IDENTIFICATION OF PROTEIN PARTNERS OF SETD3 IN MOUSE EMBRYONIC STEM CELLS

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

BY

GÖZDE GÜVEN

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

JANUARY 2021

Approval of the thesis:

IDENTIFICATION OF PROTEIN PARTNERS OF SETD3 IN MOUSE EMBRYONIC STEM CELLS

submitted by **GÖZDE GÜVEN** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Molecular Biology and Genetics, Middle East Technical University** by,

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

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

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

Examining Committee Members:

Assoc. Prof. Erkan Kiriş
Molecular Biology and Genetics, METU _____

Assist. Prof. Dr. Nihal Terzi Çizmecioğlu
Biology, METU _____

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

Date: 26.01.2021

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 : Gözde Güven

Signature :

ABSTRACT

IDENTIFICATION OF PROTEIN PARTNERS OF SETD3 IN MOUSE EMBRYONIC STEM CELLS

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

January 2021, 60 pages

Embryonic development is a highly coordinated process that contains rapid cell divisions. During this time, the cells also start differentiation, and at the end of development, the cells become fully differentiated. This differentiation commitment is regulated by epigenetic factors like histone modifiers, chromatin regulators. SETD3 is one of the histone methyltransferases that govern this pluripotency-differentiation balance. We have previously found that SETD3 is crucial for mesendoderm differentiation of mESCs. However, the partners of Setd3 in the nucleus is not well known. We have performed IP-MS and analyzed the candidate proteins deeply. Then, we performed biochemical assays (co-IP and PLA) and showed BRD2 and TFCP2L1 proteins are the protein partners of SETD3 methyltransferase.

Keywords: Embryonic Stem Cells, Pluripotency, Histone Methyltransferase

ÖZ

FARE EMBRİYONİK KÖK HÜCRELERDE SETD3 PROTEİNİYLE ETKİLEŞEN PROTEİNLERİN PROTEOMİK AÇIDAN BELİRLENMESİ

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

Ocak 2021, 60 sayfa

Embriyonik gelişim hızlı hücre bölünmelerinin bulunduğu, son derece planlı ve düzenli bir süreçtir. Bu süreçte embriyo farklılaşmaya başlar ve gelişimin sonunda hücreler tamamen farklılaşmıştır. Farklılaşmaya karar aşaması çeşitli histon ve kromatin düzenleyici kompleksler gibi birçok epigenetik faktör tarafından düzenlenir. SETD3, bu pluripotent-farklılaşma dengesinin yönetilmesinde görevli histon metiltransferazlardan biridir. Daha önceki çalışmalarımızda, SETD3'ün fare EKH'lerinin mezendoderm tabakasına farklılaşmasında gerekli bir protein olduğunu belirledik. Buna rağmen, SETD3'ün hücre çekirdeğindeki diğer protein partnerleri tam olarak bilinmemektedir. IP-MS deneyimizle belirlediğimiz aday proteinler için derinlemesine bir analiz yaptık. Ardından, biyokimyasal deneyler ile BRD2 ve TFCP2L1 proteinlerinin SETD3 metil transferazının protein partnerleri olduğunu gösterdik.

Anahtar Kelimeler: Embriyonik Kök Hücre, Pluripotent, Histon Metiltransferaz

To my devoted family

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my supervisor Assist. Prof. Dr. Nihal Terzi Çizmecioğlu for the encouragement, supports, and guidance. I have learned a lot about the scientific approach and critical thinking. The door of Assist. Prof. Dr. Terzi Çizmecioğlu's office was always open whenever I ran into a trouble or had a question about everything. I am grateful for all the things she taught me to make me a better scientist.

I would like to thank all my jury members; Assist. Prof. Dr. Bahar Değirmenci Uzun and Assoc. Prof. Erkan Kiriş for their insightful comments and encouragement.

I also would like to thank my first and almost only lab mate Ezgi Gül Keskin. She taught me teamwork and helped me from my first day of lab until the last dot of this thesis.

I would like to thank my dearest friend Hoşnaz Tuğral for her endless support, encouragement, and friendships. I cannot thank enough for always being there for me; it is a privilege to have you in my life. I would like to thank my lifelong friend Mehtap Yurdakul for her realistic approaches to every situation, even laziness..

I would like to thank Altar Özbıyık who made me an analog-photography lover and for helping me. Also, I am deeply thankful my current lab mates Emre Balbaşı who cheer me up with his songs, coffee&chocolate breaks, Dersu Sezginmert, Ceren Alganatay and Zeynep Ilgın Kılıç for making lab days so pretty and making me upset when I had to spend time at home.

I would like to thank Burak Kızıl for his support and İsmail Güderer for his knowledge about literally everything. I am also deeply grateful to Lab B59 family; Esin G. Seza, Sinem Ulusan, Aydan Torun, Çağdaş Ermiş for their support and for making 'Devrim Days' so entertaining.

I would like to thank specifically to Erson Lab, Muyan Lab, Gürsel Lab and Kiriş Lab for sharing resources. I am very grateful to each member of "Alt Kat Zombileri".

Last but not least, I would like to thank my father Eyüp Güven, my mother Melahat Güven and my older sister Gökçe Güven and my cat Yamuk 😊 who are my driving force and the only persons who will be there for me in all circumstances. My only and the greatest luck in my life is my family, and I am grateful to everything they did.

The study was supported by TÜBİTAK project 117Z953.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

ALPK3	Alpha-protein kinase 3
ARID2	AT-rich interactive domain-containing protein 2
BCL9	B-cell CLL/lymphoma 9 protein
BET	Bromodomain and Extra-Terminal
BRD2	Bromodomain-containing protein 2
COPS2	COP9 signalosome complex subunit 2
DAXX	Death domain-associated protein 6
FOXM1	Forkhead box protein M1
GLI2	Zinc finger protein GLI2
HAT	Histone Acetyltransferase
HMT	Histone Methyltransferase
HMG20A	High mobility group protein 20A
H3K4me3	H3 tri-methylation on Lysine 4
H3K27me3	H3 tri-methylation on Lysine 27
H3K36me	H3 methylation on Lysine 36
ICC	Immunocytochemistry
ICM	Inner Cell Mass
IP	Immunoprecipitation
LIF	Leukemia Inhibitory Factor

mESC	Mouse Embriyonic Stem Cells
MS	Mass Spectrometry
Oct4	Octamer-binding transcription factor 4
PcG	Polycomb Group Protein
PLA	Proximity Ligation Assay
PRIM1	DNA primase small subunit
qRT-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
RING1	E3 ubiquitin-protein ligase
SET	Su(var)3–9 Enhancer of zeste and Trithorax
SETD3	Actin-histidine N-methyltransferase
TCEA3	Transcription elongation factor A protein 3
TF	Transcription Factor
TFE3	Transcription factor E3
TFCP2L1	Transcription factor CP2-like protein 1
Trx	Trithorax

CHAPTER 1

INTRODUCTION

1.1 Early Development and Embryonic Stem Cells

Embryonic development is a highly regulated process starting from the fertilization of egg and continues with the embryo gastrulation. After fertilization of the egg, the totipotent zygote produces eight cells after three rounds of cell divisions. The genetically identical cells produce ball-like structure, which is called morula. This process is called cleavage because the aim is to increase the cell number. Then, the cells undergo compaction, and morula generates blastocyst which is a hollow sphere-like structure containing. The blastocyst has a fluid-filled cavity called blastocoel and the inner cell mass (ICM). The cells that form outer walls of the blastocyst are called as trophoblast. This is the first cell fate decision of embryonic development. Mouse embryonic stem cells (mESCs) are obtained from ICM (Takaoka and Hamada, 2012). The ESCs are pluripotent, which they can produce all tissue types and they have self-renewal ability which means they divide indefinitely (Gaspar-Maia *et al.*, 2011). The second cell fate decision occurs in the ESCs producing ICM which results in primitive endoderm and epiblast layers. The primitive germ layers; endoderm, mesoderm and ectoderm are produced as a result of a sequential lineage development of these two layers (Paranjpe and Veenstra, 2015). Owing to limitless proliferation and in vitro differentiation ability, mESCs are widely used for early development studies.

1.2 Epigenetic Mechanisms on mESCs

During embryonic development, the stemness vs. differentiation balance is regulated by multiple mechanisms providing different levels of regulation. This regulation is occurred by the chemical modification of DNA itself (DNA methylations) or chromosomal proteins that interact with genomic DNA. DNA methylation is the direct methylation of cysteine residue by DNA methyltransferases. This modification prevents the binding of chromatin regulators on the gene. On the other hand, transcription factors (TFs) and chromatin regulators are some of the chromosomal proteins interacting with genomic DNA. TFs are proteins working with epigenetic regulators to recognize specific consensus sequences on the genes for recruitment. Oct4, Nanog, Sox2 and Klf4 are some of the essential core pluripotency factors to provide and maintain pluripotency in mESCs (Orkin and Hochedlinger, 2011; Young, 2011; Liang and Zhang, 2013; Paranjpe and Veenstra, 2015).

DNA is wrapped around an octameric protein structure called histones to form nucleosomes. The N term of core histones (H2A, H2B, H3, and H4) are exposed to specific modifications that can change the accessibility of DNA. Histone methyltransferases (HMTs), histone acetyltransferases (HAT), demethylases and deacetylases are some of histone-modifying enzymes that provide these modifications. ESCs maintain pluripotency by histone modifications. The pluripotency-related genes generally have H3K4 (histone H3 lysine 4) methylation which is active mark for gene expression. On the other hand, lineage specific genes are repressed by H3K27 (histone H3 lysine 27) methylations (Bannister and Kouzarides, 2011; Young, 2011; Liang and Zhang, 2013).

Chromatin regulators are large class of proteins/complexes that contain histone modifying enzymes, chromatin remodelers etc. Polycomb group (PcG) complex is one of the major chromatin regulators to maintain pluripotency of ESCs. It mediates H3K27me3 of lineage specific genes to repress expression. Besides, Trithorax (Trx) complex provides H3K4me3 mark for activation of pluripotency related genes. Both

of these complexes contains SET (Su(var)3–9, Enhancer of zeste, and Trithorax) domain containing proteins as histone methyltransferases (Orkin and Hochedlinger, 2011; Young, 2011; Paranjpe and Veenstra, 2015). SET domain containing proteins are known as methyltransferase family and methylate different lysine residues on the histones. SeTDB1, Set1A are some SET domain containing methyltransferases necessary protein for ESC pluripotency. (Dillon *et al.*, 2005; Bilodeau *et al.*, 2009).

1.3 SETD3 Histone Methyltransferase

SETD3 was initially identified as a histone lysine methyltransferase acting on H3K4 and H3K36 (Kim *et al.*, 2014). It is also responsible for actin methylation and thus affect actin filament stability and ATP hydrolysis (Kwiatkowski *et al.*, 2018; Diep *et al.*, 2019; Guo *et al.*, 2019).

SETD3 is highly expressed in muscle cell derived tissues like heart, lungs, skeletal muscles etc. It is essential for muscle cell differentiation. Previous studies indicated that SETD3 directly interacts with muscle MyoD which is a protein having role in muscle differentiation and works as transcriptional activator for muscle specific gene expressions in myoblast cells (Eom *et al.*, 2011). Moreover, it methylates FOXM1 protein which is associated with *Vegf* expression and so it has also role in transcriptional regulation of vascular development in rats and human (Cohn *et al.*, 2016; Jiang *et al.*, 2018). It works in the activation of DNA damage induced apoptosis through the expressions of p53 targeted genes in human. As a result of these features, SETD3 is thought as potential marker for cancer prognosis (Cheng *et al.*, 2017; Abaev-schneiderman, Admoni-elisha and Levy, 2019; Nourhan *et al.*, 2020)

1.4 Preliminary Data

In the previous study, shRNA screen was performed in mESCs to identify epigenetic factors that are necessary for mesendoderm differentiation (Terzi Cizmecioglu *et al.*, 2020). It was found that SETD3 is one of the essential epigenetic factors. Then, CRISPR/Cas9 mediated genomic *Setd3* deletion in mESCs was generated and the differentiation ability of wild-type (WT) and setd3Δ mESCs into each germ layer was examined (Figure 1.1).

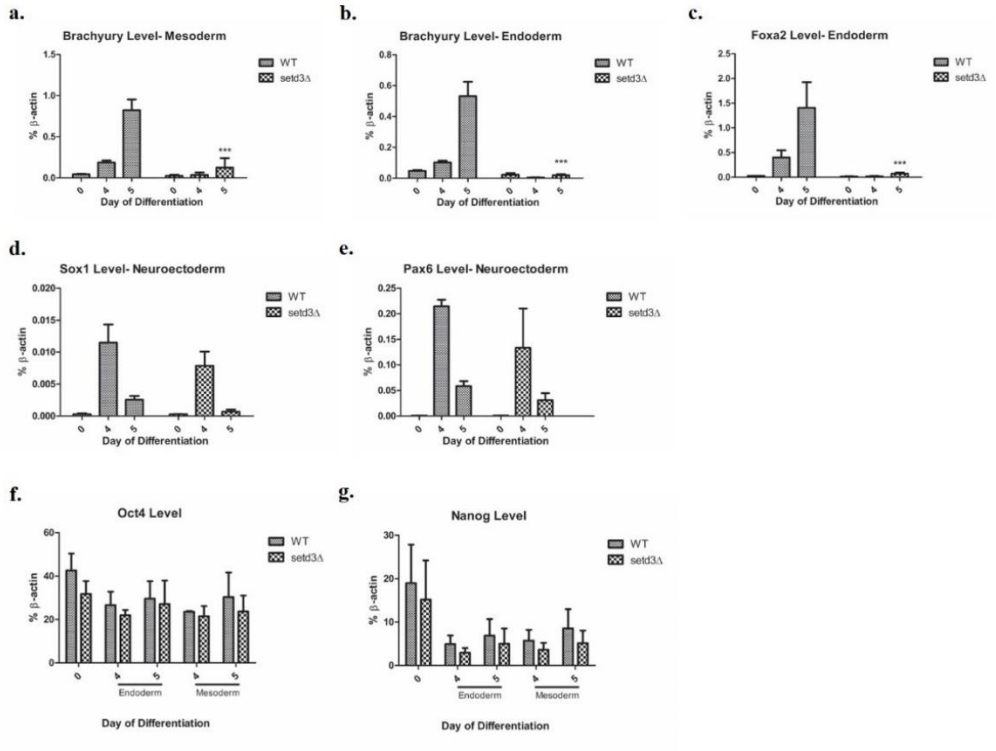

Figure 1.1. SETD3 loss results in defects in mESC differentiation towards meso-endoderm. RT-qPCR of selected transcripts during a, f-g. mesoderm, b-c. endoderm, d-e. neuroectoderm differentiation time-course of wild-type (WT) and setd3Δ mESCs. Results representative of at least 3 replicates. Statistical analysis (two-way ANOVA or t-test) was performed using Graphpad Prism software and statistically significant alterations were depicted on graphs by ** (p-value<0.01) or *** (p-value<0.001).

Consistent with shRNA screen results, setd3Δ mESCs fail to express mesendoderm specific transcription factor *Brachyury* (Figure 1.1.a-b) or endoderm specific

transcription factors *Foxa2* (Figure 1.1.c), *Sox17*, *Cxcr4* and *Cer1* (data not shown) during endoderm or mesoderm differentiations. However, expression of neuroectoderm-specific transcription factors *Sox1* and *Pax6* were similar in WT and setd3Δ mESCs (Figure 1.1.d-e). The ESC-specific transcription factors *Oct4* and *Nanog* were expressed similarly in setd3Δ mESCs compared to WT mESCs and were suppressed with similar kinetics during meso/endoderm differentiation, suggesting SETD3 does not alter pluripotency related gene network. These results demonstrate a critical function for SETD3 specifically on meso/endoderm differentiation.

1.5 Aim of Study

ESCs have ability to differentiate into all tissue types of the body. However, our previous results demonstrate that mESCs lacking SETD3 unable to differentiate into meso-endoderm germ layers. Although, recent reports focus on the cytoplasmic function of SETD3, we believe that the differentiation inability might be resulted from nuclear function of SETD3. We hypothesized that SETD3 regulates mESC differentiation and pluripotency by interacting epigenetic factors. We aimed at identifying proteomic identification of SETD interacting partners in mESCs. The streptavidin-based immunoprecipitation and mass spectrometry (IP-MS) was performed to identify protein partners of SETD3. Among thousand of candidate proteins, we focused on the nuclear proteins and selected some of them to perform biochemical analysis. We validated two important protein for ESC pluripotency as SETD3 partner. Overall, our data suggest that nuclear located SETD3 drives pluripotency by interacting with important epigenetic factors for pluripotency.

CHAPTER 2

MATERIALS AND METHODS

2.1 Cell Culture

CJ9 (WT) and mESCs that *Setd3* gene deleted by CRISPR-Cas9 system (setd3Δ) were obtained from the laboratory of Prof. Stuart Orkin (Boston Children Hospital). The J1BirA, BioSet3, CJ9 and setd3Δ mESCs were expanded on the gelatinized tissue culture (TC) plates lacking MEFs. 2i4 media was composed by 50% Neurobasal (Gibco, Thermo Fisher Scientific), 50% DMEM-F12 (Gibco, Thermo Fisher Scientific), 0.5% N-2 Supplement (100X) (Thermo Fisher Scientific), 1% 50X B-27 Supplement (Thermo Fisher Scientific), 0.5% BSA (Sigma Aldrich), 1% GlutaMAX Supplement (Gibco, Thermo Fisher Scientific), 1% Penicillin-Streptomycin (10.000 U/mL) (Thermo Fisher Scientific), monothioglycerol (1.5x10-4 M final) (Sigma Aldrich) and 4% FBS (Thermo Fisher Scientific). It was supplemented with 3 µM CHIR99021 (Selleckchem), 1 µM PD0325901 (Selleckchem) and 1% LIF (Millipore) to keep mESCs pluripotent. The cells were incubated in 5% CO2 at 37°C. When the cells were reached to 70-80% confluency, the mESCs were split by using 1X TrypLE Express Enzyme (Thermo Fisher Scientific). The detailed media recipe was written in the appendix part (Table A.1).

2.2 Immunocytochemistry

The 12mm round coverslips (Isolab; 126.02.024) were placed in the 12 well tissue culture plates and gelatinized with 1% gelatin solution. $1.5x10^5$ CJ9 mESCs were seeded into each well. The media was completed to 1ml with 2i4 medium. The cells

were incubated around 36 hours at 5% CO2 at 37°C TC incubator. The cells were washed with 250 µl PBS. They were fixed with 250 µl 4% PFA for 15 minutes and then permeabilized with 0.1% TritonX-100 TBS solution for 5 minutes two times. The remaining solutions were removed by filter paper and the coverslips were covered by 40µl 10% BSA (in PBS) solution for in minutes incubation at 37°C TC incubator. To avoid from drying, the wet papers were placed near the coverslips. The primary and secondary antibodies were diluted with 1% BSA solution. The antibody names and dilutions used in ICC and PLA experiments were given in Table B.1. After blocking, the solution was removed, and the cells were covered with 40µl primary antibody solution. The cells were incubated at 4°C for overnight incubation. The cells were washed with 250µl PBS three times and the secondary antibody incubation was performed for 1 hour in the 37°C TC incubator. After the incubation, the cells were washed with 250µl PBS three times. 40µl DAPI containing mounting medium (Sigma; DUO82040) was used for mounting. The edges of the coverslips were closed with nail polish. After they were dried for 15 minutes, the slides were stored at -20°C.

2.3 Phalloidin Staining

For phalloidin staining, $1.5x10^5$ CJ9 and setd3Δ mESC seeding on coverslips. After 24 hours for incubation, the fixation and permeabilization procedures were performed same as mentioned in previously. ActinGreen 488 ReadyProbes™ Reagent (Thermo Fischer; R37110) was used for staining. The cells were washed three times after permeabilization. The dye was diluted 10µl/ml with PBS. The cells were covered with 40 µl dye solution and incubated for 45 minutes at room temperature. Then the cells were washed with PBS for three times, and the mounting was performed by using DAPI containing mounting medium.

2.4 RNA Expression Studies

2.4.1 Sample Preparation, RNA Isolation, cDNA Synthesis

CJ9 mESCs were expanded on the gelatinized TC plates. The cells were collected and washed with ice cold PBS three times and centrifuged at 300g at 4°C for 5 minutes. Then, the supernatant was removed, and the cells were resuspended in 1 ml TRIzol (Thermo Fisher) for each 1x106 cells. The samples were stored at -20°C until RNA isolation was performed.

To obtain total RNA, Qiagen RNeasy Plus Mini Kit (Catalog no: 74134) was used based on the kit's instructions. With Biorad iScript cDNA sythesis kit (Catalog no: 1708890), complementary DNAs (cDNA) were synthesized from total RNA by using oligo dT and random hexamers. The cDNAs were stored at -20°C.

2.4.2 Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)

To check the specificity of each primer, the melting curve was constructed in qRT-PCR. 10 µl of PCR reaction volume was used and forward and reverse primer concentrations were 10 µM. 5µl of 2X SsoAdvanced universal SYBER Green (Bio-Rad). qRT-PCR was performed in 96 well qPCR plates by using Bio-Rad CFX Connect (Bio-Rad) machines. 18-32 quantification cycle (Cq) values were chosen for DNA dilutions, so Cq values were on the log-linear phase of amplification. Each reaction has two replicates and the replicates having difference more than one was eliminated. The Cq normalization was performed by the Cq values of β-actin and the fold change represented in percentage.

The qPCR primers of candidate proteins were obtained from previously used studies. The primer sequence and PubMed ID of the articles were given in the appendix part (Table C.1).

2.5 Protein Studies

2.5.1 Sample Preparation for Nuclear Extraction

The mESCs were collected into 15 ml falcon tubes. They were washed with ice cold PBS three times. The cell pellets were undergone snap freeze (99% ethanol and dry ice) and stored at -80°C.

The nuclear extracts were obtained with Universal Magnetic Co-IP Kit (Active Motif; catalog no: 54002). The protocol suggested by kit was followed. For the localization of Setd3 experiment the cytoplasm and the nucleus coming from 20x106 J1BirA and BioSetd3 mESCs were separated and stored. The nuclear protein concentration was calculated by using BCA protein assay (Thermo Fischer; catalog no: 23227). As a result of trying different concentration of nuclear lysate, we decided that 20-25 µg input was enough. After the inputs were separated, the remaining proteins were used for coimmunoprecipitation (co-IP).

2.5.2 Streptavidin-based Coimmunoprecipitation

J1BirA and BioSetd3 cell lines were used for streptavidin coimmunoprecipitation. After the nuclear extractions were obtained and the proteins were quantified, 25 µg proteins were separated as input. The inputs were resuspended in 4X Laemmli solution (Biorad) containing β-mercaptoethanol (BME) and boiled at 95°C for 10 minutes.

For the streptavidin IP Kim's protocol was followed (Kim *et al.,* 2009). 300 µg nuclear extracts of each sample were mixed with 20 µl streptavidin beads slurry (NEB; catalog no. S1420S). As the wash buffer IP350 buffer (0.3% (vol/vol) NP-40) was used. The beads and the nuclear extracts were incubated at 4°C for overnight incubation. Next day, the beads were washed with IP350 buffer for three times for 15 minutes on the rotator at 4°C, resuspended in 25µl of 2X Leammli solution containing BME. The beads were boiled at 95°C for 10 minutes.

2.5.3 Antibody-based Coimmunoprecipitation

BioSetd3, CJ9 and setd3Δ mESCs were used for antibody-based co-IP. 25 µg protein solution was separated as input. The inputs were resuspended in 4X Laemmli solution (Biorad) containing β-mercaptoethanol (BME) and boiled at 95°C for 10 minutes. Normal rabbit IgG IP was used for the negative control for SETD3 and BRD2 IP samples, goat IgG IP was added for TFCP2L1 IP sample. 400 µg of nuclear extract were used for each co-IP sample. 20 µl magnetic beads conjugated Protein G were mixed with the extract samples. After the volumes were equilibrated with Universal Magnetic Co-IP kit's Co-IP wash buffer, 5 µg of rabbit anti-SETD3, rabbit anti-BRD2, goat anti-TFCP2L1, normal rabbit and goat IgG antibodies were added to each extract samples. The information about antibodies were written in Table B.2. in the appendix part. The samples were incubated at 4°C for overnight incubation. Next day, the beads were washed three times with Co-IP wash buffer for 15 minutes on the rotator at 4°C. They were resuspended in 25µl of 2X Leammli solution containing BME. The beads were boiled at 95°C for 10 minutes.

2.5.4 Western Blot

12% polyacrylamide gel was prepared for electrophoresis. After the inputs and the co-IP samples were boiled at 95°C, they were cooled for 1 minute on ice. The samples were loaded into the gels and run for 2 hours at 100 Volt. Biorad Trans-blot Turbo Transfer System was used for transferring proteins from gel to nitrocellulose membrane. TBS-T containing 5% skimmed milk (Sigma) were used for blocking and the membranes were incubated in this solution for 1 hour at room temperature (RT). The primary antibody incubation was performed overnight at 4°C. After the primary antibody incubations, the membranes were washed with TBS-T three times, 10 minutes for each. Then, secondary antibodies that is linked to horseradish peroxidase (HRP) were put on the membranes and membranes were incubated one hour at RT. The name and dilutions of the primary and secondary antibodies were written in Table B.3. in the appendix part. After the secondary antibody incubations, the membranes were washed with TBST three times, 10 minutes for each. For visualization, Biorad Clarity Western ECL, Biorad Clarity Max Western ECL and Thermo SuperSignal™ West Dura Extended Duration Substrate were used with Biorad ChemiDoc MP system.

The molecular weight of TFCP2L1 is 54 kD. When we performed co-IP, the heavy chain of IgG gives signal at 50 kD where it overlaps with TFCP2L1 signal. Therefore, we used Veriblot, but it detects only goat IgG2.

2.6 Proximity Ligation Assay

The 12 mm round coverslips (Isolab, cat. no. 126.02.024) were put on the wells of 12 well plates. The coverslips were gelatinized with 1% gelatin solution. 1.5x105 BioSetd3 and CJ9 cells were seeded into each well and the solution was completed to 1 ml with 2i4 media. The cells were incubated around 36 hours in 37°C TC incubator. The cells were washed with 250 µl PBS. They were fixed with 250 µl 4%

PFA and permeabilized with 0.1% TritonX-100 TBS solution for 5 minutes for two times. After this step, blocking solution in Duolink PLA Kit (Sigma, cat. no. DUO92008) was used for 30 minutes incubation at 37°C TC incubator. To avoid the cells from drying, wet towel was put around the coverslips. The primary antibody mixes were diluted in antibody dilution solution. Dilution ratios were optimized with immunocytochemistry. The cells were covered with 40 µl primary antibody solutions and incubated overnight at 4°C. After that, Duolink PLA Kit's protocol was followed. Rabbit-minus probe was used for SETD3 (1:250), and BRD2 antibodies (1:250). Mouse-plus probe was used for flag antibody (1:250) and goat-plus probe was used for TFCP2L1 antibody (1:100). The coverslips were washed with wash buffer A and incubated for 1 hour at 37°C TC incubator in probe solutions. Then, again the cells were washed with wash buffer A 2 times and incubated for ligation for 30 minutes at 37°C. After washing steps, 2 hours incubation for amplification was performed at 37°C. Finally, the cells were washed with wash buffer B. 40µl DAPI containing mounting medium (Sigma; DUO82040) was used for mounting. The edges of the coverslips were closed with nail polish. After they were dried for 15 minutes, the slides were stored at -20°C.

CHAPTER 3

RESULTS

3.1 Cellular Localization of SETD3 in mESCs

Previous studies indicated that SETD3 has both nuclear and cytoplasmic functions. His73 of actin methylation by SETD3 increases actin filament contraction in the cytoplasm (Kwiatkowski *et al.*, 2018; Zheng, 2020). SETD3 also works as H3K4/K36 methyltransferase and regulates muscle cell differentiation (Eom *et al.*, 2011). To understand the function of SETD3 in mESC differentiation, we decided to focus on nuclear localized SETD3. Immunocytochemistry (ICC) analysis of SETD3 shows that it localizes both in the cytoplasm and the nucleus in mESCs (Figure 3.1.a). Previously, an mESC line (BioSetd3) that expresses BirA enzyme along with flag tagged SETD3 and biotinylating sequence was generated. J1BirA cell line is the negative control also expressing BirA enzyme without having flag tagged SETD3. The Western blot result shows that similar to endogenous SETD3, biotinylated-SETD3 (bioSETD3) also localized in both cytoplasm and the nucleus (Figure 3.1.b).

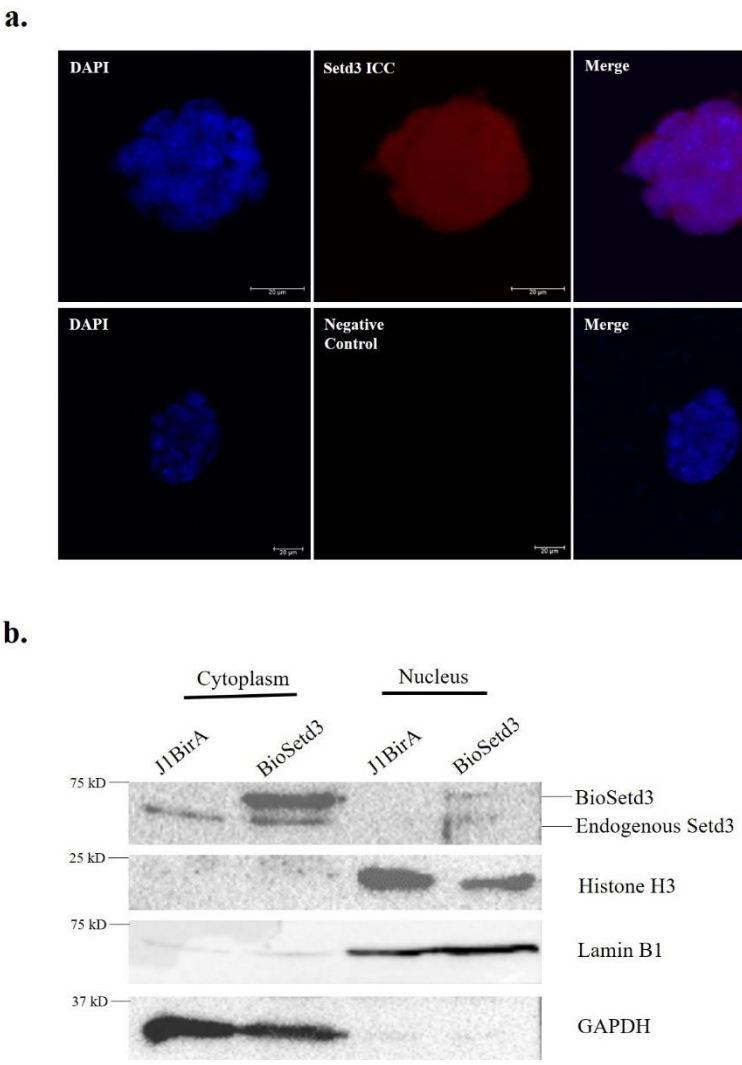

Figure 3.1. Cellular Localization of SETD3 in mESCs. a. ICC result of endogenous SETD3 localization in WT cells and its negative control. The first channel shows DAPI stained nucleus, the second channel shows Setd3 protein signal. The last panel is the merged image. b. Western blot result of the amount and the localization of SETD3 in J1BirA and BioSetd3 ESCs. BioSetd3 and endogenous Setd3 represent SETD3 proteins having biotinylation sequence and naturally translated SETD3 proteins, respectively. Histone H3 and Lamin B1 are used for nuclear controls, GAPDH is used for cytoplasmic control.

3.2 Effect of SETD3 on Actin Filament Stabilization

Previous researches represent that SETD3 also methylates His73 of actin in the cytoplasm. The methylation of actin increases the stability of actin filament in HAP1 cells (Kwiatkowski *et al.*, 2018). To understand the effect of SETD3 in mESCs, we have performed phalloidin staining (Figure 3.2.). The results show that WT and setd3Δ mESCs show similar filamentous actin. This means that absence of SETD3 does not affect the actin filament formation and stabilization in mESCs.

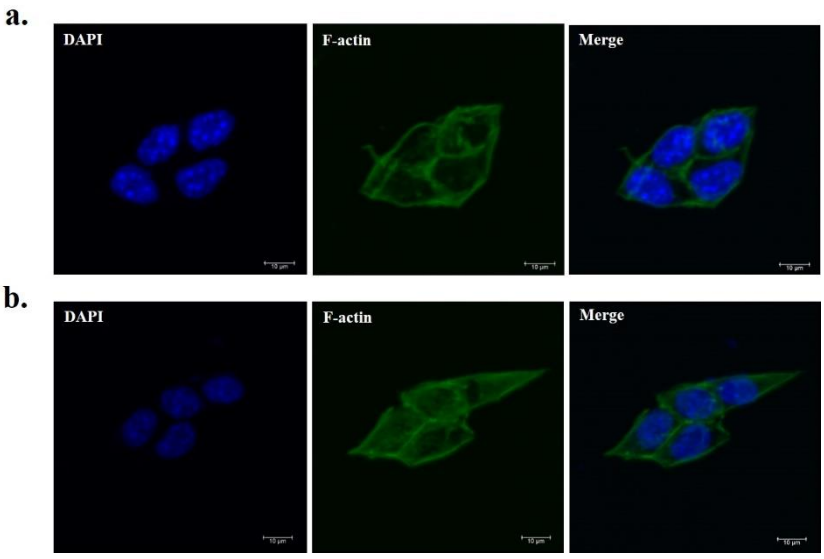

Figure 3.2. Phalloidin staining result in a. WT, b. setd3Δ mESCs. The first panels indicate the nucleus stained by DAPI, the second panels show the F-actin. The third panels are the merged images.

3.3 Identification of Candidate Proteins from Mass Spectrometry Analysis

In order to figure out how SETD3 regulates mESC differentiation, identification of the proteomic partners of SETD3 was aimed. To do this, two replicates of streptavidin-based immunoprecipitation and mass spectrometry (IP-MS) was performed previously by using BioSetd3 and J1BirA cell lines. Because there are also proteins having biotinylation sequence naturally in mammalian cells,

biotinylated proteins can be pulled down by streptavidin interaction in both cell lines. We have focused on the SETD3 specific interaction. Therefore, the first stage was to selection of the proteins only pulled down in BioSetd3 cell line. The number of proteins pulled down in both cell lines for two replicates were given in Table 3.1.

Table 3.1. The amount of proteins precipitated in IP-MS.

	J1BirA	BioSetd3
Replicate 1	2730	2270
Replicate 2	2997	2167

As seen in the Table 3.1. there are several proteins containing biotinylation sequence. Firstly, the proteins only pulled down in BioSetd3 cells were selected, then, the common proteins found in both replicates were chosen for the analysis. 79 proteins were chosen for analysis which are given in Table D.1. in the appendix part. Elimination steps of the candidate proteins are written in the Figure 3.3.

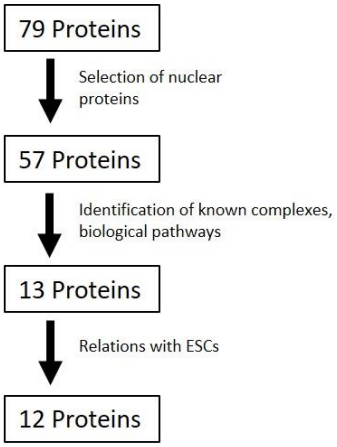

Figure 3.3. Scheme of the elimination of candidate proteins.

3.3.1 Selection of Nuclear Proteins

SETD3 has methyltransferase function, so the protein interacting with SETD3 can also have epigenetic function (Eom *et al.*, 2011; Kim *et al.*, 2014). Therefore, the nuclear localized proteins were selected for analysis. To do this, a database developed by the Jensen lab that combines the information from Psort and Yloc databases was used (http://compartments.jensenlab.org/Search). If a protein is completely cytoplasmic, it was discarded from the list. Both cytoplasm and nuclear localized proteins were kept in the list. 57 proteins, which were given in Table D.2. in the appendix part, were selected for further analysis.

3.3.2 Identification of Complex Knowledge of the Candidate Proteins

If SETD3 interacts with one of the complex members in the nucleus, other protein members could also find in MS analysis. Therefore, a complex analysis was performed for 57 proteins using CORUM database (http://mips.helmholtz-muenchen.de/genre/proj/corum). This database gives the information about the protein members of complexes in mammalian cells. The organism was selected as mouse, and if a protein found in more than one complex, the two best-known nuclear complexes were selected. Complex analysis did not return any results for some of the proteins, presumably because not all proteins are part of a complex. These proteins were kept in the list. On the other hand, some proteins are a member of complexes that are abundant in the nucleus, like spliceosome related proteins, nucleolar proteins etc. Because of high abundance, they can be possible contaminants during experimental procedure. Therefore, these proteins were also eliminated from the list. The complex knowledge of the candidate proteins were written in Table D.3. in the appendix part.

Next, the interaction of the nuclear candidate proteins was analyzed using STRING database (http://string-db.org). This database gives both known (by experimental and

from other databases) and predicted (coexpression, gene neighborhood) interactions (Figure 3.4.). By this way, physically interacted candidate proteins were focused on.

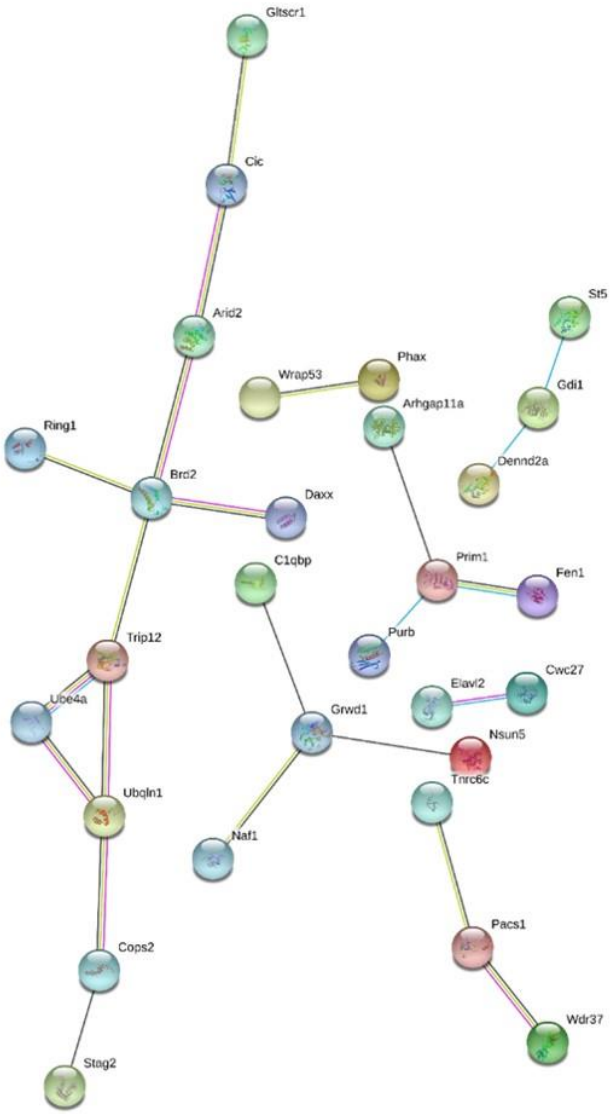

Figure 3.4. The candidate proteins that have at least one interaction on the STRING database. Blue and pink lines show curated databases and experimentally determined known interactions; green, red and dark-blue lines show predicted interactions by gene neighborhood, gene-fusion and gene co-occurrence, respectively. Yellow, dark blue, and light-blue lines show possible interactions by textmining, co-expression and protein homology, respectively. The proteins that do not have any interactions were eliminated from the figure.

3.3.3 Identification of Biological Pathways of Candidate Proteins

As a histone methyltransferase, SETD3 regulates gene expression by transcription regulation. If a protein interacts with SETD3, it could also have similar function in the cells. Therefore, the biological pathways of 57 nuclear candidate proteins were analyzed in DAVID (https://david.ncifcrf.gov/) database by considering Gene Orthology (GO) and biological pathways. The analysis result is given in Table 3.2. The proteins having function in transcription regulation, and chromatin modification were selected for literature research.

Table 3.2. Biological pathways and number of proteins found in each pathway.

Biological Pathways	**Number of Proteins Found in the Pathway**
negative regulation of transcription from RNA polymerase II promoter	8
rRNA processing	4
transcription, DNA-templated	13
regulation of transcription, DNA-templated	13
positive regulation of nuclear-transcribed mRNA poly(A) tail shortening	2
maturation of 5.8S rRNA	2
covalent chromatin modification	4
positive regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	2
regulation of protein ubiquitination	2

miRNA mediated inhibition of translation	2
negative regulation of stress fiber assembly	2
apoptotic process	5

The 13 proteins that are found in the DNA transcription, regulation and chromatin modifications pathways were selected for literature research. The proteins were listed in Table 3.3. The knowledge about expression levels, working partners, research areas etc. in mESCs of these proteins were examined.

Table 3.3. The candidate proteins selected for biochemical analysis.

UniProt ID	**Gene Symbol**
Q924C5	Alpk3
E9Q7E2	Arid2
Q9D219	Bcl9
Q7JJ13	Brd2
P61202	Cops2
O35613	Daxx
Q0VGT2	Gli2
Q9DC33	Hmg20a
P20664	Prim1
O35730	Ring1
P23881	Tcea3

Q64092	Tfe3
Q3UNW5	Tfcp2l1

3.3.4 Literature Knowledge About the Candidate Proteins

ALPK3 (Alpha-protein kinase 3) is a nuclear kinase protein. Previous researches reveal that deletion or suppression of *Alpk3* gene results in cardiomyopathy which is related with heart muscle. The ventricular heart muscles of these mice became thicker and this decreases muscle contraction (Van Sligtenhorst *et al.*, 2012). The function of ALPK3 protein is not known exactly, but it is important for expression of specific transcription factors related with cardiomyocyte development (Çağlayan *et al.*, 2017).

ARID2 (AT-rich interactive domain-containing protein 2) is a member of chromatin remodeling complex (SWI-SNF complex). This complex changes the interaction between transcription factors and chromatin by modifiying nucleosomes. Therefore, it changes the expressions of different genes in different stages of development (Xu, Flowers and Moran, 2012). Previous experiments reveal that when Baf200/Arid2 gene expression is suppressed by siRNA in NG4 (Nanog:GFP reporter cell line) mouse cells, promoter activity of *Nanog* is upregulated during mESC differentiation (Schaniel *et al.*, 2018). This result shows direct role of ARID2 protein on specific gene expression in differentiation. In addition, when *Arid2* is suppressed in mouse pre-osteoblast cells (MC3T3-E1 cell line) by shRNA, the cells cannot complete osteoblast differentiation and *bglap* gene expression which is a specific gene in osteoblast cells is reduced (Xu, Flowers and Moran, 2012).

BCL9 (B-cell CLL/lymphoma 9 protein) protein is known as the activator of Wnt/β-catenin pathway which is important for cell division and pluripotency. In fact, when muscle stem cells produce differentiated cells, Wnt/β-catenin pathway is activated, β-catenin enters the nucleus and works as TF for expression of muscle specific genes. On the other hand, if β-catenin gene (*Ctnnb1*) is downregulated, the

amount of β-catenin entering the nucleus will not become enough and this results in the low yield of muscle cell differentiation (Brack *et al.*, 2009).

Brd2 (Bromodomain-containing protein 2) protein is a member of BET (Bromodomain and extraterminal domain) family. It has two bromo domains which are important for binding of highly acetylated regions of DNA. This binding changes the accessibility and binding strength of these regions by TFs (Taniguchi, 2016). Previous results reveal that during differentiation, BRD2 has a function in upregulations of bivalent genes suppressed by H2A.Z.1 monoubiquitination in mESCs (Surface *et al.*, 2016). When these activated genes are analyzed in DAVID database, it is shown that they are the genes working in mesoderm differentiation and embryo development. In addition, BRD2-BRD4 proteins bind Nodal response elements in mESCs to provide pluripotency. When the cells begin differentiation through mesendoderm BRD4 gives place to BRD2 protein and in the ESCs exposed to BET inhibitor (JQ1), mesendodermal specific gene expressions were reduced (Fernandez-Alonso *et al.*, 2017).

COPS2 (COP9 signalosome complex subunit 2) protein is a member of COP9 signalosome (CSN) complex which work in protein degradation by ubiquitination. However, each complex member can have different roles in ESCs because deletion of each complex member results in embryo death in different developmental stage (Zhang *et al.*, 2016). *Cops2* deleted mouse embryos cannot form inner cell mass (ICM) during development (Lykke-Andersen *et al.*, 2003). In addition, when *Cops2* expression is suppressed, expressions of *Oct4* and *Nanog* important pluripotency factors are also downregulated in mESCs (Zhang *et al.*, 2016).

DAXX (Death domain-associated protein 6) protein is a histone chaperone in ATRX:DAXX complex. It works in placement of H3.3 histone to telomeric regions of chromosome in ESCs. These regions are heterochromatic regions, that is, the genes in these areas are suppressed constantly (Lewis *et al.*, 2010). When *Daxx* is deleted in embryos, they cannot differentiate into different tissue types and because

of disrupted global chromatin organization, the embryos cannot survive (Rapkin *et al.*, 2015).

GLI2 (Zinc finger protein) protein is associated with sonic hedgehog signal transduction which has important role in skeletal system of vertebrates (Mo *et al.*, 1997). In addition, *Gli2* expression is upregulated during pregnancy and milk production and helps mammary gland development in mice (Lewis *et al.*, 2001).

HMG20A (High mobility group protein 20A) protein is the inhibitor of BRAF35/HMG20b corepressor complex. This complex decreases the expression of neuronal genes in non-neuronal tissues. As an inhibitor, HMG20A protein inhibits BRAF and activates neuron related genes. In addition, it helps the binding of MLL protein which is a H3K4 methyltransferase (Ceballos-chávez *et al.*, 2012).

PRIM1 (DNA primase small subunit) is a subunit of DNA primase heteromeric enzyme. This enzyme works in synthesis of RNA primer essential for DNA replication (Job *et al.*, 2018).

RING1 (E3 ubiquitin-protein ligase) mono-ubiquitinates histone H2A at K119. It is essential for neuronal differentiation of mESCs. When it is deleted, H3K27me3 and H2AK1109ub decrease on neuronal genes (Yao *et al.*, 2018).

TCEA3 (Transcription elongation factor A protein 3) protein is highly produced in mESC. Nodal ligands bind TGFβ receptors and activate Nodal signaling pathway. This pathway activates specific SMAD proteins and TFs and increases the expression of endoderm-mesoderm related genes in mESCs (Cha *et al.*, 2013). TCEA3 activates LEFTY1 which is a negative regulator of Nodal pathway. Therefore, it promotes pluripotency. When *Tcea3* is downregulated, the expression of mesoderm marker genes is upregulated. As a result of this reason, we can say that TCEA3 protein has a direct function on pluripotency of mESCs (Park *et al.*, 2014).

TFE3 (Transcription factor E3) is a transcription factor. Its amount is important for keeping mESCs pluripotent. It is kept in the nucleus and helps the expression of Esrrb gene which is important for pluripotency. However, during differentiation,

TFE3 is removed from the nucleus and *Esrrb* is downregulated (Betschinger *et al.*, 2013).

Tfcp2l1 (Transcription factor CP2-like protein 1) is a target protein of Wnt/β-catenin pathway which is important for pluripotency. When it is activated, pluripotency related genes' expressions are also upregulated i.e. *Esrrb* (Qiu *et al.*, 2015). Previous experiments show that two terms of the protein are important for different differentiation lineages. N-term suppresses mesoderm and trophectoderm differentiations, CP-like domain suppresses endoderm differentiation and keeps the cells pluripotent (Liu *et al.*, 2017).

After these literature researches, 12 proteins were selected for biochemical analysis. ALPK3 was removed from the list because it does not have any known specific function in mESCs.

3.4 Transcriptional Expression Levels of the Candidate Proteins

Before coimmunoprecipitation (co-IP) experiments, we aimed to validate the gene expression levels of the candidate proteins in wild type WT mESCs (Figure 3.5.). We performed qRT-PCR experiment and the expression levels were normalized to the expression level of *β-actin*. *Brd2*, *Arid2*, *Cops2*, *Daxx*, *Prim1*, *Tcea3* and *Tfcp2l1* are expressed at a high level in WT mESCs, so these proteins were selected for co-IP experiments. *Bcl9*, *Hmg20a*, *Ring1*, *Gli2*, *Tfe3* and *Tnrc6c* were expressed at low level, so they are not analyzed further.

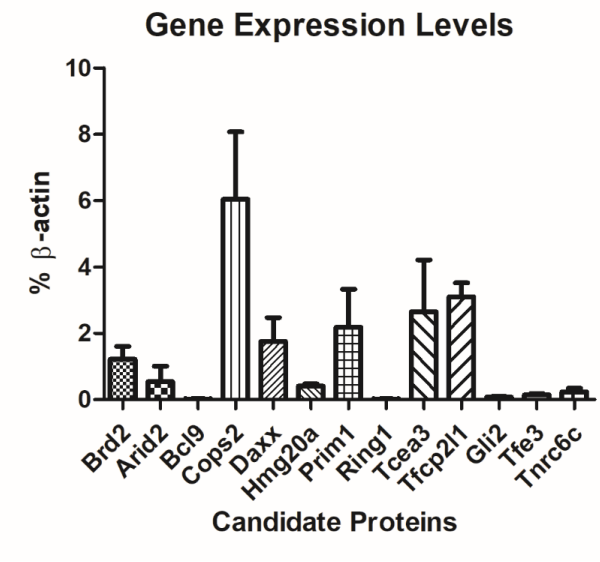

Figure 3.5. Gene Expression of Candidate Transcription Factors in WT mESCs. Gene expression in mRNA levels were measured by qRT-PCR. Expression was normalized to β-actin expression level. Data are shown as mean ±SEM of three biological replicates.

3.5 Determination of Physical Interaction between Setd3 and the Candidate Proteins by Co-Immunoprecipitation

After the examination of gene expression levels of the 12 candidate proteins, 7 proteins were selected for the confirmation by co-IP experiments. The co-IP experiments have three steps according to the cell lines and the baits (Table 3.4.).

Table 3.4. The list of performed co-IP methods.

IP Method	Bait	Cell Line
Streptavidin-based Co-IP	Biotinylated Setd3	BioSetd3
Antibody-based Co-IP	Endogenous Setd3	BioSetd3
Antibody-based Co-IP	Endogenous Setd3	CJ9 and setd3Δ

3.5.1 Streptavidin-based Co-Immunoprecipitation

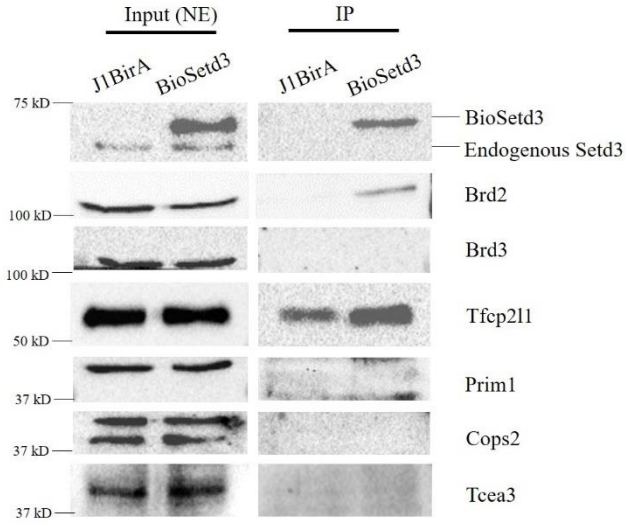

Figure 3.6. Streptavidin-based Co-Immunoprecipitation (co-IP) Results in J1BirA and BioSetd3 mESCs. 25 µg of input which is the nuclear extracts of J1BirA and BioSetd3 mESCs are shown on the left side. 300 µg NEs were used for each coIP sample. To visualize inputs and SETD3, BRD2 and TFCP2L1 IP samples Biorad Clarity ECL substrate was used. The other proteins were visualized with Biorad Clarity Max ECL. Signals from inputs and IP samples are on a line at the same exposure time. The results were obtained from three independent experiments. (IP; immunoprecipitation, NE; nuclear extract).

BioSetd3 cell line has both endogenous and biotinylated SETD3. In the first part of co-IP experiment, interaction of bioSetd3 and the candidate proteins were examined by streptavidin-based co-IP experiment (Figure 3.6). BRD2, COPS2, PRIM1, TCEA3 and TFCP2L1 were chosen to check whether there is physical interaction with bioSETD3 protein. We eliminated ARID2 and DAXX proteins because of the high nonspecific interactions caused by their antibodies. Alignments of BRD2 and BRD3 proteins are similar to each other (Leroy, Rickards and Flint, 2008). As a result of this reason, BRD3 was also used to check interaction specificity. J1BirA cell line was used as negative control to eliminate nonspecifically immunoprecipitated proteins. Inputs lanes represents that the expressions of the candidate proteins in the nuclear extracts of J1BirA and BioSetd3 mESCs.

IP lane shows that bioSETD3 was specifically immunoprecipitated. The molecular weight of bioSETD3 is consistent with the one seen in inputs as can be seen in Figure 3.5. BRD2 protein was expressed in both mESC lines. Low amount of BRD2 was immunoprecipitated through the interaction with bioSetd3. J1BirA lane was clear which shows that BRD2 was precipitated specifically. In fact, even though BRD2 has similar sequences with BRD3, it did not precipitate with bioSETD3. This shows the interaction between BRD2 and bioSETD3 was specific. Although, TFCP2L1 was precipitated in both cell lines, the amount of protein precipitated in BioSetd3 mESCs was much higher than in J1BirA cells. The reason of this TFCP2L1 may also have biotinylation sequence, but it also interacts with bioSetd3. Even though, COPS2, PRIM1 and TCEA3 proteins were expressed in both mESCs, they did not precipitate with bioSETD3. This shows there is no interaction with bioSETD3. These proteins were not examined for further experiments.

3.5.2 Antibody-based Co-Immunoprecipitation in BioSetd3 mESCs

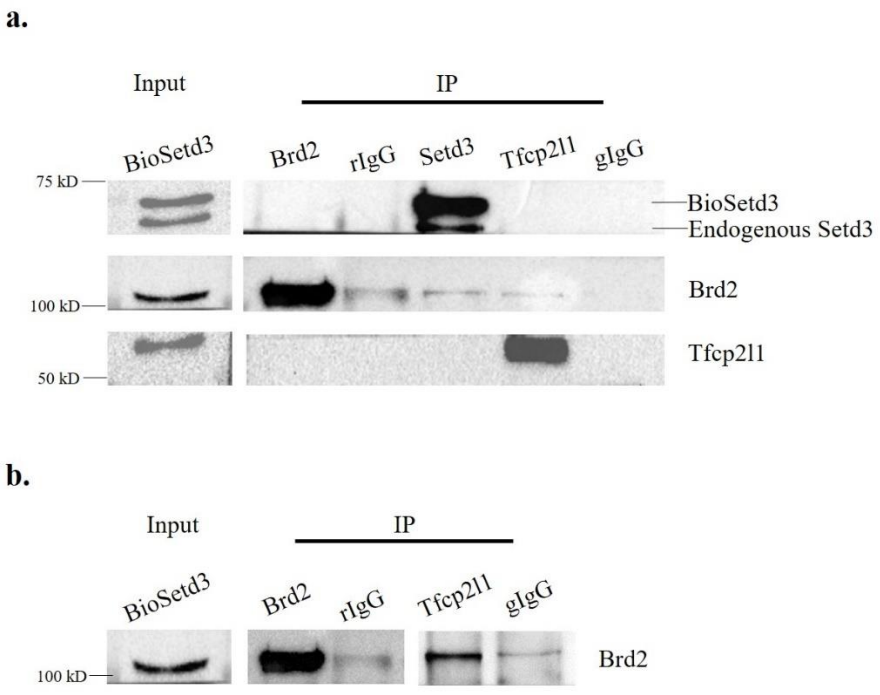

Figure 3.7. Antibody-based Co-IP Results of a. SETD3, BRD2, TFCP2L1, normal rabbi IgG and goat IgG in BioSetd3 mESCs. 25 ug of input was loaded and shown on left side. 500 ug nuclear lysate was used for each immunoprecipitation. To visualize input proteins Biorad Clarity Max ECL substrate, to visualize IP samples Pierce ECL WB substrate were used. b. BRD2 WB band exposing longer exposure. The results were obtained from two independent experiments. (IP, immunoprecipitation; WB, western blot; NE, nuclear extract).

Strong interaction between biotin and streptavidin may also lead to precipitation of naturally biotinylated proteins other than SETD3. Therefore, after the identification of BRD2 and TFCP2L1 interactions with bioSETD3 by streptavidin-based co-IP experiment in BioSetd3 mESCs, we decided to indicate the interactions of BRD2 and TFCP2L1 proteins with endogenous SETD3 (Figure 3.7.). Input lane show the expressions of these three proteins. IP lanes show the coimmunoprecipitation of each protein by using specific antibodies. Rabbit IgG was a negative control for BRD2 and SETD3 proteins, because their host of the antibodies were rabbit. Goat IgG was

used for negative control of TFCP2L1 co-IP. IP lane of SETD3 shows that as well as bioSETD3 and endogenous SETD3 was also precipitated. BRD2 and TFCP2L1 proteins were also successfully precipitated (Figure 3.7.a.). However, we did not see BRD2 and TFCP2L1 interaction in SETD3 co-IP. Because of the oversaturation in BRD2 WB band, it was hard to evaluate the other signals. Therefore, only TFCP2L1 and gIgG bands were cut and exposed to longer exposure. Interestingly, longer time exposure BRD2 and TFCP2L1 co-IP results show that BRD2 interacts with TFCP2L1. There are less amounts of precipitated Brd2 proteins compared to input. This means that not all BRD2 proteins interact with TFCP2L1.

3.5.3 Antibody-based Co-Immunoprecipitation in WT and setd3Δ mESCs

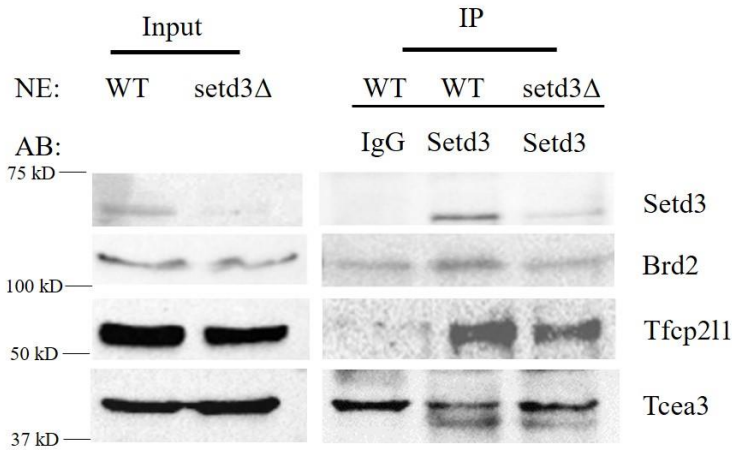

Figure 3.8. SETD3 and Normal Rabbit IgG Co-Immunoprecipitation Results in Wildtype (WT) and setd3Δ mESCs. 20 ug of input which is nuclear lysate from WT and setd3Δ mESCs shown on left side. 300 ug nuclear lysate was used for each co-IP experiments. To visualize BRD2, TFCP2L1 and TCEA3 proteins for input Biorad Clarity Max ECL substrate was used. The others were visualized with Biorad Clarity Max ECL. Signals from inputs and IP samples are on a line at the same exposure time. The results were obtained from one experiment. (IP, immunoprecipitation; WB, western blot; NE nuclear extract; AB, antibody used for immunoprecipitation).

BioSetd3 mESCs have high amounts of SETD3 proteins (both biotinylated and endogenous SETD3), so this could be formation of new interactions. This can be a reason for nonspecific interactions. Therefore, we would like to confirm the

streptavidin-based co-IP experiment to validate the SETD3 interaction with BRD2 and TFCP2L1 proteins in WT and setd3Δ mESCs (Figure 3.8). We performed SETD3 co-IP using a Setd3 antibody and rabbit IgG as a negative control. setd3Δ mESCs was used as an additional negative control because it is expected that if a protein has physically interacted with Setd3 should not be seen in the setd3Δ cells.

The input lane of setd3Δ mESCs confirms the loss of SETD3. Also, SETD3 was successfully immunoprecipitated in WT cells. setd3Δ cells have very low amount of Setd3 protein, so it is expected to see very weak or no Setd3 band in setd3Δ IP. In addition, IgG band also shows the Setd3 precipitation is specific.

Figure 3.8. input shows that BRD2 was expressed in both WT and setd3Δ cells. Slightly more BRD2 was precipitated with SETD3 compared to IgG control in the nuclear extract of WT cells. Also, less amounts of precipitated BRD2 can also be seen in setd3Δ mESCs, this is expected because of the presence of low expression of SETD3. Therefore, it can be said that BRD2 interacts with SETD3.

Figure 3.8. input shows that TFCP2L1 was highly expressed in both WT and setd3Δ mESCs. Even though, IgG lane was clear, TFCP2L1 was precipitated in both WT and setd3Δ cells. This result reveals that the interaction did not come from Fc regions of the antibodies. Tfcp2l1 proteins may be precipitated through the low amounts of SETD3 in setd3Δ cells. This is expected because SETD3 IP lane of WT mESCs has thicker TFCP2L1 band than the band obtained from setd3Δ cells. Therefore, it can be concluded that TFCP2L1 interacts with SETD3.

Previously, the SETD3-TCEA3 interaction was not obtained from the streptavidin-based co-IP of bioSETD3 in BioSetd3 mESCs. Figure 3.7. input shows that TCEA3 was expressed in both WT and setd3Δ mESCs. However, it was precipitated in both IgG control and setd3Δ cells as well. Therefore, we concluded that TCEA3 co-IP was nonspecific. However, we have obtained these results from one experiment. This can be the result of loss of unstable, dynamic interactions.

3.6 Protein Levels of BRD2 and TFCP2L1

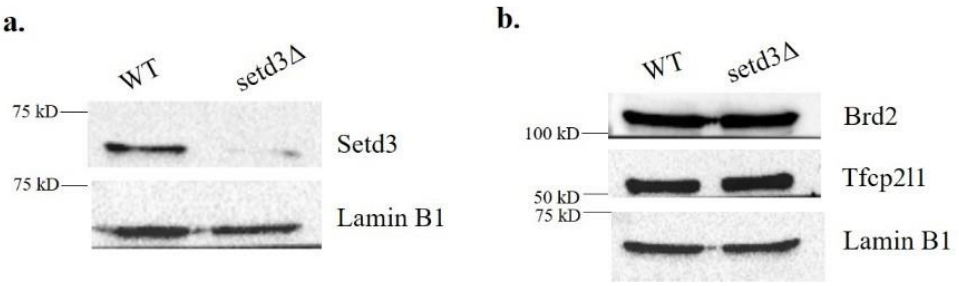

Figure 3.9. Protein levels of BRD2 and TFCP2L1 in WT and setd3Δ mESCs. Western blot is used for the measurement of protein levels of a. SETD3, b. BRD2 and TFCP2L1. Lamin B1 was used as loading control. Three biological replicates were done for each protein.

After we found the interactions, we investigated whether the protein levels of BRD2 and TFCP2L1 depend on SETD3 presence or not. To understand this, Western bot experiment was performed in WT and setd3Δ mESCs (Figure 3.9). The Figure 3.9.a shows the *Setd3* deletion is continuous, and Figure 3.8.b represents that SETD3 does not change the expressions of BRD2 and TFCP2L1.

3.7 Confirmation of the Interactions by Proximity Ligation Assay

Because of the long experimental procedure of co-IP, some unstable, dynamic and weak interactions could be lost. On the other hand, some proteins start interacting each other because of compact co-IP lysates. Proximity ligation assay (PLA) is a method to validate in situ detection of protein interactions with the use of antibody-bound DNA oligos that enable fluorescent signal amplification when the distance between two proteins is shorter than 40 nm (Alam, 2018). When the physical interaction obtained, the hybridized oligonucleotides start hybridization to make circular DNA. This DNA can be amplified and visualized by fluorescently labeled complementary oligonucleotide probes. Therefore, the interaction between two

proteins can be seen under confocal microscope as stained with red dots (Texas Red). We decided to confirm the SETD3-BRD2, SETD3-TFCP2L1 and BRD2-TFCP2L1 interactions that are indicated by co-IP experiments.

3.7.1 SETD3-BRD2 Interaction

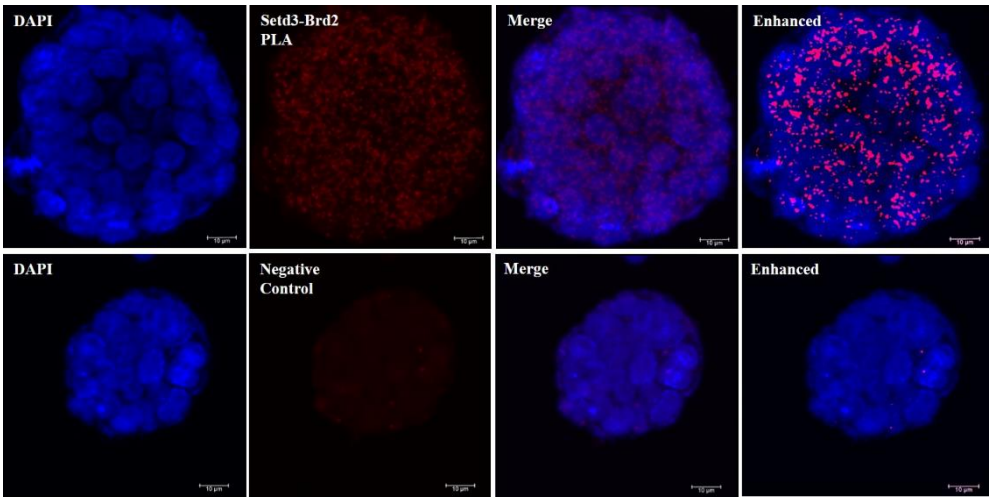

Figure 3.10. Proximity ligation assay result of bioSETD3-BRD2 interactions in BioSetd3 mESCs. The upper panels show the experimental sample, the lower panels show the negative control without using primary antibodies. The first panels indicate the nucleus stained by DAPI, the second panels show the bioSETD3-BRD2 interaction. The third panels are the merged images. The last panels were obtained from ImageJ software by enhancing the signals. The results were obtained from two independent experiments.

We decided to validate SETD3-BRD2 interactions in WT mESCs. However, the mouse anti-BRD2 antibody (Santa Cruz; sc-514103) that we ordered for PLA did not work well. The exogenous Setd3 contains mouse-flag tag as well as biotinylation sequence in bioSETD3 mESCs. Therefore, we could use mouse anti-flag antibody (1:250) and rabbit anti-BRD2 antibody (1:250) to confirm the SETD3-BRD2 interaction (Figure 3.10.). Although, the negative control has red background, the brighter red dots represent the SETD3-BRD2 interactions. The signals were also enhanced with ImageJ software. The result confirms the SETD3-BRD2 interaction in BioSetd3 mESCs.

3.7.2 SETD3-TFCP2L1 Interaction

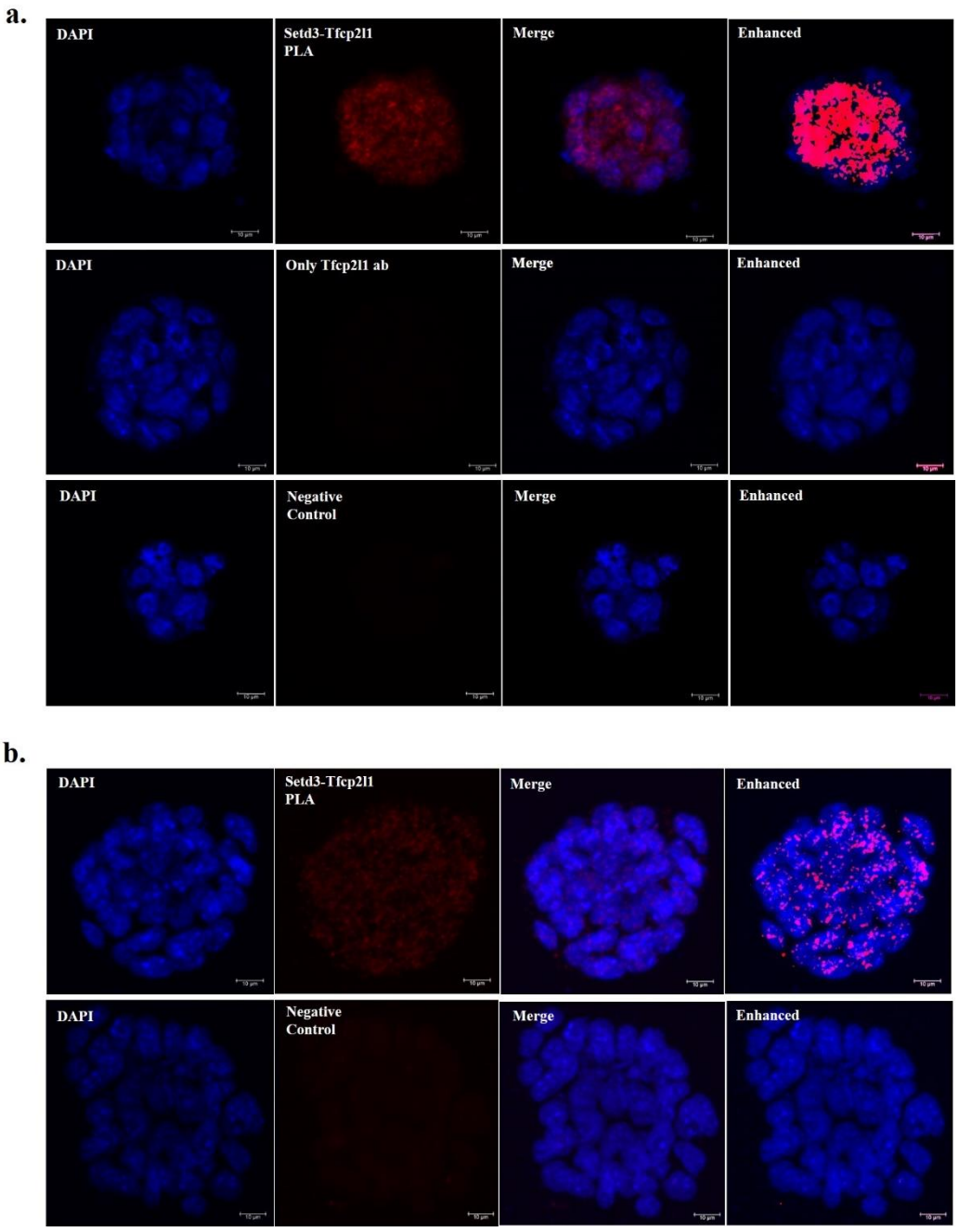

Figure 3.11.a. Proximity ligation assay result of SETD3-TFCP2L1 interactions in BioSetd3 mESCs. The upper panels show the experimental sample. The panels presence on the second lane are the negative control performed by using only goat anti-TFCP2L1 antibody. The lower panels show the negative control without using any primary antibodies. b. Proximity ligation assay result of SETD3-TFCP2L1 interactions in WT mESCs. The upper panels are the experimental samples. The lower panels show the negative control without using any primary antibodies. The first panels

indicate the nucleus stained by DAPI, the second panels show the SETD3-TFCP2L1 interaction. The third panels are the merged images. The last panels were obtained from ImageJ software by enhancing the signals. The results were obtained from two independent experiments.

Firstly, BioSetd3 mESC line was used to confirm the interaction between SETD3 and TFCP2L1 that we indicated by streptavidin-based co-IP (Figure 3.11). The rabbit anti-SETD3 (1:250) and goat anti-TFCP2L1 (1:100) was used for validation. Presence of high numbers of red dots shows the SETD3-TFCP2L1 interactions. Before PLA was performed, we optimized the antibody dilutions by immunocytochemistry (ICC). Because of the quality of anti-TFCP2L1 antibody, the antibody was used concentrated. High concentration of antibody also could be a reason for nonspecific binding. Therefore, we decided to investigate the nonspecific binding resulted from high concentration of anti-TFCP2L1 antibody. The panels at the second lane is another negative control by containing only anti-TFCP2L1 antibody. Neither the negative control that have only anti-TFCP2L1 antibody nor the one consisting any antibody give any nonspecific signal. The result shows the SETD3-TFCP2L1 interactions seen in the first lane is specific.

The SETD3-TFCP2L1 interaction was confirmed in BioSetd3. Because of the presence of overexpressed SETD3 in BioSetd3 mESCs, it may create new interaction in the cells. Therefore, we wanted to confirm the interaction with endogenous SETD3 (Figure 3.11.b). Although, the negative control gives very weak signal, the brighter red dots show the presence of SETD3-TFCP2L1 interactions. Amount of the red signals is fewer in WT mESCs compared to the number of the red signal found in BioSetd3 mESCs. The result could be the higher amount of SETD3 increases the number of interactions with TFCP2L1.

3.7.3 BRD2-TFCP2L1 Interaction

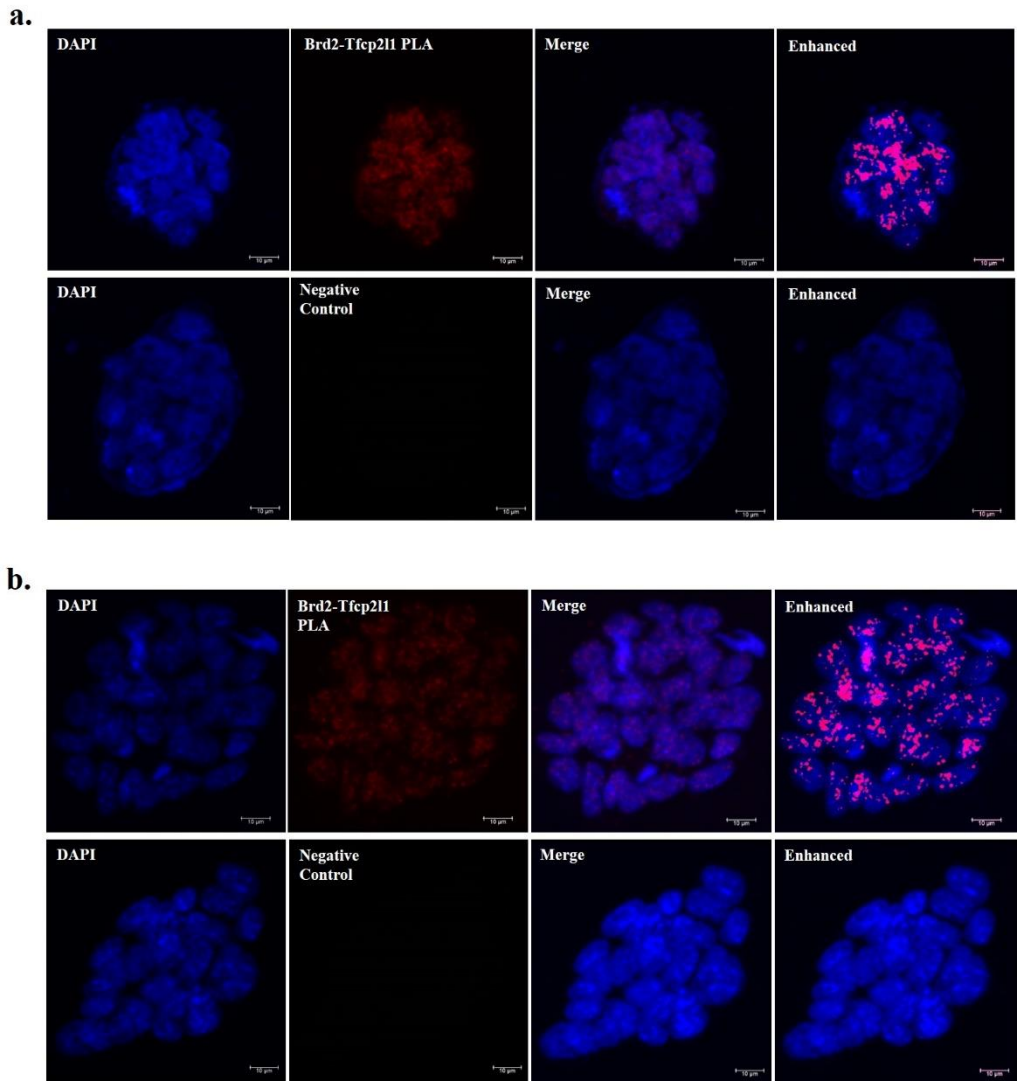

Figure 3.12. Proximity ligation assay result of Brd2-Tfcp2l1 interactions in a. BioSetd3 and b. WT mESCs. The upper panels show the experimental samples. The lower panels show the negative controls without using any primary antibodies. The first panels indicate the nucleus stained by DAPI, the second panels show the Brd2-Tfcp2l1 interaction. The third panels are the merged images. The last panels were obtained from ImageJ software by enhancing the signals. The results were obtained from two independent experiments.

The BRD2-TFCP2L1 interaction was observed by co-IP experiment in WT mESCs (Figure 3.7.b). This interaction also confirmed in BioSetd3 and WT mESCs (Figure 3.12.). Firstly, BRD2-TFCP2L1 interaction was confirmed in BioSetd3 mESCs

(Figure 3.12.a). The bright red dots indicate the presence of interactions. In addition, both Brd2 and Tfcp2l1 proteins are present in the nucleus of the cells. Presence of the red signals only in the nucleus (merged with DAPI) strengthen the specificity of the interaction between Brd2 and Tfcp2l1. Similarly, the nuclear Brd2-Tfcp2l interaction was also observed in WT mESCs (Figure 3.12.b). The negative controls for both samples were clear which is important for nonspecific binding of the oligonucleotides.

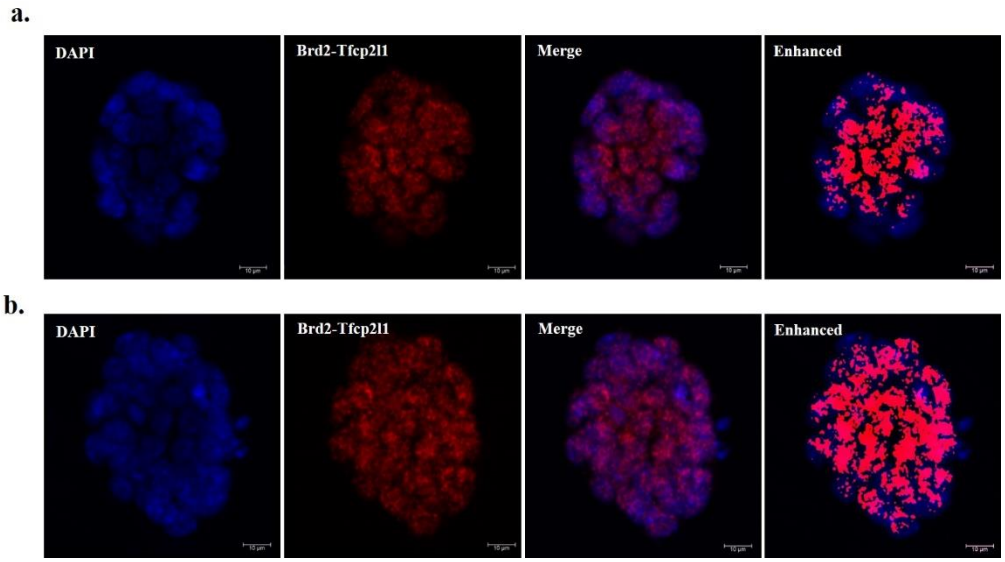

Figure 3.13. Proximity ligation assay result of Brd2-Tfcp2l1 interactions in a. WT and b. setd3Δ mESCs. The first panels indicate the nucleus stained by DAPI, the second panels show the Brd2-Tfcp2l1 interaction. The third panels are the merged images. The last panels were obtained from ImageJ software by enhancing the signals.

Also, we would like to confirm if BRD2-TFCP2L1 interaction depends on SETD3 presence or not. Figure 3.13 shows that BRD2-TFCP2L1 interactions in WT and setd3Δ mESCs. There is no significant difference between two cell lines. That shows that SETD3 does not necessary for BRD2-TFCP2L1 interaction.

CHAPTER 4

DISCUSSION

In this study, we demonstrated a critical role for SETD3 in meso/endodermal differentiation and identified BRD2 and TFCP2L1 as its physical partners in mESCs. Though recent reports focus on the cytoplasmic function of SETD3, we believe our data supports a model that its nuclear function might play an important role in mESCs: PLA experiments demonstrated the majority of SETD3-BRD2 and SETD3-TFCP2L1 interactions occur within the nucleus (Figures 3.10, and Figure 3.11). Consistent with our mESC differentiation data on SETD3, both TFCP2L1 and BRD2 are critical proteins for maintenance of ESC state and regulation of differentiation (Hancock, no date; Shang *et al.*, 2009; Surface *et al.*, 2016; Fernandez-Alonso *et al.*, 2017; Liu *et al.*, 2017).

Brd2 is a bromodomain and extra-terminal (BET) domain containing protein. It binds to regions with high level of acetylated histones. It is essential for embryonic development and neural tube development in mouse (Gyuris *et al.*, 2009; Shang *et al.*, 2009). Although closely similar, BRD2 and BRD4 occupy different regions. Brd4 is found mostly in enhancer regions whereas BRD2 binds mostly in promoters (Surface *et al.*, 2016; Fernandez-Alonso *et al.*, 2017). BRD2 is important for the activation of lineage-specific genes during mESC differentiation by exclusion of PcG complexes (Surface *et al.*, 2016). It transcriptionally activates Nodal during mesendodermal differentiation but is dispensable for mESC self-renewal (Fernandez-Alonso *et al.*, 2017). Consistent with the meso/endodermal defect in differentiation of setd3Δ mESCs, the interaction of SETD3 with BRD2 might be critical for BRD2 function at pluripotency exit. Further studies will focus on how SETD3-BRD2 interaction might regulate its function in mESCs.

TFCP2L1 is part of the core pluripotency TF network in ESCs (Chen *et al.*, 2008), downstream of both LIF-STAT3 and Wnt-β catenin pathways (Martello, Bertone and Smith, 2013; Onishi and Zandstra, 2015; Sun *et al.*, 2018). TFCP2L1 directly binds OCT4 and co-regulates its target genes (Berg *et al.*, 2010; Ye *et al.*, 2013; Kim, Jang and Park, 2021). Its critical role on the maintenance of ESC self-renewal is through direct regulation of Nanog expression (Ye *et al.*, 2013). We did not observe any changes in Nanog level in setd3Δ mESCs, suggesting SETD3-TFCP2L1 interaction is not part of this regulatory loop. SETD3 loss did not alter TFCP2L1 level either (Figure 3.9.).

As a pluripotency TF, TFCP2L1 prevents expression of lineage specific genes in mESCs (Ye *et al.*, 2013; Liu *et al.*, 2017). The rapid decline in TFCP2L1 level upon differentiation suggests its function is limited to mESCs and/or right at the exit from pluripotency. Along with SETD3, our data also shows an interaction between TFCP2L1 and BRD2 in mESCs, suggesting three proteins might work together (Figure 3.12. and Figure 3.13.). TFCP2L1 target genes in mESCs might also be bound by Brd2 since Brd2 is recruited to active genes in mESCs (Surface *et al.*, 2016). Further studies are needed to uncover the role of SETD3 in gene regulatory functions of TFCP2L1 and BRD2 in mESCs.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

In our previous studies, it was identified that SETD3 is essential for meso-endodermal differentiation of mESCs (Figure 1.1). Although, the recent studies focused on the cytoplasmic function of SETD3, which is methylation of actin, we did not see any effect of SETD3 in F-actin formation and stabilization in mESCs (Figure 3.2.). The reason could be the low stiffness of ESCs. After the ESCs continue differentiation, the stiffness of the cells increases, so the filamentous actin becomes more visible (Mote *et al.*, 2020). Therefore, we hypothesized that as an epigenetic factor, the effect of SETD3 on mESC meso-endoderm differentiation could be related with its nuclear function (Cohn *et al.*, 2016; Kwiatkowski *et al.*, 2018) In this study, we aimed to identify proteomic partners of SETD3 in mESCs to understand the function of SETD3 itself. We have selected around a dozen of candidate proteins among thousands of proteins that are obtained from IP-MS studies. Br considering different criteria, we performed a deep bioinformatic analysis. Firstly, we selected nuclear proteins and then we search for the known complex knowledge about these proteins. We assume that if a complex member interacting with SETD3, other members could also be present in MS data. Then, we investigated the biological pathways that these proteins belong to and we focused on the pathways that are related with chromatin modification, regulation etc. After we performed a deep literature research about the proteins, we obtained around a dozen of candidate proteins for biochemical analysis. We have showed the interaction of SETD3 with two proteins, BRD2 and TFCP2L1, both are highly related with mESC pluripotency and differentiation (Figure 3.6, Figure 3.10 and Figure 3.11). In addition, we also showed that BRD2 and TFCP2L1 are interacting with each other independent from SETD3 (Figure 3.12 and Figure 3.13). These results can suggest that these proteins

could work together. Because of these interactions mainly located in the nucleus, this may show the nuclear importance of SETD3.

SETD3 is a methyltransferase, BRD2 is a reader protein and TFCP2L1 is a TF. Both of these three proteins contain a DNA binding domain and it is already known that they are specifically bind the genes on the genome. (Eom *et al.*, 2011; Ye *et al.*, 2013; Surface *et al.*, 2016) Therefore, as a future direction, we can investigate the common binding regions of these proteins in the genome. Then, we could analyze whether this interaction is functional in mESC pluripotency or not.

REFERENCES

Abaev-schneiderman, E., Admoni-elisha, L. and Levy, D. (2019) 'SETD3 is a positive regulator of DNA- damage-induced apoptosis', *Cell Death and Disease*. Springer US. doi: 10.1038/s41419-019-1328-4.

Alam, M. S. (2018) 'Proximity Ligation Assay (PLA)', pp. 1–8. doi: 10.1002/cpim.58.

Bannister, A. J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Research*. Nature Publishing Group, 21(3), pp. 381–395. doi: 10.1038/cr.2011.22.

Berg, D. L. C. Van Den *et al.* (2010) 'Resource An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells', *Stem Cell*. Elsevier Ltd, 6(4), pp. 369–381. doi: 10.1016/j.stem.2010.02.014.

Betschinger, J. *et al.* (2013) 'Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3', *Cell*. Elsevier Inc., 153(2), pp. 335–347. doi: 10.1016/j.cell.2013.03.012.

Bilodeau, S. *et al.* (2009) 'SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state', pp. 2484–2489. doi: 10.1101/gad.1837309.Cole.

Brack, A. S. *et al.* (2009) 'BCL9 is an essential component of canonical Wnt signaling that mediates the differentiation of myogenic progenitors during muscle regeneration', 335, pp. 93–105. doi: 10.1016/j.ydbio.2009.08.014.

Çağlayan, A. O. *et al.* (2017) 'ALPK3 gene mutation in a patient with congenital cardiomyopathy and dysmorphic features', *Cold Spring Harbor molecular case studies*, 3(5), pp. 1–9. doi: 10.1101/mcs.a001859.

Ceballos-chávez, M. *et al.* (2012) 'Control of neuronal differentiation by sumoylation of BRAF35 , a subunit of the LSD1-CoREST histone demethylase

complex', 109(21), pp. 8085–8090. doi: 10.1073/pnas.1121522109.

Cha, Y. *et al.* (2013) 'Tcea3 regulates the vascular differentiation potential of mouse embryonic stem cells', *Gene Expression*, 16(1), pp. 25–30. doi: 10.3727/105221613X13776146743343.

Chen, X. *et al.* (2008) 'Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells', *Cell*, 133(6), pp. 1106–1117. doi: 10.1016/j.cell.2008.04.043.

Cheng, X. *et al.* (2017) 'Cell cycle-dependent degradation of the methyltransferase SETD3 attenuates cell proliferation and liver tumorigenesis', 292, pp. 9022–9033. doi: 10.1074/jbc.M117.778001.

Cohn, O. *et al.* (2016) 'Chromatin associated SETD3 negatively regulates VEGF expression', *Scientific Reports*. Nature Publishing Group, 6(October), pp. 1–10. doi: 10.1038/srep37115.

Diep, J. *et al.* (2019) 'prevents primary dystocia', *Nature*. Springer US. doi: 10.1038/s41586-018-0821-8.

Dillon, S. C. *et al.* (2005) 'The SET-domain protein superfamily: Protein lysine methyltransferases', *Genome Biology*, 6(8). doi: 10.1186/gb-2005-6-8-227.

Eom, G. H. *et al.* (2011) 'Histone Methyltransferase SETD3 Regulates Muscle', 286(40), pp. 34733–34742. doi: 10.1074/jbc.M110.203307.

Fernandez-Alonso, R. *et al.* (2017) 'Brd4-Brd2 isoform switching coordinates pluripotent exit and Smad2-dependent lineage specification', *EMBO reports*, 18(7), pp. 1108–1122. doi: 10.15252/embr.201643534.

Gaspar-Maia, A. *et al.* (2011) 'Open chromatin in pluripotency and reprogramming', *Nature Reviews Molecular Cell Biology*, 12(1), pp. 36–47. doi: 10.1038/nrm3036.

Guo, Q. *et al.* (2019) 'Structural insights into SETD3-mediated histidine methylation on b -actin', 73, pp. 1–20.

Gyuris, A. *et al.* (2009) 'Biochimica et Biophysica Acta The chromatin-targeting protein Brd2 is required for neural tube closure and embryogenesis', *BBA - Gene Regulatory Mechanisms*. Elsevier B.V., 1789(5), pp. 413–421. doi: 10.1016/j.bbagrm.2009.03.005.

Hancock, G. V (no date) 'Divergent roles for KLF4 and TFCP2L1 in Naive Ground State Pluripotency and Human Primordial Germ Cell Development CURRENT STATUS : POSTED', pp. 1–30. doi: 10.21203/rs.3.rs-24873/v1.

Jiang, X. *et al.* (2018) 'SETD3 negatively regulates VEGF expression during hypoxic pulmonary hypertension in rats', *Hypertension Research*. Springer US, 41(9), pp. 691–698. doi: 10.1038/s41440-018-0068-7.

Job, A. *et al.* (2018) 'Inactivation of PRIM1 Function Sensitizes Cancer Cells to ATR and CHK1 Inhibitors', *Neoplasia (United States)*. The Authors, 20(11), pp. 1135–1143. doi: 10.1016/j.neo.2018.08.009.

Kim, C. M., Jang, T. and Park, H. H. (2021) 'Functional Analysis of CP2-Like Domain and SAM-Like Domain in TFCP2L1 , Novel Pluripotency Factor of Embryonic Stem Cells', *Applied Biochemistry and Biotechnology*. Applied Biochemistry and Biotechnology, (2016), pp. 650–658. doi: 10.1007/s12010-016-2021-z.

Kim, D. *et al.* (2014) 'Characterization of a Novel Histone H3K36 Methyltransferase setd3 in Zebrafish Characterization of a Novel Histone H3K36 Methyltransferase setd3', 8451(May). doi: 10.1271/bbb.100648.

Kim, J. *et al.* (2009) 'Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells', *Nature Protocols*, 4(4), pp. 506–517. doi: 10.1038/nprot.2009.23.

Kwiatkowski, S. *et al.* (2018) 'SETD3 protein is the actin-specific histidine N - methyltransferase', 3(2016), pp. 1–42.

Leroy, G., Rickards, B. and Flint, S. J. (2008) 'Article The Double Bromodomain

Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription', pp. 51–60. doi: 10.1016/j.molcel.2008.01.018.

Lewis, M. T. *et al.* (2001) 'The Gli2 transcription factor is required for normal mouse mammary gland development', *Developmental Biology*, 238(1), pp. 133–144. doi: 10.1006/dbio.2001.0410.

Lewis, P. W. *et al.* (2010) 'Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres', *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), pp. 14075–14080. doi: 10.1073/pnas.1008850107.

Liang, G. and Zhang, Y. (2013) 'Embryonic stem cell and induced pluripotent stem cell: An epigenetic perspective', *Cell Research*. Nature Publishing Group, 23(1), pp. 49–69. doi: 10.1038/cr.2012.175.

Liu, K. *et al.* (2017) '*TFCP2L1* represses multiple lineage commitment of mouse embryonic stem cells through MTA1 and LEF1', *Journal of Cell Science*, 130(22), pp. 3809–3817. doi: 10.1242/jcs.206532.

Lykke-Andersen, K. *et al.* (2003) 'Disruption of the COP9 Signalosome Csn2 Subunit in Mice Causes Deficient Cell Proliferation, Accumulation of p53 and Cyclin E, and Early Embryonic Death', *Molecular and Cellular Biology*, 23(19), pp. 6790–6797. doi: 10.1128/mcb.23.19.6790-6797.2003.

Martello, G., Bertone, P. and Smith, A. (2013) 'Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor', *EMBO Journal*. Nature Publishing Group, 32(19), pp. 2561–2574. doi: 10.1038/emboj.2013.177.

Mo, R. *et al.* (1997) 'Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development.', *Development*, 124(1), pp. 113–23. doi: 10.1097/QAD.0b013e328359b99b.

Mote, R. D. *et al.* (2020) 'JBC Papers in Press . Published on October 21 , 2020 as Manuscript AC120 . 014343 The latest version is at

https://www.jbc.org/cgi/doi/10.1074/jbc.AC120.014343'. doi: 10.1074/jbc.AC120.014343.

Nourhan, H. *et al.* (2020) 'SETD3 acts as a prognostic marker in breast cancer patients and modulates the viability and invasion of breast cancer cells', pp. 1–16. doi: 10.1038/s41598-020-59057-5.

Onishi, K. and Zandstra, P. W. (2015) 'LIF signaling in stem cells and development', pp. 2230–2236. doi: 10.1242/dev.117598.

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

Paranjpe, S. S. and Veenstra, G. J. C. (2015) 'Biochimica et Biophysica Acta Establishing pluripotency in early development', *BBA - Gene Regulatory Mechanisms*. Elsevier B.V., 1849(6), pp. 626–636. doi: 10.1016/j.bbagrm.2015.03.006.

Park, K. *et al.* (2014) 'Tibia fractuur (gebroken scheenbeen)', 31(2), pp. 282–292. doi: 10.1002/stem.1284.

Qiu, D. *et al.* (2015) 'Klf2 and Tfcp2l1, Two Wnt/β-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency', *Stem Cell Reports*. The Authors, 5(3), pp. 314–322. doi: 10.1016/j.stemcr.2015.07.014.

Rapkin, L. M. *et al.* (2015) 'The histone chaperone DAXX maintains the structural organization of heterochromatin domains', *Epigenetics and Chromatin*. BioMed Central, 8(1), pp. 1–14. doi: 10.1186/s13072-015-0036-2.

Schaniel, C. *et al.* (2018) 'HHS Public Access', 27(12), pp. 2979–2991. doi: 10.1002/stem.223.Smarcc1/Baf155.

Shang, E. *et al.* (2009) 'Double Bromodomain-Containing Gene Brd2 Is Essential for Embryonic Development in Mouse', (March), pp. 908–917. doi: 10.1002/dvdy.21911.

Van Sligtenhorst, I. *et al.* (2012) 'Cardiomyopathy in α-kinase 3 (Alpk3)-deficient mice', *Veterinary Pathology*, 49(1), pp. 131–141. doi: 10.1177/0300985811402841.

Sun, H. *et al.* (2018) 'Tfcp2l1 safeguards the maintenance of human embryonic stem cell self-renewal', (January), pp. 6944–6951. doi: 10.1002/jcp.26483.

Surface, L. E. *et al.* (2016) 'H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain Poised Chromatin in ESCs', *Cell Reports*. The Authors, 14(5), pp. 1142–1155. doi: 10.1016/j.celrep.2015.12.100.

Takaoka, K. and Hamada, H. (2012) 'Cell fate decisions and axis determination in the early mouse embryo', *Development*, 139(1), pp. 3–14. doi: 10.1242/dev.060095.

Taniguchi, Y. (2016) 'The bromodomain and extra-terminal domain (BET) family: Functional anatomy of BET paralogous proteins', *International Journal of Molecular Sciences*, 17(11). doi: 10.3390/ijms17111849.

Terzi Cizmecioglu, N. *et al.* (2020) 'Arid4b is critical for mouse embryonic stem cell differentiation towards mesoderm and endoderm, linking epigenetics to pluripotency exit', *Journal of Biological Chemistry*, p. jbc.RA120.015534. doi: 10.1074/jbc.ra120.015534.

Xu, F., Flowers, S. and Moran, E. (2012) 'Essential role of ARID2 protein-containing SWI/SNF complex in tissue-specific gene expression', *Journal of Biological Chemistry*, 287(7), pp. 5033–5041. doi: 10.1074/jbc.M111.279968.

Yao, M. *et al.* (2018) 'PCGF5 is required for neural differentiation of embryonic stem cells', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-03781-0.

Ye, S. *et al.* (2013) 'Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1', *The EMBO Journal*. Nature Publishing Group, 32(19), pp. 2548–2560. doi: 10.1038/emboj.2013.175.

Young, R. A. (2011) 'Review Control of the Embryonic Stem Cell State', *Cell*. Elsevier Inc., 144(6), pp. 940–954. doi: 10.1016/j.cell.2011.01.032.

Zhang, W. *et al.* (2016) ‘Cops2 promotes pluripotency maintenance by Stabilizing Nanog Protein and Repressing Transcription’, *Scientific Reports*. Nature Publishing Group, 6(May), pp. 4–12. doi: 10.1038/srep26804.

Zheng, Y. (2020) ‘Molecular basis for histidine N3-speci fi c methylation of actin H73 by SETD3’, *Cell Discovery*. Springer US, pp. 4–7. doi: 10.1038/s41421-019-0135-5.

APPENDICES

A. Materials Used for Cell Culture

Table A.1. *The recipe of 2i4 medium and its supplement.*

Reagent	***Brand/ Cat. No***	***Final concentration/ percentage***	***Volume for needed for 100 ml medium***
Neurobasal	Thermo Fisher / 21103049	50%	50 ml
DMEM/F12	Thermo Fisher / 11320074	50%	50 ml
N2	Thermo Fisher / 17502048	0.5%	500 µl
B27 with RA	Thermo Fisher / 17504044	1%	1 ml
10% BSA (in PBS)	Sigma / A3311-50G	0.5%	500 µl
GLUTAMAX	Thermo Fisher / 35050038	1%	1 ml
PEN/STREP	Thermo Fisher / 15140122	1%	1 ml
MTG	Sigma / M6145-25ML	1.5 x 10-4 M	1.3 µl
FBS	Gibco/10500064	4%	4 ml
Supplement for 100 ml			
LIF	Millipore / ESG1107	10^5 units/ml	10 µl
CHIR-99021	S1263-5MG	3 µM	30 µl
PD0325901	Selleckchem / S1036-5MG	1 µM	10 µl

B. The Information of Antibodies Used in the Experiments

Table B.1. *The information of antibody used in ICC and PLA.*

Protein Name	***Brand/ Cat. No***	***Host***	***Reactivity***	***Dilution***
Setd3	Novus/ NBP2-32136	rabbit	mouse, human	1:250
Flag	Sigma/ F1804	mouse	all	1:250
Brd2	CST/ #5848	rabbit	mouse, human	1:250
Tfcp2l1	R&D/ AF5726	goat	human	1:100
Anti-rabbit IgG H&L (Alexa Fluor® 594)	Abcam/ ab150080	goat	rabbit	1:1000

Table B.2. *The information of antibody used in co-IPs*

Protein Name	***Host***	***Cat. No.***	***Stock concentration***	***Amount of Antibody used in IP***
Setd3	rabbit	NBP2-32136	1 mg/ml	5 µg / 400 µg
Brd2	rabbit	#5848	1 mg/ml	5 µg / 400 µg
Tfcp2l1	goat	AF5726	200 µg /0.5 ml	5 µg / 400 µg
Normal IgG	rabbit	12-370	1 mg/ml	5 µg / 400 µg
Normal IgG	goat	AB-108-C	1 mg/ml	5 µg / 400 µg

Table B.3. *The information of antibody used in western blots*

Protein Name	***Brand/ Cat. No***	***Host***	***Reactivity***	***Dilution***
Setd3	Novus/ NBP2-32136	rabbit	mouse, human	1:2000
Brd2	CST/ #5848	rabbit	mouse, human	1:1000
Brd3	Santa Cruz/ sc-81202	mouse	mouse, rat, human	1:500
Tfcp2l1	R&D/ AF5726	goat	human	1:1000
Tcea3	Santa Cruz/ sc-365894	mouse	mouse, rat, human	1:500
Cops2	Santa Cruz/ sc-136446	mouse	mouse, rat, human	1:500
Prim1	Santa Cruz, sc-390265	mouse	mouse, rat, human	1:500
Histone H3	Santa Cruz/ sc-10809	mouse	mouse, rat, human	1:1000
Lamin B1	Abcam/ ab133741)	rabbit	mouse, rat, human	1:1000
GAPDH	CST/ 2118S	rabbit	all	1:3000
Anti-rabbit IgG H&L (HRP)	Abcam/ab97051	goat	rabbit	1:5000
Anti-mouse IgG H&L (HRP)	Abcam/ab97023	goat	mouse	1:5000
Veriblot	Abcam/ ab131366		all	1:200

C. The Information of Primers Used in qRT-PCR

Table C.1. *Forward and reverse designed sequences of primers specific to candidate proteins for qRT-PCR and PubMed IDs (PMIDs) of articles obtained primer sequence. *Primer sequences are represented 5' to 3' direction.*

Primer Name	***Sequence (5'-> 3')***	***PMID***
β-actin_F_qPCR	ATGAAGATCCTGACCGAGCG	14998924
β-actin_R_qPCR	TACTTGCGCTCAGGAGGAGC	14998924
Brd2_F_qPCR	GCTGTGAAGCTGGGTTTGC	30134146
Brd2_R_qPCR	AATCCTGCATACATTCTGAGGCA	30134146
Arid2_F_qPCR	GCAGCCAATTTCCACTCCTGTTG	22184115
Arid2_R_qPCR	GATTGGTGACAGGAGTCCTCTG	22184115
Bcl9_F_qPCR	AGCCCTAAGTCAAAGCAGGAGG	28735896
Bcl9_R_qPCR	CATTTCCAGCCCCATTCTTCAG	28735896
Cops2_F_qPCR	CATCCCTCACCCACTAATCATG	27226076
Cops2_R_Qpcr	TCTTGGGCTTCCTGATTCATC	27226076
Daxx_F_Qpcr	CAGCAAGAGATTTCGGAAGG	30335163
Daxx_R_qPCR	GGAGGAATCAGCGACAGAAG	30335163
Gli2_F_Qpcr	CAACGCCTACTCTCCCAGAC	22199256
Gli2_R_qPCR	GAGCCTTGATGTACTGTACCAC	22199256
Hmg20a_F_qPCR	AACCAACCCAGAGTTTGTGG	29449530
Hmg20a_R_qPCR	TTGCTCATCTTCAGGCCTTT	29449530
Prim1_F_qPCR	AGGCGCAGTATATTCTCACAGA	20661276
Prim1_R_qPCR	CATGTCGATGTCAAAGACCAGT	20661276
Ring1_F_qPCR	CCTGGACATGCTGAAGAACA	27181215
Ring1_R_qPCR	GCAGTCCGAGCAGAACCTAT	27181215
Tcea3_F_qPCR	CAGGAACCCAGGCCTGAGGCG	23169579
Tcea3_R_qPCR	GCTACACCGCAAGAGGTCAGTGG	23169579
Tfe3_F_qPCR	GAACGACGCAGGCGATTCAACATT	25736533
Tfe3_R_qPCR	ATCCACAGATGCCTTCAGGATGGT	25736533
Tfcp2l1_F_qPCR	AGGTGCTGACCTCCTGAAGA	23942238
Tfcp2l1_R_qPCR	GTTTTGCTCCAGCTCCTGAC	23942238

D. List of Candidate Proteins Obtained from MS Analysis

Table D.1. *Common proteins found in two biological replicates*

Uniprot ID	Gene Symbol	Uniprot ID	Gene Symbol	Uniprot ID	Gene Symbol
Q91WC0	Setd3	P61202	Cops2	Q91ZT5	Fgd4
Q9EQJ9	Magi3	Q80Y83	Dixdc1	P23881	Tcea3
O35638	Stag2	Q32P12	C17orf53 homolog	Q80VD1	Fam98b
Q9D219	Bcl9	O35613	Daxx	Q5UBV8	Tnfsf15
Q91X43	Sh3d19	Q3UW53	Fam129a	Q810D6	Grwd1
Q6DFV3	Arhgap21	Q06335	Aplp2	Q4PZA2	Ece1
Q8BXK8	Agap1	Q3UZ01	Rnpc3	Q8R5L1	C1qbp
Q9DC33	Hmg20a	Q64092	Tfe3	P50396	Gdi1
E9Q7E2	Arid2	Q8BTJ4	Enpp4	P12367	Prkar2a
Q924W7	St5	Q7JJ13	Brd2	B2RXS4	Plxnb2
Q3UHC0	Tnrc6c	F8VPZ9	Gltscr1	O35730	Ring1
Q80Y19	Arhgap11a	E9Q735	Ube4a	Q8VC51	Wrap53
Q99KH8	Stk24	Q6P9J5	Kank4	A0A087WNW3	Ktn1
Q9CQJ2	Pih1d1	Q9JI08	Bin3	O55106	Strn
P20664	Prim1	E9Q9C6	Fcgbp	Q91Z50	Fen1
Q3UMQ8	Naf1	Q61249	Igbp1	A0A087WNZ7	Trip12

Q8C4S8	Dennd2a	Q924A2	Cic	Q921I9	Exosc4
Q3TKY6	Cwc27	F6ZGN3	Myo18a	Q8K4F6	Nsun5
O35127	Grcc10	O88792	F11r	Q60899	Elavl2
E9Q842	Nav2	Q8CBE3	Wdr37	D3Z0K6	Rsbn1l
Q8CBY8	Dctn4	Q7TN74	Ppp1r9a	Q80TT8	Cul9
Q924C5	Alpk3	Q9JJT0	Rcl1	E9QN52	Lrrfip2
Q0VGT2	Gli2	Q8K212	Pacs1	Q62179	Sema4b
O35295	Purb	P12815	Pdcd6	Q9R1Q9	Atp6ap1
Q8BKI2	Tnrc6b	Q76LL6	Fhod3	Q9Z247	Fkbp9
Q8R317	Ubqln1	E9Q6I3	Shroom3		
A0A087WRB8	Dst	Q8BHR2	Enox1		

Table D.2. *Nuclear localized candidate proteins*

Uniprot ID	**Gene Symbol**	**Uniprot ID**	**Gene Symbol**	**Uniprot ID**	**Gene Symbol**
Q9EQJ9	Magi3	Q3UZ01	Rnpc3	A0A087WRB8	Dst
Q91WK0	Lrrfip2	Q7JJ13	Brd2	P23881	Tcea3
O35638	Stag2	Q8C4S8	Dennd2a	F8VPZ9	Gltscr1
Q8BXK8	Agap1	E9Q7E2	Arid2	Q80VD1	Fam98b
Q80Y19	Arhgap11a	Q810D6	Grwd1	Q76LL6	Fhod3
Q9D219	Bcl9	E9Q842	Nav2	A0A087WNZ7	Trip12
Q99KH8	Stk24	Q924C5	Alpk3	Q8R5L1	C1qbp
Q9CQJ2	Pih1d1	Q3UMQ8	Naf1	P50396	Gdi1
Q80Y83	Dixdc1	O35127	Grcc10	Q91Z50	Fen1
Q924W7	St5	Q64092	Tfe3	Q8VC51	Wrap53
Q3UHC0	Tnrc6c	O35295	Purb	O35730	Ring1
Q91X43	Sh3d19	Q8BKI2	Tnrc6b	Q9JJT9	Phax
Q0VGT2	Gli2	Q8K212	Pacs1	Q8CBE3	Wdr37
P20664	Prim1	E9Q735	Ube4a	D3Z0K6	Rsbn1l
Q9DC33	Hmg20a	Q06335	Aplp2	E9QN52	Lrrfip2
Q8CBY8	Dctn4	Q921I9	Exosc4	P39749	Fen1
Q7TN74	Ppp1r9a	Q6P9J5	Kank4	Q80TT8	Cul9
Q3TKY6	Cwc27	P61202	Cops2	Q60899	Elav2
O35613	Daxx	P12815	Pdcd6	Q8K4F6	Nsun5

Table D.3. *Complex knowledge of the candidate proteins*

UniProt ID	Gene Symbol	Complex Name	UniProt ID	Gene Symbol	Complex Name
Q8BXK8	Agap1	Agap11-AP3 Complex	Q3UNW5	Tfcp2l1	No Known Nuclear Complex
O35638	Stag2	SNF2h-cohesin-NuRD Complex	Q06335	Aplp2	No Known Nuclear Complex
Q9D219	Bcl9	Beta-catenin-B-cell lymphoma 9 protein Complex	Q3UMQ8	Naf1	No Known Nuclear Complex
P20664	Prim1	DNA centrosome Complex	Q64092	Tfe3	No Known Nuclear Complex
E9Q7E2	Arid2	PBAF Complex	Q7JJ13	Brd2	No Known Nuclear Complex
Q9DC33	Hmg20a	LSD1 Complex	F8VPZ9	Gltscr1 (Bicra)	No Known Nuclear Complex
Q3TKY6	Cwc27	C Spliceosome Complex	E9Q735	Ube4a	No Known Nuclear Complex
Q8CBY8	Dctn4	Actin Related Protein Complex	Q6P9J5	Kank4	No Known Nuclear Complex
Q0VGT2	Gli2	Gli2-Kif7 Complex	Q8C4S8	Dennd2a	No Known Nuclear Complex
P61202	Cops2	CSA-POLIIa Complex	Q8CBE3	Wdr37	No Known Nuclear Complex
Q80Y83	Dixdc1	Ccd1-Dvl2 complex	Q7TN74	Ppp1r9a	No Known Nuclear Complex
O35613_	Daxx	ATRX-DAXX Complex, DAXX-AXIN Complex	Q91X43	Sh3d19	No Known Nuclear Complex
Q3UZ01	Rnpc3	18S U11/U12 snRNP	O35127	Grcc10	No Known Nuclear Complex
Q924A2	Cic	Ataxin-1-Capicua Complex	E9Q842	Nav2	No Known Nuclear Complex

Q9JJT9	Phax	PHAX-CBC Complex	Q76LL6	Fhod3	No Known Nuclear Complex
Q8K212	Pacs1	PKD2-PACS1 Complex	Q91ZT5	Fgd4	No Known Nuclear Complex
P12815	Pdcd6	p23 Protein Complex	P23881	Tcea3	No Known Nuclear Complex
Q80VD1	Fam98b	tRNA Splicing Ligase Complex	Q924W7	St5	No Known Nuclear Complex
Q810D6	Grwd1	CUL4B-DDB1-GRWD1 Complex	Q924C5	Alpk3	No Known Nuclear Complex
Q8R5L1	C1qbp	Emerin Complex	Q80Y19	Arhgap11a	No Known Nuclear Complex
O35730	Ring1	RING1-Polycomb Repressive Complex	P50396	Gdi1	No Known Nuclear Complex
Q8VC51	Wrap53	Telomerase Holoenzym	O35295	Purb	No Known Nuclear Complex
Q91Z50	Fen1	9-1-1-FEN1 Complex	Q8BKI2	Tnrc6b	No Known Nuclear Complex
Q8R317	Ubqln1	No Known Nuclear Complex	Q9EQJ9	Magi3	No Known Nuclear Complex
A0A087WRB8	Dst	No Known Nuclear Complex	A0A087WNZ7	Trip12	No Known Nuclear Complex
Q99KH8	Stk24	No Known Nuclear Complex	Q921I9	Exosc4	No Known Nuclear Complex
Q9CQJ2	Pih1d1	No Known Nuclear Complex	Q8K4F6	Nsun5	No Known Nuclear Complex
Q3UHC0	Tnrc6c	No Known Nuclear Complex	Q60899	Elavl2	No Known Nuclear Complex
Q80TT8	Cul9	No Known Nuclear Complex	D3Z0K6	Rsbn1l	No Known Nuclear Complex
E9QN52	Lrrfip2	No Known Nuclear Complex			