bulgular, H3K36me2 düzenlemesinin soy seçimi ve pluripotentlikten çıkış için önemini vurgulamaktadır; düzenleyici proteinler, farklılaşma aşamasına veya soya özgü ifade değişikliklerini göstermektedir. Bu modifikasyonun ve ilişkili proteinlerin nöroektoderm farklılaşmasını etkilediği mekanizmaları ortaya çıkarmak için gelecekte çalışmalara ihtiyaç vardır.

Anahtar Kelimeler: Epigenetik, H3K36 Metilasyonu, Fare Embriyonik Kök Hücreleri, Nöroektoderm Farklılaşması.

To my ever-supportive parents

ACKNOWLEDGMENTS

I would like to express my deep gratitude for my supervisor Assist. Prof. Dr. Nihal Terzi Çizmecioğlu, for her constant guidance, invaluable insights, and consistently challenging and nurturing my academic growth.

I deeply appreciate my lab mates Ceren Alganatay, Deniz Ak and Emre Balbaşı for their friendship and support. I extend my thanks to our lab interns, Alp Aydın Şipal, Duru Demiröz, Nisan Zühre Daduk and Ecesu Özdal for their assistance. I especially appreciate Alp Aydın Şipal for his experimental support in this study.

I would like to thank my dear friends Yağmur Öykü Carus and Esra Başaran. Our bond has stood the test of time and distance, and I'm grateful for the unconditional love, understanding, and support they have provided.

I am truly thankful to Emre Balbaşı for always being there for me with his constant support and love. This journey has become more manageable and less stressful with him around. I feel really lucky to have him in my life.

Last but definitely not least, a big thank you to my parents. They thought me to question the world around me, and that innate curiosity has been the driving force behind my academic pursuits. They have always stood by my decisions, providing both emotional and mental support every step of the way. I deeply value everything they have done and continue to do for me. I know that I am the person I am today largely because of them. Thank you for being such wonderful parents.

This study was partially supported by GAP Project (GAP-108-2021-10585); TEZ-YL Project (TEZ-YL-108-2002-11040) and UNESCO L'Oréal Project 2021.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Embryonic development is a multifaceted process guided by various regulatory pathways, cell fate decisions and epigenetic modifications. The early cell fate decisions are especially critical as perturbations can lead to a variety of developmental disease phenotypes, including neurodevelopmental disorders. However, it is quite a challenge to study these stages due to the small size of embryos, their inaccessibility, and associated ethical issues. Stem cell research has become a valuable tool for *in vitro* early development studies.

1.1 Early Mouse Embryogenesis

After fertilization of the oocyte, zygote undergoes several rapid cell divisions without a corresponding increase in the overall size. By 2-3 days post coitum (dpc) 8-16 identical blastomeres are formed (Loebel et al., 2003). These are totipotent cells, possessing the ability to differentiate into both embryonic and extraembryonic tissues. This totipotency is marked by the simultaneous expression of pluripotency markers, such as *Oct4* and *Nanog* (Orkin & Hochedlinger, 2011), and extraembryonic markers, including *Gata6* and *Sox17* (Niakan et al., 2010; Wamaitha et al., 2015).

Following this stage, the morula forms through a process known as compaction. Within the morula, the first apical-basal polarization takes shape alongside the formation of adherens junctions, setting the stage for differential fates of inner and outer cells (White et al., 2016). Together with compaction there is an upregulation of ion pumps. This facilitates the formation of blastocoelic cavity around 3.5 dpc through the passive diffusion of water following active transport of ions such as Na^+ and Cl^- (Barr et al., 1998). This liquid filled structure is now called a blastocyst. It

consists of two components: the outer multipotent trophectoderm layer, which forms the extraembryonic tissues; and the inner cell mass (ICM) beneath it. The ICM retains pluripotency and holds the potential to differentiate into all cell types of embryo (Loebel et al., 2003).

At 4.0 dpc, shortly before the hatching of blastocyst from the zona pellucida, a second cell fate decision is made: the cells of the ICM differentiate into either the extraembryonic primitive endoderm (hypoblast), or the pluripotent epiblast. The primitive endoderm is no longer pluripotent; it can give rise to visceral endoderm (VE) that secretes an extracellular matrix, or parietal endoderm (Wamaitha et al., 2015). In contrast, epiblast that is found in between hypoblast and polar trophectoderm retains its pluripotency (Artus et al., 2011).

Following implantation at around embryonic day 4.5, embryo exhibits a cup shaped form through cavitation known as egg cylinder (Arnold & Robertson, 2009). Within this structure, the trophectoderm layer differentiates into the extraembryonic ectoderm (EXE) and the ectoplacental cone (Goldin & Papaioannou, 2003), which are pivotal in generating supporting tissues, including the placenta. A subset of epiblast cells differentiates into a pluripotent layer known as the primitive ectoderm or 'primed epiblast', so named because it is ready for gastrulation (Nichols & Smith, 2011; Pelton et al., 2002). This layer exhibits increased cell proliferation rates with a shortened cell cycle, preparing the developing embryo to possess enough cell count for the subsequent gastrulation (Bardot & Hadjantonakis, 2020).

Primitive streak formation at around embryonic day 6.5 marks the onset of gastrulation (Rivera-Pérez & Magnuson, 2005). During gastrulation, the primitive ectoderm differentiates into the three germ layers, with their fates largely determined by their initial positions and signals they receive from adjacent cells. Through a process known as epithelial-to-mesenchymal transition (EMT), some cells migrate into the primitive streak to form mesoderm and endoderm layers (Arnold & Robertson, 2009). Remaining cells are involved in the formation of definitive ectoderm layer.

The definitive ectoderm layer is bipotential; it can differentiate into either the surface ectoderm, or the neurectoderm layer at around embryonic days 7.0-7.5 (L. Li et al., 2013) depending on the presence or absence of BMP4 activity, respectively (Davis et al., 2004). The neural plate, formed at close proximity to mesendoderm layer, is comprised of neuroectoderm cells. This structure is characterized by the expression of *Sox1*, which indicates neural commitment in mice (Aubert et al., 2003), followed by *Pax6* (Suter et al., 2009). The cells of the neural plate have the potential to differentiate into various neural cell types. Following a rapid cell division, a thickening and folding forms the neural tube, which will form the central nervous system (CNS) consisting of neurons, radial glial cells, astrocytes, and other glial cells. Additionally, cells at the border of the neural plate and non-neural ectoderm can become neural crest cells. These neural crest cells have a remarkable migratory capability and multipotency, leading to the formation of diverse cell types, including those of the peripheral nervous system, as well as facial bones and cartilage, among other derivatives (Götz & Huttner, 2005).

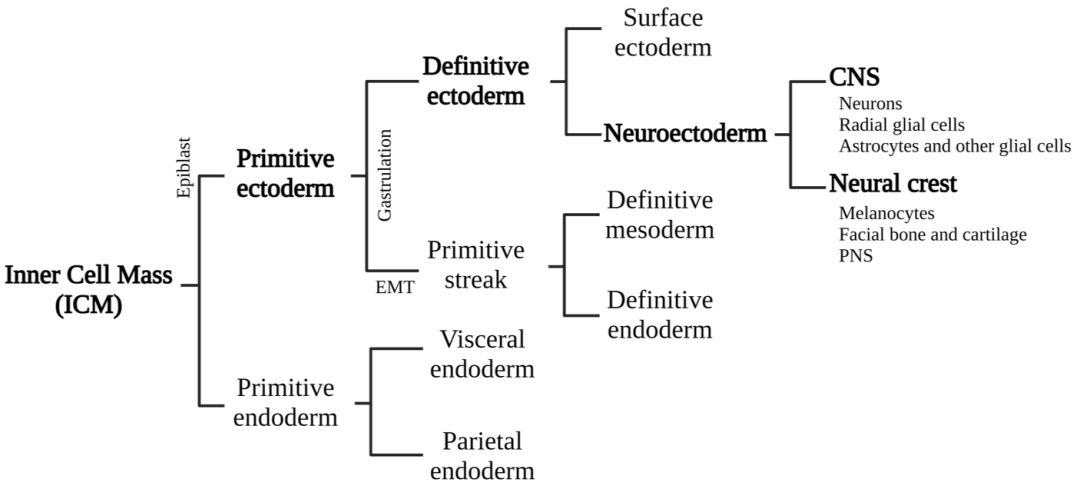

Figure 1.1 Schematic representation of early embryonic development. The processes leading to neuroectoderm formation are written in bold letters. Adapted from Shparberg et al., 2019.

1.2 Embryonic Stem Cells: Pluripotency, Maintenance and Directed Differentiation

Embryonic stem cells (ECSs) are derived from the ICM of developing blastocyst. These cells are characterized by their inherent capacity for unlimited self-renewal and their pluripotent nature (Nichols & Smith, 2011). Pluripotency refers to the ability to differentiate into all three primary germ layers; the endoderm, mesoderm and ectoderm. From these germ layers all tissues and cell types in a mature organism emerge (Jaenisch & Young, 2008). Owing to these properties, ESCs have become invaluable tools in scientific research, allowing for detailed *in vitro* exploration of early embryogenesis.

Mouse ESCs (mESCs) can be cultured and maintained in chemically defined 2i medium supplemented with LIF (Leukemia Inhibitory Factor), eliminating the need for feeder cells (Wray et al., 2010). The term "2i" refers to two inhibitors; CHIR99021 and PD0325901. Together with LIF, these inhibitors preserve the undifferentiated state of mESCs. Specifically, CHIR inhibits GSK3β, leading to stabilization and subsequent nuclear translocation of β-catenin, which then induces the expression of pluripotency-associated genes. On the other hand, PD, a MEK inhibitor, promotes pluripotency by preventing the activation of the ERK pathway involved in mESC differentiation. LIF activates the JAK-STAT signaling pathway, promoting the expression of core pluripotency network (Ying et al., 2008). mESCs successfully grown under these conditions *in vitro* display a distinctive three-dimensional and bright colony morphology when observed under the microscope. This morphology serves as a hallmark of healthy and undifferentiated mESCs, indicative of their preserved pluripotent state. However, if exposed to suboptimal culture conditions, the pluripotency of mESCs can be compromised, resulting in spontaneous differentiation. Such a transition can be observed through morphological changes, with cells typically appearing flatter and less bright (Balbasi, Guven, et al., 2022).

The pluripotency is maintained by the coordinated actions of a core network of transcription factors; OCT4, NANOG and SOX2. While these transcription factors repress the expression of differentiation-related genes, they form a self-regulatory loop to regulate each-other's expression together with numerous other genes, including those in the extended pluripotency network such as *Klf4* and *c-Myc* (Orkin & Hochedlinger, 2011). Notably , when combined with OCT4 and SOX2, these two factors constitute the Yamanaka factors used in generation of induced pluripotent stem cells (iPSCs) (Takahashi & Yamanaka, 2006).

Directing pluripotent mESCs towards specific lineages requires a precise and coordinated regulation of genetic networks. This involves not only the suppression of pluripotency network and genes tied to alternative cell fates, but also the simultaneous activation of genes associated with the desired cell fate (Orkin & Hochedlinger, 2011).

During the differentiation of mESCs into neuroectodermal lineages, distinct morphological changes become apparent. The typical dome-shaped mESC colonies transition to flatter individual cells with the initiation of neurite-like projections (Balbasi, Sezginmert, et al., 2022). Alongside these morphological shifts, the cellular gene expression profiles undergo drastic changes. A rapid decrease in expression levels of pluripotency markers such as *Oct4* and *Nanog*, is accompanied by an increase in *Sox1* expression. *Sox1* is known as the earliest expressed neuroectodermal marker in neuroectoderm differentiation of mESCs (Aubert et al., 2003; Pevny & Placzek, 2005). Following *Sox1*, there's an increase in *Pax6* expression, further indicating commitment to the neuroectodermal lineage (Suter et al., 2009).

Various methodologies have been developed to guide mESCs towards the neuroectoderm lineage. These include the treatment with retinoic acid following embryoid body (EB) formation (Wiles & Johansson, 1999), co-culturing with stromal feeder (PA6) (Kawasaki et al., 2000), and adherent monoculture (Ying et al., 2003) approaches. Each technique presents its own set of benefits and limitations. However, the adherent monoculture method with low cell seeding density and use of

defined N2B27 medium emerges as an easier and more efficient approach allowing for more homogenous differentiation towards neuroectoderm. Furthermore, it's been documented that the neural precursors derived with this method can be further differentiated into neural and glial lineages including astrocytes and oligodendrocytes (Ying et al., 2003).

Studying mESC differentiation offers valuable insights into the key mechanisms of early embryonic development and cell fate decisions. This complex process is primarily regulated and orchestrated by epigenetic mechanisms, which shift gene expression profiles from a pluripotent state to a designated cell lineage.

1.3 Epigenetic Mechanisms

In eukaryotic cells, DNA is organized into a dynamic structure known as chromatin within the nucleus. Central to this organization is the nucleosome, in which approximately 147 base pairs of DNA coil around a histone octamer (Peterson & Laniel, 2004). This octamer consists of two copies each core histones: H2A, H2B, H3, and H4. The further compaction and organization of chromatin is facilitated by the linker histone, H1 (Campos & Reinberg, 2009). The architecture of chromatin plays a crucial role beyond mere organization of large genomic DNA; it regulates gene expression through a range of histone modifications and other epigenetic mechanisms like DNA methylation and chromatin remodeling (G. Li & Reinberg, 2011).

Chromatin exists in two primary structural states, each contributing to the dynamic regulation of gene expression: euchromatin and heterochromatin (Horvath et al., 2001). Euchromatin, the less condensed form, is typically associated with transcriptionally active genomic regions. Its open structure makes it readily accessible to the transcriptional machinery, thereby promoting gene expression. Conversely, heterochromatin represents the densely packed regions of chromatin where genes are often silenced or expressed in minimal levels. Yet, the role of

heterochromatin extends beyond transcriptional silencing. It plays an indispensable role in preserving genomic stability, especially at genomic regions like telomeres and centromere that contains constitutive heterochromatin. On the other hand, another form of heterochromatin, termed facultative heterochromatin, can alternate between condensed and relaxed states which is crucial for cellular differentiation (Saksouk et al., 2015).

Histones, characterized with their basic nature, bind to the negatively charged DNA through electrostatic interactions. These proteins possess a globular domain along with a protruding N-terminal tail, rich in lysine, arginine and serine residues that are susceptible to post-translational modifications by various enzymes. These covalent modifications include, but not limited to, methylation, phosphorylation, acetylation and ubiquitylation (Strahl & Allis, 2000). The levels and distributions of these reversible modifications are regulated by opposing activities of "writer" and "eraser" proteins. While histone methyltransferases (HMTs), histone acetyltransferases (HATs) and kinases function as writers by adding modifications, histone demethylases (HDMs), histon deacetylases (HDACs) and phosphatases serve as erasers by removing them (Biswas & Rao, 2018).

Histone modifications influence the structure and function of chromatin in multiple ways. For instance, acetylation alters the charge of histone proteins, reducing their electrostatic attraction to the negatively charged DNA, which can render the chromatin more accessible to transcriptional machinery (Turner, 1991). In contrast, methylation does not change the overall charge but can serve as a signal for the recruitment of specific "reader" protein complexes, either activating or repressing transcription depending on the specific location and context of the modification. Similarly, phosphorylation, ubiquitylation, and sumoylation can also affect chromatin structure and recruit various proteins, further influencing gene expression. Thus, these modifications collectively modulate the dynamic interplay between chromatin structure and gene regulation.

Acetylation at 27th lysine residue of H3 (H3K27ac) is indicative of active transcription, whereas trimethylation at the same residue (H3K27me3) is a hallmark of gene silencing. This is regulated by opposing actions of Trithorax group complex (TrxG) and Polycomb group complex (PrG), respectively (Marinho et al., 2017). These complexes have critical importance in pluripotency and development, as they influence the balance between activation and silencing of developmental genes (Martinez & Cavalli, 2006).

In ESC state, promoters of many developmental gene promoters display a "bivalent" chromatin state, characterized by the simultaneous presence of repressive mark H3K27me2/3 and active mark H3K4me2/3 (Bernstein et al., 2006; Vermeulen et al., 2010). This unique chromatin landscape keeps RNA Polymerase II paused at these promoters, allowing for rapid gene activation or repression during differentiation depending on the specific lineage.

DNA methylation, another critical epigenetic modification, is especially vital during development. Typically, methylation at DNA promoter regions silences genes. During development, *de novo* methylation is facilitated by DNA methyltransferases DNMT3A and DNMT3B (Fukuda et al., 2021), while DNMT1 ensures its maintenance through cell divisions (Nishiyama et al., 2020). In the pluripotent state of ESCs, the global DNA methylation level is comparatively low, and the promoters of pluripotency genes often carry activating histone marks like H3K4me3 with unmethylated DNA.

1.4 H3K36 Methylation

The 36th lysine residue of histone H3 (H3K36) can undergo mono-, di-, or tri-methylation, with each methylation state having distinct genomic distributions and functions. H3K36me1, the least studied form of H3K36 methylation, is broadly distributed in the genome. While recognized as a precursor to more advanced methylation states, its specific roles are yet to be fully elucidated. The di-methylated

form, H3K36me2, is predominantly localized close to transcription start site of actively transcribed genes and in some regulatory intergenic regions. In contrast, H3K36me3 is more prevalent towards the 3' end of actively transcribed genes where H3K36me2 levels decrease (Huang & Zhu, 2018; Kuo et al., 2011). Notably, H3K36me2 is more abundant than H3K36me3, comprising 20-45% of the total H3 together with the unmethylated form (Topchu et al., 2022). H3K36 methylation, typically associated with active chromatin, plays various roles in biological processes. Its functions include transcription regulation, alternative splicing (Edmunds et al., 2008; Shin et al., 2014), DNA damage repair (Pai et al., 2014), prevention of aberrant transcription, dosage compensation (Larschan et al., 2007), regulation DNA methylation, and transmission of gene expression patterns during development (Wagner & Carpenter, 2012).

In yeast, Set2 is the sole enzyme responsible for all levels of H3K36 methylation, and its activity is coupled with RNA Polymerase II during transcription elongation (Krogan et al., 2003). However, in higher eukaryotes, including mammals, the process is more complex with multiple methyltransferases controlling H3K36 methylation. These enzymes have SET domains that use S-adenosylmethionine (SAM) to add methyl groups to Lysine residue (Dillon et al., 2005). While mono- and dimethylation of H3K36 is carried out by a diverse set of methyltransferases including NSD1-3 (Y. Li et al., 2009; Qiao et al., 2011), SETD3, ASH1L, PSIP1 and SETMAR (Lam et al., 2022; Pradeepa et al., 2012; Wagner & Carpenter, 2012); trimethylation of H3K36 (H3K36me3) is primarily catalyzed by SETD2 (Edmunds et al., 2008), and, its recently discovered homologue, SETD5 (Sessa et al., 2019).

The levels of H3K36 methylation is regulated by a dynamic interplay between methyltransferases and demethylases. In mammals, the enzymes KDM2A/2B (JHDM1A/B), KDM4A-C (JHDM2A-C) and KDM8 are involved in demethylation of H3K36. KDM2A and KDM2B preferentially target H3K36me1 and H3K36me2 (Tsukada et al., 2006), whereas KDM4A-C predominantly demethylate H3K36me2 and H3K36me3 (Klose et al., 2006) (Figure 1.2).

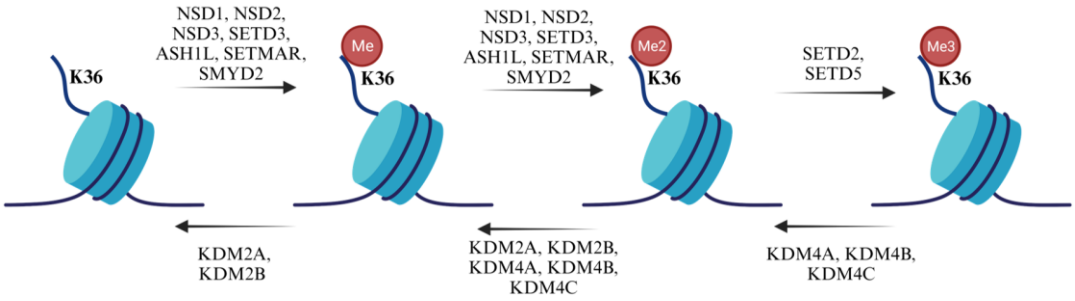

Figure 1.2 H3K36 methyltransferases and demethylases that add or remove each level of methylation. Created with BioRender.com.

The diverse functions attributed to H3K36 methylation are primarily mediated by its "readers", proteins that recognize and bind to specific methylation states with their PWWP, chromodomains or Tudor domains (J. Li et al., 2019). Notably, these readers mediate various cellular processes such as alternative splicing (ZMYND11, MRG15, PHF19, PSIP1), DNA damage repair (PHF1, MSH6), and DNA methylation (DNMT3A/3B). Among these, ZMYND11 is specific to H3.3K36me3 variant (Zaghi et al., 2020) where H3.3 histone variant is known to be deposited in the gene bodies of actively transcribed genes (Henikoff & Smith, 2015).

1.5 Preliminary Data

In a previous Mass Spectrometry analysis examining histone modifications in mESCs, and endoderm and neuroectoderm differentiated cells, we observed a notable increase in H3K36me2 levels solely in cells undergoing neuroectoderm differentiation (Figure 1.3). Interestingly, this rise was not accompanied by changes in H3K36me3 levels. Such observations suggest a unique role for regulation of H3K36 methylation during neuroectoderm differentiation distinct from endoderm differentiation. In line with this, mutations in the genes associated with H3K36 methyltransferases, demethylases, and readers have been implicated in various neurodevelopmental disorders. Conditions such as autism spectrum disorders (Zaghi et al., 2020), Sotos syndrome (Kurotaki et al., 2002), Wolf-Hirschhorn syndrome (Lucio-Eterovic et al., 2010), Tatton-Brown-Rahman syndrome (Tatton-Brown et

al., 2014) all exhibit neurodevelopmental symptoms, highlighting the pivotal role of H3K36 methylation in neural development.

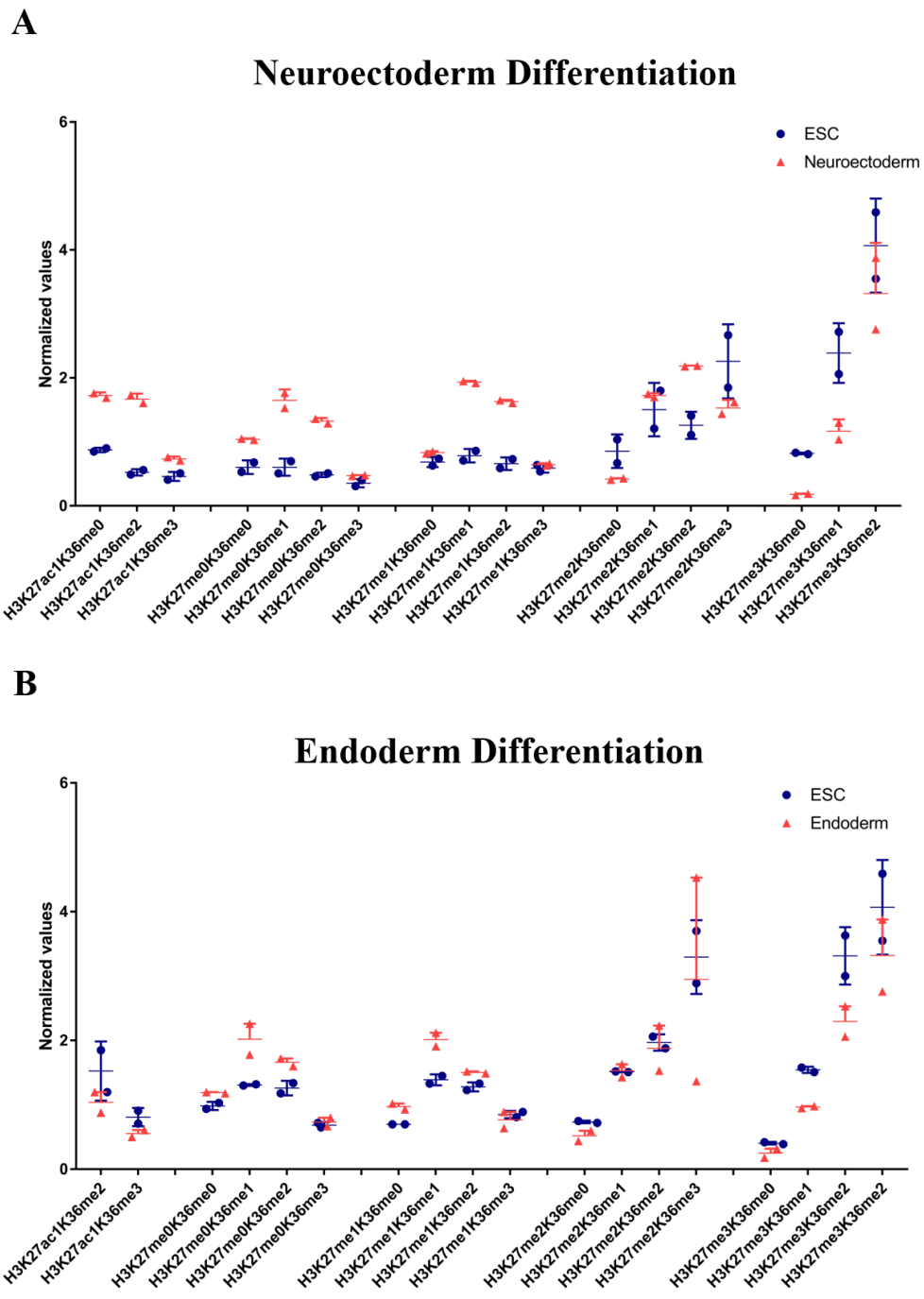

Figure 1.3 Quantitative analysis of histone 3 modifications during mESC differentiation. Proteomic analysis of H3 peptides in mESCs and **A.** neuroectoderm and **B.** endoderm committed cells. The abundance of each modified peptide was normalized to H3K4me0 peptide. Error bars represent SEM of two biological replicates.

1.6 Aim of the Study

Early embryonic development is a complex process. Utilizing mESCs and associated differentiation protocols, we can study the changes inherent to this process and uncover the underlying regulatory mechanisms. Epigenetic factors play a dominant role during the generation of specialized cells from pluripotent stem cells. Based on our preliminary data, we hypothesized that H3K36me2 and associated H3K36 methylation regulation might have roles in early cell fate decisions especially during neuroectoderm differentiation. This study aims to enlighten this regulatory system and the processes related to H3K36 methylation that might have roles in differentiation process. With RT-qPCR analysis of methyltransferases, demethylases and readers of H3K36 methylation, we determined their expression profiles during neuroectoderm differentiation. Using a previous RNA sequencing data generated in our laboratory for endoderm differentiation, we compared these profiles to those in endoderm. Furthermore, through a CRISPR knock out screening experiment done to pinpoint the epigenetic factors essential for neuroectoderm differentiation, we identified the key players of H3K36 regulation for neuroectoderm differentiation.

CHAPTER 2

MATERIALS AND METHODS

2.1 mESC Growth

CJ9 (wild-type) mESCs were obtained from Prof. Stuart Orkin's laboratory at Boston Children's Hospital. These cells were routinely grown and maintained in our laboratory using 4% serum-containing 2i medium (2i indicates the presence of two inhibitors, namely PD0325901 and CHIR-99021, together with Leukemia Inhibitory Factor (LIF) to maintain pluripotency). However, neuroectoderm differentiation medium (N2B27 medium; *see* Appendix A for the recipe) is devoid of serum, which may reduce the viability. To minimize the negative effects of serum depletion during neuroectoderm differentiation, mESCs were adapted to completely serum-free conditions prior to differentiation by reducing the serum concentration by half with each passage. Adapted mESCs were grown and maintained on gelatin-coated tissue culture plates in serum-free 2i medium (*see* Appendix A for recipe) at 37ºC, 5% CO_2 as described (Balbasi, Guven, et al., 2022). mESCs were passaged every 2-3 days using Accutase Cell Detachment Solution (Millipore, SCR005), a method that provides milder enzymatic shearing than routinely used TrypLE Express Enzyme (1X), phenol red (Gibco, 12605-010). This modification was implemented to minimize attachment problems encountered during the neuroectoderm differentiation protocol with the goal of improving cell viability and the success of differentiation.

2.2 Neuroectoderm Differentiation

For neuroectoderm differentiation of mESCs, an adherent monoculture differentiation protocol (Q. L. Ying et al., 2003) was followed with minor

modifications as described in (Balbasi, Sezginmert, et al., 2022). Preliminary experiments were conducted with the Sox1-GFP reporter cell line to assess the success of differentiation. In this cell line, the open reading frame of Sox1 gene is replaced with a GFP-IRES-Pac cassette, which allows for monitoring the expression of *Sox1*, which is the earliest marker for neuroectoderm differentiation (Q. L. Ying et al., 2003). It was observed that the majority of population displayed GFP signal by the fifth day of neuroectoderm differentiation (Appendix C, Figure S.1), which indicates that successful differentiation was achieved.

Neuroectoderm differentiation in CJ9 WT mESCs was optimized in terms of initial cell seeding density and the amount of medium used on different days. The primary criteria for evaluating differentiation success involve tracking cell morphology, viability, and attachment to the culture plate. The optimal conditions were then tested for the actual success of differentiation by measuring the expression levels of key neuroectoderm differentiation markers (*Sox1*, *Pax6*, *N-cadherin*) and pluripotency markers (*Oct4* (*Pou5f1*), *Nanog*) in the mESC state and at various days of differentiation through RT-qPCR analysis.

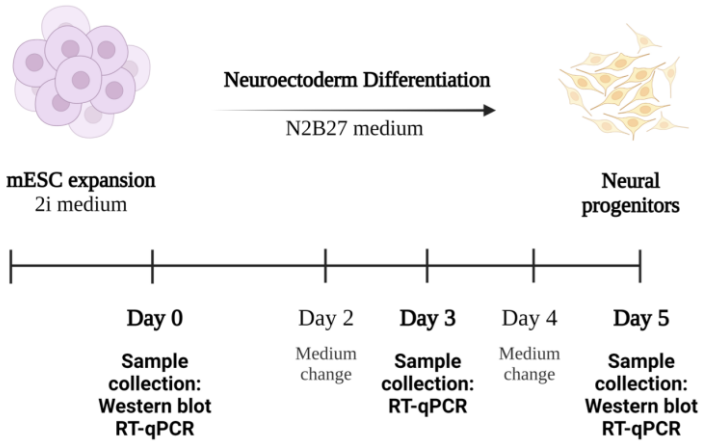

Figure 2.1. Schematic representation of neuroectoderm differentiation protocol. Created with BioRender.com

The optimized protocol is as follows; mESCs are detached from culture plates using Accutase Cell Detachment Solution and counted with Countess™ Automated Cell Counter (Invitrogen, T10282) after Trypan Blue Stain (0.4%) application. 1.1 x 10^6

cells are seeded into each well of gelatin-coated 6-well plate using a low volume (1.5 mL) N2B27 medium to aid easier attachment. The cells are then incubated at 37ºC, 5% CO_2 for 48 hours. On the second and fourth days of differentiation, the medium is gently replaced with a higher volume (3 mL) of fresh N2B27 medium without detaching the cells. On the third and fifth days, samples were collected for RT-qPCR and Western blot analyses (Figure 2.1).

2.3 Endoderm Differentiation

Endoderm differentiation of mESCs were done by Emre Balbaşı as described (Balbasi, Sezginmert, et al., 2022; Balbaşı, 2022).

2.4 Protein Sample Collection and Histone Extraction

2 x 10^6 cells were collected from mESC state and from the fifth day of neuroectoderm and endoderm differentiation. After cells were washed with ice-cold 1X PBS two times, pellets were snap-frozen in liquid nitrogen and stored at -80ºC until isolation.

Due to the highly basic nature of histone proteins, specialized methods should be employed in order to effectively separate them from other proteins and DNA (Shechter et al., 2007). In this study, an acid extraction protocol was followed (Vossaert et al., 2014). Briefly, cell pellets were thawed and resuspended in 200 µL Triton Extraction Buffer (TEB; *see* Appendix B for recipe), followed by lysis on ice through gentle stirring for 10 minutes. Samples were centrifuged at 6500 rcf for 10 minutes at 4ºC to pellet the nuclei. After supernatant was removed, pellet was washed with half volume TEB (100 µL) and centrifuged as previously described. Pellet was then resuspended in 50 µL 0.2M HCl and incubated at 4ºC overnight for acid extraction. The following day, samples were centrifuged at 6500 rcf for 10 minutes at 4ºC, and the supernatant was transferred to new Eppendorf tubes. Finally, the solution was neutralized by adding 1/10 volume (5 µL) of 2M NaOH. Pierce™

BCA Protein Assay Kit (Thermo, 23227) was employed to measure the protein concentrations of samples following the manufacturer's instructions.

2.5 SDS-PAGE and Western Blot Analysis

Isolated histones were incubated in 2X Laemmli Buffer (Bio-Rad, 1610737) supplemented with β-mercaptoethanol (Sigma, M-6250) at 95ºC for 10 minutes. After samples were incubated on ice and centrifuged at maximum speed, 1 µg of histone proteins was loaded on a 15% polyacrylamide gel (PAG) (*see* Appendix B for recipe) and run at constant 100 V. 15% PAG was chosen due to the need for high resolution of histone proteins, which have small molecular weights.

Following electrophoresis, proteins were transferred onto nitrocellulose membrane using Bio-Rad Trans-Blot Turbo Transfer System (Bio-Rad, 1704150), employing a semi-dry transfer method. In this method, the ion stacks and nitrocellulose membrane were soaked in Transfer Buffer (*see* Appendix B) and the excess solution was removed after the assembly. After the transfer, the membrane was blocked with 5% skimmed milk in 1x TBS-T (*see* Appendix B) at 4ºC overnight by shaking. Primary antibody incubations were done overnight for H3K36me2 and H3K36me3 at 4ºC by rotating. HRP-conjugated secondary antibody incubations and primary antibody incubation for loading control (histone H3) were done for 1 hour at room temperature. Following each antibody incubation, membrane was washed three times with 1x TBS-T at room temperature for 10 minutes. H3K36me2 and H3K36me3 were detected on separate blots using the identical samples and amounts. Following visualization, the membranes were stripped using a warm strip buffer (*see* Appendix B) for approximately 6-10 minutes, depending on the intensity of the bands. Once the signal was removed, the same blots were incubated with H3 loading control as described above. All antibody dilutions were prepared using 5% skimmed-milk (Sigma, 70166-500G). Further details about antibodies used for western blotting can be found in the appendices (*see* Appendix D, Table 5).

2.5.1 Visualization, Band Intensity Quantification and Statistical Analysis

For visualization, Clarity Western ECL (Bio-Rad, 1705060) was used with ChemiDoc XRS+ Imaging System (Bio-Rad, 1708265). Since H3K36me3 is much less abundant than H3K36me2 (Topchu et al., 2022), Clarity Western Max ECL (Bio-Rad, 1705062), an amplified form of ECL, was used for visualization of H3K36me3.

Band intensities were quantified by using ImageJ software. H3K36me2 and H3K36me3 modifications were normalized to H3 loading control and reported as relative intensities. Graphs were generated using GraphPad Prism Software. Statistical analysis (Student's t-test) was performed using the same software for at least two biological replicates.

2.6 RNA Sample Collection, Isolation and cDNA Conversion

5 x 10^5 cells were collected from mESC state and during neuroectoderm differentiation on days 3 and 5. The cells were pelleted by centrifugation at 300 rcf, and lysed in 500 µL TRIzol™ Reagent (Invitrogen, A33251). The resulting lysates were stored at -20ºC until subsequent RNA isolation. For each biological replicate, total RNA isolation was performed collectively only after all samples from differentiation days were collected and lysed. RNeasy Plus Micro Kit (Qiagen, 74004) was utilized according to the manufacturer's protocol. RNA concentrations were measured with NanoDrop Spectrophotometer (MaestroGen, MN-913) Complementary DNA (cDNA) was synthesized from 500 ng RNA using iScript cDNA Synthesis Kit (Bio-Rad, 1708891), following the provided protocol. The thermocycler conditions were set as follows; 5 minutes at 25ºC, 1 hour at 42ºC, 5 minutes at 85ºC, and then held at 4ºC. Resulting cDNA samples were stored at -20ºC.

2.7 Primer Design and RT-qPCR Analysis

Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR) analyses were carried out with GoTaq® qPCR Master Mix (Promega, A6002) in a 10 µL reaction volume. The reactions were run on CFX Connect Real-Time PCR Detection System (Bio-Rad, 1855201) thermal cycler with the following conditions: initial denaturation at 95.0°C for 10 minutes, 40 cycles of denaturation at 95.0°C for 30 seconds, annealing at 60.0°C for 30 seconds, and extension at 72.0°C for 30 seconds, with a plate read after each extension step. A melt curve analysis was conducted, ranging from 50.0°C to 99.0°C, with 1.0°C increments and a 5-second hold at each temperature, followed by a plate read.

qPCR primer sequences were found from literature sources or designed using NCBI Primer Blast. Several criteria were considered for primer design, including selecting primers that are 18-24 nucleotides long, a preference for GC content in the range of 40-60%, and compatibility of melting temperatures (Tm) between primer pairs. In either case, the primer sequences were blasted to the mouse genome (NCBI Primer Blast; Refseq representative genomes) and transcriptome (NCBI Primer Blast; Refseq mRNA). The primer pairs that have the lowest off-targets in the transcriptome and no targets in genome were selected for a reliable expression analysis. Also, alternative transcript variants were considered and the primers that targets the most variant were preferred. Specificity was verified by performing melt curve analysis for all primer pairs. Primer sequences can be found in appendices (Appendix E, Table 6).

2.7.1 Statistical Analysis

Data were normalized to the expression levels of *β-actin* house-keeping gene, and reported as a percentage of *β-actin* expression (% *β-actin*). Normalization was carried out using the following formula: $2^{(-\Delta Ct)}$ x 100. Graphs were created by using GraphPad Prism Software, which was also employed for performing statistical

analyses (Student's t-test). Prior to the analysis, data normality and variance homogeneity were assessed via Shapiro-Wilk Test and Levene's Test, respectively, using R.

A Spearman correlation test was conducted to analyze the relationship between the H3K36me2 levels and the expression levels of selected methyltransferases, demethylases and readers during neuroectoderm differentiation (Day 5 versus mESC state) using R. The selection of these genes was based on the expression trends and the statistical significance during neuroectoderm differentiation.

2.8 RNA Sequencing of Endoderm Differentiation

Endoderm differentiation of wild type CJ9 mESCs and following experiments were performed by Emre Balbaşı as described (Balbaşı, 2022). RNA samples were isolated during the endoderm differentiation time course and sent for RNA sequencing. Sequencing and bioinformatic analyses were conducted by Gen-Era. Log (FC) (FC: fold-change) values for D4/D0 and D3/D0 were obtained after differential expression analysis.

2.8.1 Integrative Genomics Viewer (IGV) Analysis

IGV software was utilized for the visualization of transcripts. The BAM files obtained from RNA-sequencing of the first biological replicate of endoderm differentiation were submitted to IGV software and a selected gene list was visualized using the reference genome Mus musculus 10 (GRCm38/mm10). The results are representative of three biological replicates.

2.9 CRISPR-Cas9 Mediated Pooled Screen for Key Epigenetic Factors Involved in Neuroectoderm Differentiation

A CRISPR-mediated knockout screen was performed to identify the essential epigenetic factors required for neuroectoderm differentiation. Briefly, using a sgRNA library targeting 304 different epigenetic factors, Sox1-GFP mESCs were transduced and differentiated to neuroectoderm for five days. At the end of differentiation, cells were sorted via FACS (Fluorescence Activated Cell Sorting) according to their GFP signal intensity as an indication for successful neuroectoderm differentiation (Figure 2.2). A population exhibiting the highest GFP signal (top ~7%) was collected. By comparing the sgRNAs representation between this population and the mESC state, we determined the factors required for neuroectoderm differentiation. A lower representation of sgRNAs for a given gene indicates that its knockout disrupted the GFP signal and thus, neuroectoderm differentiation. Emre Balbaşı and Deniz Ak executed the majority of the wet-lab procedures, while I was responsible for NGS library preparation and subsequent analyses, including the MAGeCK analysis.

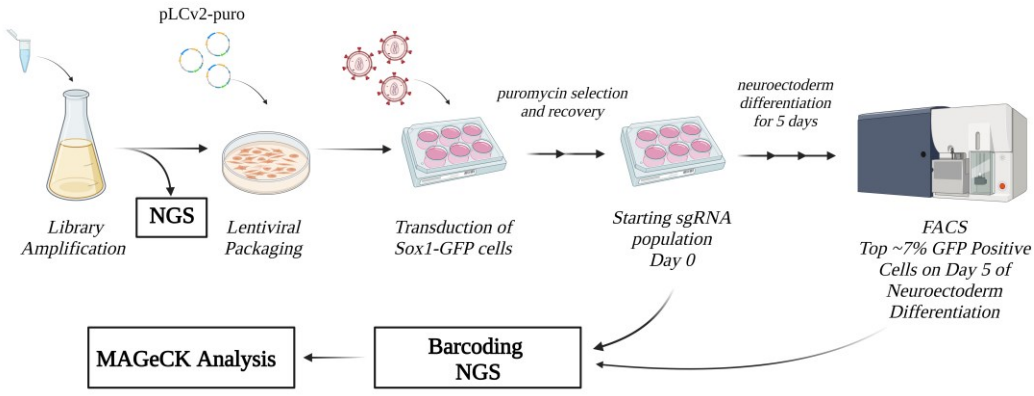

Figure 2.2. Experimental design and flow of pooled CRISPR knock-out screen of epigenetic factors. Created with BioRender.com

2.9.1 Preparation of the sgRNA Library

The library was designed to target 304 different epigenetic factors. Additionally, sgRNAs targeting the pluripotency network were included as controls. For each gene, 3 different sgRNAs were included. These sequences were taken from Mouse Gouda (Addgene, 136987) and Mouse Brie (Addgene, 73633) libraries. From the Brie Library, sgRNAs were specifically chosen based on the Rule Set 2 score, which evaluates the efficiency of the sgRNA on its target. As positive controls, 50 sgRNAs that directly target GFP were incorporated from the GeCKo library. Meanwhile, 95 non-targeting sgRNAs from the Brie library served as negative controls. After the addition of two constant sequences required for annealing to plasmid later, to each end of sgRNA sequences, the library was synthesized by LC Sciences. The library was then cloned into pLCv2-puro plasmid via Gibson assembly and given to Endura electrocompetent bacteria through electroporation for amplification.

2.9.1.1 NGS Analysis of the Amplified sgRNA Library

To verify the coverage and equal distribution of sgRNAs after library amplification, NGS analysis was done on the amplified library. For this purpose, the python script ("count_spacers.py") provided by reference protocol (Joung et al., 2017) was utilized, which was updated for Python 3.9, and subsequently modified for our needs. Successful amplification was defined by the following criteria: the percentage of perfectly matched sgRNAs should exceed 70%, the percentage of undetected sgRNAs should not surpass 0.5%, and the skew-ratio, which indicates any amplification bias, should be below 10. Our results (Appendix H, Table 9 and Figure S. 2) implies successful and unbiased amplification and that library can be used for lentivirus production.

2.9.2 Lentiviral Transduction of Sox1-GFP mESCs, Neuroectoderm Differentiation, and Cell Sorting via FACS

Lentivirus production was done using HEK293FT cells gifted by Prof. Dr. Mesut Muyan. Sox1-GFP cells grown with 2i medium were transduced with lentivirus at a low MOI and puromycin selection was applied. Sox1-GFP cells successfully transduced with the library were subjected neuroectoderm differentiation protocol as described earlier. At mESC state, 1 x 10^6 cells were collected and washed with cold 1X PBS three times. Pellet was stored at -80ºC until the day of genomic DNA isolation. On day 5 of neuroectoderm differentiation, cells were collected in FACS buffer and sorted with FACSAria II Cell Sorter (Figure 2.2). The population with the ~5-7% highest GFP signal was collected and centrifuged. The pellet was stored at -80ºC until genomic DNA isolation. The experiment was done as three biological replicates. gDNA isolation for all samples was performed simultaneously

2.9.2.1 NGS for Screening and MAGeCK Analysis

For sequencing, each sample was uniquely barcoded with 6 different indices, and staggered primers at varying lengths were added. After NGS, these adapters were removed from the raw reads using *cutadapt* function (Martin, 2011). We then mapped these cleaned reads to our reference library using the *bowtie2* (Langmead & Salzberg, 2012), allowing for a single mismatch event. The success of the mapping was visualized using MAGeCK-VISPR algorithm (W. Li et al., 2015) (Appendix I, Tables 10, 11 and Figure S. 3). The analysis for pooled screening and detection of critical epigenetic factors in neuroectoderm differentiation was done using MAGeCK algorithm (W. Li et al., 2014) on Ubuntu. This algorithm provides a robust and sensitive detection method. The default algorithm performs a median normalization. However, the inclusion of nontargeting sgRNAs as negative controls might skew the representation of sgRNAs. To account for that, control normalization method of the algorithm is used, where after normalization, non-targeting sgRNAs

are also used to generate the null distribution. We acknowledge potential increases in false positives with this method, so we've included results from both approaches for comparison. sgRNAs identified in both analyses are deemed more reliable, while unique hits from the control-normalized analysis require further validation. The sgRNAs that are found in the top 7% population at least 1.5 times ($|logFC| \geq 0.58$) less than mESC state are considered significant. The genes that have at least two sgRNAs identified as significant were considered crucial for neuroectoderm differentiation. The significant genes according to control-normalization method was visualized via STRING database.

CHAPTER 3

RESULTS

3.1 mESC Growth and Differentiation

The preliminary mass spectrometry data suggested that neuroectoderm differentiation is accompanied by elevated H3K36me2 levels However, this increase was not observed during endoderm differentiation, nor was it seen in the levels of H3K36me3. To validate and further investigate these findings, WT CJ9 mESCs were cultured in serum-free 2i medium, and differentiated into neuroectoderm for 5 days. This matches the time course used in the initial mass spectrometry analysis.

In mESC state, colonies were round and three-dimensional, while as neuroectoderm differentiation progressed, colonies transitioned to a non-spherical and monolayer morphology (Figure 3.1) This morphological change indicates a successful neuroectoderm differentiation (Q. L. Ying et al., 2003).

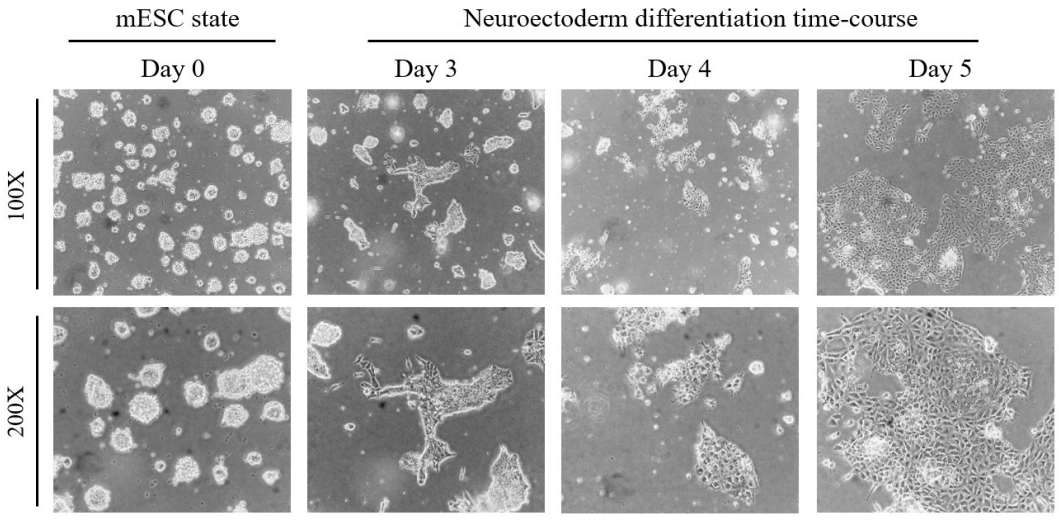

Figure 3.1. Neuroectoderm differentiation of mESCs. Light microscopy images were taken from mESC state and neuroectoderm differentiation time-course using Olympus CKX-53 microscope at 100X and 200X magnifications.

Four independent biological replicates were subjected to the same differentiation protocol and samples were collected on mESC state and the third and fifth day of differentiation for RT-qPCR analysis.

The success and efficacy of neuroectoderm differentiation were assessed through RT-qPCR analysis of key pluripotency markers (*Nanog*, *Pouf51* (*Oct4*)) (Figure 3.2 A, B), as well as genes specific to early neuroectoderm differentiation (*Sox1*, *Pax6*, *N-cadherin*) (Figure 3.2 C-D). In line with previous literature (Suter et al., 2009), expression level of *Sox1*, the earliest known neural lineage marker, reached its highest point at the third day of neuroectoderm differentiation. Its downstream targets, *Pax6* and *N-cadherin*, displayed a gradual increase through the fifth day of differentiation. These trends were accompanied by a sharp decline in the expression levels of both pluripotency markers, showing the transition from pluripotency to lineage commitment. Collectively, these results indicate a successful neuroectoderm differentiation process.

Endoderm differentiation was performed by Emre Balbaşı following an established protocol (Balbasi, Sezginmert, et al., 2022). Success of endoderm differentiation was assesed by the reducton of pluripotency markers (*Oct4*, *Sox2*, *Nanog*, *Klf4*) and the emergence and upregulation of endoderm related genes (*Foxa2*, *Gsc*, *Sox17*, *Gata6*) (*see* Appendix G, Table 8). It is apparent that downstream targets of *Foxa2*; *Gsc*, *Sox17* and *Gata6*, followed its the expression trend showing the functional outcome of *Foxa2* expression. To validate preliminary mass spectrometry results, samples were collected from mESC state and the fifth day of endoderm differentiation for histone extraction and Western blot analysis.

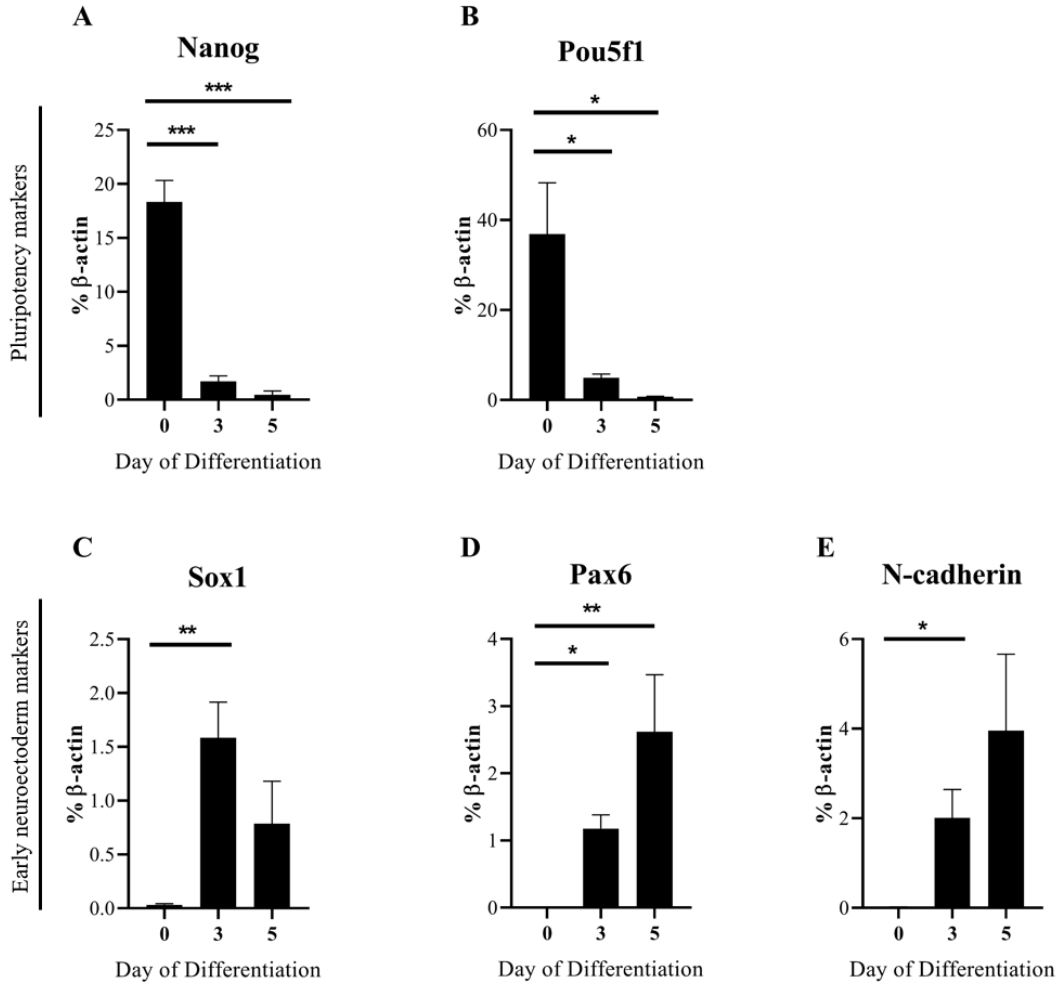

Figure 3.2. RT-qPCR analysis of **A-B** pluripotency and **C-E** early neuroectoderm lineage markers at mESC state (day 0) and during neuroectoderm differentiation time-course. Results were normalized to *β-actin* levels. The error bars represent the SEM of at least three biological replicates. Statistical analysis (Student's t-test) was performed using GraphPad Prism software. Statistical significance is denoted as follows; *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001.

3.2 Validation of Mass Spectrometry Results for Levels of H3K36 Methylation during Neuroectoderm and Endoderm Differentiation

In order to ensure the robustness and reliability of the initial findings on H3K36 methylation derived from mass spectrometry, validation studies were conducted using Western blotting. Cells were collected at the mESC state and on the fifth day of both neuroectoderm and endoderm differentiation. Following histone extraction,

Western blot analysis was performed to examine the levels of H3K36me2 and H3K36me3 levels. The intensities of bands were quantified for at least two biological replicates and modifications were reported as fold over H3 for a better comparison. The results demonstrated a substantial increase in H3K36me2 level on the fifth day of neuroectoderm differentiation compared to mESC state (Figure 3.3 A, C), while H3K36me3 levels remained unchanged (Figure 3.3 B, D). On the other hand, during endoderm differentiation, both H3K36me2 (Figure 3.3 A, C) and H3K36me3 (Figure 3.3 B, D) levels remained similar.

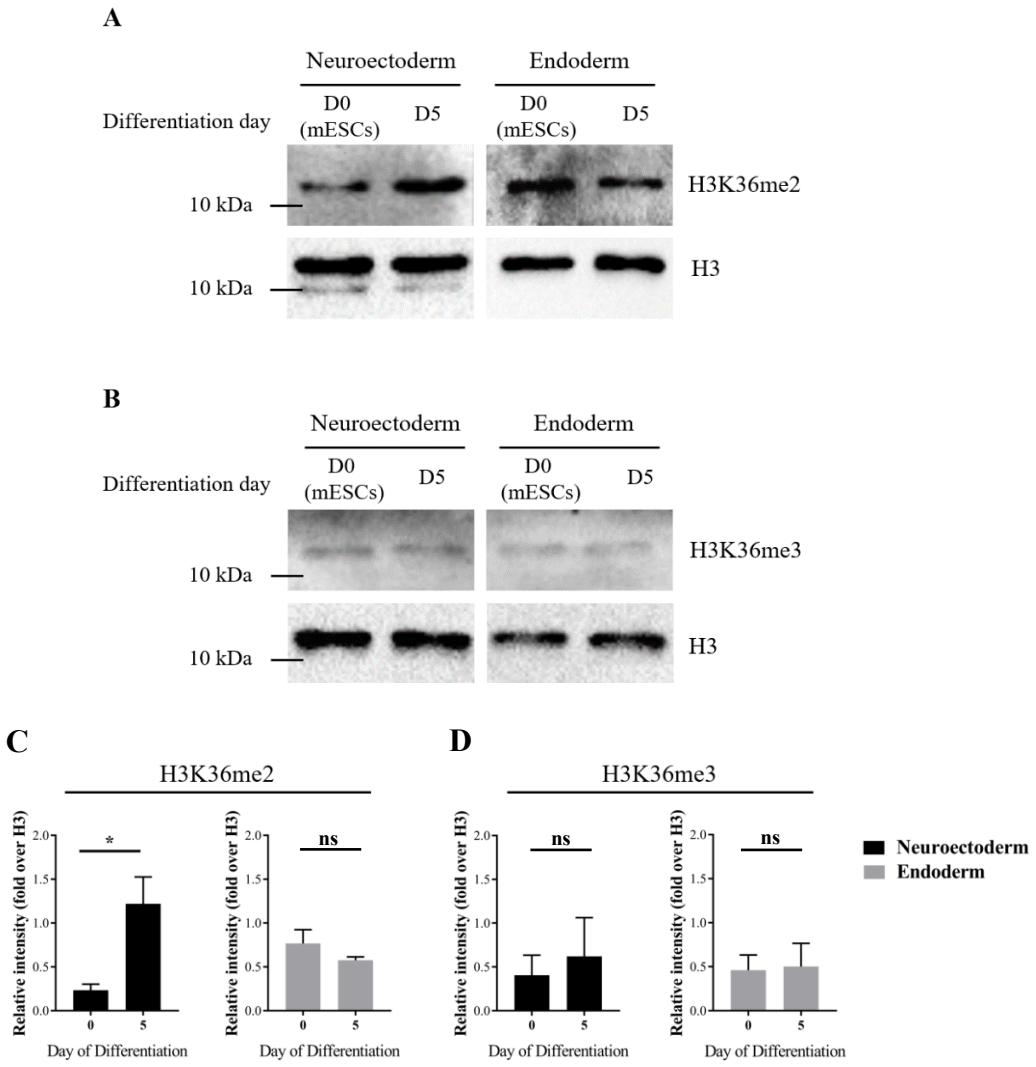

Figure 3.3. Validation of mass spectrometry findings in mESCs and neuroectoderm and endoderm differentiation. The levels of **A.** H3K36me2 and **B.** H3K36me3 were assessed via Western blot using histone extracts from mESCs and neurectoderm and endoderm directed cells. Histone 3 (H3) was used as loading control. The results are representative of at least two biological replicates. Intensities of H3K36me2 and H3K36me3 were quantified using ImageJ software and normalized to H3. Relative intensities of **C.** H3K36me2 and **D.** H3K36me3 are presented in bar graphs, shown as fold over H3. Error bars represent SEM of at least two biological replicates. Student's t-test was done by using GraphPad Prism software. *: p-value < 0.05. Nonsignificant values were denoted as "ns".

Taken together, these results confirmed the findings obtained from mass spectrometry for the levels of H3K36 methylations during endoderm and neuroectoderm differentiation. The data suggest that H3K36me2 might have a

distinctive role during neuroectoderm lineage choice rather than being primarily involved in the exit from pluripotency.

3.3 Transcriptional Levels of H3K36 Methyltransferases during Neuroectoderm Differentiation

The initial step for investigating the potential role of H3K36me2 in neuroectoderm differentiation was to understand its regulation during the differentiation process. To this end, the expression levels of the methyltransferases responsible for H3K36 methylations were examined using RT-qPCR analysis. Among these methyltransferases, only *Setd2* and *Setd5* are capable of adding a third methyl group to an already dimethylated 36th lysine residue (Sessa et al., 2019; Wagner & Carpenter, 2012), while the others can only add up to two methyl groups to H3K36.

A significant decrease in *Nsd1* expression level was observed during the course of neuroectoderm differentiation (Figure 3.4 A), while a contrasting increase was noted in *Nsd2* expression (Figure 3.4 B). These expression trends were found to correlate strongly with H3K36me2 levels during neuroectoderm differentiation in a negative and positive manner, respectively (*see* Appendix F, Table 7). Although the level of *Nsd3* expression showed a rising trend, it was found to be statistically insignificant (Figure 3.4 C). Moreover, the overall expression of *Nsd3* much lower compared to *Nsd1* and *Nsd2*. These differential trends could suggest unique roles for these enzymes at different time points. Among the other mono- and di-methyltransferases, only *Setd3* expression displayed a declining trend (Figure 3.4 E), while expression levels of *Aslh1l*, *Setmar* and *Smyd2* did not reveal any apparent trend (Figure 3.4 G-I).

Regarding trimethyltransferases, no clear trend was noted for *Setd2* expression (Figure 3.4 D), while the expression level of *Setd5* exhibited a gradual decline that was statistically significant on day 5 compared to mESC state (Figure 3.4 F). This decline could potentially contribute to observed increase in H3K36me2 levels, as

dimethylated K36 residue is used as a substrate for trimethyltransferases. Indeed, a strong negative correlation was found between *Setd5* expression and H3K36me2 level during neuroectoderm differentiation (*see* Appendix F, Table 7). Though SETD2 is considered as the main methyltransferase responsible for H3K36me3, SETD5 might have more prominent roles in neuroectoderm differentiation.

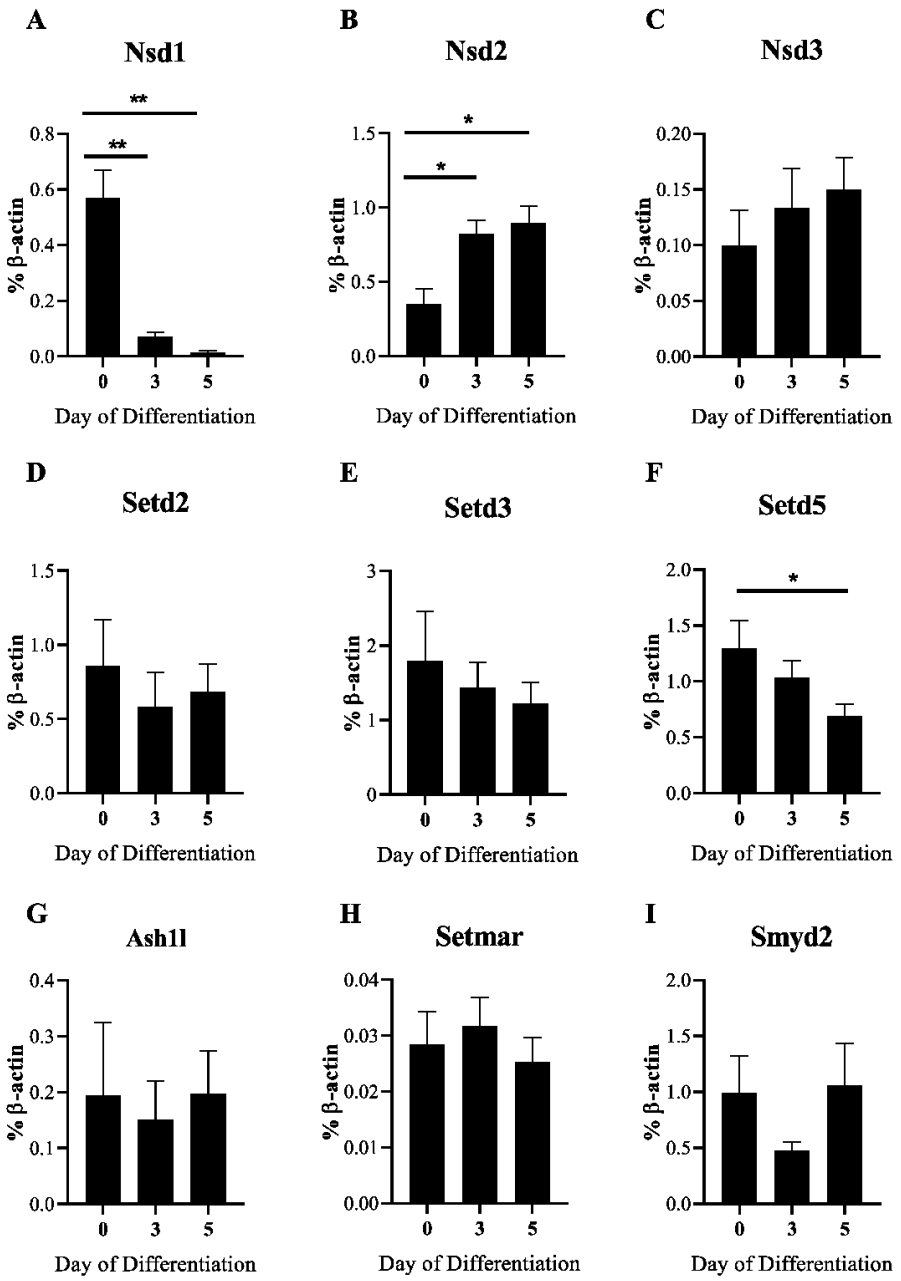

Figure 3.4. RT-qPCR analysis of H3K36 methyltransferases at mESC state (day 0) and during neuroectoderm differentiation time-course. Results were normalized to *β-actin* levels. The error bars represent the SEM of at least three biological replicates. Statistical analysis (Student's t-test) was performed using GraphPad Prism software. Statistical significance is denoted as follows; *: p-value < 0.05, **: p-value < 0.01. Nonsignificant values are not shown.

3.4 Transcriptional Levels of H3K36 Demethylases during Neuroectoderm Differentiation

It is critical to consider that histone methylation levels are dynamically regulated not only by methyltransferases but also by demethylases. Thus, to gain a more comprehensive understanding on the regulation of H3K36 methylations during the course of neuroectoderm differentiation, the expression trends of H3K36 demethylases were next explored.

H3K36 demethylation is primarily carried out by two families of enzymes; KDM2 and KDM4. While demethylases from the KDM2 family act on H3K36me1/2, the KDM4 family demethylase H3K36me2/3 and H3K9me2/3 (Klose et al., 2006). RT-qPCR analysis did not reveal discernable trends for *Kdm2a*, *Kdm2b*, *Kdm4a* and *Kdm4b* throughout the differentiation (Figure 3.5 A-D). Nevertheless, the *Kdm4c* expression level was observed to decline gradually, reaching statistically significance on day 5 compared to mESC state (Figure 3.5 E). It also very strongly correlated with H3K36me2 level during neuroectoderm differentiation in a negative manner (*see* Appendix F, Table 7). Expression levels of *Kdm8* were quite low both in mESC state and during differentiation (data not shown). This is not unexpected since it is associated with already differentiated proliferating Schwann cells (Fuhrmann et al., 2018).

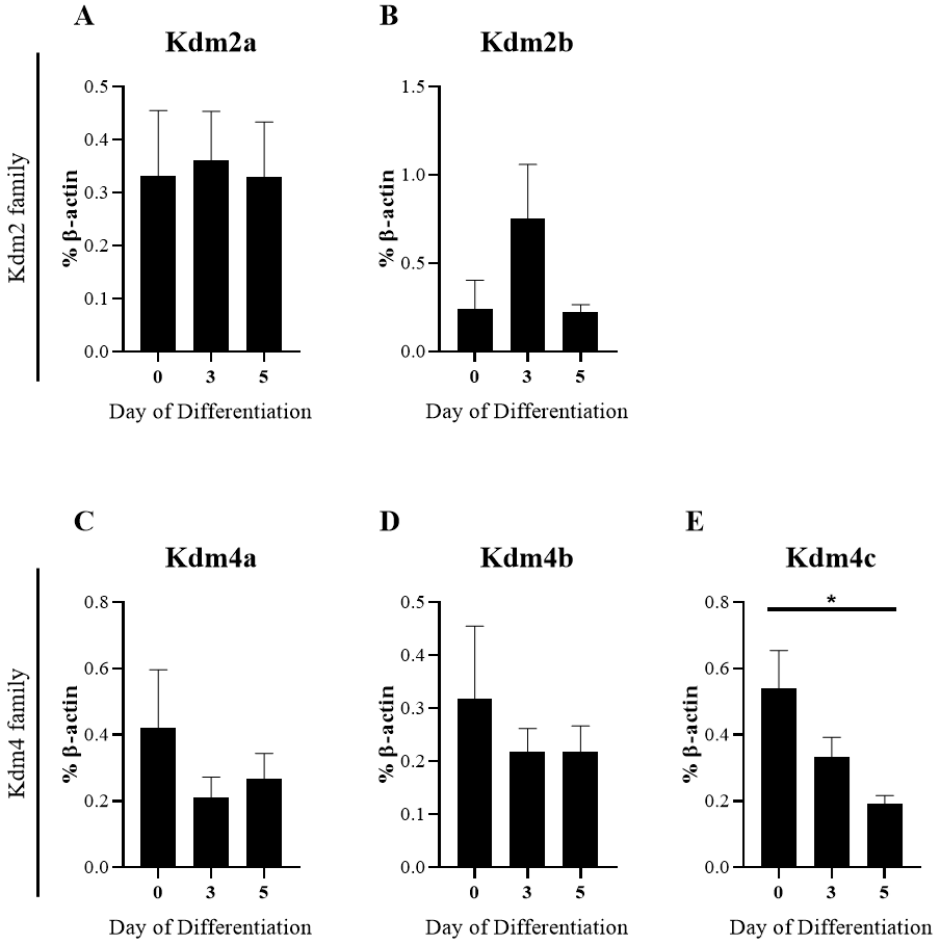

Figure 3.5. RT-qPCR analysis of H3K36 demethylases at mESC state (day 0) and during neuroectoderm differentiation time-course. Results were normalized to *β-actin* levels. The error bars represent the SEM of at least three biological replicates. Statistical analysis (Student's t-test) was performed using GraphPad Prism software. Statistical significance is denoted as follows; *: p-value < 0.05. Nonsignificant values are not shown.

3.5 Transcriptional Levels of H3K36me2/3 Readers during Neuroectoderm Differentiation

In addition to studying the regulation of H3K36 methylation during neuroectoderm differentiation, understanding how these modifications are interpreted within the cellular and differentiation context is equally crucial. This interpretation is facilitated by reader proteins that can recognize certain modifications via conserved domains, and recruit downstream proteins to the site to exert the function. Therefore, the next

step of this study was to examine the expression patterns of H3K36 readers during the neuroectoderm differentiation time course.

Among H3K36me3 readers, *Phf1* (PCL1) and *Phf19* (PCL3) displayed relatively low levels of expression (data not shown) both at the mESC state and during neuroectoderm differentiation. As such, these readers were not included in the analysis of the remaining replicates.

Two other readers of H3K36me2/3 are DNMT3A and DNMT3B; *de novo* DNA methyltransferases that are involved in establishing DNA methylation patterns during early embryonic development. Both proteins contain PWWP domains that specifically recognize H3K36me2/3. At any time-point, the expression level of *Dnmt3a* was observed to be much lower than *Dnmt3b*. While *Dnmt3a* displayed quite a visible downward trend, it didn't reach statistical significance (Figure 3.6 A). *Dnmt3b*, in contrast, showed a significant increase in expression on day 3 compared to mESC state, and then decreased on day 5 compared to day 3 (Figure 3.6 B). These different patterns could suggest that these two readers might have distinct yet overlapping roles during the course of early neuroectoderm differentiation.

Expression of *Msh6*, an H3K36me2/3 reader involved in DNA mismatch repair (MMR) machinery, declined significantly throughout differentiation (Figure 3.6 C). A very strong correlation was found between this decline and the H3K36me2 level during neuroectoderm differentiation (*see* Appendix F, Table 7). A steady decrease was observed in the expression levels of *Mrg15* (*Morf4l1*) (Figure 3.6 D), while *Psip1* showed a continuous increase (Figure 3.6 E), though neither changes were statistically significant.

Lastly, another reader that is specific to lysine 36 trimethylation on H3.3 histone variant, *Zmynd11*, displayed significant gradual increase in expression with neuroectoderm differentiation (Figure 3.6 F).

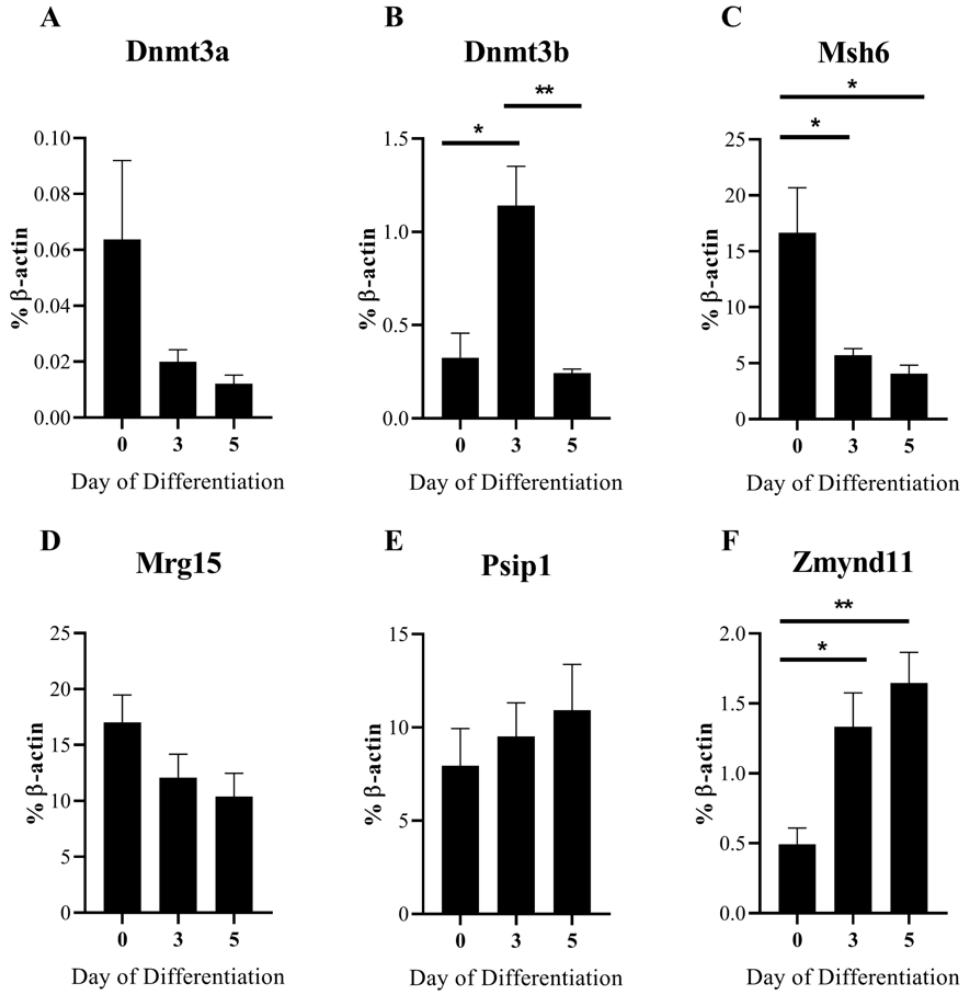

Figure 3.6. RT-qPCR analysis of H3K36 readers at mESC state (day 0) and during neuroectoderm differentiation time-course. Results were normalized to *β-actin* levels. The error bars represent the SEM of at least three biological replicates. Statistical analysis (Student's t-test) was performed using GraphPad Prism software. Statistical significance is denoted as follows; *: p-value < 0.05, **: p-value < 0.01. Nonsignificant values are not shown.

3.6 Transcriptomic Analysis of Selected Genes Involved in H3K36 Methylation During Endoderm Differentiation

The previously conducted RT-qPCR analyses done in the previous sections highlighted a list of genes related to H3K36 methylations that are significantly altered during neuroectoderm differentiation. Additionally, a number of genes displayed noticeable trends, even if not reaching statistical significance. The next

critical point to consider is whether these changes are exclusive to neuroectoderm differentiation or are observed more generally during the exit from pluripotency. To address this query, examining the expression changes of these genes during endoderm differentiation can provide valuable insights given that neuroectoderm and mesoderm/endoderm branch off early in embryonic development. For this purpose, transcriptomics data of endoderm differentiation time-course generated by Emre Balbaşı was utilized.

Both the visualization via IGV and differential expression analysis of the RNA-seq data indicated significant increases in *Nsd2* (Figure 3.7 A, Table 1) and *Zmynd11* (Figure 3.7 C, Table 3) expression levels with endoderm differentiation. This finding aligns with their behavior during neuroectoderm differentiation. Conversely, expression levels of *Kdm4c* (Figure 3.7 D, Table 2) and *Msh6* (Figure 3.7 C, Table 3) displayed a significant decline during both endoderm and neuroectoderm differentiation. Yet, the significant decrease observed in *Nsd1* expression was confined to neuroectoderm differentiation with no corresponding significant change during endoderm differentiation (Figure 3.7 A, Table 1).

The expression pattern of *Dnmt3b* differed between the two differentiation processes. While *Dnmt3b* reached its peak on day 3 of neuroectoderm differentiation followed by a decline, its levels rose and remained high during endoderm differentiation (Figure 3.7 C, Table 3). Setd5 expression trend also diverged across differentiation processes. A notable decrease was observed during neuroectoderm differentiation, whereas an upward trend, though not statistically significant, was seen during endoderm differentiation. (Figure 3.7 A, Table 1). This difference in *Setd5* behavior might have a contribution to the observed increase in accumulated H3K36me2 levels, which was specific to neuroectoderm differentiation but not evident in endoderm differentiation.

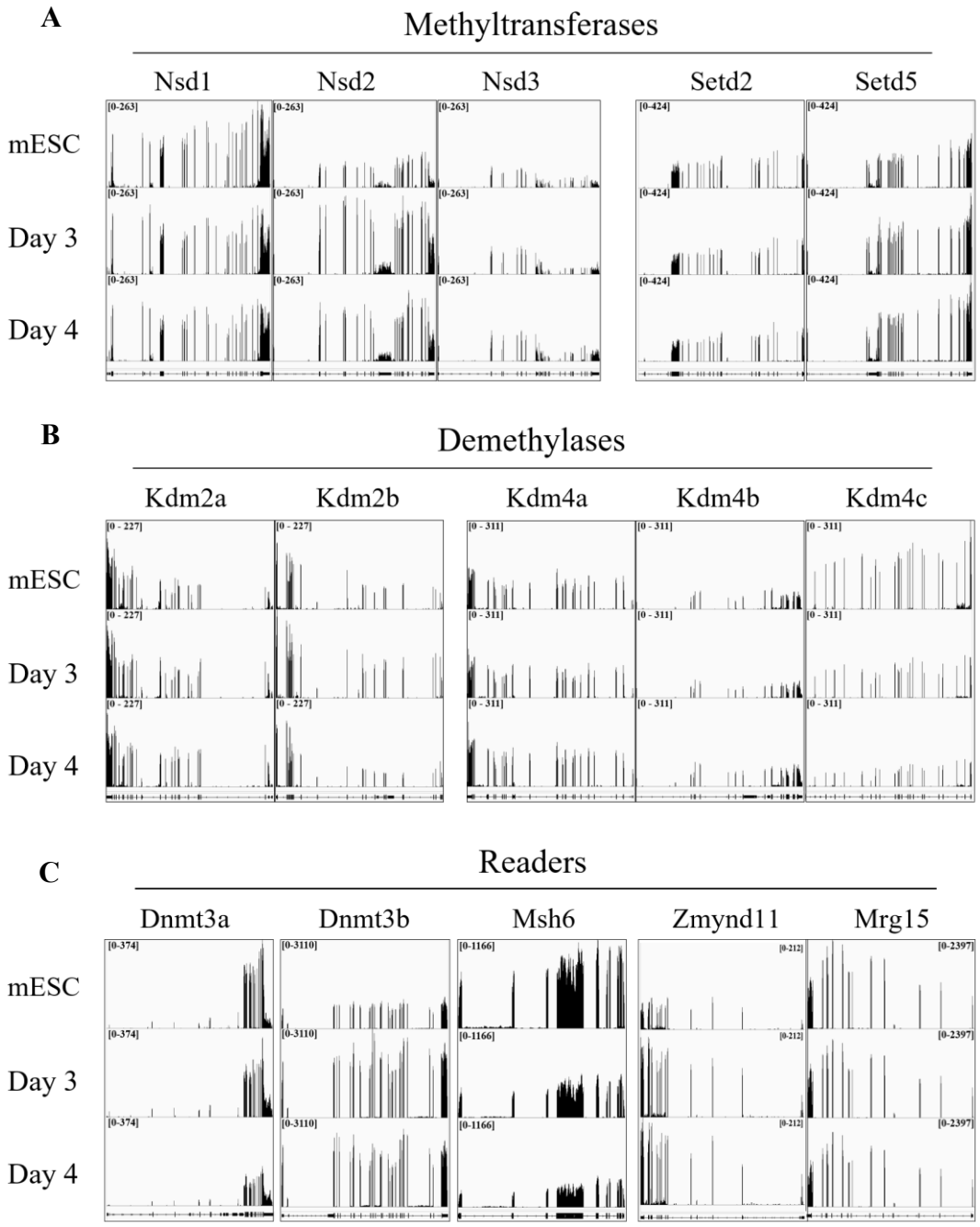

Figure 3.7. Expression levels of selected genes during the course of endoderm differentiation. Integrative Genomics Viewer (IGV) was utilized for visualization. The results are representative of three biological replicates.

Table 1. Relative expression levels of selected H3K36 methyltransferases during endoderm differentiation. Genes demonstrating an absolute fold change (|FC|) of ≥ 1.5 (corresponding to |logFC| ≥ 0.58) and a false discovery rate (FDR) of < 0.05 are considered significantly altered during endoderm differentiation, and are highlighted in the table.

Gene	*Ratio*	*Log(FC)*	*FDR*
Nsd1	ESC/Day 3	0.18	0.001
	ESC/Day 4	0.17	0.002
Nsd2	ESC/Day 3	-1.36	0.000
	ESC/Day 4	-1.09	0.000
Nsd3	ESC/Day 3	-0.20	0.021
	ESC/Day 4	-0.50	0.000
Setd2	ESC/Day 3	0.08	0.263
	ESC/Day 4	0.14	0.032
Setd5	ESC/Day 3	-0.37	0.000
	ESC/Day 4	-0.38	0.000

Table 2. Relative expression levels of selected H3K36 demethylases during endoderm differentiation. Genes demonstrating an absolute fold change (|FC|) of ≥ 1.5 (corresponding to |logFC| ≥ 0.58) and a false discovery rate (FDR) of < 0.05 are considered significantly altered during endoderm differentiation, and are highlighted in the table.

Gene	*Ratio*	*Log(FC)*	*FDR*
Kdm2a	ESC/Day 3	-0.16	0.0201
	ESC/Day 4	-0.11	0.101
Kdm2b	ESC/Day 3	-0.52	0.000
	ESC/Day 4	0.08	0.527
Kdm4a	ESC/Day 3	0.14	0.040
	ESC/Day 4	-0.14	0.042
Kdm4b	ESC/Day 3	0.42	0.002
	ESC/Day 4	-0.06	0.627
Kdm4c	ESC/Day 3	0.70	0.000
	ESC/Day 4	1.28	0.000

Table 3. Relative expression levels of selected H3K36 readers during endoderm differentiation. Genes demonstrating an absolute fold change (|FC|) of ≥ 1.5 (corresponding to |logFC| ≥ 0.58) and a false discovery rate (FDR) of < 0.05 are considered significantly altered during endoderm differentiation, and are highlighted in the table.

Gene	*Ratio*	*Log(FC)*	*FDR*
Dnmt3a	ESC/Day 3	-0.40	0.002
	ESC/Day 4	0.31	0.012
Dnmt3b	ESC/Day 3	-1.83	0.000
	ESC/Day 4	-1.53	0.000
Msh6	ESC/Day 3	0.87	0.000
	ESC/Day 4	1.47	0.000
Zmynd11	ESC/Day 3	-1.01	0.000
	ESC/Day 4	-1.06	0.000
Psip1	ESC/Day 3	-0.71	0.000
	ESC/Day 4	-0.25	0.000
Mrg15	ESC/Day 3	0.22	0.000
	ESC/Day 4	0.27	0.000

3.7 CRISPR-Cas9 Mediated Knock-out Screening of Epigenetic Factors Revealed A List of H3K36 Methylation-Related Genes Required for Neuroectoderm Differentiation

Through control-normalization method, the MAGeCK analysis identified several epigenetic factors as crucial for neuroectoderm differentiation. Notably, these encompass members of the well-characterized PRC2 complex, as well as factors associated with DNA methylation, H3K4 methylation, H3K9 methylation, and H3K36 methylation (Figure 3.8, highlighted). The role of PRC2 in differentiation and pluripotency has been comprehensively investigated in previous studies (Richter et al., 2009; Shan et al., 2017; Walker et al., 2010), underscoring the credibility of our findings.

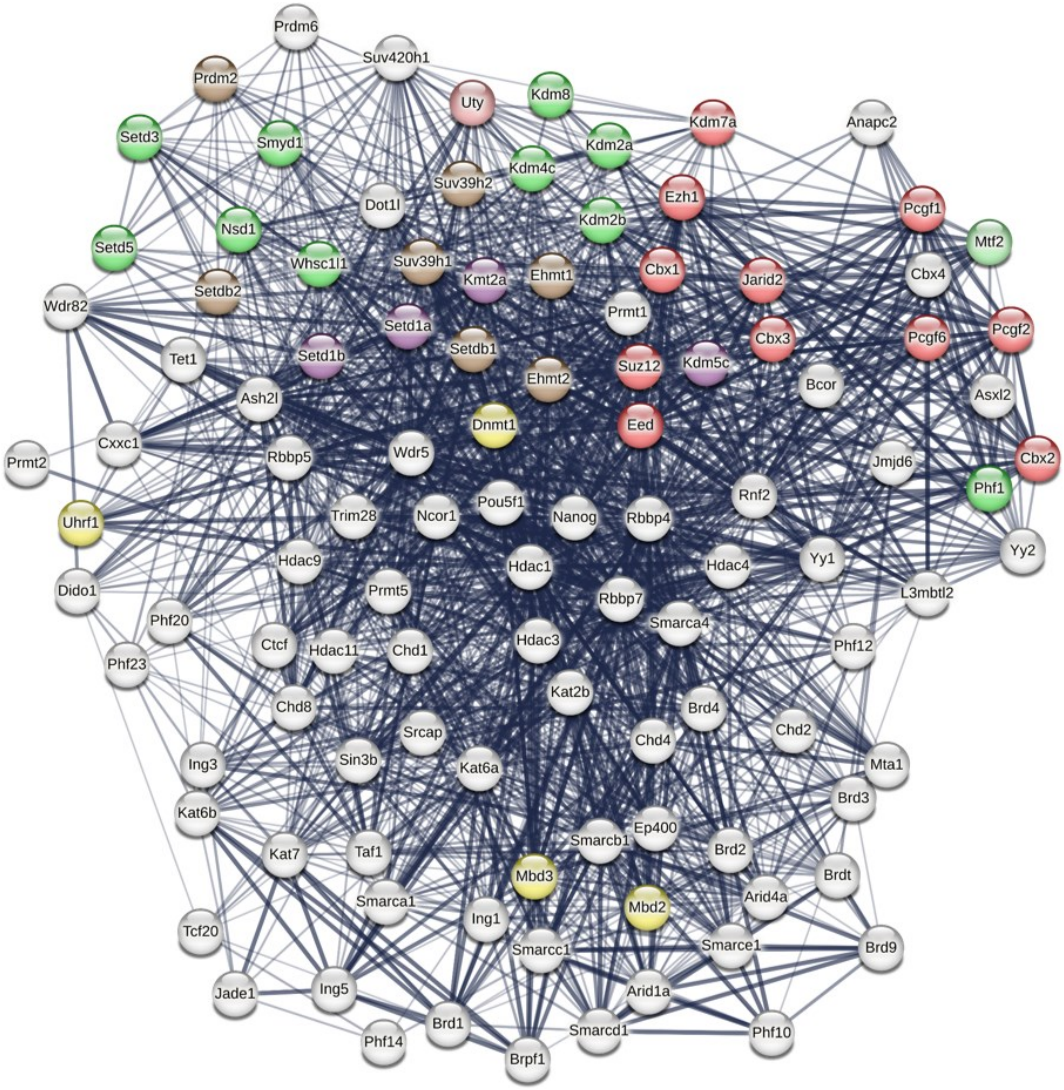

Figure 3.8. STRING analysis (Szklarczyk et al., 2015) of all factors found in modified MAGeCK analysis after control normalization. The ones related to H3K36 methylation (Green), PRC complexes and H3K27 methylation (Red), H3K9 methylation (Brown), DNA methylation (Yellow) and H3K4 methylation (Purple) are highlighted.

Among the methyltransferases, demethylases and readers of H3K36 methylation, the control-normalization method detected eight genes (*Nsd1, Setd5, Setd3, Symd1, Kdm2a, Kdm2b, Kdm4c* and *Phf1)* as essential for neuroectoderm differentiation. In contrast, the default model detected only three: *Nsd1, Setd5* and *Kdm2b* (Table 4). While the genes identified by both methods can be approached with higher

confidence, the others should be regarded cautiously due to the potential for false positives.

Table 4. H3K36 methyltransferases, demethylases and readers found to be essential for neuroectoderm differentiation according to CRISPR-based knockout screening results. Default MAGeCK analysis with median normalization and MAGeCK analysis after control normalization was done. *Nsd1, Setd5,* and *Kdm2b* were found in both analyses, while *Setd3, Smyd1, Kdm2a, Kdm4c, and Phf1* appeared only after control-normalization.

Class	*Gene*	*Default MAGeCK (Median Normalization)*	*Modified MAGeCK (Control Normalization)*
Methyltransferases	*Nsd1*	✓	✓
	Nsd3		✓
	Setd5	✓	✓
	Setd3		✓
	Smyd1		✓
Demethylases	*Kdm2a*		✓
	Kdm2b	✓	✓
	Kdm4c		✓
Readers	*Phf1*		✓

CHAPTER 4

DISCUSSION

Embryonic development represents a finely orchestrated and multifaceted process. Investigating this critical window is essential both for unraveling the mechanisms governing development and for providing insights into developmental disorders. Central to this complex process is the role of epigenetics. Epigenetic modifications, including histone modifications and DNA methylation, act as regulatory switches. These epigenetic factors guide pluripotent stem cells down specific developmental pathways, forming the diverse array of specialized cells essential for complex organisms.

To gain insights into these developmental pathways, we employed mESCs and differentiation protocols as our models. These provide an opportunity to unravel particular epigenetic shifts accompanying earliest cell fate decisions and loss of pluripotency. Notably, we detected a discernable increase in the global H3K36me2 levels, specifically during neuroectoderm differentiation (Figure 1.3 and Figure 3.3). Such an upregulation suggests a potential role of H3K36 methylation in the neuroectoderm differentiation process.

We assessed the expression levels of H3K36 methyltransferases, demethylases, and readers during both endoderm and neuroectoderm differentiation time courses. Our aim was to shed light on the network involved in regulating H3K36 methylation during the two pivotal arms of early embryonic development. This analysis allowed us to group the observed changes into three possible categories: those associated with maintaining pluripotency, those guiding the transition from pluripotency to differentiation, and those involved in directing differentiation towards a specific lineage. Finally, using a pooled CRISPR knock-out screening experiment for epigenetic factors, we determined the critical players of H3K36 methylation essential

for neuroectoderm differentiation. We found elevated expression levels of *Nsd2* and *Zmynd11* during differentiation to both lineages (Figure 3.4 B, Figure 3.6 F, and Figure 3.7 A, C), whereas *Kdm4c* and *Msh6* demonstrated higher expression in mESCs compared to differentiated states (Figure 3.5 E, Figure 3.6 C, and Figure 3.7 B, C). On the other hand, *Nsd1* and *Setd5* displayed lineage-specific expression patterns (Figure 3.4 A, F, and Figure 3.7 A). Intriguingly, these two methyltransferases also emerged in our CRISPR screening results, indicating their crucial roles in neuroectoderm differentiation (Figure 3.8, Table 4).

Among the readers, *Msh6* exhibited significantly elevated expression levels in mESCs during the differentiation to both lineages (Figure 3.6 C, Figure 3.7 C, and Table 3). Possessing a PWWP (Pro-Trp-Trp-Pro) domain, this reader is capable of specifically recognizing H3K36me3. Together with MSH2, MSH6 participates in DNA damage repair, constituting one of the main pathways for mismatch repair (MMR). The recruitment of this heterodimer, termed MutSα, is critical to safeguard the genome from mutations during replication (Li et al., 2013). Notably, H3K36me3-mediated MMR is reported to preferentially protect actively transcribed genes from mutations induced by mismatches (Huang & Zhu, 2018; Hyun et al., 2017). This protection is facilitated by the deposition of H3K36me3 in these gene bodies, mainly by RNA Polymerase II-coupled SETD2 (Sun et al., 2005). Considering ESCs represent very early stages of embryonic development, any unrepaired mutation can have far-reaching consequences for the entire organism. Thus, MMR becomes especially critical for rapidly dividing ESCs with an open chromatin and active transcription, where mismatches may occur at a higher frequency (Fortini et al., 2013). Additionally, it is reported that with differentiation, somatic mutations can become more acceptable to some extent (Frosina, 2010). These findings are consistent with our observations where MMR component *Msh6* showed high expression levels in mESCs followed by a decline, though still present, with differentiation.

In the context of DNA repair, it is worth noting that while H3K36me3 is associated with homologous recombination (HR), H3K36me2 is implicated in non-homologous

end joining (NHEJ) (Fnu et al., 2011; Pai et al., 2014; Pfister et al., 2014). Among these two double-stranded break (DSB) repair systems, HR operates with high precision during the S-phase of the cell cycle, utilizing a template to ensure accuracy. In contrast, NHEJ acts throughout the cell cycle, offering flexibility at the expense of potential in/del mutations (Fortini et al., 2013). One of the mechanisms directing HR is through an H3K36me3 reader, PSIP1 (LEDGF). Despite consistently high *Psip1* levels during both differentiation processes (Figure 3.6 E, data not shown), our CRISPR screening results did not identify it as essential for neuroectoderm differentiation (Figure 3.8). This suggests that, although H3K36me3-PSIP1 mediated HR may be active during both differentiation processes, it is not indispensable and other mechanisms might be compensating without any detrimental consequences.

Strikingly, a shift from a preference for HR to NHEJ repair with differentiation has been reported (Choi et al., 2020; Stambrook & Tichy, 2010). There are two systems shown to be involved in H3K36me2-mediated NHEJ; one through SETMAR (also known as Metnase) and the other by NSD2 (also known as MMSET, WHSC1). SETMAR gets recruited to DSB sites, leading to H3K36 dimethylation, which in turn facilitates the binding and stabilization of NBS1 and Ku70 at DSBs, enhancing NHEJ repair (Fnu et al., 2011; Hyun et al., 2017). However, in the context of early embryonic development, this mechanism might not be dominant since the expression levels of *Setmar* were relatively low in both lineages (Figure 3.4 H, data not shown) and it did not appear in CRISPR knockout screen as well (Figure 3.8).

The other mechanism involves the recruitment of NSD2 to DSBs. NSD2, a SET domain-containing methyltransferase, facilitates the deposition of H3K36me2, which in turn aids in NHEJ repair. However, the exact mechanism of this process has yet to be fully elucidated (de Krijger et al., 2020). We observed significant upregulation of *Nsd2* expression levels during both differentiation time courses (Figure 3.4 B, Figure 3.7 A and Table 1). This upregulation, along with the downregulation of *Msh6*, might be reflective of the shift from H3K36me3-mediated HR to H3K36me2-mediated NHEJ repair. The increased H3K36me2 levels observed

in neuroectoderm differentiation could be also attributed to the increase in *Nsd2* to some extent. However, this expression trend was observed in both lineage differentiations without a corresponding rise in H3K36me2 level in endoderm differentiation; and NSD2 was absent in our CRISPR screening results. These observations imply that NSD2 may not be the primary contributor to the observed H3K36me2 increase during neuroectoderm differentiation. While it is recognized as an H3K36 methyltransferase, its role may be more pronounced in the dimethylation of H3K4, H3K7, and H4K20 (Wagner & Carpenter, 2012) during this developmental phase.

The maintenance of ESC pluripotency involves chromatin factors and key pluripotency-related transcription factors *Nanog, Oct4 (Pou5f1)* and *Sox2,* as well as the extended pluripotency network members *Klf4* and *c-Myc* (Ding et al., 2012). To achieve proper differentiation, the pluripotency network must be suppressed, and the genes related to the desired cell fate must be simultaneously activated (Orkin & Hochedlinger, 2011). Deposition of K36me3 on the H3.3 histone variant is considered a mark for actively transcribed genes (Guo et al., 2014). This modification is recognized by the reader protein ZMYND11 (Wen et al., 2014). We observed that while the expression level of *Zmynd11* remains relatively low in the mESC state, it significantly increases during differentiation towards both lineages (Figure 3.6 F, Figure 3.7 C and Table 3). This common trend might be indicative of a role in the pluripotency exit. *Zmynd11* was reported to be a tumor suppressor gene due to its role in transcriptional suppression of actively transcribed *c-Myc* (Wen et al., 2014). The mechanism behind this repression is not fully understood, but it was suggested to be through the regulation of RNA Polymerase II elongation and pausing (Wen et al., 2014). Given that c-Myc is one of the four "Yamanaka factors" essential for induced pluripotent stem cell generation (Takahashi & Yamanaka, 2006), and is also considered a pluripotency factor, it is plausible to speculate that ZMYND11 might be involved in repression of other highly expressed pluripotency factors marked with H3.3K36me3 for exit from pluripotency and differentiation of mESCs.

The role of H3K36 methylation in preventing intragenic cryptic transcription has been extensively studied in yeast. In this organism, Set2 is the sole methyltransferase responsible for all three levels of H3K36 methylation. It is recruited by Ser2-phosphorylated elongating form of RNA polymerase, which results in the deposition of H3K36me3 towards the 3' end of actively transcribed genes. This subsequently facilitates the recruitment of histone deacetylases (HDACs), such as Rpd3, to gene bodies where they remove histone acetylation. Consequently, a more condensed chromatin structure forms after RNA polymerase passage, both preventing cryptic transcription initiation and promoting efficient transcription elongation (Carrozza et al., 2005). In mammals, RNA polymerase II-coupled recruitment of SETD2 and H3K36me3 addition to gene bodies are conserved, just as the role of H3K36me3 in preventing spurious intragenic transcription initiation is. However, the underlying mechanisms may be more diverse and not necessarily involve HDACs' recruitment (Mccauley & Dang, 2022).

One proposed mechanism involves the recruitment of MRG15, an H3K36me3 reader, to gene bodies. This then leads to the recruitment of KDM5B, an H3K4-specific demethylase establishing a less "active" intragenic chromatin environment (Hayakawa et al., 2007). Studies have reported that the depletion of KDM5B or MRG15 results in increased levels of H3K4me3 within gene bodies, leading to a surge in cryptic transcription (Mccauley & Dang, 2022; Xie et al., 2011). This cryptic transcription initiation can, in turn, interfere with proper transcriptional elongation, resulting in decreased expression levels of KDM5B targets, which include self-renewal-associated genes. In this study, we noted that *Mrg15* expression was high in mESCs but diminished over the course of differentiation, though not statistically significant (Figure 3.6 D, Figure 3.7 C and Table 3). This decrease might suggest a corresponding reduction of *Kdm5b*, a trend we observed during endoderm differentiation (data not shown). These downward trends might be involved in the repression of self-renewal genes by increased cryptic transcription. Indeed, one study has shown that *Kdm5b* knockdown enhances the expression of lineage-specific genes (Xie et al., 2011), presumably by inhibiting the pluripotency network.

A similar mechanism has been proposed for DNMT3B, a *de novo* DNA methyltransferase capable of recognizing H3K36me3. Upon binding to H3K36me3 within gene bodies, DNMT3B promotes increased DNA methylation, which helps safeguard against spurious transcription (Baubec et al., 2015). Supporting this, a study found that in the absence of DNA methylation, mESCs showed elevated levels of intragenic transcriptional initiation (Neri et al., 2017). However, it is essential to note that DNMT3B, in collaboration with DNMT3A, primarily establishes DNA methylation patterns during early embryogenesis, and their expression profiles should be interpreted in that context. Interestingly, DNMT3B knock-out cells were found to be incapable of differentiating into meso-endodermal lineages, yet the neuroectodermal differentiation remained unaffected (Lauria et al., 2023). This aligns with our data showing a consistent increase in *Dnmt3b* during endoderm differentiation (Figure 3.7 C and Table 3); whereas in neuroectoderm differentiation, it just peaked at day 3 before decreasing (Figure 3.6 B). The aforementioned study might also explain why we did not encounter *Dnmt3b* in our CRISPR screen to be essential for neuroectoderm differentiation.

During both differentiation processes, we observed significantly decreased levels of *Kdm4c* (Figure 3.5 E, Figure 3.7 B and Table 2). This demethylase specifically targets and removes the methyl group from H3K36me3, and thereby facilitates the recruitment of PRC2 which adds repressive H3K27me3 to its targets. This dynamic is an example of the antagonism between the two histone modifications where the presence of H3K36me2/3 acts to limit H3K27me3 (Yuan et al., 2011). A previous study demonstrated that the reduction in *Kdm4c* is essential for differentiation, a process during which lineage-specific genes are indirectly activated (Das et al., 2014). In line with this, the same study identified mesendoderm and neuroectoderm related genes as targets of KDM4C. These targets include neuronal differentiation marker gene *Pax6* (Suter et al., 2009) and mesendoderm differentiation marker *Brachyury* (Lolas et al., 2014). Furthermore, *Kdm4c* appeared in our CRISPR screening results (Figure 3.8 and Table 4), supporting the role of H3K36me2 and *Kdm4c* in differentiation. This could be attributed to the possibility that the genes

linked to alternative cell fates are not properly silenced in the absence of KDM4C during differentiation towards a specific lineage, or even within the mESC state.

The interplay between H3K36 methylation and PRC complexes can also be exemplified by the roles of KDM2A/B. These H3K36me2-specific demethylases have been shown to recruit PRC1 complexes leading to the repression of lineage-specific genes in mESCs (Farcas et al., 2012; J. Li et al., 2019). Our CRISPR screening results align with these findings, highlighting the importance of KDM2A/B in neuroectoderm differentiation (Figure 3.8 and Table 4). However, sgRNAs targeting these genes appear to be underrepresented in the mESC state with approximately 1,900 reads compared to the overall average of 3,500 reads. This could potentially suggest a pluripotency defect presumably through an insufficient repression of lineage-specific genes.

During early embryonic development, alternative splicing (AS) diversifies the transcriptome to facilitate lineage specification (Fiszbein & Kornblihtt, 2017; Xu et al., 2018). This process is especially pronounced in the central nervous system and neural development, where AS events are abundant (Furlanis et al., 2019; Raj & Blencowe, 2015). H3K36 methylation is intricately linked to AS through two adapter systems: PSIP1/SRSF1 and MRG15/PTBP1. These have opposing outcomes: while recognition by PSIP1 promotes the inclusion of targeted exons, MRG15 favors their exclusion (Luco et al., 2010; Pradeepa et al., 2012). Interestingly, neither of these systems appeared on our CRISPR screen. This absence could be attributed to the early differentiation stage we examined. The AS process might become more pivotal during later neural specification. The observed increase in H3K36me2, being a precursor to H3K36me3, might hint at a genomic preparation for extensive AS events in subsequent developmental phases though such a mechanism has not yet been studied.

Among the three NSD family methyltransferases responsible for H3K36me2, our CRISPR screen identified *Nsd1* and *Nsd3 (Whsc1l1)* as crucial factors for neuroectoderm differentiation (Figure 3.8 and Table 4). Interestingly, sgRNAs

targeting *Nsd3* were underrepresented in the mESC state, suggesting its significance in pluripotency and self-renewal, a role that requires further exploration. Although the significant upregulation of *Nsd2* coinciding with increased H3K36me2 levels during neuroectoderm differentiation suggests a link, its absence in CRISPR screen results implies that the role of NSD2 is not indispensable.

While NSD1 acts as a *de novo* methyltransferase establishing methylation patterns, NSD2 primarily functions as a propagator, recognizing already dimethylated H3K36 (Kuo et al., 2011; Martinez-Garcia et al., 2011; Sankaran et al., 2016; Waggoner et al., 2005). They also differ in their genomic localization; while NSD1 depletion diminishes H3K36me2 levels globally, NSD2 predominantly occupies active transcripts and elongating regions (Streubel et al., 2018). Although *Nsd1* expression levels decreased with neuroectoderm differentiation (Figure 3.4 A), this does not necessarily indicate an abrupt decrease in protein levels. As a *de novo* methyltransferase, NSD1 might be critical for establishing methylation patterns early in development; whereas other H3K36me2 methyltransferases might be compensating the propagator role of NSD2 in the absence of it. Thus, the deletion of *Nsd1* might have more drastic effects on H3K36me2 distribution and potentially its levels. The methylation patterns laid down by NSD1 are crucial to counteract PRC2 activity, preventing excessive deposition of the repressive H3K27me3 mark (J. Li et al., 2019; Streubel et al., 2018). This is critical for obtaining an accessible chromatin for development (Ritchie & Lizarraga, 2023). Additionally, NSD1 plays a role in transcriptional repression by recruiting HDAC1 to enhancers of developmental genes in mESCs. This might also be critical during neuroectoderm differentiation ensuring other lineage-specific genes are silenced during differentiation. Furthermore, through the recruitment of the de novo DNA methyltransferase DNMT3A, NSD1 facilitates shaping DNA methylation patterns. These multifaceted roles of NSD1 possibly account for its identification in our CRISPR screen; yet, more research is needed to clarify its association with H3K36me2 levels during neuroectoderm differentiation. Given that NSD1 is implied in the neurodevelopmental disorder Sotos syndrome (Kurotaki et al., 2002), and also is

involved in the development of the brain structure later on in the development (Ritchie & Lizarraga, 2023; Speir et al., 2021) understanding its early influence on neuroectoderm differentiation and chromatin preparation is pivotal for neurodevelopment insights.

Intriguingly, our CRISPR results revealed that SETD5 is crucial for neuroectoderm differentiation. This methyltransferase has recently been identified to perform trimethylation of H3K36 (Sessa et al., 2019; Zaghi et al., 2020), a role long has been attributed only to SETD2. We noted a decreasing trend in SETD5 expression during neuroectoderm differentiation (Figure 3.4 F), which might account for the rising levels of H3K36me2, given its substrate role for SETD5-mediated H3K36me3. SETD5 has already been reported in diverse physiological functions. It is shown to play a significant role in early nervous system development, with pronounced expression in the cerebral cortex during various developmental stages (Deliu et al., 2018; Kuechler et al., 2015). Additionally, mutations in the SETD5 gene have been linked to a range of neurodevelopmental disorders, including intellectual disability, autism spectrum disorder, and KBG syndrome (Pascolini et al., 2022; Rawlins et al., 2017; Szczałuba et al., 2016). Notably, these mutations often correlate with developmental delays in areas such as speech, language, and motor functions (Moore et al., 2019). Furthermore, studies indicate that loss of SETD5 function leads to embryonic fatality between embryonic days 10.5 and 11.5 (M. Li et al., 2023; Osipovich et al., 2016). Given this, understanding the implications of SETD5 absence during early neuroectoderm differentiation is intriguing, especially since its disruptions seem to initiate from this stage.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Cell fate decisions at the earliest developmental stages form the foundation for the cellular differentiation and the development of a complex organism. During early development, the neuroectoderm lineage diverges from meso/endodermal fate. Studying the mechanisms that govern this early cell fate decision is crucial for a comprehensive understanding of neurodevelopmental disorders.

This study identified a marked increase in H3K36me2 levels specifically during neuroectoderm differentiation; a change not observed during endoderm differentiation. Given that H3K36 methylation and its associated regulatory system have been previously linked to neural disorders (Zaghi et al., 2020), our findings are indicative that its significance might be originating early in the development, as early as the initial cell fate decisions leading to neuroectoderm lineage. Our CRISPR-mediated knock out screening results highlighted NSD1, SETD5 and KDM2B as crucial factors for neuroectoderm differentiation involved in H3K36 methylation regulation. Among these, *Nsd1* and *Setd5* also showed lineage specific expression trends making them primary candidates for future study. Additionally, shared expression patterns between the two lineages suggested that NSD2 and ZMYND11 might be associated with the transition from pluripotency to differentiation, while KDM4C and MSH6 might be involved in maintaining pluripotency. As these genes are involved in various processes, ranging from DNA damage repair to contrasting the action of repressive PRC2 complex, future studies delving into their specific roles will further elucidate the multifaceted nature of their involvement.

To uncover the changes in the distribution and levels of H3K36me2 and H3K36me3, this study needs ChIP-seq experiments performed on the mESCs and neuroectoderm differentiated cells. The observed increase in H3K36me2 during neuroectoderm differentiation may imply a broader distribution of the modification and/or elevated

levels. Conversely, seeing no change in overall H3K36me3 levels does not necessarily mean this modification remained the same throughout the genome during differentiation. Identifying which genomic regions and subsets of genes these methylations are deposited on, and how these patterns change with neuroectoderm differentiation would provide invaluable insights into their roles. For this purpose, we initially attempted to use the CUT&RUN technique, which offers several advantages over traditional ChIP-seq, particularly in its requirement for fewer cells (Skene & Henikoff, 2017). However, given that the protocol and associated kits were in their nascent stages of development and had not been extensively validated, we encountered technical challenges that prevented the derivation of meaningful results. Despite this setback, the necessity of such an analysis remains clear, and future work should prioritize employing ChIP-seq to gain deeper insights into H3K36me2/me3 distribution during neuroectoderm differentiation.

Given the intricate interplay among H3K36 methylations, H3K27 methylation established by PRC2, and DNA methylation catalyzed by DNMT3A/3B, it would be highly informative to map their distributions during neuroectoderm differentiation. For this purpose, techniques such as ChIP-Seq for H3K27me3 and DNMT3A/3B, and Bisulfite Sequencing for DNA methylation can be employed. Aligning these distribution profiles with those of H3K36me2/3 would give us a much more comprehensive perspective on the roles H3K36 methylation play during differentiation. While the main focus is on performing these experiments during neuroectoderm differentiation, extending them to the endoderm would offer a valuable comparison, helping to pinpoint the roles and changes specific to the neuroectoderm.

Another critical follow-up study is to validate the CRISPR knock out results and delve deeper into their implications. Our screen, together with insights from the preexisting literature, offers a variety of experiments that can be performed to further investigate the roles of these genes play during neuroectoderm differentiation and within the mESC state. Specifically, *Nsd1, Setd5* and *Kdm2b* emerge as primary subjects of interest due to their presence in the results of both our analysis methods.

Among these, sgRNAs targeting *Kdm2b* were found to be underrepresented in the mESCs compared to the differentiated state, suggesting a possible role in the maintenance of pluripotency. Thus, after the establishment of kdm2bΔ cell line, experiments should first be focused on the possible defects in the mESC state and accompanying epigenetic changes. Initially, the expression levels of pluripotency markers such as *Oct4, Nanog,* and *Sox2*; and differentiation related genes such as *Brachyury, Foxa2, Pax6* and *Sox1* should be compared between kdm2bΔ and WT mESCs via RT-qPCR analysis. If a defect exists, one would expect to observe lower levels of pluripotency related genes and a corresponding elevation in the expression of lineage specific genes in kdm2bΔ mESCs. Following this, the overall levels of H3K36me2 should be examined with Western blot experiment given that KDM2B act as a demethylase of H3K36me2/1. It would also be interesting to investigate how the levels of H3K36me2 and H3K27me3 change in the absence of KDM2B, particularly focusing on developmental genes, using ChIP-qPCR experiments. Since KDM2B is involved in the recruitment of PRC complexes and the consequent repression of developmental genes, an anticipated observation would be elevated H3K36me2 levels and lower H3K27me3 levels, particularly on genes that exhibit increased expression. This can also give us an idea about whether the defect disrupts differentiation to both lineages or is specific to neuroectoderm. Accordingly, kdm2bΔ mESCs could be differentiated into neuroectoderm and endoderm; followed by an RT-qPCR analysis of lineage marker expression levels. Alternatively, the Kdm2b deletion can be introduced into reporter cell lines, such as Sox1-GFP, allowing us to monitor the proportion of cells that differentiate successfully using Flow Cytometry.

The other two genes identified in our CRISPR screen, *Nsd1* and *Setd5,* displayed also lineage-specific expression patterns, suggesting potential roles as primary regulators of H3K36 methylation during neuroectoderm differentiation. First, the neuroectoderm differentiation defect needs to be validated. After establishing the nsd1Δ and setd5Δ cell lines, both should undergo neuroectoderm differentiation. The success of differentiation should then be assessed as previously described. If these

are indeed the main methyltransferases regulating the global levels of H3K36me2 during neuroectoderm differentiation, their absence should cause a discernible change detectable via Western blot analysis. If no changes in H3K36 methylation levels are observed during differentiation in these knockouts, subsequent knockouts of NSD2 and NSD3 may be necessary to identify the primary regulator. Observing a significant shift in H3K36me2 levels in nsd2Δ mESCs without a corresponding neuroectoderm differentiation defect would imply that the increase in H3K36me2 is not pivotal for this developmental phase. Then, further differentiation to specific neuronal cells might be necessary to uncover the onset of defect.

Along with the experiments suggested before, it is crucial to conduct further experiments to specifically target the various roles attributed to NSD1. Since it is reported to be a *de novo* methyltransferase, one would expect to see a drastic change in the distribution of H3K36me2 even in the mESC state. Additionally, H3K27me3 distributions might be altered since NSD1 reported to counteract PRC2 activity. Through ChIP-Seq experiments, determining the subset of genes and genomic regions these changes occur could give us insight on the neuroectoderm defect we observed. Furthermore, an HDAC1-ChIP-Seq would be informative for the repressive action of NSD1. For a subset of developmental genes, one would expect to see a decline in HDAC1 recruitment to enhancers. This could explain the differentiation defect if the genes related to alternative cell fates are not properly silenced in the absence of NSD1.

If we can pinpoint the main methyltransferases responsible for H3K36me2 elevation during neuroectoderm differentiation, then, further experiments could be conducted to understand which processes this alteration is involved in. Examining its effect on alternative splicing would be particularly insightful, especially given that no components of the alternative splicing machinery linked to H3K36 methylation were detected in our CRISPR screen. This may suggest that alternative splicing is not yet critical and/or other mechanisms might be compensating. However, later in neural development, alternative splicing is known to become crucial. Thus, it would be informative to observe alternative splicing events through RT-qPCR in the absence

of elevated H3K36me2 levels compared to WT. Furthermore, cryptic transcription initiation is reported to be prevented by H3K36-mediated mechanisms as discussed earlier. Thus, one could expect to see increased cryptic transcription initiation when a defect in H3K36me2 levels occur.

SETD5 is a newly identified H3K36 methyltransferase capable of adding the third methyl group. This enzyme is linked with neurodevelopmental diseases and was highlighted in our CRISPR screen results as being critical for neuroectoderm differentiation. Thus, it would be critical to see where in the genome H3K36me2/3 changes occur in the absence of SETD5, together with any possible change in the localization of SETD5 in WT cells during differentiation. If there is a shift in the subset of genes or genomic regions SETD5 is bound to with differentiation, the next step could be to detect the dynamic interaction partners of SETD5 through Co-IP Mass Spectrometry to gain insights into roles of SETD5 during neuroectoderm differentiation.

REFERENCES

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*, 91–103. https://doi.org/10.1038/nrm2618

Artus, J., Piliszek, A., & Hadjantonakis, A. (2011). The primitive endoderm lineage of the mouse blastocyst : Sequential transcription factor activation and regulation of differentiation by Sox17. *Developmental Biology*, *350*(2), 393–404. https://doi.org/10.1016/j.ydbio.2010.12.007

Aubert, J., Stavridis, M. P., Tweedie, S., O'Reilly, M., Vierlinger, K., Li, M., Ghazal, P., Pratt, T., Mason, J. O., Roy, D., & Smith, A. (2003). Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *100 Suppl*(Suppl 1), 11836–11841. https://doi.org/10.1073/pnas.1734197100

Aubert, J., Stavridis, M. P., Tweedie, S., Reilly, M. O., Vierlinger, K., Li, M., Ghazal, P., Pratt, T., Mason, J. O., Roy, D., & Smith, A. (2003). *Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1 – gfp knock-in mice*. https://doi.org/10.1073/pnas.1734197100

Balbasi, E., Guven, G., & Terzi Cizmecioglu, N. (2022). Mouse Embryonic Stem Cell Culture in Serum-Containing or 2i Conditions. *Methods in Molecular Biology*, *2520*, 275–294. https://doi.org/10.1007/7651_2021_438

Balbasi, E., Sezginmert, D., Alganatay, C., & Terzi Cizmecioglu, N. (2022). Directed Differentiation of Mouse Embryonic Stem Cells to Mesoderm, Endoderm, and Neuroectoderm Lineages. *Methods in Molecular Biology*, *2520*, 295–307. https://doi.org/10.1007/7651_2021_439

Balbaşı, E. (2022). *SETD3-Dependent Gene Expression Changes During Endoderm Differentiation of Mouse Embryonic Stem Cells* [Middle East Technical University]. https://open.metu.edu.tr/bitstream/handle/11511/96679/10385264.pdf

Bardot, E. S., & Hadjantonakis, A.-K. (2020). Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. *Mechanisms of Development*, *163*, 103617. https://doi.org/10.1016/j.mod.2020.103617

Barr, K. J., Garrill, A., Jones, D. H., Orlowski, J., & Kidder, G. M. (1998). Contributions of Na+/H+ Exchanger Isoforms to Preimplantation Development of the Mouse. *Molecular Reproduction and Development*, *153*(September 1997), 146–153.

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A., & Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, *520*(7546), 243–247. https://doi.org/10.1038/nature14176

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

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

Campos, E. I., & Reinberg, D. (2009). Histones: Annotating chromatin. *Annual Review of Genetics*, *43*, 559–599. https://doi.org/10.1146/annurev.genet.032608.103928

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., & Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, *123*(4), 581–592. https://doi.org/10.1016/j.cell.2005.10.023

Choi, E. H., Yoon, S., Koh, Y. E., Seo, Y. J., & Kim, K. P. (2020). Maintenance of genome integrity and active homologous recombination in embryonic stem cells. *Experimental and Molecular Medicine*, *52*(8), 1220–1229. https://doi.org/10.1038/s12276-020-0481-2

Das, P. P., Shao, Z., Beyaz, S., Apostolou, E., Pinello, L., Angeles, A. D. L., O'Brien, K., Atsma, J. M., Fujiwara, Y., Nguyen, M., Ljuboja, D., Guo, G., Woo, A., Yuan, G. C., Onder, T., Daley, G., Hochedlinger, K., Kim, J., & Orkin, S. H. (2014). Distinct and Combinatorial Functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in Mouse Embryonic Stem Cell Identity. *Molecular Cell*, *53*(1), 32–48. https://doi.org/10.1016/j.molcel.2013.11.011

Davis, S., Miura, S., Hill, C., Mishina, Y., & Klingensmith, J. (2004). BMP receptor IA is required in the mammalian embryo for endodermal morphogenesis and ectodermal patterning. *Developmental Biology* 270, 47–63. https://doi.org/10.1016/j.ydbio.2004.01.048

de Krijger, I., van der Torre, J., Peuscher, M. H., Eder, M., & Jacobs, J. J. L. (2020). H3K36 dimethylation by MMSET promotes classical non-homologous end-joining at unprotected telomeres. *Oncogene*, *39*(25), 4814–4827. https://doi.org/10.1038/s41388-020-1334-0

Deliu, E., Arecco, N., Morandell, J., Dotter, C. P., Contreras, X., Girardot, C., Käsper, E.-L., Kozlova, A., Kishi, K., Chiaradia, I., Noh, K.-M., & Novarino,

G. (2018). Haploinsufficiency of the intellectual disability gene SETD5 disturbs developmental gene expression and cognition. *Nature Neuroscience*, *21*(12), 1717–1727. https://doi.org/10.1038/s41593-018-0266-2

Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology*, *6*(8), 227. https://doi.org/10.1186/gb-2005-6-8-227

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., & Wang, J. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Research*, *22*(1), 155–167. https://doi.org/10.1038/cr.2011.179

Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO Journal*, *27*(2), 406–420. https://doi.org/10.1038/sj.emboj.7601967

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., Mcgouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W., Koseki, H., Brockdorff, N., Ponting, C. P., Kessler, B. M., & Klose, R. J. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *ELife*, *1*, 1–26. https://doi.org/10.7554/eLife.00205

Fiszbein, A., & Kornblihtt, A. R. (2017). Alternative splicing switches: Important players in cell differentiation. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *39*(6). https://doi.org/10.1002/bies.201600157

Fnu, S., Williamson, E. A., De Haro, L. P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S. H., Nickoloff, J. A., & Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(2), 540–545. https://doi.org/10.1073/pnas.1013571108

Fortini, P., Ferretti, C., & Dogliotti, E. (2013). The response to DNA damage during differentiation: pathways and consequences. *Mutation Research*, *743–744*, 160–168. https://doi.org/10.1016/j.mrfmmm.2013.03.004

Frosina, G. (2010). The bright and the dark sides of DNA repair in stem cells. *Journal of Biomedicine and Biotechnology*, *2010*(Table 1). https://doi.org/10.1155/2010/845396

Fuhrmann, D., Mernberger, M., Nist, A., Stiewe, T., & Elsässer, H. P. (2018). Miz1 controls schwann cell proliferation via H3K36 me2 demethylase Kdm8 to prevent peripheral nerve demyelination. *Journal of Neuroscience*, *38*(4), 858–877. https://doi.org/10.1523/JNEUROSCI.0843-17.2017

Fukuda, A., Hazelbaker, D. Z., Motosugi, N., Hao, J., Limone, F., Beccard, A., Mazzucato, P., Messana, A., Okada, C., San Juan, I. G., Qian, M., Umezawa,

A., Akutsu, H., Barrett, L. E., & Eggan, K. (2021). De novo DNA methyltransferases DNMT3A and DNMT3B are essential for XIST silencing for erosion of dosage compensation in pluripotent stem cells. *Stem Cell Reports*, *16*(9), 2138–2148. https://doi.org/10.1016/j.stemcr.2021.07.015

Furlanis, E., Traunmüller, L., Fucile, G., & Scheiffele, P. (2019). Landscape of ribosome-engaged transcript isoforms reveals extensive neuronal-cell-class-specific alternative splicing programs. *Nature Neuroscience*, *22*(October), 1709–1717. https://doi.org/10.1038/s41593-019-0465-5

Goldin, S. N., & Papaioannou, V. E. (2003). Paracrine Action of FGF4 During Periimplantation Development Maintains Trophectoderm and Primitive Endoderm. *Genesis*, *36*(1), 40–47. https://doi.org/10.1002/gene.10192

Götz, M., & Huttner, W. B. (2005). The cell biology of neurogenesis. *Nature Reviews Molecular Cell Biology*, *6*(10), 777–788. https://doi.org/10.1038/nrm1739

Guo, R., Zheng, L., Park, J. W., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J., Wang, Z., Yang, P., Wu, F., Hui, J., Fu, X., Shi, X., Shi, Y. G., Xing, Y., Lan, F., & Shi, Y. (2014). BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Molecular Cell*, *56*(2), 298–310. https://doi.org/10.1016/j.molcel.2014.08.022

Hayakawa, T., Ohtani, Y., Hayakawa, N., Shinmyozu, K., Saito, M., Ishikawa, F., & Nakayama, J. (2007). RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, *12*(6), 811–826. https://doi.org/10.1111/j.1365-2443.2007.01089.x

Henikoff, S., & Smith, M. M. (2015). Histone variants and epigenetics. *Cold Spring Harbor Perspectives in Biology*, *7*(1), a019364. https://doi.org/10.1101/cshperspect.a019364

Horvath, J. E., Bailey, J. A., Locke, D. P., & Eichler, E. E. (2001). Lessons from the human genome: Transitions between euchromatin and heterochromatin. *Human Molecular Genetics*, *10*(20), 2215–2223. https://doi.org/10.1093/hmg/10.20.2215

Huang, C., & Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophysics Reports*, *4*(4), 170–177. https://doi.org/10.1007/s41048-018-0063-1

Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations [Nature Publishing Group]. In *Experimental and Molecular Medicine* (Vol. 49, Issue 4). https://doi.org/10.1038/emm.2017.11

Jaenisch, R., & Young, R. (2008). Stem Cells, the Molecular Circuitry of

Pluripotency and Nuclear Reprogramming. *Cell*, *132*(4), 567–582. https://doi.org/10.1016/j.cell.2008.01.015

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*, *12*(4), 828–863. https://doi.org/10.1038/nprot.2017.016

Kawasaki, H., Mizuseki, K., Nishikawa, S., Kaneko, S., Kuwana, Y., Nakanishi, S., Nishikawa, S., & Sasai, Y. (2000). *Induction of Midbrain Dopaminergic Neurons from ES Cells by Stromal Cell – Derived Inducing Activity*. *28*, 31–40.

Klose, R. J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., & Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature*, *442*(7100), 312–316. https://doi.org/10.1038/nature04853

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., & Greenblatt, J. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Molecular and Cellular Biology*, *23*(12), 4207–4218. https://doi.org/10.1128/MCB.23.12.4207-4218.2003

Kuechler, A., Zink, A. M., Wieland, T., Lüdecke, H.-J., Cremer, K., Salviati, L., Magini, P., Najafi, K., Zweier, C., Czeschik, J. C., Aretz, S., Endele, S., Tamburrino, F., Pinato, C., Clementi, M., Gundlach, J., Maylahn, C., Mazzanti, L., Wohlleber, E., … Engels, H. (2015). Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *European Journal of Human Genetics : EJHG*, *23*(6), 753–760. https://doi.org/10.1038/ejhg.2014.165

Kuo, A. J., Cheung, P., Chen, K., Zee, B. M., Kioi, M., Lauring, J., Xi, Y., Park, B. H., Shi, X., Garcia, B. A., Li, W., & Gozani, O. (2011). NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. *Molecular Cell*, *44*(4), 609–620. https://doi.org/10.1016/j.molcel.2011.08.042

Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y., Sugimoto, T., Sonoda, T., Hasegawa, T., Chinen, Y., Tomita Ha, H., Kinoshita, A., Mizuguchi, T., Yoshiura Ki, K., Ohta, T., … Matsumoto, N. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. *Nature Genetics*, *30*(4), 365–366. https://doi.org/10.1038/ng863

Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X., & Chen, E. S. (2022). Structural and functional specificity of H3K36 methylation. *Epigenetics and Chromatin*, *15*(1), 1–20. https://doi.org/10.1186/s13072-022-00446-7

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.

Nature Methods, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larschan, E., Alekseyenko, A. A., Gortchakov, A. A., Peng, S., Li, B., Yang, P., Workman, J. L., Park, P. J., & Kuroda, M. I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Molecular Cell*, *28*(1), 121–133. https://doi.org/10.1016/j.molcel.2007.08.011

Lauria, A., Meng, G., Proserpio, V., Rapelli, S., Maldotti, M., Polignano, I. L., Anselmi, F., Incarnato, D., Krepelova, A., Donna, D., Levra Levron, C., Donati, G., Molineris, I., Neri, F., & Oliviero, S. (2023). DNMT3B supports meso-endoderm differentiation from mouse embryonic stem cells. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-35938-x

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., & Li, G. (2013). The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutS a. *Cell*, *153*(3), 590–600. https://doi.org/10.1016/j.cell.2013.03.025

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, J., Ahn, J. H., & Wang, G. G. (2019). Understanding histone H3 lysine 36 methylation and its deregulation in disease. In *Cellular and Molecular Life Sciences* (Vol. 76, Issue 15, pp. 2899–2916). Birkhauser Verlag AG. https://doi.org/10.1007/s00018-019-03144-y

Li, L., Liu, C., Biechele, S., Zhu, Q., Song, L., Lanner, F., Jing, N., & Rossant, J. (2013). Location of transient ectodermal progenitor potential in mouse development. *Development*, *140*(22), 4533–4543. https://doi.org/10.1242/dev.092866

Li, M., Hou, Y., Zhang, Z., Zhang, B., Huang, T., Sun, A., Shao, G., & Lin, Q. (2023). Structure, activity and function of the lysine methyltransferase SETD5. *Frontiers in Endocrinology*, *14*(February), 1–6. https://doi.org/10.3389/fendo.2023.1089527

Li, W., Köster, J., Xu, H., Chen, C. H., Xiao, T., Liu, J. S., Brown, M., & Liu, X. S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biology*, *16*(1), 1–13. https://doi.org/10.1186/s13059-015-0843-6

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., Irizarry, R. A., Liu, J. S., Brown, M., & Liu, X. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*, *15*(12), 554. https://doi.org/10.1186/preaccept-1316450832143458

Li, Y., Trojer, P., Xu, C.-F., Cheung, P., Kuo, A., Drury, W. J. 3rd, Qiao, Q., Neubert,

T. A., Xu, R.-M., Gozani, O., & Reinberg, D. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *The Journal of Biological Chemistry*, *284*(49), 34283–34295. https://doi.org/10.1074/jbc.M109.034462

Loebel, D. A. F., Watson, C. M., De Young, R. A., & Tam, P. P. L. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Developmental Biology*, *264*(1), 1–14. https://doi.org/10.1016/S0012-1606(03)00390-7

Lolas, M., Valenzuela, P. D. T., Tjian, R., & Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4478–4483. https://doi.org/10.1073/pnas.1402612111

Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C., & Carpenter, P. B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3with RNApolymerase II function. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39), 16952–16957. https://doi.org/10.1073/pnas.1002653107

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-, O. M., & Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. *Science*, *February 2010*, 996–1001. https://doi.org/10.1126/science.1184208

Marinho, L. S. es R., Rissi, V. B., Lindquist, A. G., Seneda, M. M., & Bordignon, V. (2017). Acetylation and methylation profiles of H3K27 in porcine embryos cultured in vitro. *Zygote*, *25*(5), 575–582. https://doi.org/10.1017/S0967199417000405

Martin, M. (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBO Journal*, *17*(1), 10–12.

Martinez-Garcia, E., Popovic, R., Min, D.-J., Sweet, S. M. M., Thomas, P. M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L. M., Levens, D. L., Kelleher, N. L., & Licht, J. D. (2011). The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood*, *117*(1), 211–220. https://doi.org/10.1182/blood-2010-07-298349

Martinez, A. M., & Cavalli, G. (2006). The role of polycomb group proteins in cell cycle regulation during development. *Cell Cycle*, *5*(11), 1189–1197. https://doi.org/10.4161/cc.5.11.2781

Mccauley, B. S., & Dang, W. (2022). Loosening chromatin and dysregulated transcription : a perspective on cryptic transcription during mammalian aging. *Briefings in Functional Genomics*, *21*(1), 56–61.

Moore, S. M., Seidman, J. S., Ellegood, J., Gao, R., Savchenko, A., Troutman, T. D., Abe, Y., Stender, J., Lee, D., Wang, S., Voytek, B., Lerch, J. P., Suh, H., Glass, C. K., & Muotri, A. R. (2019). Setd5 haploinsufficiency alters neuronal network connectivity and leads to autistic-like behaviors in mice. *Translational Psychiatry*, *9*(1), 24. https://doi.org/10.1038/s41398-018-0344-y

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77. https://doi.org/10.1038/nature21373

Niakan, K. K., Ji, H., Maehr, R., Vokes, S. A., Rodolfa, K. T., Sherwood, R. I., Yamaki, M., Dimos, J. T., Chen, A. E., Melton, D. A., McMahon, A. P., & Eggan, K. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes and Development*, *24*(3), 312–326. https://doi.org/10.1101/gad.1833510

Nichols, J., & Smith, A. (2011). The origin and identity of embryonic stem cells. *Development*, *138*(1), 3–8. https://doi.org/10.1242/dev.050831

Nishiyama, A., Mulholland, C. B., Bultmann, S., Kori, S., Endo, A., Saeki, Y., Qin, W., Trummer, C., Chiba, Y., Yokoyama, H., Kumamoto, S., Kawakami, T., Hojo, H., Nagae, G., Aburatani, H., Tanaka, K., Arita, K., Leonhardt, H., & Nakanishi, M. (2020). Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. *Nature Communications*, *11*(1), 1–7. https://doi.org/10.1038/s41467-020-15006-4

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

Osipovich, A. B., Gangula, R., Vianna, P. G., & Magnuson, M. A. (2016). Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation. *Development (Cambridge)*, *143*(24), 4595–4607. https://doi.org/10.1242/dev.141465

Pai, C. C., Deegan, R. S., Subramanian, L., Gal, C., Sarkar, S., Blaikley, E. J., Walker, C., Hulme, L., Bernhard, E., Codlin, S., Bähler, J., Allshire, R., Whitehall, S., & Humphrey, T. C. (2014). A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nature Communications*, *5*(May). https://doi.org/10.1038/ncomms5091

Pascolini, G., Gnazzo, M., Novelli, A., & Grammatico, P. (2022). Clinical refinement of the SETD5-associated phenotype in a child displaying novel features and KBG syndrome-like appearance. In *American journal of medical genetics. Part A* (Vol. 188, Issue 5, pp. 1623–1625). https://doi.org/10.1002/ajmg.a.62679

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J., & Rathjen, P. D. (2002). Transient pluripotent cell populations during primitive ectoderm formation : correlation of in vivo and in vitro pluripotent cell development . *Journal of Cell Science*, *115*(2), 329–339.

Peterson, C. L., & Laniel, M. A. (2004). Histones and histone modifications. *Current Biology : CB*, *14*(14), 546–551. https://doi.org/10.1016/j.cub.2004.07.007

Pevny, L., & Placzek, M. (2005). SOX genes and neural progenitor identity. *Current Opinion in Neurobiology*, *15*(1), 7–13. https://doi.org/10.1016/j.conb.2005.01.016

Pfister, S. X., Ahrabi, S., Zalmas, L. P., Sarkar, S., Aymard, F., Bachrati, C. Z., Helleday, T., Legube, G., LaThangue, N. B., Porter, A. C. G., & Humphrey, T. C. (2014). SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell Reports*, *7*(6), 2006–2018. https://doi.org/10.1016/j.celrep.2014.05.026

Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R., & Bickmore, W. A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genetics*, *8*(5). https://doi.org/10.1371/journal.pgen.1002717

Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., & Xu, R.-M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *The Journal of Biological Chemistry*, *286*(10), 8361–8368. https://doi.org/10.1074/jbc.M110.204115

Raj, B., & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*, *87*(1), 14–27. https://doi.org/10.1016/j.neuron.2015.05.004

Rawlins, L. E., Stals, K. L., Eason, J. D., & Turnpenny, P. D. (2017). De novo SETD5 nonsense mutation associated with diaphragmatic hernia and severe cerebral cortical dysplasia. *Clinical Dysmorphology*, *26*(2), 95–97. https://doi.org/10.1097/MCD.0000000000000144

Richter, G. H. S., Plehm, S., Fasan, A., Rössler, S., Unland, R., Bennani-Baiti, I. M., Hotfilder, M., Löwel, D., Von Luettichau, I., Mossbrugger, I., Quintanilla-Martinez, L., Kovar, H., Staege, M. S., Müller-Tidow, C., & Burdach, S. (2009). EZH2 is a mediator of EWS/FLI1 driven tumor growth and metastasis blocking endothelial and neuro-ectodermal differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5324–5329. https://doi.org/10.1073/pnas.0810759106

Ritchie, F. D., & Lizarraga, S. B. (2023). The role of histone methyltransferases in neurocognitive disorders associated with brain size abnormalities. *Frontiers in Neuroscience*, *17*, 989109. https://doi.org/10.3389/fnins.2023.989109

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

Saksouk, N., Simboeck, E., & Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics and Chromatin*, *8*(1), 1–17. https://doi.org/10.1186/1756-8935-8-3

Sankaran, S. M., Wilkinson, A. W., Elias, J. E., & Gozani, O. (2016). A PWWP Domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated lys-36 of histone H3 and regulates NSD2 function at chromatin. *Journal of Biological Chemistry*, *291*(16), 8465–8474. https://doi.org/10.1074/jbc.M116.720748

Sessa, A., Fagnocchi, L., Mastrototaro, G., Massimino, L., Zaghi, M., Indrigo, M., Cattaneo, S., Martini, D., Gabellini, C., Pucci, C., Fasciani, A., Belli, R., Taverna, S., Andreazzoli, M., Zippo, A., & Broccoli, V. (2019). SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron*, *104*(2), 271-289.e13. https://doi.org/10.1016/j.neuron.2019.07.013

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y., Huang, K., Liu, Y., Wang, X., Chen, Q., Zhang, J., Shang, B., Li, S., Shi, X., Liao, B., … Pan, G. (2017). PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00668-4

Shechter, D., Dormann, H. L., Allis, C. D., & Hake, S. B. (2007). Extraction, purification and analysis of histones. *Nature Protocols*. https://doi.org/10.1038/nprot.2007.202

Shin, J., Ming, G., & Song, H. (2014). Decoding neural transcriptomes and epigenomes via high-throughput sequencing. *Nature Neuroscience*, *17*(11), 1463–1475. https://doi.org/10.1038/nn.3814

Shparberg, R., Glover, H., & Morris, M. B. (2019). Modelling mammalian commitment to the neural lineage using embryos and embryonic stem cells. *Frontiers in Physiology*, *10*, 705. https://doi.org/10.3389/fphys.2019.00705

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *ELife*. https://doi.org/10.7554/eLife.21856

Speir, M. L., Bhaduri, A., Markov, N. S., Moreno, P., Nowakowski, T. J., Papatheodorou, I., Pollen, A. A., Raney, B. J., Seninge, L., Kent, W. J., & Haeussler, M. (2021). UCSC Cell Browser: visualize your single-cell data. *Bioinformatics*, *37*(23), 4578–4580. https://doi.org/10.1093/bioinformatics/btab503

Stambrook, P. J., & Tichy, E. D. (2010). Preservation of Genomic Integrity in Mouse Embryonic Stem Cells. In E. Meshorer & K. Plath (Eds.), *The Cell Biology of Stem Cells* (pp. 59–75). Springer US. https://doi.org/10.1007/978-1-4419-7037-4_5

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. https://doi.org/10.1038/47412

Streubel, G., Watson, A., Jammula, S. G., Scelfo, A., Fitzpatrick, D. J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G. L., Dillon, E., Wynne, K., Pasini, D., Krogan, N. J., Bracken, A. P., & Cagney, G. (2018). The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Molecular Cell*, *70*(2), 371-379.e5. https://doi.org/10.1016/j.molcel.2018.02.027

Sun, X. J., Wei, J., Wu, X. Y., Hu, M., Wang, L., Wang, H. H., Zhang, Q. H., Chen, S. J., Huang, Q. H., & Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *Journal of Biological Chemistry*, *280*(42), 35261–35271. https://doi.org/10.1074/jbc.M504012200

Suter, D. M., Tirefort, D., Julien, S., & Krause, K.-H. (2009). A Sox1 to Pax6 Switch Drives Neuroectoderm to Radial Glia Progression During Differentiation of Mouse Embryonic Stem Cells. *Stem Cells*, *27*(1), 49–58. https://doi.org/10.1634/stemcells.2008-0319

Szczałuba, K., Brzezinska, M., Kot, J., Rydzanicz, M., Walczak, A., Stawiński, P., Werner, B., & Płoski, R. (2016). SETD5 loss-of-function mutation as a likely cause of a familial syndromic intellectual disability with variable phenotypic expression. *American Journal of Medical Genetics. Part A*, *170*(9), 2322–2327. https://doi.org/10.1002/ajmg.a.37832

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(Database issue), D447-52. https://doi.org/10.1093/nar/gku1003

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *2*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., Baralle, D., Dabir, T., Gener, B., Goudie, D., Homfray, T., Kumar, A., Pilz, D. T., Selicorni, A., Temple, I. K., … Rahman, N. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nature Genetics*, *46*(4), 385–388. https://doi.org/10.1038/ng.2917

Topchu, I., Pangeni, R. P., Bychkov, I., Miller, S. A., Izumchenko, E., Yu, J., Golemis, E., Karanicolas, J., & Boumber, Y. (2022). The role of NSD1, NSD2, and NSD3 histone methyltransferases in solid tumors. *Cellular and Molecular Life Sciences*, *79*(6), 1–19. https://doi.org/10.1007/s00018-022-04321-2

Tsukada, Y. I., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., & Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, *439*(7078), 811–816. https://doi.org/10.1038/nature04433

Turner, B. M. (1991). Histone acetylation and control of gene expression. *Journal of Cell Science*, *99*(1), 13–20. https://doi.org/10.1242/jcs.99.1.13

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

Vossaert, L., Meert, P., Scheerlinck, E., Glibert, P., Van Roy, N., Heindryckx, B., De Sutter, P., Dhaenens, M., & Deforce, D. (2014). Identification of histone H3 clipping activity in human embryonic stem cells. *Stem Cell Research*, *13*(1), 123–134. https://doi.org/10.1016/j.scr.2014.05.002

Waggoner, D. J., Raca, G., Welch, K., Dempsey, M., Anderes, E., Ostrovnaya, I., Alkhateeb, A., Kamimura, J., Matsumoto, N., Schaeffer, G. B., Martin, C. L., & Das, S. (2005). NSD1 analysis for Sotos syndrome: Insights and perspectives from the clinical laboratory. *Genetics in Medicine*, *7*(8), 524–533. https://doi.org/10.1097/01.GIM.0000178503.15559.d3

Wagner, E. J., & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nature Reviews. Molecular Cell Biology*, *13*(2), 115–126. https://doi.org/10.1038/nrm3274

Walker, E., Chang, W. Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N. J., Reiter, J. F., & Stanford, W. L. (2010). Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell*, *6*(2), 153–166. https://doi.org/10.1016/j.stem.2009.12.014

Wamaitha, S. E., del Valle, I., Cho, L. T. Y., Wei, Y., Fogarty, N. M. E., Blakeley, P., Sherwood, R. I., Ji, H., & Niakan, K. K. (2015). Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes and Development*, *29*(12), 1239–1255. https://doi.org/10.1101/gad.257071.114

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., & Ren, Y. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature*. https://doi.org/10.1038/nature13045

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

Wiles, M. V, & Johansson, B. M. (1999). Embryonic Stem Cell Development in a Chemically Defined Medium. *Experimental Cell Research*, *247*, 241–248.

Wray, J., Kalkan, T., & Smith, A. G. (2010). Revolutionizing Drug Discovery with Stem Cell Technology The ground state of pluripotency. *Biochemical Society Transactions*, *38*(4), 1027–1032. https://doi.org/10.1042/BST0381027

Xie, L., Pelz, C., Wang, W., Bashar, A., Varlamova, O., Shadle, S., & Impey, S. (2011). KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO Journal*, *30*(8), 1473–1484. https://doi.org/10.1038/emboj.2011.91

Xu, Y., Zhao, W., Olson, S. D., Prabhakara, K. S., & Zhou, X. (2018). Alternative splicing links histone modifications to stem cell fate decision. *Genome Biology*, *19*(1), 1–21. https://doi.org/10.1186/s13059-018-1512-3

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

Ying, Q., Wray, J., Nichols, J., Batlle-morera, L., Doble, B., Woodgett, J., Cohen, P., & Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature*, *453*(May), 519–524. https://doi.org/10.1038/nature06968

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., & Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *Journal of Biological Chemistry*, *286*(10), 7983–7989. https://doi.org/10.1074/jbc.M110.194027

Zaghi, M., Broccoli, V., & Sessa, A. (2020). H3K36 Methylation in Neural Development and Associated Diseases. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.01291

APPENDICES

A. Media Recipes for Cell Culture

2i Medium (Serum-free Medium, 500 mL): 250 mL Neurobasal Medium (Cat no.: 21103049, Gibco), 250 mL DMEM/F-12 (Gibco, 11320074), 2.5 mL N-2 Supplement (100X) (Gibco, 17502048), 5 mL B-27™ Supplement (50X), serum free (Gibco, 17504044), 5 mL Pen/Strep (Gibco, 15140122), 2.5 mL 10% BSA, 5 mL GlutaMAX I (Gibco, 35050061), 6.5 µL MTG (1-Thioglycerol) (Sigma, M6145-25ML). The medium is supplemented with 3 µM CHIR-99021 (Selleckchem, S2924), 1 µM PD0325901 (Selleckchem, S1036) and 1% Leukemia Inhibitory Factor (LIF) (Millipore, ESG1107).

N2B27 Medium (Neuroectoderm Differentiation Medium, 500 mL): 250 mL Neurobasal Medium (Cat no.: 21103049, Gibco), 250 mL DMEM/F-12 (Gibco, 11320074), 5 mL N-2 MAX Media Supplement (100X) (R&D Systems, AR009), 10 mL B-27™ Supplement (50X), serum free (Gibco, 17504044), 250 µL 10% BSA, 19.5 µL MTG (1-Thioglycerol) (Sigma, M6145-25ML). MTG, a reducing agent, was added to decrease the negative effects of oxidative stress.

B. Solution Recipes

10% BSA (w/v): Prepared by dissolving 5 g Bovine Serum Albumin (BSA) (Sigma, A3311-50G) in 50 mL Dulbecco's Phosphate Buffered Saline (DPBS) (Biological Industries, 02-023-1A). Filter-sterilized and stored at 4°C.

TEB (Triton Extraction Buffer, 100 mL): 500 µL Triton X-100 (Amresco, 0694-1L) and 0.02 g NaN3 in 100 mL 1X PBS to prepare the stock buffer. 2 mM phenylmethylsulfonyl fluoride (PMSF) (Selleck Chemicals, S3025) is added to aliquot before use.

15% polyacrylamide gel (resolving): 2.75 mL Acrylamide/Bis Solution (37.5:1) (Serva, SE1068801), 1.4 mL 1M Tris-HCl, pH 8.8, 1.35 mL distilled water are mixed. 5.5 µL TEMED (AMRESCO, 0761-25ml) and 55 µL 10% APS (Serva, AI0486-25) are added fresh and mixed well before pouring the gel.

4% polyacrylamide gel (stacking): 194 µL Acrylamide/Bis Solution (37.5:1) (Serva, SE1068801), 372 µL 1M Tris-HCl, pH 6.8, 920 µL distilled water are mixed. 1.5 µL TEMED (AMRESCO, 0761-25ml) and 15 µL 10% APS (Serva, AI0486-25) are added fresh and mixed well before pouring the gel.

Transfer Buffer: 2.88 g Tris-base (Serva, 37190.02) and 1.475 g Glycine (Serva, SE2339004) are dissolved in 200 mL distilled water. After 100 mL Methanol (Merck, 106009) is added, buffer is completed to 500 mL with distilled water. It is stored at 4°C.

Strip buffer: 3 g Glycine and 0.2 g SDS (Serva, SE2076503) are dissolved in distilled water. After 2 mL Tween-20 (AMRESCO, AI0777-1) is added, pH is adjusted to 2.2 and completed to 200 mL with distilled water. It is stored at room temperature and an aliquot is warmed before use.

TBS: 20X TBS is obtained by dissolving 48.4 g Tris-base (Serva, 37190.02) and 160 g NaCl (Merck, 1.06404.1000) in 1 L distilled water. It is diluted to 1X with distilled water before use.

TBS-T: Prepared by mixing 10 mL %10 Tween-20 (VWR, 0777-1L) stock and 50 mL 20X TBS in 1 L distilled water.

C. Flow Cytometry Analysis for Sox1-GFP Neuroectoderm Differentiation

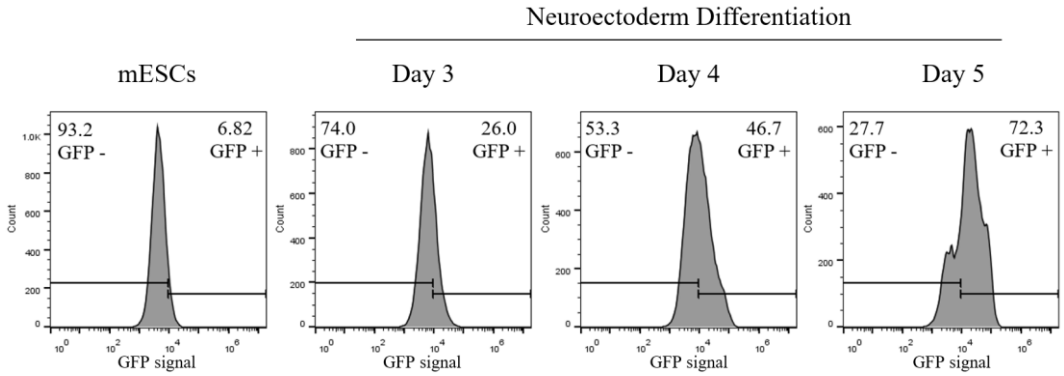

Figure S. 1. Flow cytometry results depicting the Green Fluorescent Protein (GFP) signal from the neuroectoderm differentiation time course of the Sox1-GFP cell line. The x-axis represents the intensity of the GFP signal, while the y-axis shows the corresponding cell count. The gating was set based on the GFP signal in the undifferentiated mESC state, and was applied consistently for subsequent differentiation days. This gating was used to determine the percentage of GFP-positive cells in the population, providing an indication of Sox1 expression and neuroectoderm differentiation.

D. Details About Antibodies Utilized in Western Blotting

Table 5. Information about the antibodies used in this study for Western blot experiments.

Antibody	*Host*	*Cat. No.*	*Brand*	*Dilution*
Anti-H3K36me2	Rabbit	ab176921	Abcam	1:5000
Anti-H3K36me3	Rabbit	d5a7	CST	1:1000
Anti-H3	Mouse	sc-517576	Santa Cruz	1:1000
Anti-rabbit IgG H&L (HRP)	Goat	ab97051	Abcam	1:5000
Anti-mouse IgG H&L (HRP)	Goat	ab97023	Abcam	1:5000

E. RT-qPCR Primer Sequences

Table 6. Primer sequences used in this study for RT-qPCR analyses.

Transcript	*Primer orientation*	*Primer Sequence (5' – 3')*
β-actin	Forward	ATGAAGATCCTGACCGAGCG
	Reverse	TACTTGCGCTCAGGAGGAGC
Nanog	Forward	ATGAAGTGCAAGCGGTGGCAGAAA
	Reverse	CCTGGTGGAGTCACAGAGTAGTTC
Oct4 (Pou5f1)	Forward	CTGAGGGCCAGGCAGGAGGACGAG
	Reverse	CTGTAGGGAGGGCTTCGGGCACTT
Pax6	Forward	AGTGAATGGGCGGAGTTATG
	Reverse	ACTTGGACGGGAACTGACAC
Sox1	Forward	CCTCGGATCTCTGGTCAAGT
	Reverse	GCAGGTACATGCTGATCATCT
N-cadherin	Forward	CAGGGTGGACGTCATTGTAG
	Reverse	AGGGTCTCCACCACTGATTC
Nsd1	Forward	ATTCTCGGGGCCGTCCAATA
	Reverse	CTGTGCCTGCATCAACCTCA
Nsd2	Forward	TGCCAAAAAGGAGTACGTGTG
	Reverse	CTTCGGGAAAGTCCAAGGCAG
Nsd3	Forward	AGGTGCCAGCGAGATTTCAG
	Reverse	GCAGTAGCTGACGGGCTATC
Setmar	Forward	TGGAGAACTTGCCTGTGAGC
	Reverse	GCAACATGATCAGGAGTATACTGG
Smyd2	Forward	GGAGGGCCAAACACTACAAA
	Reverse	TGAGGGAGTACACGGGGTAG
Setd2	Forward	TAAGGGCTGCTAAGGATCTTCC
	Reverse	GTGGCATCTATTATCTCGTCATTTT

Table 5 (continued)

Transcript	*Primer orientation*	*Primer Sequence (5' – 3')*
Setd3	Forward	CACCGCCATTGATACTATGTAGCCG
	Reverse	ACTGTGGATCACAAACTCTGC
Setd5	Forward	GAGCCATTGAGCTCAGCACTC
	Reverse	GGACGAACTCTGCTGAAGGAG
Ash1l	Forward	AGTGAAAGGGCAATACAGTCGT
	Reverse	TCTGGAAGGAACTCCATTCACT
Kdm2a	Forward	CAACAGCGATCCCAAGTTAGC
	Reverse	TGGCCGAGTGGGGAATTTAAG
Kdm2b	Forward	AGCAGACAGAAGCCACCAAT
	Reverse	AGGTGCCTCCAAAGTCAATG
Kdm4a	Forward	AAGAAAGCCATGACCGTTCGTG
	Reverse	AAATTCACTGTATCGCGGGGTG
Kdm4b	Forward	TACTGTACCCCACGCCATCA
	Reverse	TCAGATTGCCGATGTTCCAC
Kdm4c	Forward	TGTGAAGCAGCAGGTAGCGAGT
	Reverse	GTCTGCCAAAGGTGGATGAGAG
Kdm8	Forward	AAAGTTGTCGCAGTCCTCCA
	Reverse	CGTCACACTTTGCCTTCTTGG
Phf1	Forward	TGAGAAGTGTCGCCATGCTTA
	Reverse	CATAGGGACCTTTCTTCAGTGC
Phf19	Forward	CTATGCAGGTGGACAGACGG
	Reverse	TCAACTGCGATGGGGATGTG
Dnmt3a	Forward	CCTACTACATCAGCAAACGGAAAC
	Reverse	GTTCTCTTCCACAGCATTCATTACT
Dnmt3b	Forward	ACCAGTGGTTAATAAGTCGAAGGT
	Reverse	GACTCAGAAGCAGCAGAGTCATT

Table 5 (continued)

Transcript	*Primer orientation*	*Primer Sequence (5' – 3')*
Msh6	Forward	CGGCGCTTGTTCTGTAACTTCG
	Reverse	GCGACACATAGGACCATCTCCT
Psip1	Forward	ATGGTGACTGCAGCAACCGCTT
	Reverse	TCACCGTCTGAAGGACAAGGCT
Zmynd11	Forward	GGCAGAATCCAGTATCTCCTCC
	Reverse	CTCACTGCTTCCGTTTCTGGCT
Mrg15	Forward	AGTGAGACACCTCAGCCTCCTA
	Reverse	CAGTCATCCACAAGCCAGGGTT

F. Correlation Analysis

Table 7. The correlation between H3K36me2 levels, as quantified by Western Blot, and the expression of selected genes, as quantified by RT-qPCR, during the differentiation of neuroectoderm (comparing day 5 to day 0). Genes with a statistically significant correlation are highlighted in grey. A Spearman's rho value less than zero ($\rho<0$) indicates a negative correlation, while a value greater than zero ($\rho>0$) indicates a positive correlation between the expression of the selected gene and H3K36me2 levels.

Transcript	*Spearman's Test*	
	Spearman's ρ	*p-value*
Nsd1	-0.92857	0.002232
Nsd2	0.761905	0.03676
Setd5	-0.78571	0.02793
Kdm4c	-0.90476	0.004563
Dnmt3a	-0.59524	0.1323
Dnmt3b	-0.38095	0.3599
Msh6	-0.85714	0.01071
Mrg15	-0.66667	0.08309
Zmynd11	-0.09759	0.8182

G. Relative Expression Levels of Pluripotency and Early Endoderm Markers During the Course of Endoderm Differentiation

Table 8. Relative expression levels of pluripotency (*Oct4, Sox2, Nanog, Klf4*) and endoderm differentiation (*Foxa2, Gsc, Sox17,* Gata6) related genes during endoderm differentiation.

Transcript	*FC Ratio*	*Log(FC)*	*FDR*
Oct4	ESC/Day 3	0.07	0.607
	ESC/Day 4	1.65	0.000
Sox2	ESC/Day 3	3.16	0.000
	ESC/Day 4	3.87	0.000
Nanog	ESC/Day 3	0.23	0.025
	ESC/Day 4	2.00	0.000
Klf4	ESC/Day 3	6.44	0.000
	ESC/Day 4	5.77	0.000
Foxa2	ESC/Day 3	-5.41	0.000
	ESC/Day 4	-5.44	0.000
Gsc	ESC/Day 3	-4.77	0.000
	ESC/Day 4	-5.81	0.000
Sox17	ESC/Day 3	-4.33	0.000
	ESC/Day 4	-7.86	0.000
Gata6	ESC/Day 3	-2.85	0.000
	ESC/Day 4	-5.94	0.000

H. NGS Analysis After Library Amplification

Table 9. Assessment of library amplification success post-NGS analysis.

Parameter	*Result*
Percentage of guides that matched perfectly	74.8
Percentage of undetected guides	0
Skew ratio of top 10% to bottom 10%	3.32

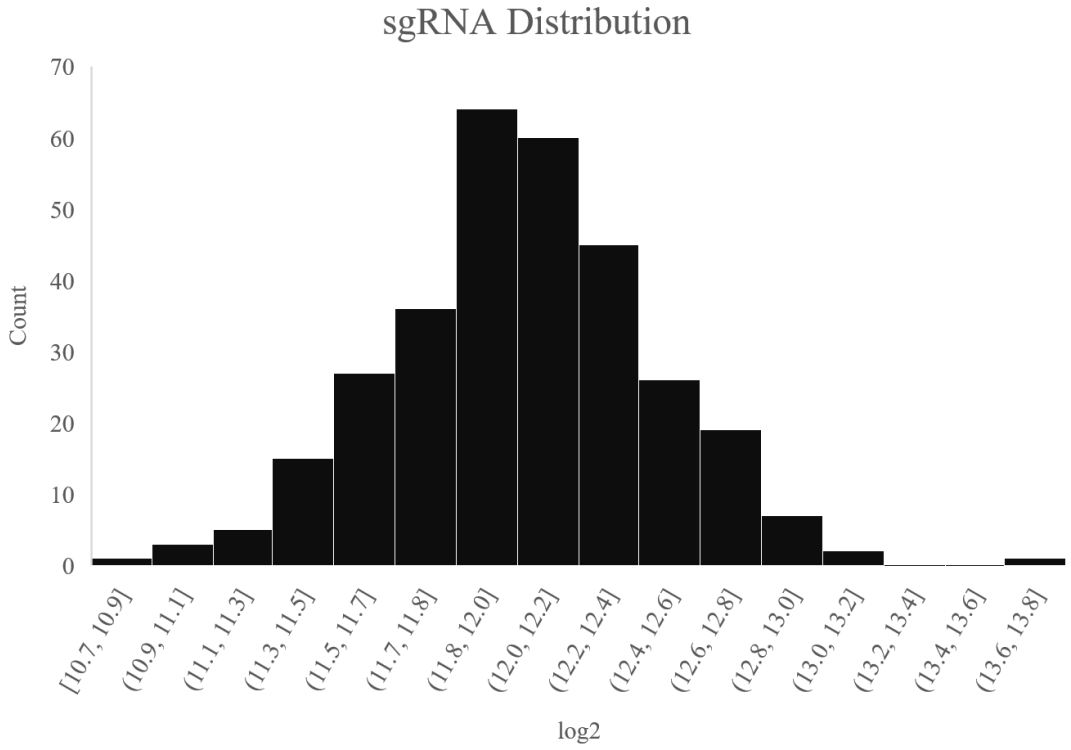

Figure S. 2 Bar graph illustrating the distribution of sgRNAs after library amplification. The x-axis categorizes sgRNA groups based on their log-transformed counts, while the y-axis quantifies the number of sgRNAs within each group. The graph demonstrates a fairly equal distribution of sgRNAs without any visible skew.

I. NGS Analysis and Mapping Success

Table 10. The sequencing data and mapping results of mESC state (day0) and neuroectoderm differentiated state (day5) samples of three biological replicates.

Sample	*Reads*	*Mapped*	*Percentage*
day0_rep1	4160140	3858826	92.76
day0_rep2	3851497	3593509	93.3
day0_rep3	4334594	4037778	93.15
day5_rep1	3944003	3625336	91.92
day5_rep2	5917053	5494884	92.87
day5_rep3	5269186	4824229	91.56

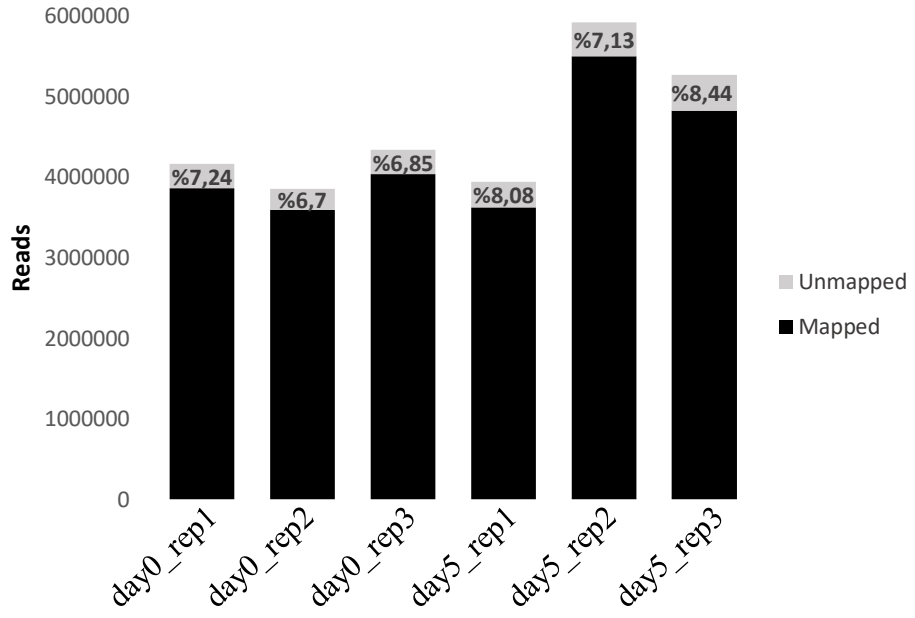

Figure S. 3. Bar graph representing the comparison between the percentages of mapped vs unmapped reads for each sample.

Table 11. Analysis of sgRNA distribution for each sample from the mESC state (day0) and neuroectoderm differentiated state (day5) over three biological replicates.

Sample	*Total sgRNAs*	*Zero Counts*	*Gini Index*
day0_rep1	1075	0	0.07
day0_rep2	1075	0	0.05
day0_rep3	1075	0	0.04
day5_rep1	1075	1	0.11
day5_rep2	1075	1	0.09
day5_rep3	1075	0	0.08