Understanding How Loss of Histone H3 Lysine 36 Trimethylation Enhances Somatic Cell Reprogramming

by

Büşra Bayırbaşı

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Molecular Biology and Genetics

December 23, 2021

Understanding How Loss of Histone H3 Lysine 36 Trimethylation Enhances Somatic Cell Reprogramming

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Büşra Bayırbaşı

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Prof. Tamer Önder (Advisor)

Assoc. Prof. Tolga Emre

Asst. Prof. Gözde Korkmaz

Date: 23.12.2021

ABSTRACT

Understanding How Loss of Histone H3 Lysine 36 Trimethylation Enhances Somatic Cell Reprogramming

Büşra Bayırbaşı

Master of Science in Molecular Biology and Genetics

23.12.2021

Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) using a set of transcription factors (OCT4, SOX2, KLF4 and c-MYC). However, reprogramming of somatic cells to iPSCs suffers from a very low efficiency indicating the presence of barriers against reprogramming. Epigenetic factors regulating post-translational modifications of DNA and histones are among such barriers of reprogramming. In a preliminary study, SETD2 which is the only enzyme catalyzing histone H3 lysine 36 tri-methylation (H3K36me3) in mammals was shown to be a barrier of reprogramming. H3K36me3 is a histone mark deposited throughout gene bodies of actively transcribed genes. In this thesis, I investigated the molecular mechanisms of how H3K36me3 loss enhances iPSC generation.

I proposed several hypotheses to understand how H3K36me3 loss increases reprogramming efficiency. First, by analyzing previously generated RNA-sequencing data, I considered four transcription factors (CITED2, EBF3, SMAD3 and FOSL1) downregulated by SETD2 knockdown as candidate SETD2-downstream genes. I tested whether overexpression of these transcription factors could reverse the increased reprogramming phenotype of SETD2 knockdown. I showed that none of these transcription factors could alone decrease the increased reprogramming efficiency by SETD2 knockdown.

In a second line of experiments, I carried out chromatin immunoprecipitation (ChIP) for active H3K36me3 and repressive H3K27me3 marks, which are deposited in a mutually exclusively manner on the tails of histone H3. I analyzed the distribution of H3K27me3 on the genes downregulated upon SETD2 knockdown to investigate whether H3K36me3 loss could lead to H3K27me3 deposition. However, ChIP-qPCR results did not indicate a reciprocal relationship between these two marks on the selected genes such as SMAD3, LPAR1, EBF3 and SFRP1.

Lastly, I performed a CRISPR/Cas9-based knockout screen targeting all known H3K36me3 readers in mammals to define the roles of these readers in somatic cell reprogramming. H3K36me3 readers are involved in diverse cellular processes such as alternative splicing, DNA repair, transcription elongation and DNA and histone methylation. Therefore, I aimed to find which H3K36me3 reader could phenocopy SETD2 depletion in reprogramming. CRISPR-based screen demonstrated that PSIP1, MRG15, MSH6, MSL3, NSD2 and NSD3 are barriers against reprogramming as their knockout led to increased reprogramming efficiency. These H3K36me3 readers were validated as barriers to reprogramming and MRG15, the H3K36me3 reader associated with regulation of alternative splicing was chosen for function-related assays. Therefore, I analyzed the effect of SETD2 knockdown on alternative splicing of key factors known to be significant for reprogramming with the goal of understanding whether SETD2 inhibition promotes reprogramming by regulating alternative splicing. Through RT-qPCR analysis, I observed that both SETD2 knockdown and MRG15 knockout individually led to a significant increase in the pluripotent cell-specific variant of MBD2 expression. Taken together, these findings identified the factors serving as barrier against somatic cell reprogramming through reading trimethylation of H3K36 residue and alternative splicing of MBD2 gene as an underlying mechanism for enhanced reprogramming by H3K36me3 loss.

ÖZETÇE

Histon H3 Lizin 36 Üç-metil Kaybının Somatik Hücre Yeniden Programlamasını Nasıl Geliştirdiğinin Anlaşılması

Büşra Bayırbaşı

Moleküler Biyoloji ve Genetik, Yüksek Lisans

23 Aralık 2021

Somatik hücreler, bir dizi transkripsiyon faktörü (OCT4, SOX2, KLF4 ve c-MYC) kullanılarak uyarılmış pluripotent kök hücrelere (UPKH) yeniden programlanabilir. Bununla birlikte, somatik hücrelerin UPKH'lere yeniden programlanması, yeniden programlamaya karşı engellerin varlığını gösteren çok düşük verimlilikten mustariptir. DNA ve histonların translasyon sonrası modifikasyonlarını düzenleyen epigenetik faktörler, yeniden programlamanın bu tür engelleri arasındadır. Ön çalışmamızda, histon H3 lizin 36 üç-metillenmesini (H3K36me3) katalize eden memelilerdeki tek enzim SETD2'nin yeniden programlamanın bir bariyeri olduğu gösterilmiştir. H3K36me3, aktif transkripsiyon gösteren genlerin gen gövdeleri boyunca biriken bir histon işaretidir. Bu tezde, H3K36me3 kaybının UPKH oluşumunu nasıl geliştirdiğinin moleküler mekanizmalarını araştırdım.

H3K36me3 kaybının yeniden programlama verimliliğini nasıl artırdığını anlamak için birkaç hipotez önerdim. İlk olarak, ön çalışmamızdaki RNA dizileme analiziyle belirlenen SETD2 baskılanmasıyla ekspresyonu azalan genler arasından dört transkripsiyon faktörünü (CITED2, EBF3, SMAD3 ve FOSL1) yeniden programlamada test ettim. Bu transkripsiyon faktörlerinin aşırı ekspresyonunun, SETD2 baskılanmasıyla artan yeniden programlama fenotipini tersine çevirip çeviremeyeceğini test ettim. Bu transkripsiyon faktörlerinin hiçbirinin SETD2 baskılanmasıyla artan yeniden programlama verimliliğini tek başına azaltamayacağını gösterdim.

Deneylerin ikinci kısmında, birbirini dışlayarak histon H3'ün kuyruklarında biriken aktif H3K36me3 ve baskılayıcı H3K27me3 işaretleri için kromatin immünopresipitasyon (ChIP) gerçekleştirdim. H3K36me3 kaybının H3K27me3 birikimine yol açıp açmayacağını araştırmak için, SETD2 baskılanması ile ekspresyonu azalan genler üzerindeki H3K27me3 dağılımını analiz ettim. Ancak ChIP-qPCR sonuçları, SMAD3, LPAR1, EBF3 ve SFRP1 gibi seçilen genler üzerindeki bu iki işaret arasında karşıt bir ilişki olduğunu göstermedi.

Son olarak, yeniden programlamadaki rollerini anlamak amacıyla memelilerde bilinen tüm H3K36me3 okuyucularını hedefleyen CRISPR/Cas9 tabanlı bir tarama gerçekleştirdim. Hangi H3K36me3 okuyucusunun SETD2 tükenmesini feno-kopyalayacağını bulmayı amaçladım. Bu tarama, PSIP1, MRG15, MSH6, MSL3, NSD2 ve NSD3'ün yeniden programlamaya engel olduğunu gösterdi, çünkü bunların nakavtı yeniden programlama verimliliğini arttırdı. Bu okuyucular, yeniden programlamanın önündeki engeller olarak doğrulandı ve işleve yönelik testler için alternatif uç-birleştirme ile ilişkili H3K36me3 okuyucusu MRG15 seçildi. Bu nedenle, MRG15 baskılanmasının alternatif uç-birleştirme olaylarını etkileyerek yeniden programlamayı geliştirmiş olabileceğini test etmek amacıyla, yeniden programlama için önemli olduğu bilinen faktörlerin alternatif uç-birleştirmesi üzerinde SETD2 baskılanmasının etkisini analiz ettim. RT-qPCR analizi, hem SETD2 baskılanmasının hem de MRG15 nakavtının ayrı ayrı MBD2'nin pluripotent hücrelere özgü varyantının ekspresyonunda önemli bir artışa yol açtığını gösterdi. Bu sonuçlar, H3K36 üç-metillenmesini okuyarak somatik hücre yeniden programlamasına karşı bariyer oluşturan faktörleri ve H3K36me3 kaybının somatik hücre yeniden programlamasını nasıl geliştirdiğine altta yatan bir mekanizma olarak MBD2 geninin alternatif uç-birleştirmesini göstermiştir.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Prof. Tamer Önder for giving me the chance to work with him. I am grateful to him for everything he has patiently taught me in practice and theory, starting from the very basics, and for giving me the ability to think scientifically. He calmly accepted all the mistakes I made and encouraged me to keep going. In addition, I always felt his support by my side in the problems I faced in the lab and in my private life. It was not possible for me to complete my master's degree and thesis without his support and enthusiasm.

Also, I would like to thank my thesis committee members Assoc. Prof. Tolga Emre and Asst. Prof. Gözde Korkmaz for accepting to be my jury members and for their constructive and supportive comments on my data and future perspectives of the study.

I would like to thank all past and current Önder Lab members that I have worked with for two years. I would like to thank alumni Dr. Burcu Özçimen, who previously worked on this project, for all the work she had done for this project and for providing a comfortable working environment by leaving everything to me in a very organized and understandable way. I would like to thank Dr. Kenan Sevinç, Gülben Gürhan-Sevinç and Dr. Deniz Uğurlu-Çimen, from whom I learned most of the molecular biology and cell culture techniques, for their time and effort. I am especially grateful to Kenan, who always patiently and satisfactorily answered my questions. In addition, I cannot ignore the impact of our pleasant conversations in cell culture with Gülben on my better work. I would also like to thank my dear friend Dilşad Arabacı, for whom I started this path together, overcame the first difficulties together, and whose support I always felt. I would also like to thank Alperen Yılmaz, who has just joined us, but for whom I consulted many times while writing the thesis.

Three of my colleagues I worked with became like my family in the lab and they got very special place in my heart. I am grateful to Arda Odabaş, one of these people, who motivated me whenever I needed it, and always tried to find solutions to the problems I faced in the lab. I also learned a lot from all the scientific discussions we had as the result of the weird questions that came to his mind. Another member of this family, Ceyda Nur Altınışık, was the positive energy source of our lab with her personality even more colorful than her colorful socks. I would like to thank Ceyda for providing me with a synchronized and collaborative environment in the lab that I enjoyed a lot. Beyond our lab work, Ceyda's warm heart and always being by my side are very precious to me. And

the last but not least member of this family, Berfin Yeşil, is my dear friend with whom I shared a lot in a short time. She is the hero of protecting my psychological and mental health during my thesis writing period. I thank Berfin very much for being such a good friend.

I would like to thank Ayşenur Er, who has been my best friend since high school, for never leaving me alone even though we are far from each other. We chose Bilkent University together, so being by my side through all the difficulties I faced since the beginning of my academic life she has an important role in my success. I would like to thank my dear friend Gizem Sunar, for whom we overcame all the tasks such as exams, presentations, and lab reports during our undergraduate education. She also supported me a lot by sharing her experiences while writing my thesis. I am grateful to another friend of mine, Büşra Nur Ata, whom I met in my undergraduate life, for supporting me a lot by providing an environment in her house that motivated me especially during the times when I had difficulty in writing the thesis.

And of course, I cannot thank my family enough. My father, Adnan's endless support, respect, and faith in me had a very important place in my success throughout my academic life. Besides all her sacrifices and efforts, my mother, Saadet, always listened and supported me while I talked about my lessons and experiments even though she had difficulty in understanding the scientific language I spoke. And my brother, Ömer, supported me a lot with our chats and the activities we did to relieve my tiredness and stress when I went home at the end of the day.

Finally, I would like to thank the people whose support I felt especially during the difficult three months of thesis writing after my family moved to England. I am grateful to Güneş Family, with whom I stayed during this period, who made me feel like a part of their family rather than a guest. I would thank the couple Gülben and Kenan again, whose support I felt very much as they guided me like an older sister/brother throughout this process. I am very thankful to Arda for his helps and efforts to answer even my simplest questions every time during this period. And most of all, I am grateful to Ceyda and Berfin, who gave me strength with their moral support and being with me during this process. I also thank myself for not giving up and holding on to my work despite all the difficulties that life brings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

iPSC	Induced Pluripotent Stem Cell
ChIP	Chromatin Immunoprecipitation
SETD2	Set Domain Containing 2
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
RT-qPCR	Real Time Quantitative Polymerase Chain Reaction
ICM	Inner Cell Mass
ESC	Embryonic Stem Cells
EC	Embryonic Carcinoma
mESC	Mouse Embryonic Stem Cell
hESC	Human Embryonic Stem Cell
MEF	Mouse Embryonic Fibroblast
SCNT	Somatic Cell Nuclear Transfer
TF	Transcription Factor
OSKM	OCT4, SOX2, KLF4 and c-MYC
MET	Mesenchymal-to-Epithelial Transition
HDAC	Histone Deacetylase
NHEK	Neonatal Human Epidermal Keratinocyte
HAT	Histone Acetyltransferase
PRC2	Polycomb Repressive Complex 2
DOT1L	Disruptor of Telomeric Silencing 1-Like
P300/CBP	EP300/CREB-Binding Protein
SUMO	Small Ubiquitin-Related Modifier
NHDF	Normal Human Dermal Fibroblast
TET	Ten-Eleven Translocation
m^6A	N^6-methyladenosine
LPS	Lipopolysaccharide
SET	Associated with SET
AWS	Suppressor of variegation 3-9, Enhancer of Zeste, and Trithorax
SRI	Set2 Rpb1 Interacting
CTD	C-Terminal Domain

RNAPII	RNA Polymerase II
FACT	Facilitates Chromatin Transcription
SHI	SETD2-hnRNP Interaction
ccRCC	Clear Cell Renal Cell Carcinoma
DNMT3A/B	DNA (cytosine-5) Methyltransferase 3 Alpha/Beta
PHF1/19	PHD Finger Protein 1/19
MTF2	Metal Response Element Binding Transcription Factor 2
BRPF1	Bromodomain and PHD Finger Containing 1
MSH6	MutS Homolog 6
MUM1	Multiple Myeloma Oncogene 1
NSD1/2/3	Nuclear Receptor Binding SET Domain Protein 1/2/3
MSL3	Male-Specific Lethal Complex Subunit 3
GLYR1	Glyoxylate Reductase 1 Homolog
PSIP1	PC4 and SFRS1 Interacting Protein 1
MRG15	MORF-Related Gene 15
ZMYND11	Zinc Finger MYND-Type Containing 11
sgRNA	Single Guide RNA
DMSO	Dimethyl Sulfoxide
IGV	Integrative Genomics Viewer
DMEM	Dulbecco's Modified Eagle Medium
PBS	Phosphate-Buffered Saline
FBS	Fetal Bovine Serum
Ct	Threshold Cycle
cDNA	Complementary DNA
HRP	Horseradish Peroxidase
TPM	Transcripts Per Million
NT1	Non-Targeting 1
CITED2	Cbp/P300 Interacting Transactivator With Glu/Asp Rich Carboxy-Terminal Domain 2
EBF3	EBF Transcription Factor 3
SMAD3	SMAD Family Member 3
FOSL1	FOS Like-1, AP-1 Transcription Factor Subunit
MBD2	Methyl-CpG Binding Domain Protein 2

Chapter 1: INTRODUCTION

1.1 Embryonic Development

Embryonic development begins with fertilization of an egg with a sperm. The fertilized egg, also known as zygote, constitutes the first entity of life. Zygote has capacity to develop into the whole organism, so it is characterized as a totipotent cell (Taei et al., 2020). Zygote turns into morula by undergoing mitotic divisions. Totipotency is retained until the eight-cell stage of the morula. Subsequently, blastocyst is formed through cell differentiation. A blastocyst has two main parts: outer trophoblast cells and inner cell mass (ICM). ICM is composed of undifferentiated cells which are no longer totipotent but can form all types of embryonic tissues. This potency is further restricted in later developmental stages. Adult stem cells residing in specific tissues possess multipotency or unipotency and they can maintain the integrity of these tissues by differentiating into certain types of cells (Ratajczak et al., 2012). Eventually, fully differentiated somatic cells are formed.

1.2 Embryonic Stem Cells

ICM is the origin of human and mouse embryonic stem cells (ESCs). ESCs are able to differentiate into all three germ layers (endoderm, mesoderm and ectoderm) when properly grown in the cell culture. This developmental potential of ESCs is called pluripotency (Taei et al., 2020). In addition, ESCs have unlimited self-renewal capacity. Therefore, they can perpetuate themselves by dividing into more stem cells and maintaining their undifferentiated states (Shenghui et al., 2009).

1.2.1 History of Embryonic Stem Cells

Pluripotent stem cell research began in early 1970s with the establishment of embryonic carcinoma (EC) cell lines. EC cells are undifferentiated pluripotent cells found in teratocarcinomas, a malignant tumor derived from primordial germ cells consisting of multiple embryonic tissues (Stevens and Little, 1954; Stevens, 1970). When transplanted into extrauterine sites of mouse, these tumors led to benign teratomas (Stevens, 1970).

These studies indicated that EC cells residing in teratocarcinomas retained their pluripotency and could give rise to all three germ layers.

The first mouse embryonic stem cell (mESC) lines were generated in the early 1980s (Evans and Kaufman, 1981). The isolated early mouse embryos (blastocysts) were cultivated on mitomycin C-treated MEF feeder cells. Then, the ICMs were isolated by enzymatic dissociation (Evans and Kaufman, 1981). Cells that grew out under these culture conditions displayed similar properties to EC cells as they could proliferate indefinitely in culture. In addition, they led to teratoma formation when injected into the immunodeficient mice and germ-line chimeras when transplanted into blastocysts which indicate that their differentiation ability into all three lineages was retained (Martin, 1981; Bradley et al., 1984). Therefore, such embryo-derived cells were named embryonic stem cells. Furthermore, some fundamental characteristics were defined for mESC lines such as the expression of cell surface antigen SSEA-1 and transcription factor OCT3/4 along with the enzymatic activity for telomerase and alkaline phosphatase (Zhao et al., 2012).

The first attempt for human ES cell (hESC) derivation was in 1994 (Bongso et al., 1994). The ICM cells isolated from human embryos were cultured on human tubal epithelial monolayer. The isolated cells possessed stem cell-like properties such as alkaline phosphatase activity. However, these cells differentiated after continued culture *in vitro* (Bongso et al., 1994). A few years later, in 1998, the first successful hESC lines were generated from the human embryos that were produced by *in vitro* fertilization. From the total 36 embryos donated by patients, five hESC lines were generated. Similar to mESCs, the resulting hESCs had OCT3/4 expression as well as telomerase activity, and they generated teratomas when injected into mice. In addition, hESC lines possessed the expression of proteoglycans, TRA-1-60 and TRA-1-81 and cell surface antigens, SSEA-3 and SSEA-4, differently from mESCs (Thomson, 1998).

With the capacity of unlimited self-renewal and pluripotency, hESCs gained great importance in basic and applied science. They can be used as a powerful tool to study human development, and the molecular and genetic basis of specific developmental processes (Dvash 2006). hESCs can also be used to understand the mechanisms of many diseases and test novel therapeutical approaches. In addition, hESCs are a limitless source of somatic cells for treatment of the diseases due to their differentiation capacity (Siller et al., 2013). However, there are a number of important concerns about the use of ESCs. The potential of generating benign teratoma could be retained in ESCs-derived

differentiated cells (Kim et al., 2002). Therefore, they can lead to teratomas at the injection sites as the undifferentiated mouse and human ESCs form teratomas when injected into the mice. Immunological incompatibility is another concern about the use of ESCs-derived cells in the treatment of diseases (Taei et al., 2020). Besides these technical concerns, the derivation of hESCs from the early human embryos causes ethical and moral concerns since it requires destruction of human embryos (Alison et al., 2002).

1.3 Cellular Reprogramming

The technical and ethical concerns about ESCs have led researchers to find new methods to obtain pluripotent cells without harvesting embryos. The concept of cellular programming pioneered the studies conducted for this purpose. The findings in cellular reprogramming originated from the discovery of somatic cell nuclear transfer (SCNT) technique in 1938 (Spemann, 1938). This technique is based on microinjection of the nucleus isolated from a fully differentiated cell into an enucleated egg. Later the reprogramming by nuclear transfer is referred to as “cloning”.

The first successful cellular reprogramming was achieved by John Gurdon using the SCNT technique. He injected nuclei of intestinal epithelium cells into enucleated oocytes of *Xenopus laevis* (Gurdon, 1962). In this way, he was able to generate tadpoles from somatic cells. Around the same time, in 1957, Conrad Waddington proposed a metaphor in which a cell was shown as a rolling ball through an epigenetic landscape from undifferentiated state to a final stable differentiated state. According to this doctrine, the changes in cell fate during development are strictly restricted indicated as the ridges limiting the movement of the ball between different valleys at the end of the landscape (Creighton and Waddington, 1958). The landmark experiments of John Gurdon, in contrast to Waddington’s doctrine, demonstrated that terminal differentiation is not irreversible.

Ian Wilmut and his colleagues performed the first successful mammalian cloning via microinjection of the nucleus of a somatic cell from Finn Dorset sheep into enucleated egg of Scottish Blackface sheep in 1996. The generated embryo was transplanted into the uterus of a surrogate mother sheep, and the resulting first mammalian live offspring was named Dolly (Wilmut et al., 1997). This work demonstrated that fertilized eggs possess certain factors that could reprogram differentiated nuclei and the cellular reprogramming

of somatic cells is also possible in mammals. 5 years after this work, the reprogramming of somatic cells into the pluripotent state was achieved by fusing thymocytes with mouse ESCs (Tada et al., 2001). The resulting hybrid cells show stem-cell like properties which suggests that pluripotent cells contain some factors that are capable of reprogramming differentiated cells.

1.3.1 Induced Pluripotent Stem Cells

Besides the concept of nuclear reprogramming, the discovery of master regulator transcription factors has been another avenue of research leading researchers to find new methods to obtain pluripotent cells. In 1987, two transcription factors that regulate the cell fate were defined. The ectopic expression of Antennapedia, a transcription factor in *Drosophila*, induced the leg formation instead of antennae at the site of expression (Schneuwly et al., 1987). On the other hand, the expression of MyoD which is a mammalian transcription factor could convert mouse embryonic fibroblast cells into myoblasts (Davis et al., 1987). Therefore, researchers sought to find the master regulators of pluripotent state to obtain pluripotent cells.

In 2006, Yamanaka and Takahashi generated pluripotent cells from mouse fibroblasts by simultaneous expression of four transcription factors (TFs). These TFs are OCT4, SOX2, KLF4, c-MYC and are collectively known as OSKM or Yamanaka factors. The resulting de-differentiated cells resembled ESCs in terms of morphology, pluripotency, self-renewal, teratoma formation ability and marker expression, and they are named induced pluripotent stem cells (iPSCs). The Yamanaka factors were identified through a genetic screen of 24 genes which were previously shown to be highly expressed and important in ESCs. For this screen, they engineered mouse fibroblasts in a way that they became resistant to the geneticin (G418) upon reactivation of the ESC-specific gene, Fbx15 (Takahashi and Yamanaka, 2006). Hence, the reprogrammed cells could be isolated via antibiotic treatment. Following the delivery of the chosen 24 genes into the mouse fibroblasts via retroviral vectors, G418-resistant colonies were selected. In follow-up experiments, single TFs were removed one-by-one from the starting cocktail and the minimal set of TFs sufficient to reactivate Fbx15 gene were specified. With this approach, they demonstrated that OSKM could sufficiently induce ESC-like colonies from the mouse fibroblasts (Takahashi and Yamanaka, 2006). In 2007, the same group reported

that human iPSCs could be also generated from human fibroblasts using the same TFs (Takahashi et al., 2007). After a very short time, Thomson group published their study showing generation of human iPSCs from a newborn foreskin fibroblast with four different TFs: OCT4, SOX2, NANOG and LIN28 (Yu et al., 2007).

With the discovery of iPSC technology, a new scientific period began. Unlike ESCs, the simplicity and reproducibility of the iPSCs make them highly accessible for regenerative medicine and biomedical research. iPSC generation enables derivation of patient-specific pluripotent cells from somatic cells of individual patients. Patient-specific iPSCs can then be differentiated into various cell types such as adipocytes, cardiomyocytes, neuronal cells and pancreatic beta-cells (Amabile and Meissner, 2009; Paik et al., 2020). It was shown that patient-derived iPSCs can be utilized for patient-specific disease modeling, toxicity tests, drug screening and three-dimensional organoid production (Shi et al., 2017). Therefore, iPSCs have become an important source for both analysis of disease mechanisms and personalized therapies in regenerative medicine.

1.3.2 Phases of Reprogramming Process

The reprogramming process is divided into three phases as initiation, maturation and stabilization (Samavarchi-Tehrani et al., 2010). At initiation phase, somatic cells undergo morphological changes primarily characterized as mesenchymal-to-epithelial transition (MET) (David and Polo, 2014). MET occurs via the repression of mesenchymal transcription factors such as Snai1/2 and Zeb1/2 and the expression of epithelial genes such as E-cadherin, Cdh1 (David and Polo, 2014). In addition, cells gain enhanced proliferation capacity and resistance to apoptosis and senescence (David and Polo, 2014). Furthermore, cells obtain some pluripotency markers such as SSEA1 and alkaline phosphatase during the initiation phase (Samavarchi-Tehrani et al., 2010). It was reported that most of the starting somatic cells could not transit into maturation phase during reprogramming which indicates that progression into maturation phase is a bottleneck for reprogramming (Tanabe et al., 2013). The major event observed at maturation phase is the expression of some pluripotency-related genes such as OCT4, SALL4 and NANOG (David and Polo, 2014). Stabilization phase includes epigenetic changes leading to telomere elongation and X-chromosome reactivation occur after the cells gain pluripotency (David and Polo, 2014). Besides these changes, the major event of

stabilization phase is the gain of ability to self-renew independently from transgene expression (Golipour et al., 2012).

1.3.3 Limitations of iPSC Technology

The delivery method of OSKM factors is one of the main problems of this approach. The lentiviral or retroviral vectors used for reprogramming have ability to fully integrate into the host cells' genome which can lead to reactivation of cancer genes. It was reported that 20% of mice obtained from mouse iPSCs developed tumors due to c-MYC transgene reactivation (Okita et al., 2007). Therefore, retroviral delivery of c-MYC gene restricts the use of iPSCs for clinal applications. Alternatively, adenoviral plasmids may be used since their genome-integration capacity is at very low frequency rates. Episomal vectors are another alternative as integration-free delivery system of transgenes, however the efficiency of reprogramming with such vectors are also low (Yu et al., 2009; Stadtfeld et al., 2008). The use of single plasmid expressing all necessary genes for reprogramming is another focus to lower the number of genomic integrations. In 2009, a group achieved to derive iPSCs using a single lentiviral plasmid containing four transcription factors (OSKM) from postnatal fibroblasts. They observed ten times higher reprogramming efficiency than the standard approach (Sommer et al., 2009). Furthermore, different methods including Sendai virus, synthesized RNAs and proteins were also implemented as a way of integration-free iPSC generation (Fusaki et al., 2009; Warren et al., 2010; Kim et al., 2009).

The identification of synthetic molecules as reprogramming enhancers is a major topic of focus since they can provide safe, rapid and more efficient approach for iPSC generation. The inhibitors of DNA methyltransferases and histone deacetylase (HDAC) were tested to enhance reprogramming efficiency. The combination of OSKM expression with 5'-azacytidine, an inhibitor of DNA methyltransferase DNMT1, resulted in 4-fold higher reprogramming efficiency (Mikkelsen et al., 2008). There have also been studies to find substitutes for some of the reprogramming factors with small molecule inhibitors. In one study, BIX-01294 (DNA methyltransferase inhibitor) and BayK8644 (L-type calcium agonist) were used with the expression of only two transcription factors OCT4 and KLF4 to reprogram MEFs (Shi et al., 2008). In another one, iPSC generation was achieved from MEFs and adult mouse fibroblasts using the expression of only OCT4 with

a combination of small molecules (Li et al., 2011). Accordingly, the combination of VPA (an HDAC inhibitor), tranylcypromine (an H3K4 demethylation inhibitor) and CHIR99021 (a GSK3-β inhibitor) could replace SOX2, KLF4 and c-MYC for somatic cell reprogramming (Li et al., 2011). Also, reprogramming of neonatal human epidermal keratinocytes (NHEKs) into iPSC was achieved with a small molecule cocktail with exogenous OCT4 expression only (Zhu et al., 2010). However, a combination of synthetic molecules to reprogram human somatic cells independently from exogenous induction of a transcription factor has not been discovered to date.

The inefficiency of reprogramming process is another limitation that still needs to be overcome. The efficiency of iPSC generation is typically less than 1%, which means that only 1% of the starting somatic cells can yield an iPSC colony. However, different somatic cells possess different potential of iPSC generation (Yamanaka, 2012). It is suggested that the reprogramming process begins in more than 1% of the initial somatic cells but most of them cannot complete this process. Therefore, somatic cells undergo subsequent stochastic events along with transgene induction to continue reprogramming process efficiently. When the genomic sequences of the somatic cells and iPSCs derived from them were compared, no major differences were reported (Takahashi, 2014). However, it is known that reprogramming process requires many transcriptional changes involving inactivation of somatic cells-specific genes and activation of the genes related with pluripotency and self-renewal. Epigenetic changes become an important consideration to better understand and enhance reprogramming process since epigenetics is the mechanism of how different cell types exhibit differential gene expressions and phenotype while retaining the same genomic sequences (Holliday, 2006).

1.4 Epigenetics

The word of “epigenetics” was first used by Waddington to explain the connections between developmental process and genetics (Creighton and Waddington, 1958). Epigenetics was then defined as the heritable information of reversible changes controlling gene expression without altering the genome sequences. Hence, many cellular processes other than development are also controlled by epigenetic mechanisms (Waterland, 2006). These epigenetic changes can occur at the level of DNA, chromatin or RNA through chemical modifications.

1.4.1 DNA Methylation

The epigenetic changes at the level of DNA are seen as 5-methyl cytosine through the addition of a methyl group (CH3) to the 5'-position of cytosine residues. The effect of DNA methylation on gene expression and its heritability were reported in 1975 (Holliday, 2006). DNA methylation is restricted to CpG sites, which means that cytosine residues located at 5' of a guanosine, in mammals and 70-80% of CpG sites are found to be methylated in vertebrates. DNA methylation can lead to either inhibiting or facilitating the binding of DNA-recognizing proteins to the methylated sequences of DNA. On the other hand, DNA methylation at promoter regions is generally associated with gene repression (Gibney and Nolan, 2010).

The enzymes catalyzing the methylation of DNA are called as DNA methyltransferases (DNMTs) and four different DNMTs have been found in mammals as DNMT1, DNMT2, DNMT3A and DNMT3B (Weber and Schübeler, 2007). Also, there is a regulatory factor DNMT3L from DNMT3 family proteins (Cheng and Blumenthal, 2008). DNMT1 is the enzyme responsible for inheritance of DNA methylation pattern into the newly synthesized DNA strand during replication process while DNMT3A/B are *de novo* DNA methyltransferases. DNMT3L is the factor known to enhance *de novo* DNA methylation catalyzed by DNMT3A/B (Cheng and Blumenthal, 2008). DNMT2 shows only weak DNA methylation *in vitro* (Gibney and Nolan, 2010). There are various proteins such as MBD1, MBD2, MBD3 and MECP2 recognizing methylated DNA through their methyl-binding domains and have repressive effects on gene expression through recruitment of different repressors (Wade, 2001; Razin, 1998). Therefore, DNA methylation is generally considered as a repressive modification for transcription. Gene body DNA methylation is not associated with repression as promoter methylation, rather it is associated with proper splicing and compact chromatin structure to prevent spurious transcription at active genes (Baubec et al., 2015).

1.4.2 Histone Modifications

In the nucleus, DNA is associated with histone proteins which are basic proteins providing condensation of DNA into chromatin. Histones are positively charged proteins so they can associate with negatively charged DNA. Nucleosome, the structure containing DNA and histones, is composed of 147bp DNA wrapped around a histone octamer. Histone octamers are spoon-like formation containing eight histone proteins: two copies

of each histone H2A, H2B, H3 and H4 (Luger and Richmond, 1998). In the early 1960s, Vincent Allfrey's pioneering studies showed post-translational modifications of histones (Allfrey et al., 1964). Then, in 1997, the high-resolution X-ray structure of the nucleosome demonstrated that amino (N)-terminal tails of histones protrude from the nucleosome (Luger et al., 1997). The post-translational modifications occur on these tails can regulate chromatin structure and mediate many biological processes related with DNA-templated processes. Different histone modifications have been described such as acetylation, methylation, phosphorylation and ubiquitylation.

1.4.3 Histone Acetylation

Histones are acetylated through the Lysine(K) residues on their N-terminal tails and acetylation is dynamically regulated by opposing activities of two group of enzymes: histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Bannister and Kouzarides, 2011). Acetylation of lysine residues neutralizes the positive charge of lysine thereby weakening the DNA-histone interactions. In this way, acetylation leads to looser nucleosome structure resulting in more accessible DNA for binding of transcription factors. Histone acetylation occurs at histone H3 Lysine 9, 14, 18 and 23 residues and histone H4 Lysine 5, 8, 12 and 16 residues in many species (Grunstein, 1997). The higher level of histone acetylation is generally associated with transcriptional activation and lower level is associated with transcriptional silencing (Wade, 2001).

1.4.4 Histone Methylation

Lysine (K) and Arginine (R) residues of histone tails are the major sites for histone methylation through the function of histone methyltransferases. There are many residues of methylation have been found such as H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20 (Bannister et al., 2002). Unlike acetylation, methylation does not affect the charge of histones and can occur in more complex forms. Lysine residues can be mono-, di- or tri-methylated while arginine residues are mono-, symmetrically or asymmetrically di-methylated (Bannister and Kouzarides, 2011). Hence, different levels of histone methylation enable binding of many different chromatin remodelers with various functions. In this way, histone methylation involves in diverse biological processes from

heterochromatin formation, X-chromosome inactivation, genomic imprinting to DNA repair, cell cycle, and stress response (Greer and Shi, 2012). Tri-methylation of H3K4, H3K36 and H3K79 residues are considered as active histone marks while tri-methylation of H3K9 and H3K27 are regarded as repressive histone marks for transcription (Bannister et al., 2002).

1.5 Epigenetic Regulators of Reprogramming

Somatic cell reprogramming needs to reprogram somatic cell identity into a pluripotent state. This requires resetting chromatin landscape of somatic cells in order to silence somatic cell-specific genes and reactivate pluripotency-related genes. Therefore, reprogramming process is widely limited by functioning of chromatin-based modifiers. Their effects on reprogramming are associated with their either repressive or activating functioning in gene expression. The reactivation of pluripotency genes is restricted by repressive epigenetic modifications such as DNA methylation, histone deacetylation, H3K27 and H3K9 methylation (Gao et al., 2013; Mali et al., 2010; Mansour et al., 2012; Ma et al., 2008) (Figure 1.1). On the other hand, the expression of somatic cell-specific genes is maintained by active chromatin modifications such as H3K79 methylation and SUMOylation (Onder et al., 2012; Kolundzic et al., 2018; Cossec et al., 2018) (Figure 1.1). Furthermore, H3K4 methylation is an active histone mark which acts as an enhancer of reprogramming by enhancing the expression of pluripotency markers (Koche et al., 2011).

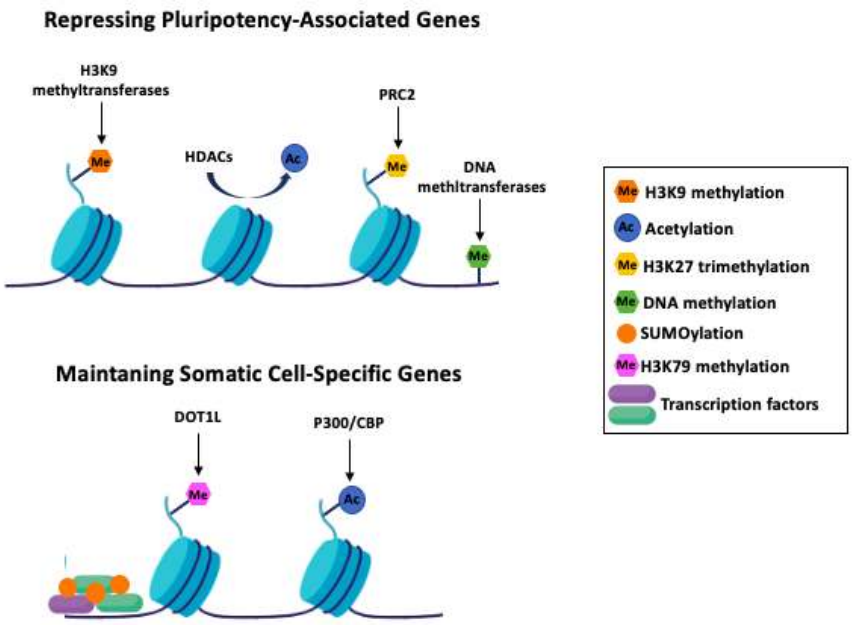


Figure 1.1: Epigenetic regulators of reprogramming acting as barriers against somatic cell reprogramming. The regulators with repressive functions are involved in repression of pluripotency-associated genes shown *at the top*. The regulators functioning in active transcription maintain the expression of somatic cell-specific genes shown *at the bottom*. HDACs; Histone deacetylases, PRC2; Polycomb repressive complex 2, DOT1L; Disruptor of telomeric silencing 1-like, P300/CBP; EP300/CREB-binding protein, SUMO; Small ubiquitin-related modifier. Adapted from "*Going up the Hill: Chromatin-based Barriers to Epigenetic Reprogramming*" by H.D. Arabacı, 2021.

DNA methylation is reported as one of the major barriers against somatic cell reprogramming. shRNA-mediated suppression of DNMT3A and DNMT3B enhanced reprogramming efficiency while their overexpression resulted in opposite effect (Gao et al., 2013). Similarly, DNMT1 inhibition by 5-Azacitidine or its knockdown improved the partial reprogramming process upon OSKM induction (Mikkelsen et al., 2008). On the other hand, the requirement of ten-eleven translocation (TET) proteins for reprogramming was demonstrated. TET proteins are responsible for hydroxylation of DNA 5-methylcytosine, so they regulate DNA demethylation (Melamed et al., 2018). TET1 was shown to promote reactivation of endogenous OCT4 gene by its demethylation and could replace the exogenous OCT4 expression during OSKM-mediated

reprogramming (Gao et al., 2013). These findings indicate that DNA methyltransferases serve important barriers in somatic cell reprogramming and dynamic regulation of DNA methylation is essential for obtaining pluripotency.

There are several reports showing that deacetylation of histones by HDACs blocks reactivation of pluripotency-related genes during reprogramming (Mali et al., 2010). Therefore, HDACs are regarded as barriers against iPSC generation and broad HDAC inhibitors have been tested to enhance reprogramming efficiency (Arabacı et al., 2021).

Polycomb repressor complex 2 (PRC2) is responsible for mono/di/tri-methylation of H3K27 residue which is a vital epigenetic modification for development by repression of Hox-family genes (Qin et al., 2013). PRC2 functioning seems significant for TF-induced reprogramming as well, since shRNA-mediated knockdown of its components blocked human reprogramming (Onder et al., 2012). On the other hand, H3K27 methylation seems to prevent reactivation of pluripotency-related genes. Therefore, demethylation of H3K27 residue enhanced induction of pluripotency (Mansour et al., 2012).

H3K9 methylation which is catalyzed by various histone methyltransferases is mainly involved in heterochromatinization and gene silencing (Arabacı et al., 2021). Therefore, it acts as roadblock to reprogramming by repression of pluripotency related genes. Depletion of its methyltransferases enhanced reprogramming efficiency by relieving repression of the NANOG (Shi et al., 2008; Sridharan et al., 2013). H3K9me3 serves a barrier against reprogramming via recruitment of transcriptional repressor TRIM28. shRNA-mediated knockdown of TRIM28 resulted in euchromatin formation on H3K9me3-marked regions leading to higher reprogramming efficiency (Miles et al., 2017).

H3K79me3 is an active histone mark functioning in transcriptional elongation (Steger et al., 2008). This mark is enriched at somatic cell specific genes so the inhibition of its methyltransferase DOT1L resulted in enhanced TF-induced reprogramming of fibroblasts (Onder et al., 2012). In this way, H3K79me3 catalyzed by DOT1L acts as a barrier against TF-induced reprogramming.

Small ubiquitin-related modifier (SUMO) modification (SUMOylation) is another active post-translational modification acting as roadblock to TF-induced reprogramming. It can enhance stability of transcription factors that safeguard somatic cell identity so that provides cell-fate maintenance (Cossec et al., 2018). When the SUMOylation levels were

decreased on somatic enhancers, the efficiency of somatic cell reprogramming increased (Cossec et al., 2018).

Apart from all those DNA and chromatin-based barriers, mRNA methylation is also important for reprogramming. The most abundant mRNA methylation is formed as N^6-methyladenosine (m^6A) which is dynamically regulated by writers, erasers and reader proteins (Zaccara et al., 2019). The presence of m^6A on specific mRNAs regulates cellular differentiation, tumorigenesis and many other processes because m^6A methylation affects mRNA instability, translation, splicing and nuclear export (Huang et al., 2019; Zaccara et al., 2019). m^6A methylation was shown to safeguard mouse ESC identity through recruitment of its reader YTHDC1 (Xu et al., 2021; Liu et al., 2021). On the other hand, its effect on reprogramming remains controversial. While accumulation of m^6A by overexpression of METTL3, a writer of m^6A, enhanced mouse reprogramming efficiency in a study; increased level of m^6A by genetic ablation of Alkbh5, an eraser of m^6A, decreased mouse reprogramming efficiency in another study (Aguilo and Walsh, 2017; Khodeer et al., 2021).

1.6 H3K36 Methylation

Methylation of histone H3 Lysine 36 (H3K36) is a conserved histone modification known to regulate chromatin structure and diverse biological processes. H3K36 residue is found as mono-, di- or tri-methylated. Mono-methylation is generally regarded as an intermediate level for higher methylation status, while di- and tri-methylation of H3K36 have different distributions and molecular roles. H3K36 di-methylation (H3K36me2) is localized at intergenic regions on the genome but H3K36 tri-methylation (H3K36me3) is generally found at gene bodies of transcriptionally active genes (Zaghi et al., 2020). In addition, genome-wide analyses showed that H3K36me1 and -me2 are primarily present near 5'end regions while higher methylation levels (H3K36me2 and -me3) are observed towards 3'ends of genes (Wozniak and Strahl, 2014).

There is an antagonistic relationship between H3K36me2/3 and the repressive H3K27me3 as the H3K36me2/3 marks prevent deposition of H3K27me3 by the polycomb repressive complex (PRC2) (Schmitges et al., 2011; Yuan et al., 2011). Therefore, on the same histone H3 tail H3K27me3 and H3K36me2/3 are mutually exclusive. In this reciprocal relationship, H3K36me2/3 appear dominant over H3K27me3

since pre-deposited H3K27me3 cannot affect the enzymatic activity of histone methyltransferases ASH1L and SETD2 which are responsible for H3K36me2/3 (Yuan et al., 2011). EZH2, a component of PRC2 complex, needs to interact with unmethylated K36 residue on histone H3 tail for positioning and catalytic activity of PRC2 (Finogenova et al., 2020).

1.6.1 Mutations at H3K36 Residue

There are mutations found at H3K36 residue affecting the catalytic activity of SETD2 and being involved in different processes such as differentiation and tumorigenesis (Lutsik et al., 2020; Jain et al., 2020). Lysine36-to-Methionine36 (K36M) mutation on H3.3 variant was detected in 90% of chondroblastoma, which is a benign bone tumor (Zhang et al., 2017). H3.3, the most common variant of histone H3, is a replication-independent histone deposited specifically on actively transcribed genes (Zaidan et al., 2020). H3.3K36M mutation prevented SETD2 activity resulting in the lower H3K36me3 and higher H3K27me3 levels (Lu et al., 2016).

In a study investigating how histone variant H3.3 provides rapid and high-level expression for stimulation-induced genes, Armache *et al.* showed that H3.3S31 phosphorylation engaged SETD2 activity (Armache et al., 2020). In mouse macrophages, H3.3S31 residue was found to be phosphorylated upon stimulation with lipopolysaccharide (LPS) (Armache et al., 2020). To understand the underlying mechanisms, Armache *et al.* used H3.3S31E, a phosphomimic mutation, and observed that H3.3S31E increased expression of LPS-induced genes such as Bcl2a1b, Ccl4 and Cxcl2 by enhancing SETD2 catalytic activity (Armache et al., 2020).

On the other hand, H3G34W/R mutations frequently found at non-brainstem gliomas prevent SETD2 activity and lead to redistribution of H3K36me3 without altering global H3K36me3 level (Zhang et al., 2017; Nohr et al., 2017). In addition, H3G34 mutations affects chromatin structure of enhancers in mesenchymal stem cells by gain of H3K27me3 and loss of H3K27ac (Jain et al., 2020). Such changes result in impaired osteogenic differentiation of mesenchymal stem cells (Lutsik et al., 2020).

1.6.2 H3K36 Methylation and Reprogramming

There are several studies indicating H3K36 methylation as a roadblock against somatic cell reprogramming. In one of them, ascorbic acid (Vitamin C) which is a cofactor of H3K36 demethylases KDM2A/B was used to enhance iPSC generation (Esteban et al., 2010). In addition, ascorbic acid was shown to increase the expression levels of key cell cycle regulators by demethylation of H3K36me2/3 (Wang et al., 2011). Another study demonstrated that overexpression of KDM2B, H3K36me2-specific demethylase, could improve reprogramming efficiency by mediating the reactivation of pluripotency-related genes at early phase of reprogramming (Liang et al., 2012). Moreover, the effect of KDM4B catalyzing demethylation of H3K9- and K36-trimethylation on reprogramming of MEFs was analyzed (Wei et al., 2017). With KDM4B induction, the lower levels of H3K9me3 and H3K36me3 led to higher reprogramming efficiency (Wei et al., 2017).

1.7 H3K36 Methyltransferases

In yeast, Set2 (SET domain containing protein-2) is the methyltransferase catalyzing all methylation levels of H3K36 residue. In contrast, H3K36 methylation is performed by different histone methyltransferases in mammals. These methyltransferases include a common catalytic domain called as SET (Suppressor of variegation 3-9, Enhancer of Zeste, and Trithorax) to transfer methyl group to H3K36 using S-adenosylmethionine (McDaniel and Strahl, 2017).

NSD1 (nuclear receptor SET domain-containing protein 1) also known as KMT3B, is one of the mammalian methyltransferases that can catalyze mono- and di-methylation of H3K36 (Wagner and Carpenter, 2012). It can potentially methylate H4K20 residue and some non-histone substrates such as NF-kB. Another methyltransferase is NSD2 also known as WHSC1 (the product of Wolf-Hirschhorn syndrome candidate gene 1) or MMSET (Li et al., 2016). NSD2 performs mono- and di-methylation of H3K36 residue like NSD1. In addition, it can target H4K20, H3K4 and H3K27 residues. NSD3 (also known as WHSC1L1), has been shown to be specific to H3K36 residue (Wagner and Carpenter, 2012). Similar to NSD1 and NSD2, NSD3 acts as mono- and di-methyltransferase of H3K36 (Wagner and Carpenter, 2012). There are other H3K36 methyltransferases performing mono- and dimethylation besides the NSD family

enzymes in mammals. These enzymes are ASH1L (ASH1-like), SETMAR (SET domain and mariner transposase fusion gene-containing), SMYD2 (SET and MYND domain-containing 2) and SETD3 (Zaghi et al., 2020).

1.8 SETD2 and Trimethylation of H3K36

Although there are several different methyltransferases identified for mono- and di-methylation of H3K36 residue, SETD2 is the only methyltransferase catalyzing H3K36 trimethylation in mammals. SETD2 (also known as HYPB and KMT3A) contains three conserved functional domains: SET, WW and SRI domains. The SET domain is generally found as multi-domain with an AWS (Associated with SET) and a Post-SET domain in H3K36 methyltransferases (McDaniel and Strahl, 2017). This 130 amino acid-long domain is responsible for catalytic activity of SETD2 enzyme (Li et al., 2016). The WW domain is composed of two tryptophan (W) separated by 20-22 amino acids. Through this domain, SETD2 can recognize different proline-rich motifs (Macias et al., 1996; Lu et al., 1999). Thus, SETD2 involves in a variety of protein interactions and molecular processes. Finally, the SRI (Set2 Rpb1 Interacting) domain enables SETD2 to interact with phosphorylated C-terminal domain (CTD) of RNA Polymerase II (RNAPII) (Kizer et al., 2005).

The most well-described function of SETD2 and H3K36me3 is transcription of active euchromatin regions. The link between H3K36 methylation and transcription was discovered from the interaction of Set2 with phosphorylated CTD of RNAPII through its SRI domain in yeast (*Saccharomyces cerevisiae*) (Zaghi et al., 2020). Through this interaction, H3K36 methylation is involved in the coordination of transcriptional elongation, RNA splicing and inhibition of spurious intragenic transcripts. In yeast, H3K36me3 represses spurious intragenic transcription by regulating histone deacetylation, but this is not a conserved mechanism in mammals (Zaghi et al., 2020). Instead, in mammals the methyltransferase catalyzing H3K36me3, SETD2, is responsible for repression of cryptic transcription through recruitment of FACT (Facilitates Chromatin Transcription) subunits SPT16 and SSRP1 (Carvalho et al., 2013).

Besides interacting with RNA Pol II, SETD2 also interacts with a heterogeneous ribonucleoprotein, hnRNP-L which is responsible for alternative splicing (Bhattacharya et al., 2021). hnRNPs are RNA-binding proteins specifically bind to splice sites and affect

splicing of pre-mRNA transcripts (Lee and Rio, 2015). The interaction between SETD2 and hnRNP-L takes place via an undefined domain on SETD2 which was named SETD2-hnRNP interaction (SHI) domain by Bhattacharya *et al.* (Bhattacharya et al., 2021). By binding both RNA Pol II and hnRNP-L, SETD2 provides the link between mRNA splicing and transcription.

In human, DNA damage repair systems are affected by SETD2 and H3K36 methylation as well. This effect is partly the result of interaction with p53 which is a transcription factor critical for efficient repair. SETD2 can mediate expression levels of p53 downstream genes by enhancing its transcription factor activity (Xie et al., 2008). Hence, H3K36me3 loss resulted in persistence of DNA damage in cells and defective homologous recombination (Carvalho et al., 2014; Pfister et al., 2014).

The knockdown of SETD2 causes complete loss of H3K36me3 indicating that SETD2 is a non-redundant enzyme, and its main function is H3K36 trimethylation which is a stable post-translational modification of histones (Edmunds et al., 2008; Skucha et al., 2019). Through several different effector proteins recognizing H3K36me3, H3K36me3 is involved in many different cellular processes such as alternative splicing, DNA replication and repair, transcriptional repression, DNA methylation and dosage compensation (Wagner and Carpenter, 2012). Hence, other functions associated with SETD2 are mediated through these reader proteins.

1.8.1 Role of SETD2 in Stem Cells and Development

SETD2 and H3K36me3 are crucial for development and stem cell differentiation consistent with their diverse interactions. Their requirement for animal development was firstly identified in *Drosophila* as RNAi-mediated SETD2 knockdown impaired development at the larval stage (Stabell et al., 2007). Homozygous knockout of SETD2 caused lethality in mice embryos as they die at E10.5-E11.5 days due to vascular defects (Hu et al., 2010). SETD2 regulates stem cell differentiation in different aspects. Zhang *et al.* showed the importance of SETD2 catalytic activity for endodermal differentiation in mouse ESCs by engineering conditional SETD2 homozygous knock-out mice and found that SETD2 mediates endoderm differentiation via regulating Fgfr3 transcription (Zhang et al., 2014). Moreover, Zhang *et al.* analyzed the effect of SETD2 on hematopoietic stem cells (HSC) using the same conditional knockout system and showed that SETD2 is

required for self-renewal of HSC and its deficiency affected differentiation capacity (Zhang et al., 2018).

1.8.2 SETD2 and Cancer

Several mutations of SETD2 have been identified across a wide range of human cancers. Most SETD2 mutations tend to be inactivating frameshift or nonsense mutations whereas missense mutations in critical domains have also been reported (Gerlinger et al., 2012). Overall, SETD2 was found to be altered in 4.75% of all cancers (Sweeney et al., 2017). The first SETD2 mutation was found in clear cell renal cell carcinoma (ccRCC) which is the most common type of kidney cancer (van den Berg, 2013). The high frequency of inactivating SETD2 mutations in ccRCC indicates a tumor-suppressor-like function for SETD2 (Creighton et al., 2013). SETD2 mutations also occur in a variety of tumors of the central nervous system, predominantly were seen in high-grade gliomas but rarely in low-grade gliomas (Fontebasso et al., 2013). In addition, SETD2 mutations were identified in acute leukemias. They were commonly nonsense or frameshift mutations which implies SETD2 loss as a critical event in leukemia development (Zhu et al., 2014). Furthermore, SETD2 is found to be mutated in lung and breast carcinomas at substantial rates (Oshita et al., 2016; Sweeney et al., 2017).

1.9 H3K36me3 Readers

Regulation of chromatin is dynamically controlled by epigenetic modifications which are controlled by writers, readers and erasers. Writers are the enzymes catalyzing post-translational modifications of histones while erasers are the enzymes removing these modifications. Readers or effectors are the proteins recognizing epigenetic modifications. Epigenetic modifications get involved in a wide range of cellular process serving as docking sites for reader proteins.

H3K36me3 is recognized by many readers containing PWWP, chromo- or Tudor-domains in mammals (Yun et al., 2011) (Figure 1.2). The PWWP domain contains Pro-Trp-Trp-Pro sequence motif and can bind both histone and DNA methylation (Rona et al., 2016). The chromo (Chromatin Organization Modifier) domain is found in many proteins and provide binding to methylated lysine residues of histone H3 tail (DasGupta

et al., 2020). The Tudor domain is about 60 amino acids length region recognizing methylated lysine and arginine residues (Bohnsack and Pandey, 2021).

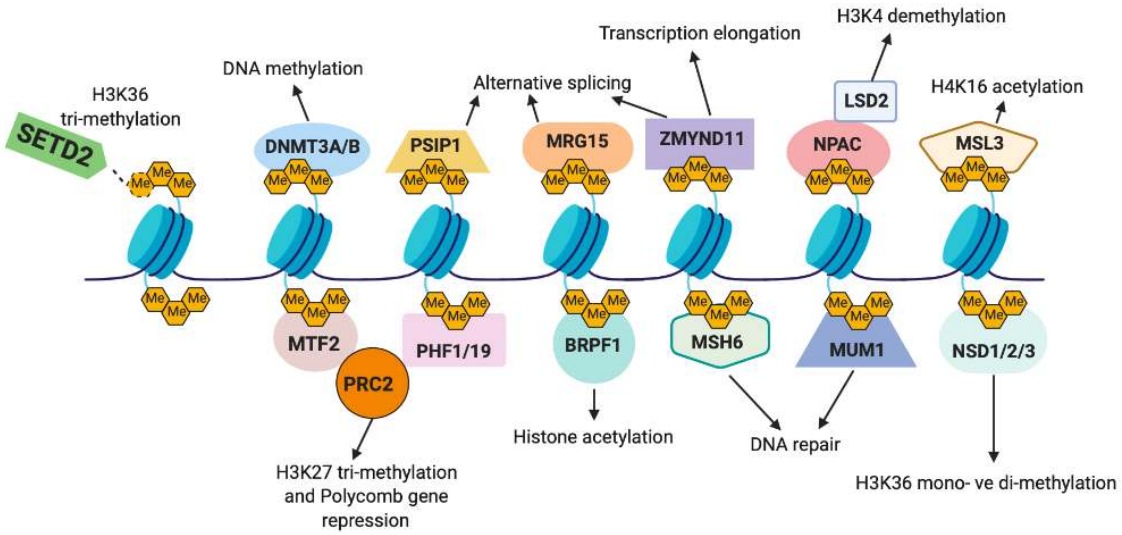


Figure 1.2: Representation of H3K36me3 recognition by mammalian readers and their cellular functions. SETD2 is the sole enzyme catalyzing tri-methylation of H3K36 residue in mammals. Tri-methylated H3K36 can be recognized by DNMT3A/B regulating DNA methylation; PSIP1 and MRG15 regulating alternative splicing; ZMYND11 regulating alternative splicing and transcriptional elongation; NPAC regulating H3K4 demethylation via LSD2; MSL3 regulating H4K16 acetylation, MTF2 and PHF1/19 parts of PRC2 complex regulating H3K27 tri-methylation; BRPF1 regulating histone acetylation; MSH6 and MUM1 involved in DNA repair and NSD1/2/3 catalyzing H3K36 methylation.

1.9.1 DNA (cytosine-5) Methyltransferase 3 Alpha and Beta (DNMT3A/B)

The DNMT3A and DNMT3B catalyze DNA methylation at CpG sites through reading H3K36me3 mark with their PWWP domain. Although DNA methylation is basically functioning in transcriptional repression of promoters, it is critical for many processes such as development, cellular differentiation, genomic imprinting and chromatin organization in mammals (Dhayalan et al., 2010). In addition, deregulation of DNMT3A and DNMT3B is involved in human developmental disorders and cancer

(Weinberg et al., 2019). As mentioned above, DNA methylation has a preventing role in the regulation of somatic cell reprogramming.

The interaction of PWWP domain and H3K36me3 enhances the activity of these methyltransferases for DNA methylation (Dhayalan et al., 2010). DNMT3A could recognize both di- and tri-methylated H3K36 *in vitro* but showed higher binding for H3K36me2 (Weinberg et al., 2019). Therefore, DNMT3A has wider localization through both genic and intergenic regions depending on the deposition of H3K36me2 (Weinberg et al., 2019). However, DNMT3B localizes within gene bodies of transcribed genes containing H3K36me3 mark. Then, it mediates DNA methylation of H3K36me3-marked gene bodies. Binding and enzymatic activity of DNMT3B at H3K36me3-deposited regions was prevented by genetic ablation of SETD2 whereas binding of DNMT3A remained unchanged (Baubec et al., 2015).

1.9.2 PHD Finger Protein 1/19 (PHF1/19) and Metal Response Element Binding Transcription Factor 2 (MTF2)

PHF1 (PCL1), PHF19 (PLC3) and MTF2 (PLC2) are Polycomblike (PLC) proteins found in sub-stoichiometric ratios in PRC2 complexes (Brien et al., 2012). PRC2 is a repressive epigenetic modifier catalyzing tri-methylation of H3K27 residue. It has important roles in development, cellular differentiation, genome maintenance and cell-fate transitions (Qin et al., 2013; Brien et al., 2012). PHF1, PHF19 and MTF2 recognize H3K36me3 through their N-terminal Tudor domains and regulate recruitment and functioning of PRC2 complex (Qin et al., 2013).

The binding of PHF1 to H3K36me3 inhibited methyltransferase activity of PRC2 and promoted localization of PHF1 at the DNA damage sites which was induced by genotoxic stress (Musselman et al., 2013). PHF19-H3K36me3 interaction recruits PRC2 complex to H3K36me3-marked sites and provided recruitment of NO66 which is a H3K36me3 demethylase (Brien et al., 2012). These two functions of PHF19 via transient binding to H3K36me3 were shown to be essential for turning active genes into Polycomb-repressed state during mouse embryonic stem cell differentiation (Brien et al., 2012). MTF2 was found to be highly expressed in ESCs where it was involved in the regulation of self-renewal and pluripotency and cell-fate transitions (Dong et al., 2020). Also, it

acted as an enhancer of mouse somatic cell reprogramming as a critical component of PRC2 complex (Zhang et al., 2011).

1.9.3 Bromodomain and PHD Finger Containing 1 (BRPF1)

BRPF1 recognizes H3K36me3 through its C-terminal PWWP domain and is enriched on actively transcribed genes (Wu et al., 2011). BRPF1 is a key subunit of MOZ/MORF histone acetyltransferase complexes by providing its recruitment, assembling and activation (Ullah et al., 2008). Through its interaction with MOZ/MORF complexes, BRPF1 regulates gene expression by promoting histone acetylation at active chromatin regions (Vezzoli et al., 2010). Furthermore, BRPF1 has a crucial role for mouse embryonic development as its genetic ablation prevented proliferation of MEFs (You et al., 2015).

1.9.4 MutS Homolog 6 (MSH6)

MSH6 interacts with H3K36me3 through its PWWP domain and this interaction provides recruitment of mismatch-recognition protein MutSa complex (a heterodimer of MSH2 and MSH6) during DNA replication (Li et al., 2013). MutSa protein functions in DNA mismatch repair and prevents microsatellite instability preferentially on H3K36me3-marked actively transcribed regions (Huang et al., 2018). Therefore, MSH6 can affect many cellular processes and it is implicated in tumorigenesis.

1.9.5 Multiple Myeloma Oncogene 1 (MUM1)

MUM1 contains a PWWP-domain and showed affinity towards H3K36me3 in *in vitro* tests (Wu et al., 2011). MUM1 is commonly known as interferon regulatory factor-4 (IRF4) that regulates B cell lymphoid differentiation (Gaidano and Carbone, 2000). However, Huen *et al.* identified MUM1 in a different context and named it EXPAND1 in this research (Huen et al., 2010). They showed that MUM1/EXPAND1 participates in chromatin decondensation upon DNA damage by directly interacting with DNA damage mediator 53BP1 which enables cell survival after DNA damage (Huen et al., 2010).

1.9.6 Nuclear Receptor Binding SET Domain Protein 1/2/3 (NSD1/2/3)

NSD1, NSD2 and NSD3 are methyltransferases catalyzing mono- and di-methylation of H3K36 residue. They all have PWWP domains through which they can interact with H3K36me2/3 (Vermeulen et al., 2010; Wu et al., 2011). Reading of H3K36me2/3 by NSD family of H3K36 methyltransferases suggests a positive feedback regulation of H3K36 methylation. However, this regulation has only been reported for NSD2 to date. Sankaran *et al.* showed that recognition of H3K36me2 stabilized NSD2 at chromatin and led to higher H3K36me2 levels which consequently enhanced cellular proliferation (Sankaran et al., 2016). Additionally, NSD family of proteins contribute to chromatin integrity and their dysregulation results in diverse human diseases especially cancer (Li et al., 2009).

1.9.7 Male-Specific Lethal Complex Subunit 3 (MSL3)

MSL3 is a subunit of MOF-containing human MSL complex having acetyltransferase activity for histone H4K16 (Moore et al., 2010). Hence, MSL complex enhances transcription and regulates cell-cycle progression, DNA repair, ESC pluripotency and self-renewal, and carcinogenesis (Kim et al., 2010; Chen et al., 2015a). MSL3 contains a conserved N-terminal chromo-barrel domain (chromodomain) which shows binding affinity towards both methylated H4K20 and H3K36 (Moore et al., 2010). It preferred binding to mono- or di-methylated H4K20 over trimethylated H3K36 *in vitro* (Moore et al., 2010).

1.9.8 Glyoxylate Reductase 1 Homolog (GLYR1/NPAC)

GLYR1/NPAC is a reader of H3K36me3 mark interacting through its PWWP domain (Vermeulen et al., 2010; Yu et al., 2021). GLYR1/NPAC is characterized as a cofactor of LSD2/KDM1B which is demethylase of H3K4 residue (Fang et al., 2013). Therefore, GLYR1/NPAC is involved in the regulation of gene expression by histone demethylation. Recently, GLYR1/NPAC was found to be required for mouse ESC pluripotency as its genetic ablation caused differentiation and decreased reprogramming efficiency of MEFs (Yu et al., 2021). It was enriched on H3K36me3-marked regions of

actively transcribed genes and shown to regulate transcriptional elongation of pluripotency genes via direct interaction with RNA Pol II (Yu et al., 2021).

1.9.9 PC4 and SFRS1 Interacting Protein 1 (PSIP1/LEDGF)

PSIP1/LEDGF contains a PWWP domain that recognizes H3K36me3 and DNA with low affinity (Van Nuland et al., 2013). PSIP1/LEDGF functions in DNA repair and alternative splicing. It can modulate homologous recombination repair pathway by recruitment of RBBP8 to the DNA double-strand breaks (DSBs) (Daugaard et al., 2012). In addition, it regulates alternative splicing by acting as chromatin-adapter for SRFS1 which is a significant regulator of alternative splicing (Pradeepa et al., 2012). PSIP1/LEDGF was also reported to enhance the inclusion levels of targeted exons during alternative splicing by binding to H3K36me3-positive regions (Xu et al., 2018). Through regulation of alternative splicing events, PSIP1/LEDGF contributes to stem cell fate determination since specific isoforms of some genes are required for stem cell identity and their proper differentiation (Xu et al., 2018).

1.9.10 MORF-Related Gene 15 (MRG15/MORF4L1)

MRG15, human homolog of MRG1 found in *Caenorhabditis elegans,* is a transcription factor essential for normal development, proliferation and cellular senescence (Zhang et al., 2006a). In addition, it controls H3K4 demethylation through its N-terminal chromodomain recognizing H3K36me3. Upon binding to H3K36me3, MRG15 recruits retinoblastoma binding protein 2 (RBP2) complex catalyzing demethylation of H3K4me3 (Luco et al., 2010). Furthermore, MRG15-H3K36me3 interaction regulates alternative splicing depending on the function of PTBP1 which reduces the inclusion levels of targeted exon (Xu et al., 2018). Similar to PSIP1/LEDGF, MRG15 is involved in stem cell fate decision due to its role in alternative splicing.

1.9.11 Zinc Finger MYND-Type Containing 11 (ZMYND11/BS69)

ZMYND11 selectively binds to trimethylated K36 residue of histone variant H3.3 via its PWWP domain (Guo et al., 2014). Binding to H3.3K36me3 enables ZMYND11 to regulate alternative splicing and transcriptional elongation processes. The main effect

of ZMYND11 in alternative splicing is intron retention (IR) by directly interacting with EFTUD2 which is a component of spliceosome (Bartels et al., 2002). On the other hand, ZMYND11 affects RNA Pol II functioning during transcriptional elongation resulting in repression of transcription (Wen et al., 2014). Specifically, oncogenes were found to be suppressed by ZMYND11, so it is critical for tumor suppression (Wen et al., 2014).

Main question addressed:

Although there are several studies that suggested H3K36 methylation as a barrier against reprogramming, the exact role of H3K36me3 in this process has not been elucidated yet. The diverse interactions and functions of H3K36me3 in the cell make it likely that it would be an important modification regulating somatic cell reprogramming. Previous work demonstrated that the genetic suppression of SETD2, the only methyltransferase catalyzing H3K36me3 in mammals, enhances OSKM-mediated reprogramming of human fibroblasts (Figure 1.3.a and Figure 1.3.b) (Ozcimen, 2018). Upon SETD2 knockdown, global H3K36me3 level significantly decreases while other major chromatin marks were unaffected (Figure 1.3.c) (Ozcimen, 2018). These findings suggested that SETD2-mediated H3K36 tri-methylation is a roadblock against iPSC generation. In this thesis, I aimed to decipher the molecular mechanisms through which SETD2 inhibition and H3K36me3 loss enhance reprogramming efficiency.

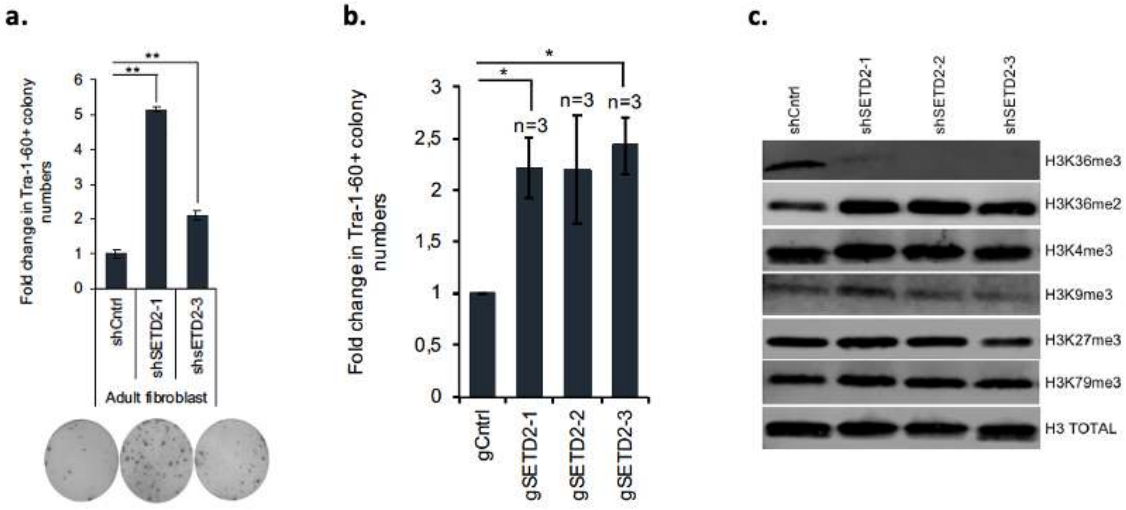


Figure 1.3: Preliminary data showing the effects of SETD2 depletion on reprogramming of dH1f cells and major histone marks. a. Fold change in TRA-1-60+ colony numbers of SETD2 knockdown cells over shCntrl-expressing cells. SETD2 was knockdown in dH1f cells with two independent shRNAs (shSETD2-1 and shSETD2-3). Representative well images were shown below. **b.** Fold change in TRA-1-60+ colony numbers of SETD2 knockout cells over gCntrl-expressing cells. SETD2 was knockout in dH1f cells using three independent gRNAs (gSETD2-1/2/3). **c.** Western blot analysis of the major histone marks H3K36me2/3, H3K4me3, H3K9me3, H3K27me3 and H3K79me3 in SETD2 knockdown lysates. Total histone H3 was used as nuclear loading control. Retrieved from "*The Role of Histone H3 Lysine 36 Methylation in Reprogramming of Fibroblasts and on Induced Pluripotent Stem Cell Generation*" by B. Ozcimen, 2018.

For this purpose, I proposed several different hypotheses. Based on the previous RNA-sequencing data which showed that SETD2 depletion leads to global changes in gene expression during reprogramming, I hypothesized that SETD2 may suppress iPSC generation via regulating the expression levels of key transcription factors. I also analyzed the distribution of H3K36me3 and H3K27me3 marks which were shown to have a reciprocal relationship to test the hypothesis that H3K36 trimethylation may act as an

anti-silencing mark on transcribed genes that should otherwise be targeted by PRC2 during reprogramming. Lastly, I hypothesized that SETD2 may impede reprogramming via the functioning of H3K36me3 readers, since H3K36me3 is recognized by several reader proteins which are involved in many different cellular processes. Thus, I tested the effect of H3K36me3 readers in reprogramming by a loss-of-function approach to identify the readers whose loss mimic the phenotype of SETD2 depletion.

Chapter 2: METHODS

2.1 Competent Bacteria Preparation

For preparation of chemically competent bacteria, Stbl3 bacteria from frozen stock was streak onto LB-agar plate without antibiotic and incubated at 37 ℃ for 16 hours. The day after, a single colony was picked from the plate and inoculated into 3 mL of LB-broth. It was again incubated at 37 ℃ for 16 hours. From this overnight culture, 1 mL was diluted into 100 mL of LB-broth media in flask. This culture was incubated at 37 ℃ until the optic density (OD_{600}) of the culture was reached to 0.4 (0.4-0.55 optimum). After the desired OD_{600} value was obtained, the culture was incubated on ice for 15 minutes. The culture was then centrifuged at 3000 rpm, 4 ℃ for 15 minutes. The pellet was resuspended with 33 mL of chilled RF1 solution (Table 2.1) and incubated on ice for 15 minutes. Then, it was centrifuged as before. The pellet was resuspended with 4 mL of chilled RF2 solution (Table 2.2) and again incubated on ice for 15 minutes. After the incubation, the bacteria were aliquoted into chilled 1.5 mL Eppendorf tubes and frozen in liquid nitrogen.

Table 2.1: Preparation of RF1 solution

Reagents	**Amount**	**Final Conc**
Rubidium Chloride (RbCl)	1.21 g	100 mM
Mangnese(II) chloride tetrahydrate ($MnCl_2 \cdot 4H_2O$)	0.99 g	50 mM
Potassium acetate	0.294 g	30 mM
Calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$)	0.148 g	10 mM
Glycerol	12 mL	15% wt/vol
dH_2O	up to 100 mL	
Final pH was adjusted to 5.8 with 0.2 M acetic acid.		

Table 2.2: Preparation of RF2 solution

Reagents	**Amount**	**Final Conc**
MOPS	0.105 g	10 mM
Rubidium Chloride (RbCl)	0.06 g	10 mM
Calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$)	0.55 g	75 mM
Glycerol	6 mL	15% wt/vol
dH_2O	up to 50 mL	
Final pH was adjusted to 6.8 with 1 M NaOH.		

2.2 Bacterial Transformation

For transformation, chemically competent *E.coli* Stbl3 bacteria were used. 50 µL of Stbl3 bacteria was mixed with 50 ng of plasmid and the mixture was incubated on ice for 15 minutes. Then, the mixture was heat-shocked at 42 ℃ for exactly 30 seconds. After that, it was immediately placed on ice and incubated for 5 minutes. 250 µL of SOC medium was added onto mixture and the whole mix was shaken at 37 ℃, 225 rpm for 15 minutes. SOC medium is a nutrient-rich microbial growth medium used for bacterial transformation (Sambrook and Russell, 2001). 100 µL of bacteria mixture was spread onto LB agar plates containing either ampicillin or kanamycin and plates were left to incubation at 37 ℃ for 16 hours.

For isolation of plasmids, NucleoSpin Plasmid mini kit (MN 740588.50) and NucleoBond Xtra midi kit (MN 740410.10) were used. The single colony picked from LB-agar plates was inoculated in 5 mL of LB broth for mini plasmid purification kit or in 200 mL of LB broth for midi plasmid purification kit. Inoculated bacteria were then shaken at 37 ℃, 225 rpm for 16-18 hours. The concentration of isolated plasmid was measured using NanoDrop spectrophotometer. For diagnostic digest, 500-1000 ng of plasmid was used in reaction of suitable restriction endonucleases at 37 ℃ for 1 hour. Then, digested samples were detected by running on 1.2% agarose gel. If bands were observed at expected sizes, the purified plasmids were used in further steps.

2.3 Cloning of shRNAs and sgRNAs targeting SETD2

Three independent shRNAs (Table 2.3) targeting SETD2 were previously amplified and cloned into pSMP backbone (Addgene #36394) (Ozcimen, 2018).

Table 2.3: Sequences of shRNAs targeting SETD2

Oligo	**Sequence (5'-3')**
shControl	TGCTGTTGACAGTGAGCGCCCGCCTGAAGTCTCTGATTAATAGTGAAGCCACAGATGTATTAATCAGAGACTTCAGGCGGTTGCCTACTGCCTCGGA
shSETD2-1	TGCTGTTGACAGTGAGCGCGCTGACTCACGGTGTTATGAATAGTGAAGCCACAGATGTATTCATAACACCGTGAGTCAGCTTGCCTACTGCCTCGGA
shSETD2-3	TGCTGTTGACAGTGAGCGCGGAATTAGACTCTTTATCTAATAGTGAAGCCACAGATGTATTAGATAAAGAGTCTAATTCCTTGCCTACTGCCTCGGA

Single guide RNAs (sgRNAs) (Table 2.4) targeting SETD2 and sgNT1 (non-targeting 1 used as control) were also previously cloned into pLentiCRISPR_version1 (pLentiCRISPR_v1) plasmid backbone (Addgene #49535) (Ozcimen, 2018).

Table 2.4: Sequences of sgRNAs cloned into pLentiCRISPR_v1 vector

Oligo		**Sequence (5'-3')**
sgNT1	Top	CACCTGAACCGCATCGAGCTGAA
	Bottom	AAACTTCAGCTCGATGCGGTTCA
sgSETD2-2	Top	CACCGATCTAACTGCTACATCTAC
	Bottom	AAACGTAGATGTAGCAGTTAGATC
sgSETD2-3	Top	CACCGAACTGCTACATCTACTGGTA
	Bottom	AAACTACCAGTAGATGTAGCAGTTC

2.4 Cloning of sgRNA oligos into pLentiCRISPR_v2 vector

All sgRNA sequences (Table 2.5) were obtained from human GeCKOv2 (Addgene#1000000048) and Brunello (Addgene#73178) CRISPR libraries (Sanjana et al., 2014; Doench et al., 2016). For each target gene, three different sgRNAs were used. The compatible ends for BsmBI cut sites ("CACCG" for 5'ends of top oligos; "AAAC" for 5'ends of bottom oligos and "C" for 3'ends of bottom oligos) were added to sgRNA sequences and designed sequences were synthesized by Macrogen (http://www.macrogen.com/en/main/index.php).

Table 2.5: Sequences of sgRNAs cloned into pLentiCRISPR_v2 vector

Oligo			**Sequence (5'-3')**
sgNT1	Control	Top	CACCGACGGAGGCTAAGCGTCGCAA
		Bottom	AAACTTGCGACGCTTAGCCTCCGTC
sg1	sgSETD2-1	Top	CACCGACATCTACTGGTAAGGGTAC
		Bottom	AAACGTACCCTTACCAGTAGATGTC
sg2	sgSETD2-2	Top	CACCGATCTAACTGCTACATCTAC
		Bottom	AAACGTAGATGTAGCAGTTAGATC
sg3	sgSETD2-3	Top	CACCGAACTGCTACATCTACTGGTA
		Bottom	AAACTACCAGTAGATGTAGCAGTTC
sg4	sgBRPF1-1	Top	CACCGAGCACAACGCCCTCAAAACA
		Bottom	AAACTGTTTTGAGGGCGTTGTGCTC
sg5	sgBRPF1-2	Top	CACCGAGGGTGACTGCAGGCAACGG
		Bottom	AAACCCGTTGCCTGCAGTCACCCTC
sg6	sgBRPF1-3	Top	CACCGATCTTGGGGGCTCGCGAAAC
		Bottom	AAACGTTTCGCGAGCCCCCAAGATC
sg7	sgDNMT3A-1	Top	CACCGATGCCCGCCATGCCCTCCAG
		Bottom	AAACCTGGAGGGCATGGCGGGCATC
sg8	sgDNMT3A-2	Top	CACCGCAACATCGAATCCATGAAAA
		Bottom	AAACTTTTCATGGATTCGATGTTGC
sg9	sgDNMT3A-3	Top	CACCGCCGCCCCACCTTCCGTGCCG
		Bottom	AAACCGGCACGGAAGGTGGGGCGGC

sg10	sgDNMT3B-1	Top	CACCGAGAGTCGCGAGCTTGATCTT
		Bottom	AAACAAGATCAAGCTCGCGACTCTC
sg11	sgDNMT3B-2	Top	CACCGATCCGCACCCCGGAGATCAG
		Bottom	AAACCTGATCTCCGGGGTGCGGATC
sg12	sgDNMT3B-3	Top	CACCGCCATGTGGACGAGTCCCCCG
		Bottom	AAACCGGGGGACTCGTCCACATGGC
sg19	sgPSIP1-1	Top	CACCGAAAGGAAACTAGTGTTTCAA
		Bottom	AAACTTGAAACACTAGTTTCCTTTC
sg20	sgPSIP1-2	Top	CACCGAAGAGCCGGATAAAAAAGAG
		Bottom	AAACCTCTTTTTTATCCGGCTCTTC
sg21	sgPSIP1-3	Top	CACCGAGATCGAAAACGCAAGCAAG
		Bottom	AAACCTTGCTTGCGTTTTCGATCTC
sg22	sgMRG15-1	Top	CACCGAAATACTTCATACATTACAG
		Bottom	AAACCTGTAATGTATGAAGTATTTC
sg23	sgMRG15-2	Top	CACCGCCACATCTCCTGAGATTATT
		Bottom	AAACAATAATCTCAGGAGATGTGGC
sg24	sgMRG15-3	Top	CACCGGCCTCTTCTTTATGAAGCAA
		Bottom	AAACTTGCTTCATAAAGAAGAGGCC
sg25	sgMSH6-1	Top	CACCGAAGGCGAAGAACCTCAACGG
		Bottom	AAACCCGTTGAGGTTCTTCGCCTTC
sg26	sgMSH6-2	Top	CACCGATCACCACCTCCACTAACGT
		Bottom	AAACACGTTAGTGGAGGTGGTGATC
sg27	sgMSH6-3	Top	CACCGCAAAATGTTGCTGTGCGCCT
		Bottom	AAACAGGCGCACAGCAACATTTTGC
sg28	sgMSL3-1	Top	CACCGATCGTACAGCACTCGCGCCT
		Bottom	AAACAGGCGCGAGTGCTGTACGATC
sg29	sgMSL3-2	Top	CACCGATTGTTACTACATTAACAGG
		Bottom	AAACCCTGTTAATGTAGTAACAATC
sg30	sgMSL3-3	Top	CACCGCAACATTTGCTGCGATTGTT
		Bottom	AAACAACAATCGCAGCAAATGTTGC
sg31	sgMTF2-1	Top	CACCGAACTGGGATAGATTGCACCC
		Bottom	AAACGGGTGCAATCTATCCCAGTTC

sg32	sgMTF2-2	Top	CACCGAGAATGCACCGGCAGTCCGC
		Bottom	AAACGCGGACTGCCGGTGCATTCTC
sg33	sgMTF2-3	Top	CACCGATATGTGACAAGTGTGGCCA
		Bottom	AAACTGGCCACACTTGTCACATATC
sg34	sgMUM1-1	Top	CACCGATTCCTACTTCAAACGAACG
		Bottom	AAACCGTTCGTTTGAAGTAGGAATC
sg35	sgMUM1-2	Top	CACCGCCAGTCTTCCGAAGAGTCCA
		Bottom	AAACTGGACTCTTCGGAAGACTGGC
sg36	sgMUM1-3	Top	CACCGCTTACCTAACGAGGAAGCAA
		Bottom	AAACTTGCTTCCTCGTTAGGTAAGC
sg37	sgNPAC-1	Top	CACCGAAGAAGCATTTCTTTCCGCG
		Bottom	AAACCGCGGAAAGAAATGCTTCTTC
sg38	sgNPAC-2	Top	CACCGATATCCTCCTTGGCCAGGAA
		Bottom	AAACTTCCTGGCCAAGGAGGATATC
sg39	sgNPAC-3	Top	CACCGCACACCCACAGACAAAAAGT
		Bottom	AAACACTTTTTGTCTGTGGGTGTGC
sg40	sgNSD1-1	Top	CACCGAAAGGGCTGTCCTTGTCTTC
		Bottom	AAACGAAGACAAGGACAGCCCTTTC
sg41	sgNSD1-2	Top	CACCGAAGCACATAAAGATGAACGG
		Bottom	AAACCCGTTCATCTTTATGTGCTTC
sg42	sgNSD1-3	Top	CACCGAATCTGTTCATGCGCTTACG
		Bottom	AAACCGTAAGCGCATGAACAGATTC
sg43	sgNSD2-1	Top	CACCGACAACGTTGGTGATTTGGTG
		Bottom	AAACCACCAAATCACCAACGTTGTC
sg44	sgNSD2-2	Top	CACCGAGGGGCCACCAGGGCAAACT
		Bottom	AAACAGTTTGCCCTGGTGGCCCCTC
sg45	sgNSD2-3	Top	CACCGCAAAGAACTGTACGTGATAC
		Bottom	AAACGTATCACGTACAGTTCTTTGC
sg46	sgNSD3-1	Top	CACCGACACTTTACACGAAAAACAT
		Bottom	AAACATGTTTTTCGTGTAAAGTGTC
sg47	sgNSD3-2	Top	CACCGACAGTGTCAACCCTCTCATT
		Bottom	AAACAATGAGAGGGTTGACACTGTC

sg48	sgNSD3-3	Top	CACCGCCAAGGATAGGTTCCCACCT
		Bottom	AAACAGGTGGGAACCTATCCTTGGC
sg49	sgPHF1-1	Top	CACCGACTGATGGGCTGCTATACTT
		Bottom	AAACAAGTATAGCAGCCCATCAGTC
sg50	sgPHF1-2	Top	CACCGACTGCGAATCATCCTCAAAC
		Bottom	AAACGTTTGAGGATGATTCGCAGTC
sg51	sgPHF1-3	Top	CACCGAGCTGTGAGAAGTGTCGCCA
		Bottom	AAACTGGCGACACTTCTCACAGCTC
sg52	sgPHF19-1	Top	CACCGAATTATCTTCGAAAGTCACG
		Bottom	AAACCGTGACTTTCGAAGATAATTC
sg53	sgPHF19-2	Top	CACCGAGCACCCCAGTGACAGATCG
		Bottom	AAACCGATCTGTCACTGGGGTGCTC
sg54	sgPHF19-3	Top	CACCGATACTGGCCCTCCGTCAGTT
		Bottom	AAACAACTGACGGAGGGCCAGTATC
sg64	sgZMYND11-1	Top	CACCGAAATCTAAGAATGAGGACCG
		Bottom	AAACCGGTCCTCATTCTTAGATTTC
sg65	sgZMYND11-2	Top	CACCGACGTTTAACAAAAAGACGAC
		Bottom	AAACGTCGTCTTTTTGTTAAACGTC
sg66	sgZMYND11-3	Top	CACCGATAAACACCCGATGTACAGG
		Bottom	AAACCCTGTACATCGGGTGTTTATC

For cloning of sgRNAs into pLentiCRISPR_version2 (pLentiCRISPR_v2) backbone (Addgene #98291), CRISPR cloning protocol of Zhang Lab was followed (Shalem et al., 2014; Sanjana et al., 2014). Firstly, pLentiCRISPR_v2 vector was prepared for cloning as 8 μg of pLentiCRISPR_v2 backbone was digested with BsmBI enzyme at 55 ℃ for 3 hours in 4 separate reactions (Table 2.6).

Table 2.6: BsmBI digestion of pLentiCRISPR_v2 vector

Reagents	**1X Reaction**
pLentiCRISPR_v2	2000 ng
BsmBI	2 µL
Buffer 3.1	4 µL
dH_2O	up to 40 µL

The digested vector was then dephosphorylated by incubation with 2 µL Antarctic phosphatase (NEB M0289S) at 37 ℃ for 30 minutes to prevent self-ligation. The vector was run on an agarose gel (0.8%). BsmBI restriction endonuclease has two cutting sites and digestion releases 2 kb part from the pLentiCRISPR_v2 vector (Figure 2.1). After the release of this 2 kb part was detected on the gel, the remaining plasmid (about 13kb) was purified from the gel using NucleoSpin Gel and PCR Clean-up kit (MN 740609.50). Concentration of purified backbone was detected with NanoDrop spectrophotometer.

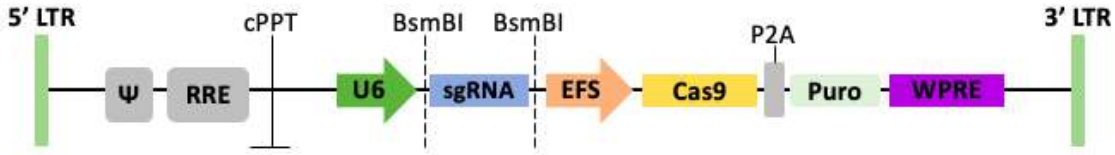


Figure 2.1: Schematic representation of pLentiCRISPR_v2 vector with inserted sgRNA.

sgRNA oligos were also prepared for cloning by phosphorylation and annealing. First, they were phosphorylated with T4 PNK (NEB M0201S). Then, the top and bottom oligos were annealed in a thermal cycler. The mix (Table 2.7) was prepared and incubated at 37 ℃ for 30 minutes, 95 ℃ for 5 minutes and then ramp down to 25 ℃ at the rate of 0.1 ℃/sec. Finally, annealed oligos were 1/100 diluted with dH_2O.

Table 2.7: Phosphorylation and annealing of sgRNA pairs

Reagents	**Volume for 1X reaction**
Top Oligo (100 μM)	1 μL
Bottom Oligo (100 μM)	1 μL
T4 PNK	0.5 μL
10X T4 DNA Ligase Buffer	1 μL
dH_2O	6.5 μL

50 ng of the digested backbone and 1 μL of diluted oligos were ligated using T4 DNA ligase (NEB #M0202). Ligation mix (Table 2.8) was incubated at room temperature (24-25 ℃) for 4 hours. Then, 5 μL of ligation product was transformed into 50 μL of Stbl3 bacteria. Plasmid was purified with mini kit (MN 740588.50) from the picked single colony. Finally, cloning was verified by Sanger sequencing with LKO.1 5' sequencing primer (5'-GACTATCATATGCTTACCGT-3', Weinberg Lab).

Table 2.8: Ligation mix for sgRNA cloning

Reagents	**1X Reaction**
Digested vector	50 ng
1:100 diluted oligo mix	1 μL
10X T4 DNA Ligase Buffer	2 μL
T4 DNA Ligase	1 μL
dH_2O	up to 20 μL

2.5 Cloning of H3.3 mutants into pWZL-Blast vector

Histone H3.3 cDNA sequence obtained from pcDNA-FLAG-H3.3-polyA plasmid (Addgene #59780) was previously cloned into pWZL-Blast vector (GFP sequence was removed, Addgene #12269) (Ozcimen, 2018). At the same time, H3.3K36M mutation was obtained with Q5 site-directed mutagenesis kit (NEB E0554) (Ozcimen, 2018). pWZL-Blast plasmid with H3.3 wild-type insert was now used to create H3.3G34 and H3.3S31 mutations as well (Figure 2.2). The suitable primers (Table 2.9) used in Q5 site-

directed mutagenesis were designed using the online tool NEBaseChanger (https://nebasechanger.neb.com).

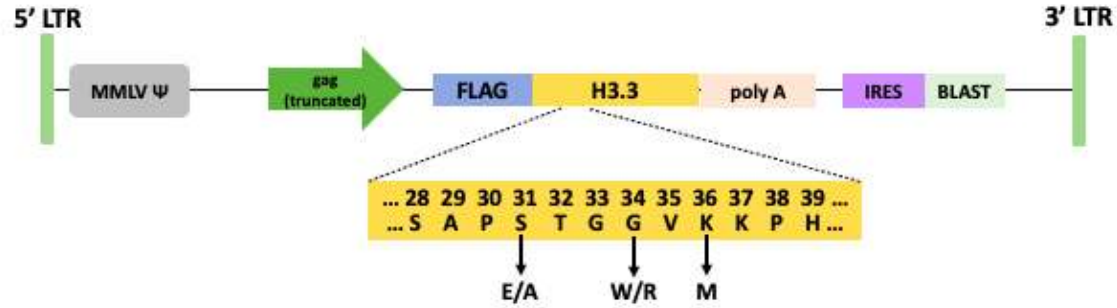


Figure 2.2: Schematic representation of pWZL-H3.3-Blast plasmid and mutations generated on H3.3 insert.

Table 2.9: Primers used in Q5 mutagenesis

Primers	**Sequences (5'-3')**
H3.3G34W Frw	CTCTACCGGCtggGTGAAGAAGC
H3.3G34W Rev	GGAGCGCTTTTCCTGGCG
H3.3G34R Frw	CTCTACCGGCaggGTGAAGAAGC
H3.3G34R Rev	GGAGCGCTTTTCCTGGCG
H3.3S31A Frw	AAGCGCTCCCgccACCGGCGGGG
H3.3S31A Rev	TTCCTGGCGGCTTTCGTGGC
H3.3S31E Frw	AAGCGCTCCCgaaACCGGCGGGG
H3.3S31E Rev	TTCCTGGCGGCTTTCGTGGC
Frw: Forward, Rev: Reverse	

In the first step of Q5 site-directed mutagenesis, the exponential amplification PCR was performed as described in Table 2.10 with defined thermocycling conditions (Table 2.11).

Table 2.10: Exponential amplification PCR reaction of Q5 mutagenesis

Reagents	**1X Reaction**	**Final Concentration**
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µL	1X
Forward primer (10 µM)	1.25 µL	0.5 µM
Reverse primer (10 µM)	1.25 µL	0.5 µM
Template plasmid (10 ng/µL)	1 µL	10 ng
dH_2O	9 µL	
	Total: 25 µL	

Table 2.11: Thermocycling conditions for PCR

STEP	**Temperature**	**Time**
Initial denaturation	98 ℃	30 seconds
25 Cycles	98 ℃	10 seconds
	50-72 ℃ (Annealing temp of primers)	30 seconds
	72 ℃	3 minutes
Final extension	72 ℃	5 minutes
Hold	4 ℃	

After generating substitutions by PCR, 1 µL of PCR product was treated with KLD (kinase, ligase & DpnI) mix at room temperature for 5 minutes to remove unmodified vectors (Table 2.12).

Table 2.12: Reaction for KLD treatment

Reagents	**1X Reaction**	**Final Concentration**
PCR product	1 µL	
2X KLD Reaction Buffer	5 µL	1X
10X KLD Enzyme Mix	1 µL	1X
dH_2O	3 µL	
	Total: 10 µL	

1 µL of KLD-treated mix was transformed into 50 µL of Stbl3 bacteria following manufacturer's instructions (NEB E0554). Single colonies were picked from LB-agar plates and inoculated for mini plasmid purification. Then, the isolated plasmid was sequenced with primer (5'- GGATGGCACACAGGTTGGTA-3') to verify mutagenesis (Figure 2.3 and Figure 2.4).

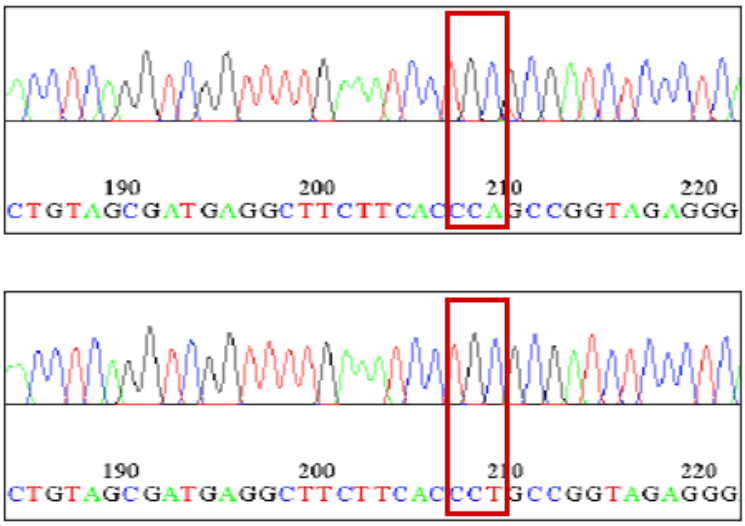


Figure 2.3: H3.3 sequence with glycine-to-tryptophan (top) and glycine-to-arginine (bottom) mutations at 34th residue.

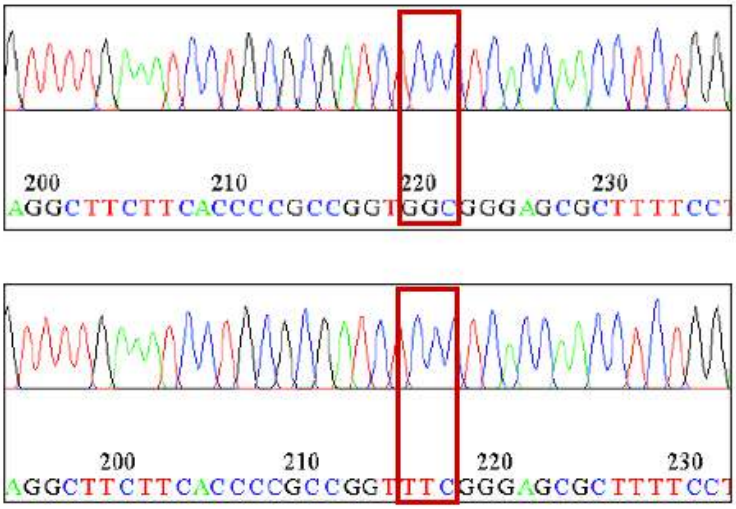


Figure 2.4: H3.3 sequence with serine-to-alanine (top) and serine-to-glutamic acid (bottom) mutations at 31th residue.

2.6 Viral Packaging

For viral packaging of lenti- and retro-plasmids, 293T cells were used which were cultured in 10% FBS containing DMEM media (called as D10, prepared with DMEM 1X, 10% heat-inactivated FBS, 1% Penicillin/Streptomycin) at 37 ℃ with 5% CO_2. The day before transfection, 293T cells were seeded as 2x10^6 cells per 10 cm plate. For transfection, the required plasmids were assembled in defined amounts (Table 2.13).

Table 2.13: Viral packaging

Plasmids	**Retroviral**	**Lentiviral**
psPAX2 (Addgene #12260)	-	2250 ng
pUMVC (Addgene #8449)	2250 ng	-
pCMV-VSV-G (Addgene #8454)	250 ng	250 ng
Vector	2500 ng	2500 ng
DMEM	up to 200 µL	up to 200 µL

Besides this plasmid mix, 15 µL of FuGENE (Promega E2312) was diluted with 185 µL of DMEM. Then, the plasmid mix was added onto FuGENE and the whole mix was incubated at room temperature for 30 minutes. After incubation, this mix was added onto 293T cells dropwise. The medium of 293T cells was changed with fresh D10 after 16 hours. At 48 and 72 hours after transfection, the medium of 293T cells was collected. The collected medium was centrifuged at 1500 rpm for 5 minutes and then filtered with 0.45 µm filter to get rid of 293T cells. Finally, the medium containing viruses was aliquoted and stored at -80 °C. For preparation of concentrated virus, PEG 8000 (polyethylene glycol, Sigma P5413) was used. 50% PEG 8000 was prepared with 1X PBS and added onto the collected and filtered medium containing virus at 1/4 of its volume. After two days incubation with PEG at 4 °C, the mix was centrifuged at 2000 rpm for 20 minutes and supernatant was discarded. The pellet was resuspended with 1X PBS at 1/100 of the initial volume. Then, it was aliquoted in small volumes (50-100 µL) and stored at -80 °C.

2.7 Culture of dH1f cells

dH1f cells were frozen in the freezing media (50% FBS, 40% D10 and 10% DMSO). Frozen dH1f cells were thawed at 37 °C for 2 minutes and freezing media was removed through centrifugation at 1400 rpm for 4 minutes. The cell pellet was resuspended with D10 medium and seeded into plates as 10^6 cells per 10 cm plate. The cells were then left at 37 °C under 5% CO_2 pressure. The medium of the cells was changed with fresh D10 every 2 days and cells were passaged when they reached at 80% confluency. For dH1f cells passaging, the cells were detached by trypsinization at 37 °C for 3 minutes and then trypsin was inactivated by D10 media. Using hemocytometer, the cells were counted and seeded at a density depending on the experiment.

2.8 Viral Infection

Viral infections were performed either at 6-well plate or 12-well plate according to the need of the experiment. For a well of 6-well plate, 10^5 dH1f cells were seeded while $5x10^4$ cells were seeded into a well of 12-well plate. The dH1f cells seeded into 6-well plate were treated with 1 mL of unconcentrated virus, 500 µL of D10 medium and 1.5 µL of protamine sulfate (8 mg/mL, Sigma P4505). The cells seeded into 12-well plate were

treated with 500 µL of unconcentrated virus, 200 µL of D10 medium and 0.7 µL of protamine sulfate. After 16 hours from viral infection, the medium of cells was changed with fresh D10. Then, viral infection was repeated at 24 hours and the medium was again changed after overnight incubation. Antibiotic selection for infected cells were performed depending on the resistance gene found at the transduced plasmid. Puromycin was used as 1 µg/mL while blasticidin was applied as 7 µg/mL. Antibiotic selection was generally completed in 3 days for dH1f cells.

2.9 Reprogramming of dH1f cells into iPSCs

Reprogramming of dH1f cells was performed in 12-well plates so dH1f cells were seeded as $5x10^4$ per well. For transfer of Yamanaka factors, the OCT4 and SOX2 expressing plasmid pSIN4-EF2-O2S (Addgene #21162) and KLF4 and c-MYC expressing plasmid pSIN4-CMV-K2M (Addgene #21164) were used together. The cells were infected with unconcentrated viruses of pSIN4-EF2-O2S and pSIN4-CMV-K2M plasmids the day after seeding onto 12- well plates. The cells were left for 16 hours incubation. Then, the viral transduction medium was changed with fresh D10 medium. This change was accepted as the first day medium change of 21-day reprogramming process and medium change with fresh D10 was repeated at third and fifth days as well. By the sixth day, 1/8 of the cells were transferred onto MEF-feeder cells via trypsinization. Mitomycin-inactivated MEF-feeder preparation was performed as described in Ozcimen, 2018. On the next day of transfer onto MEF-feeder, the medium of the cells was changed with hES medium. hES medium was prepared with 384 mL DMEM-F12, 100 mL knockout serum (KOSR), 5 mL L-Glutamine, 5 mL NEAA, 5 mL Penicillin/Streptomycin, 500 µL 2-mercaptoethanol and 500 µL of 10 µg/mL FGF2. Until 14^{th} day of reprogramming, the media of cells was changed in every two days with fresh hES medium. After that, the medium was renewed every day.

At the end of 21-day reprogramming process, the cells were fixed with 4% paraformaldehyde at room temperature for 15 minutes. Then, plates were washed three times with PBS. For quantification of the generated iPSC colonies, the staining was performed based on the stem cell marker TRA-1-60. Fixed cells were incubated with the biotin-conjugated TRA-1-60 first antibody (BioLegend 330604) overnight at 4 °C. The antibody was 1:200 diluted with the staining buffer containing 3% Trition-X and 3% FBS

in PBS. After washing 5 times with PBS, the cells were treated with streptavidin-conjugated HRP (BioLegend 405210) for 2 hours at room temperature. The HRP enzyme was 1:500 diluted also with the staining buffer. Following the washing steps 5 times with PBS, the cells were exposed to DAB solution, the substrate of HRP enzyme, for 15-20 minutes at room temperature. DAB solution was prepared as 0.05% DAB (Sigma-Aldrich D8001), 0.05% Nickel Ammonium Sulfate (Sigma-Aldrich 574988) and 0.015% hydrogen peroxide (Sigma Aldrich H1009) in PBS. HRP-bound iPSC cells form brown-precipitates upon DAB treatment. Thus, the plates were scanned and TRA-1-60+ iPSC colony numbers were determined by ImageJ analysis.

2.10 RNA isolation and cDNA synthesis

For RNA isolation, NucleoSpin RNA kit (MN 740955.50) was followed. RNA isolation was performed with direct-zol RNA miniprep kit (Zymogen R2050) if the isolated RNA was further planned to be sequenced. Then, 1000 ng of RNA was used for cDNA synthesis. In the first step of cDNA synthesis, RNA, dNTP and random hexamers were assembled in PCR tubes with the defined amounts (Table 2.14).

Table 2.14: cDNA synthesis

Reagents	**Catalog No**	**1X Reaction**
dNTP (2 mM)	Thermo Scientific R0192	2.5 µL
Random Hexamer (50 µM)	Jena Bioscience PM-301S	1 µL
RNA	-	1000 ng
dH_2O	-	up to 16.5 µL

The mixture was incubated at 65 °C for 5 minutes, immediately after it was placed on ice. In the next step, reverse transcriptase mix was prepared (Table 2.15) and added onto the first mix. The whole mixture was incubated at room temperature for 10 minutes.

Table 2.15: Reverse transcriptase mix

Reagents	Catalog No	1X Reaction
5X First Strand Buffer	Invitrogen P/N Y02321	5 µL
DTT (0.1 M)	Invitrogen P/N Y00147	2 µL
Rnasin	Promega N251B	0.5 µL

After 10 minutes incubation, 1 µL of MMLV RT enzyme (Invitrogen 28025-013) was added onto each sample and reaction was performed at 37 ℃ for 1 hour and inactivated at 70 ℃ for 15 minutes. Finally, the cDNA products were ¼ diluted with 75 µL of DNase/RNase free water and stored at -20 ℃.

2.11 Real-Time Quantitative PCR (RT-qPCR)

RT-qPCR was performed to analyze relative mRNA expression levels from cDNA samples. For RT-qPCR, SYBR mixture was prepared and assembled with cDNA in 96-well PCR plates (Table 2.16).

Table 2.16: RT-qPCR mix

Reagents	1X Reaction
2X SYBR Green (Roche 12054420)	10 µL
Primer mix (2.5 µM)	2 µL
dH_2O	6 µL
cDNA	2 µL

Reaction was performed using Roche Light Cycler 480II. After Ct data obtained from LightCycler, ΔΔCT method was followed and Beta-Actin expression was used as the internal control. The primers used for RT-qPCR was listed at Table 2.17 and Table 2.18.

Table 2.17: The list of RT-qPCR primers

Primers	**Direction**	**Sequences (5'-3')**
BETA-ACTIN	Forward	TGAAGTGTGACGTGGACATC
	Reverse	GGAGGAGCAATGATCTTGAT
SETD2	Forward	GGAAAACACAACAACTGAACGAG
	Reverse	TGAGTTTGCTGGTCTGGGTCT
CITED2	Forward	TTCCCTCACTTTCTCCAGTGCTCA
	Reverse	ATGAAGCGAGATGGCAGTTTGTGC
EBF3	Forward	GACCCGTCAGAAGCCACTC
	Reverse	TGCAGCCCGTCAAAGAAG
SMAD3	Forward	GCGTGCGGCTCTACTACATC
	Reverse	GCACATTCGGGTCAACTGGTA
FOSL1	Forward	TGACCACACCCTCCCTAACTC
	Reverse	CTGCTGCTACTCTTGCGATGA
NANOG	Forward	TGATTTGTGGGCCTGAAGAAA
	Reverse	TGGTGGTAGGAAGAGTAAAG
LIN28A	Forward	GCGGGCATCTGTAAGTGG
	Reverse	GGAACCCTTCCATGTGCAG
PSIP1 (LEDGF)	Forward	GGGCCAGAAGGAAGAAGATAAG
	Reverse	TCAGAATCGGAGGTTGAAGTAAC
MRG15	Forward	AGCAATGTTGGCTTATACACCTC
	Reverse	AGCTTTCCGATGGTACTCAGG
MSH6	Forward	CCAAGGCGAAGAACCTCAAC
	Reverse	ACCAGGGGTAACCCTCCATC
MSL3	Forward	TGTGCCTGACAATTACCCC
	Reverse	CCAAGGATTTCTGGAAGTTTCAC
NSD2	Forward	TGTAAACCACTGAAGAAGCGA
	Reverse	GTCCGAGACCTCATTCTCAG
NSD3	Forward	CAGACGTTTCTGATGTGCAGTCC
	Reverse	CTCCAGGTGAAAGTGTTTGCAGC

Table 2.18: The list of RT-qPCR primers for alternative splicing analysis

Primers	**Direction**	**Sequences (5'-3')**
MBD2a	Forward	AGCAAGCCTCAGTTGGCAAGGT
	Reverse	TGTTCATTCATTCGTTGTGGGTCTG
MBD2c	Forward	AGCAAGCCTCAGTTGGCAAGGT
	Reverse	TGAAAGCGCATGCCATGGTGCA
ZNF207A	Forward	CCAGGCATGCCACCTGGAATGA
	Reverse	GAGCTTGTGACAGGTGTCCCCAT
ZNF207B	Forward	TTGCATCATCAGAGAAAATAC
	Reverse	GAGCTTGTGACAGGTGTCCCCAT
ZNF207C	Forward	CCACCTGTTCCACGTCCT
	Reverse	CTGAGCCTGTCCAGCACT
PRDM14_short	Forward	GTGGTGTCGGAGGATCTGA
	Reverse	GCCCAAACCTGACTCCTTTG
PRDM14_long	Forward	CTCTTGGAGGTGGTGTCGGA
	Reverse	GCCAGGCTGTTTCTGTAGTGT
FOXP1a	Forward	TCAGTGGTAACCCTTCCCTT
	Reverse	CTGGCTAAGTTGCCCAGAGT
FOXP1b	Forward	CACGTGGAAGGGTGCCAT
	Reverse	GAGGACGGCCTCGTTGAATA
TCF3a	Forward	CCCGGACCAGCCCAGA
	Reverse	CAGCTCCTTAAAGGCCTCGT
TCF3b	Forward	CCGGACCAGCAGTACGGA
	Reverse	AGGATGAGCAGCTTGGTCTG
GRHL1_long	Forward	CAGACTGAGATCCGAGGGGA
	Reverse	CAGACTGAGATCCGAGGGGA
GRHL1_short	Forward	GAAGGCGTTCAGGAGAGTGT
	Reverse	TGCATCTTTGTTTAGCGGTGTG

2.12 Western Blotting

For protein isolation, different protocols were followed depending on the type of lysate needed. However, the part to obtain cell pellets was done as standard by centrifugation of fibroblasts after viral infection and antibiotic selection. Histone extraction was performed as described in the protocol of Abcam website (https://www.abcam.com/protocols/histone-extraction-protocol-for-western-blot). The fibroblast pellets were lysed with Triton Extraction Buffer (TEB: 0.5% Trition X, 2 mM PMSF, 0.02% (w/v) NaN_3 in PBS) on ice for 10 minutes by gentle shaking. Then, cells were centrifuged at 2000 rpm for 10 minutes and supernatant was discarded. The cell pellet was washed with the half volume of TEB and centrifuged again. The pellet was resuspended with 0.2 N HCl and it was left at 4 ℃ overnight. During this incubation, acid extracts histones into supernatant. On the next day, the sample was centrifuged at 2000 rpm for 10 minutes and supernatant containing histones was collected. To neutralize the samples, 0.1 M NaOH up to 1/5 volume of HCl used was added into supernatant. The extracted histones were stored at -80 ℃ for long-term or at -20 ℃ for short-term until western blotting.

To extract whole cell protein, cell pellets were lysed on ice for 60 minutes with the lysis buffer composed of the reagents listed in Table 2.19. During the lysis, the samples were vortexed at each 15 minutes interval. Then, they were centrifuged at 13000 rpm for 10 minutes to pellet cell debris. The supernatant containing proteins was collected and stored at -80 ℃ for further experiments.

Table 2.19: Whole cell lysis buffer composition

Reagents	**Stock Conc.**	**Final Conc.**
TRIS (pH: 7.4)	1 M	50 mM
NaCl	1 M	250 mM
EDTA	10X	1X
NaF	1 M	50 mM
Noidet P40 (NP40)	10%	1%
NaN_3	1%	0.025%
PMSF	100 mM	1 mM
Protease Inhibitor Cocktail	10X	1X
dH_2O	required volume	

To extract nuclear proteins, nuclear/cytosolic fractionation protocol was followed. Cell pellets were firstly resuspended with cytosolic lysis buffer (Table 2.20) and lysed on ice for 15 minutes by shaking. Then, the samples were centrifuged at 3000 g for 3 minutes and supernatant was discarded. The pellet was washed with the half volume of cytosolic buffer used and centrifuged again. In the second step, the pellet was resuspended with nuclear lysis buffer (Table 2.21). The samples were sonicated as 4 cycles of 10 seconds on and 30 seconds off using Bioruptor Sonicator (Diagenode B01020001). The sonicated samples were then centrifuged at 15000 g for 5 minutes and supernatant containing nuclear proteins was collected. The proteins were stored at -80 °C for long-term usage.

Table 2.20: Cytosolic lysis buffer composition

Reagents	**Stock Conc.**	**Final Conc.**
HEPES (pH: 7.9)	500 mM	10 mM
KCl	500 mM	10 mM
EDTA	25 mM	0.1 mM
Noidet P40 (NP40)	10%	0.4%
Protease Inhibitor Cocktail	10X	1X
dH_2O	required volume	

Table 2.21: Nuclear lysis buffer composition

Reagents	**Stock Conc.**	**Final Conc.**
HEPES (pH: 7.9)	500 mM	20 mM
NaCl	5 mM	0.4 mM
EDTA	25 mM	1 mM
Glycerol	50%	10%
Protease Inhibitor Cocktail	10X	1X
dH_2O	required volume	

Concentrations of the extracted proteins were determined with BCA protein assay kit (Pierce 23227). The color development through BCA reaction was quantified with plate reader at 562 nm. Then, concentrations were calculated according to absorbance values of BSA standards. BSA protein (supplied with Pierce 23227) was used in 7 different concentrations (2, 1.5, 1, 0.75, 0.5, 0.37 and 0.25 µM) as standards in BCA analysis.

After protein concentrations were calculated, the proteins were prepared for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Firstly, protein loading buffer was prepared as 4X mixing 900 µL of 4X Laemmli sample buffer (Biorad 1610747) with 100 µL of 5 M 2-mercaptoethanol. Then, protein concentrations were adjusted in the desired volume with 1X loading buffer and dH_2O, and proteins were boiled at 95 °C for 15 minutes. The boiled samples were immediately put and kept on ice. If they were used in further experiments, they were stored at -20 °C.

Prepared proteins were loaded into polyacrylamide gel (Mini-Protean 4568084) and run with 1X TGS running buffer at 20 mA per gel for 45-60 minutes. 1X TGS (Tris/Glycine/SDS) buffer was prepared by dilution of 10X TGS buffer with distilled water (BioRad 161-0772). Proteins were transferred onto methanol-activated PVDF membrane from the gel after running was completed. Transfer was performed using Trans-Blot Turbo transfer system (BioRad 1704150) as semi-dry. For transfer, 1X transfer buffer was prepared from 10X transfer buffer TG (Tris/Glycine) with 70% H_2O and 20% methanol. The stacks used in the transfer system were wet with 1X transfer buffer. After transfer, PVDF membrane was blocked in 5% non-fat dry milk prepared

with TBS-T (24 g Trizma Base, 88 g NaCl in 1L dH_2O with pH 7.6 and supplemented with 1 mL Tween-20) for 1 hour at room temperature by shaking. Then, membrane was left overnight at 4 ℃ for first antibody incubation by shaking. First antibodies were used as 1:1000 diluted with primary antibody buffer (2% BSA in TBS-T). After 1^o antibody incubation, membranes were washed with TBS-T several times by shaking. Then, they were incubated with 1:5000 diluted HRP-conjugated secondary antibody at room temperature for 1 hour by shaking. The membrane was washed again with TBS-T several times. Consequently, the membranes were visualized by Li-cor Odyssey Fc imaging system using enhanced chemiluminescence (ECL, Thermo Scientific 32106) as the substrate of HRP. The antibodies used in western blot experiments were listed in Table 2.22.

Table 2.22: Antibodies used in western blot experiments

Antibodies	**Catalog No**	**Source**
Histone H3 (D1H2) XP	Cell Signaling 4499S	Rabbit
Tri-Methyl-Histone H3 (Lys36) (D5A7)	Cell Signaling 4909S	Rabbit
Tri-Methyl-Histone H3 (Lys27) (C36B11)	Cell Signaling 9733S	Rabbit
Di-Methyl-Histone H3 (Lys36) (C75H12)	Cell Signaling 2901S	Rabbit
Monoclonal ANTI-FLAG M2	Sigma F1804	Mouse
Beta-tubulin	Abcam ab6046	Rabbit
Anti-PSIP1/LEDGF [EPR11890]	Abcam ab177159	Rabbit
MORF4L1/MRG15 (D2Y4J)	Cell Signaling 14098	Rabbit
Anti-WHSC1/NSD2 [29D1]	Abcam ab75359	Mouse
Anti-NSD3 [EPR13813]	Abcam ab180500	Rabbit
Goat Anti-Rabbit IgG H&L (HRP)	Abcam ab95071	Goat
Goat Anti-Mouse IgG H&L (HRP)	Abcam ab97023	Goat

2.13 Flow Cytometry Analysis

The titer of concentrated OCT4 and SOX2 viruses was determined by flow cytometry analysis. The plasmids of OCT4 (Addgene #17225) and SOX2 (Addgene #17226) contain IRES-GFP, so titration was performed based on GFP percentage (Park

et al., 2008). 293T cells infected with the increasing amount of OCT4 and SOX2 viruses were detached by trypsinization, and trypsin was inactivated with PBS supplemented with 10% FBS. Then, detached cells were centrifuged at 1500 rpm for 5 minutes. The pellets were resuspended in chilled PBS supplemented with 1% FBS. After cells were passed through cell strainer, they were subjected to BD Accuri C6 flow cytometry. Uninfected 293T cells were used as control for gating. In each reading, 10.000 events were recorded. According to GFP percentage detected by flow cytometry, the amount of viruses needed for infection of dH1f cells was determined.

2.14 Culture of Sf9 insect cell line

Sf9 insect cell culture was optimized from the protocol of Thermo Fisher Scientific(https://tools.thermofisher.com/content/sfs/manuals/Insect_Cell_Lines_UG.pdf). Sf9 cells were seeded at a density of $3x10^5$–$5x10^5$ viable cells/mL for suspension culture. They were incubated in a non-humidified incubator at 27 °C shaking at 90 rpm. They were grown in Insect-XPRESS media (BELN12-730Q) supplemented with 10% FBS. The cells were passaged and expanded every four days.

2.15 Chromatin Immunoprecipitation-coupled qPCR

Chromatin immunoprecipitation (ChIP) sequencing data of normal human dermal fibroblast (NHDF) and iPSC was obtained from ENCODE and visualized using Integrative Genomics Viewer (IGV) tool based on Human GRCh37/hg19 reference genome. qPCR primers (Table 2.23) used were designed according to the data of H3K36me3 and H3K27me3 from NHDF and iPSC cell lines. The efficiency of ChIP-qPCR primers was checked by serial dilutions.

Table 2.23: The list of ChIP-qPCR primers

Primers	**Direction**	**Sequences (5'-3')**
LPAR1-3'end	Forward	AGGGGGAGGCTGTTTATCCT
	Reverse	CCTCCCTCAACCACACCATC
SFRP1-Pro	Forward	CCTACTCAAGTCCTAGCCCG
	Reverse	TGTGTGCCTGAGTGATGGAC
SMAD3-3'end	Forward	CAGCTCTGAGACCCAGGAAC
	Reverse	ATAGGATCGTGCAGCAGGTG
NANOG-Pro	Forward	CACACACACCCACACGAGAT
	Reverse	TTGTCTATCCCTCCTCCCAGG
EBF3-3'end-1	Forward	TCCAAGCAATGCACACACTC
	Reverse	CTCCGCTAACTCTCCCTACG
EBF3-3'end-2	Forward	CGTGCGGTCATCTCTCTGTA
	Reverse	TGGGCTGTGAGTGGCTTT
EBF3-3'end-3	Forward	ACATAGACCTCACAGTAGCCT
	Reverse	AGCAAGTGTGATCTGCAACA
EBF3-Pro1	Forward	AGTGGGAAGGAGGTCAGAGG
	Reverse	CGTGCTGTTTCCACCCTATC
EBF3-Pro2	Forward	CACTGGTGTAGGAGGGAACA
	Reverse	CAAAGCGTTTCTCAACCCCA
EBF3-5'end	Forward	TTATGTCCTTTCACCAGCGC
	Reverse	GTACAGCAACGGTAAGTGCC

For ChIP experiments, ChIP-IT High Sensitivity kit (Active Motif 53040) was utilized. Firstly, dH1f cells infected with shCntrl and shSETD2-3 viruses were collected at both before (pre-OSKM) and after 6 days (post-OSKM) OSKM-transduction. In each condition, Sf9 insect cells were added at 1/5 amount of dH1f cells as spike-in. Then, the cells were crosslinked with fixation buffer (supplied with kit) supplemented with 37% formaldehyde at room temperature for 10 minutes by shaking. After cell lysis using a dounce homogenizer (146), nuclei were sonicated by Diagenode Bioruptor Sonicator to fragment chromatin. The sonication conditions were determined for optimal shearing of

dH1f cells. According to the time course and cell number optimization data, we proceed sonication with 3 million dH1f cells for 6 rounds of 10 cycles (30 sec ON/30 second OFF) (Figure 2.5).

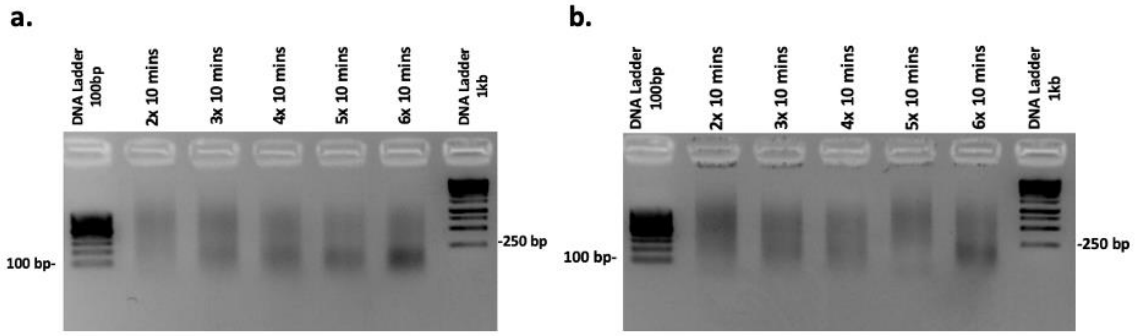


Figure 2.5: Time course and cell number optimization for sonication of dH1f cells. dH1f cells were crosslinked with paraformaldehyde (for 10 mins at room temperature). Nuclei were isolated from 3 million cells (**a.**) and 6 million cells (**b.**) per condition using buffers from ChIP-IT High Sensitivity kit (Active Motif 53040). Before sonication, samples were resuspended in 300 µL ChIP buffer. Samples were sonicated for 1, 2, 3, 4, 5 and 6 rounds of 10 cycles of 30 sec ON/30 sec OFF with Diagenode Bioruptor Sonicator. 1500 ng of each sample was analyzed on a 1.5% agarose gel.

For each ChIP reaction, the sheared chromatin of 4 million cells was incubated with 4 µg of antibody. H3 total (Abcam ab1791), H3K36me3 (Active Motif 61101) and H3K27me3 (Millipore 07-449) antibodies were used, and the antibody incubation was performed as three technical replicates at 4 °C on end-to-end rotator overnight. The sheared chromatin of $4x10^5$ cells (1/10 amount of immunoprecipitated chromatin) was used for input preparation following manufacturer's instructions. Then, antibody-bound chromatin fragments were pulled-down by Protein G agarose beads (supplied with kit). Cross-links were reversed to remove histone proteins from chromatin fragments and DNA fragments were purified for analysis. Purified DNA and prepared input were analyzed through qPCR to test enrichment of H3K36me3 and H3K27me3 marks on primer-targeted regions.

Chapter 3: RESULTS

3.1 The additive effect of H3.3K36M mutation and DOT1L inhibition on reprogramming efficiency

Lysine-to-Methionine mutation at the 36th residue of histone H3.3 variant (H3.3K36M) is a mutation detected in several cancer types and causes a global decrease in H3K36me3 levels (Fang et al., 2016; Lewis et al., 2013). Previously, it was shown that H3.3K36M mutation phenocopied the increasing effect of SETD2 inhibition on reprogramming efficiency (Ozcimen, 2018). In addition, the additive effect of SETD2 knockdown and DOT1L inhibition on reprogramming was shown (Ozcimen, 2018). In this thesis, I first repeated these experiments. Therefore, I overexpressed H3.3 wild-type (H3.3WT) and H3.3K36M mutant in dH1f cells using pWZL-H3.3-Blast and pWZL-H3.3K36M-Blast vectors which were previously cloned (Figure 3.1) (Ozcimen, 2018). Infected cells were then selected by blasticidin. H3.3WT and H3.3K36M mutant-infected cells were treated with DMSO or 3 μM EPZ4777 (iDOT1L), which is an inhibitor of DOT1L, during the first 7 days of reprogramming. Experiment was performed as 4 biological repeats, each contained 3 technical replicates. The increase in reprogramming was detected by both overexpression of H3.3K36M mutant histone and iDOT1L treatment in H3.3wt-expressed cells as expected. Furthermore, the additive effect of DOT1L inhibition on reprogramming when combined with SETD2 knockdown was similarly observed when combined with H3.3K36M mutation (Figure 3.2). These results indicate that the additive effect of SETD2 knockdown observed with DOT1L inhibition is through its methyltransferase activity for H3K36 residue.

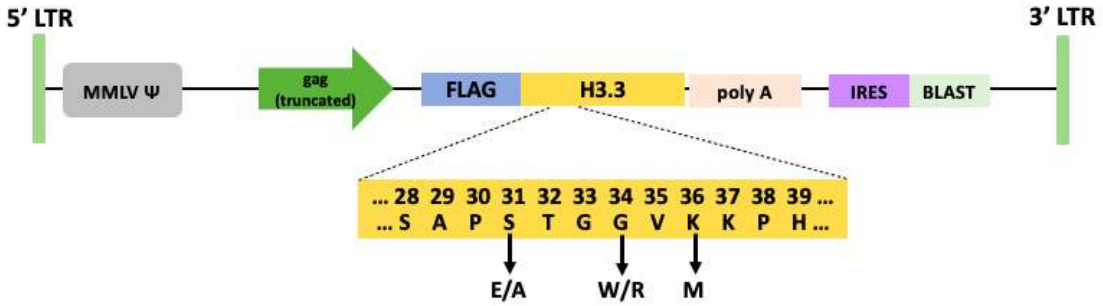


Figure 3.1: Schematic representation of pWZL-H3.3-Blast plasmid and mutations generated on H3.3 insert

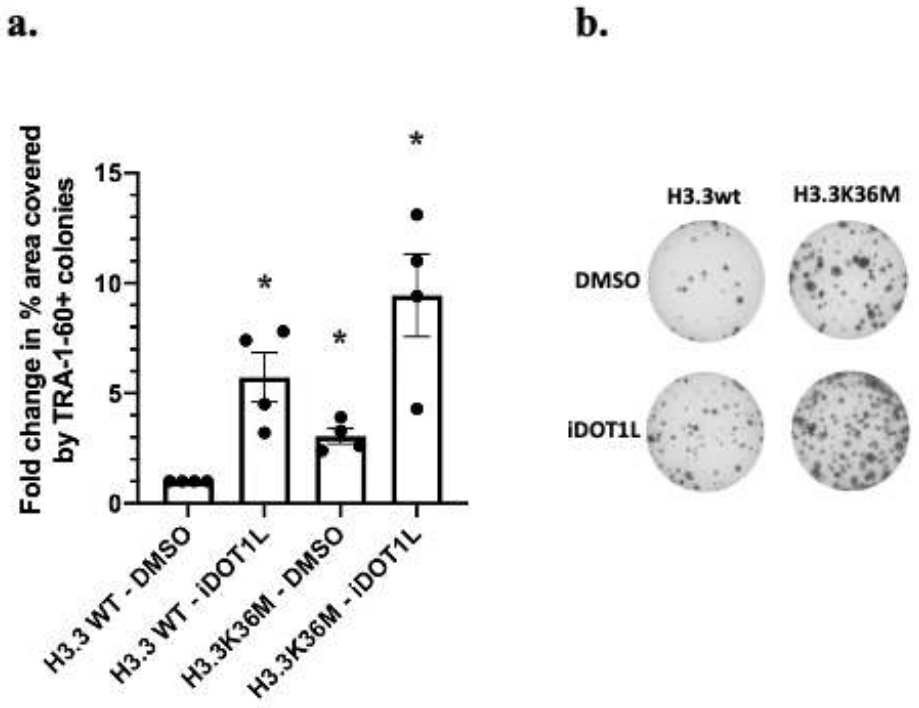


Figure 3.2: The additive effect of H3.3K36M mutation and DOT1L inhibition on OSKM-mediated reprogramming. a. Fold change in % area covered by TRA-1-60-positive colonies generated via OSKM induction over H3.3WT-DMSO control group. Bar graphs indicate means and error bars represent standard error of means. n=4, each biological replicates contained 3 technical replicates. P-values were calculated by two-tailed t-test. * denotes $p<0.05$. **b.** Representative well images of each condition.

The same experiment was conducted with OS-reprogramming excluding *KLF4* and *c-MYC* expression to understand whether decreased H3K36me3 levels could replace these 2 factors of OSKM-mediated reprogramming. The obtained data demonstrated that decreased H3K36me3 level through H3.3K36M mutation could eliminate the requirement of *KLF4* and *c-MYC* for reprogramming (Figure 3.3). When combined with DOT1L inhibition, the further increase in reprogramming efficiency was observed in OS-reprogramming as well (Figure 3.3).

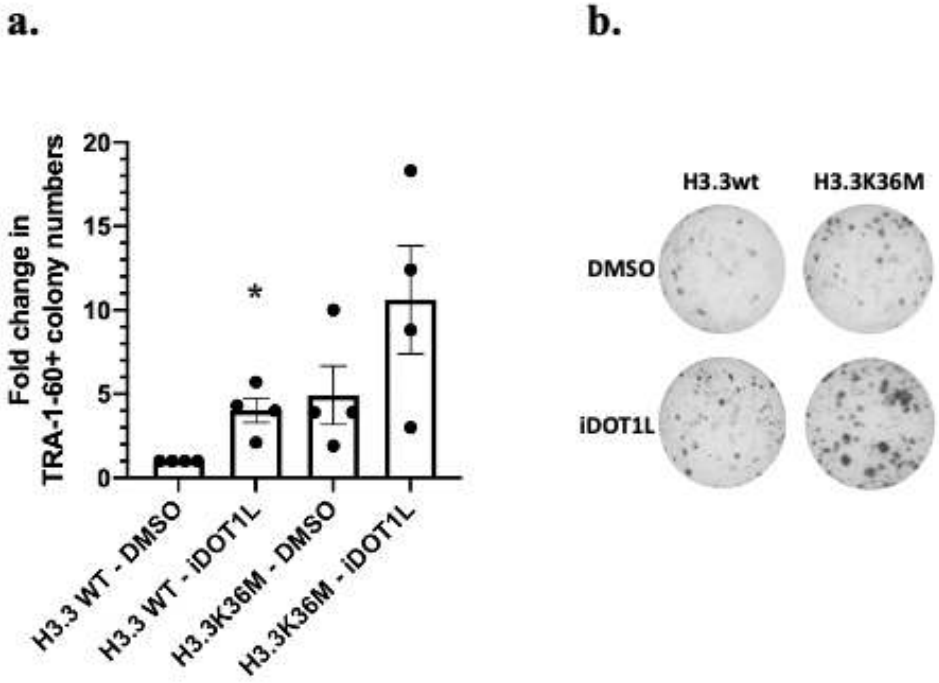


Figure 3.3: H3.3K36M mutation and DOT1L inhibition enable iPSC generation without *KLF4* and *c-MYC*. a. Fold change in the number of TRA-1-60 positive colonies generated by *OCT4* and *SOX2* (OS) induction over H3.3WT-DMSO control group. Bar graphs indicate means and error bars represent standard error of means. n=4, each biological replicates contained 3 technical replicates. P-values were calculated by two-tailed t-test. * denotes $p<0.05$. **b.** Representative well images of each condition.

3.2 Testing histone H3.3 mutations regulating SETD2 activity in reprogramming

A number of mutations on histone H3 has been shown to regulate the catalytic activity of SETD2. While the mutations at H3.3S31 residue have been reported to enhance SETD2 activity, the mutations at H3G34 residue impeded SETD2 function (Armache et al., 2020; Nohr et al., 2017). In the light of this knowledge, I asked if I could observe a similar effect in reprogramming with these mutations as SETD2 depletion. I introduced glycine-to-tryptophan and glycine-to-arginine substitutions at 34th residue on pWZL-H3.3-Blast vector via site-directed mutagenesis (Figure 2.2). On the other hand, I performed serine-to-alanine and serine-to-glutamic acid substitutions at 31th residue on

pWZL-H3.3-Blast vector (Figure 2.3). Then, I overexpressed these mutant histones and wild-type histone H3.3 as control into dH1f cells. In addition, I infected cells with H3.3K36M mutant histone as a positive control. After antibiotic selection with blasticidin, dH1f cells were reprogrammed by OSKM factors. This experiment was performed as 3 biological replicates, each contained 3 technical replicates. As expected, I observed significantly increased iPSC generation by H3.3K36M mutation compared to H3.3 wild-type infected control cells (Figure 3.4). On the other hand, among tested histone mutants only H3.3S31A led to a significant change in reprogramming efficiency as it reduced the number of iPSC colonies (Figure 3.4).

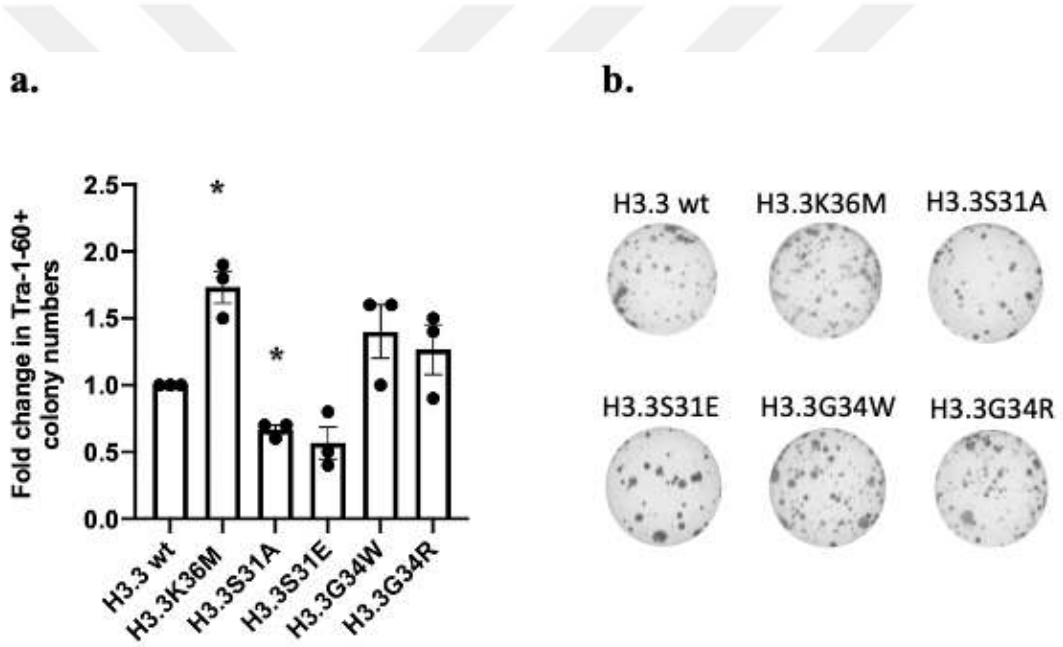


Figure 3.4: The effects of different histone H3.3 mutations regulating SETD2 activity on OSKM-reprogramming. a. Fold change in the number of TRA-1-60-positive colonies generated by OSKM induction over H3.3wt control group. Bar graphs indicate means and error bars represent standard error of means. n=3, each biological replicates contained 3 technical replicates. P-values were calculated by two-tailed t-test. * denotes $p<0.05$. **b.** Representative well images of each condition.

After testing these mutants in reprogramming, I checked the levels of H3K36me3, H3K36me2 and H3K27me3 upon their overexpression by western blotting. The expression of mutant histones was checked by FLAG immunoblotting since H3.3 insert was cloned with FLAG tag into pWZL-Blast backbone (Figure 3.1) (Ozcimen, 2018). Consistent with the previous data, H3.3K36M mutation led to decrease in global H3K36me2 and H3K36me3 levels and increase in H3K27me3 levels compared to H3.3wt control group (Figure 3.5). In H3.3S31 mutant samples, an increase was observed in H3K36me3 levels that is consistent with their enhancing effect on SETD2 activity reported by Armache *et al.* (Figure 3.5) (Armache et al., 2020). While H3.3G34W mutation appears to increase H3K36me3 levels, H3.3G34R mutation did not cause a change (Figure 3.5).

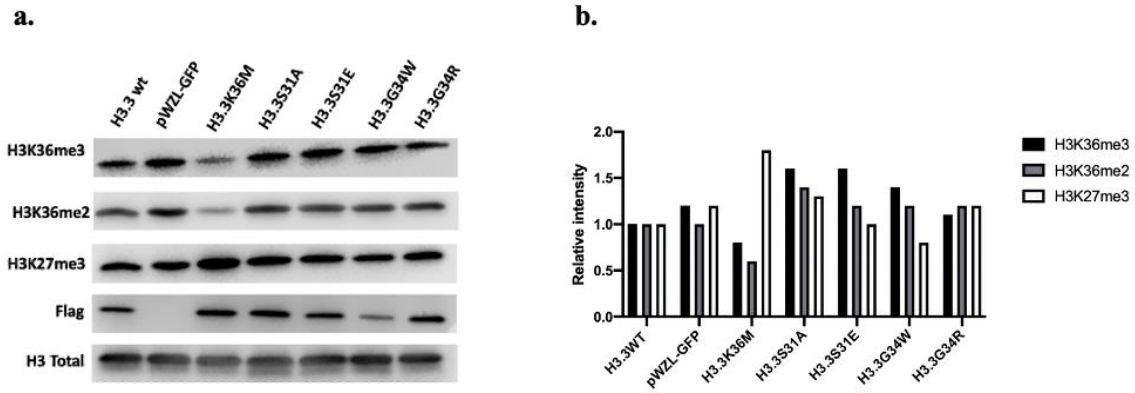


Figure 3.5: The effects of different H3.3 mutations regulating SETD2 activity on the levels of H3K36me2/3 and H3K27me3. a. Western blot analysis of H3K36me2/3 and H3K27me3 marks in H3.3wt control and different histone H3.3 mutants-expressing cells. The expression of histone H3.3 mutants were confirmed by immunoblotting of Flag. Total histone H3 was used as loading control. **b.** Quantitative image analysis of western blot showing fold change in the levels of H3K36me2/3 and H3K27me3 over H3.3wt control group. The intensity of each modification was normalized to its corresponding total histone H3 level.

3.3 Testing candidate SETD2-downstream transcription factors in reprogramming

In the preliminary study, RNA sequencing on SETD2 knockdown dH1f cells before and after OSKM-induction was performed to understand how H3K36me3 loss affects gene expression networks (Ozcimen, 2018). It was observed that SETD2 inhibition resulted in global changes in gene expressions during reprogramming (Ozcimen, 2018). Among 274 genes commonly downregulated by two independent SETD2 shRNAs, four transcription factors (*CITED2, EBF3, SMAD3, FOSL1*) were selected for further analysis (Figure 3.6.a.). The decrease in their TPM (transcript per million) expression levels is shown in Figure 3.6.b. These four transcription factors are putative SETD2-downstream genes. *SMAD3* and *FOSL1* are already known barriers against human somatic cell reprogramming (Toh et al., 2016; Xing et al., 2020). In addition, Zhang *et al*. showed direct interaction of SETD2 with *SMAD3* through its WW domain in hepatic cells (Zhang et al., 2020). Ectopic expression of *CITED2* was shown to enhance reprogramming of MEFs into iPSC (Charneca et al., 2017).

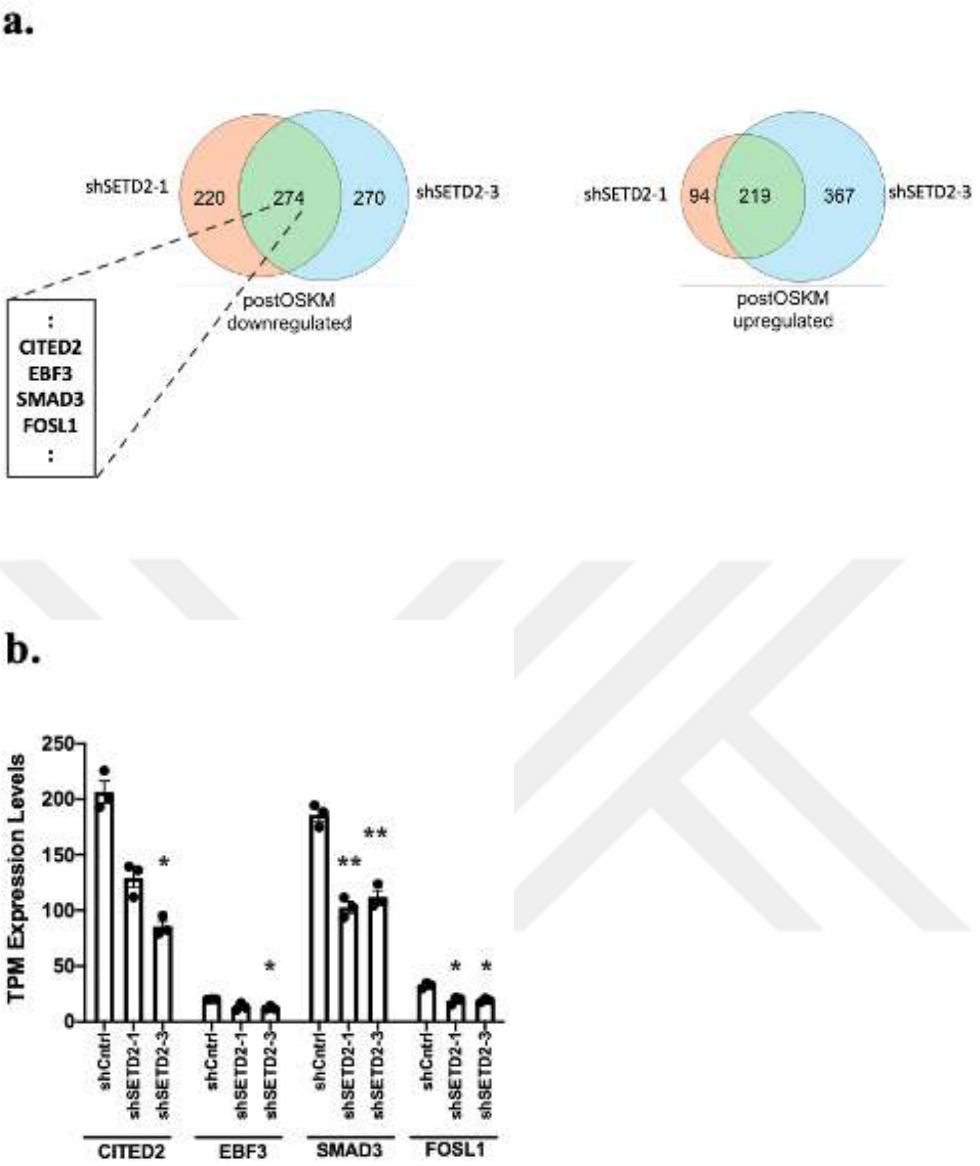


Figure 3.6: SETD2 knockdown has global effect on gene expression during reprogramming shown by preliminary RNA-seq. a. Venn diagrams representing gene numbers down- and up-regulated by SETD2 knockdown in post-OSKM samples. The numbers at intersections show the number of genes that were commonly down- or up-regulated by two independent SETD2 shRNAs (shSETD2-1 and shSETD2-3) compared to shControl-expressing cells. They were determined with p-value ≤0.001 and log2 fold change ≥0.5 (Ozcimen, 2018). Four genes shown in a panel separated with dashed lines from main image are the chosen transcription factors from the list of commonly downregulated genes upon SETD2 knockdown in post-OSKM samples. **b.** Average expression levels of *CITED2, EBF3, SMAD3* and *FOSL1* in transcripts per million (TPM)

across SETD2 knockdown and control fibroblasts. Bar graphs indicate means and error bars represent standard error of means. n=3, biological replicates. P-values were calculated by two-tailed t-test. * denotes p<0.05, ** denotes p<0.005. FeatureCounts data of RNA-seq was converted to TPM by Arda Odabaş.

I first knocked-down SETD2 in dH1f cells via two independent shRNAs along with shControl expression (Figure 3.7). Then, the four transcription factors were individually overexpressed in SETD2 knockdown and shControl- expressing cells, and these cells were subjected to OSKM-mediated reprogramming (Figure 3.7).

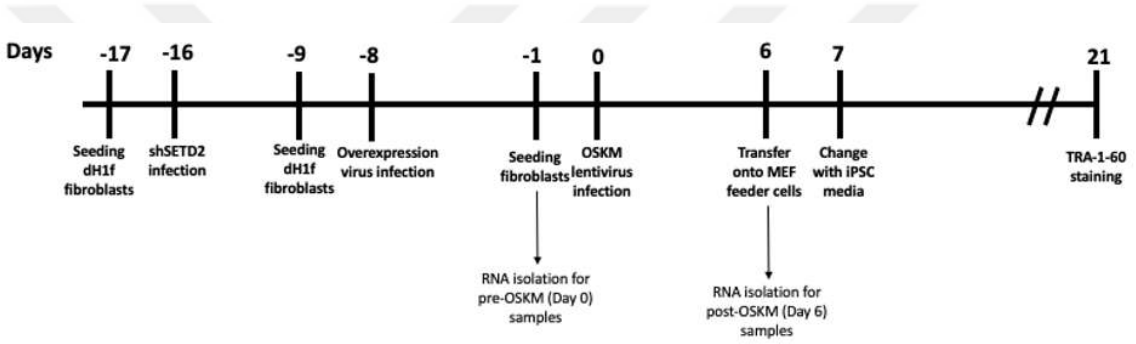


Figure 3.7: Timeline for the experiments of testing *CITED2, EBF3, SMAD3* and *FOSL1* the putative SETD2-downstream transcription factors in reprogramming.

Overexpression of *CITED2, EBF3, SMAD3* and *FOSL1* was induced using retroviral plasmids HsCD00329572, HsCD00443139, HsCD00330086 and HsCD00329450 respectively. This experiment was carried out as 4 biological replicates, each composed of 3 technical replicates and reprogramming data was quantified by TRA-1-60 staining. I hypothesized that if SETD2 suppresses iPSC generation via regulating anyone of these 4 transcription factors, then the overexpression of this transcription factor should reverse the increasing effect of SETD2 inhibition on reprogramming. According to the reprogramming data I obtained, dH1f cells infected by shSETD2-1 showed a significant increase in the number of iPSC colonies compared to shControl cells (Figure 3.8). In contrast, this increase was not detected by shSETD2-3 infected cells (Figure 3.8). In SETD2 knockdown cells generated via shSETD2-1, only *EBF3* overexpression appeared to cause a significant decrease in reprogramming efficiency (Figure 3.8). However, it could not completely reverse the increase induced by SETD2 knockdown.

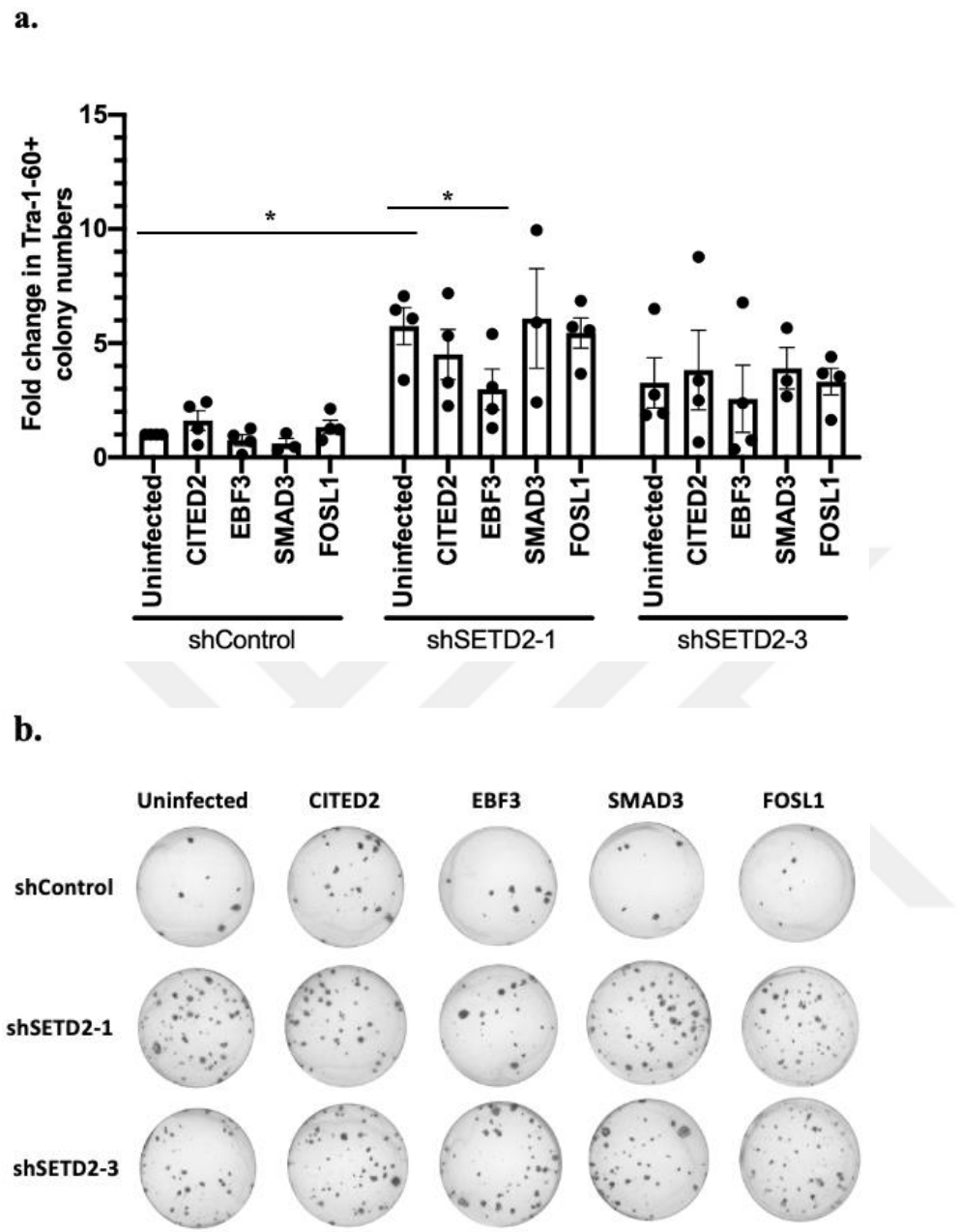


Figure 3.8: The effects of candidate SETD2-downstream transcription factors in combination with SETD2 knockdown on OSKM-reprogramming. a. Fold change in the number of TRA-1-60-positive colonies generated by OSKM induction over uninfected shControl cells. Bar graphs indicate means and error bars represent standard error of means. n=4, each biological replicates contained 3 technical replicates. P-values were calculated by two-tailed t-test. * denotes $p<0.05$. **b.** Representative well images of each condition.

I assessed knockdown of SETD2 and overexpression of *CITED2, EBF3, SMAD3* and *FOSL1* in pre-OSKM samples collected before OSKM-transduction and post-OSKM samples obtained after OSKM-transduction by RT-qPCR analysis (Figure 3.7). I detected almost 50% decrease in relative mRNA levels of SETD2 with both shSETD2-1 and shSETD2-3 mediated knockdown (Figure 3.9). Both *SMAD3* and *FOSL1* overexpression led to an increase of 5 and 10-fold while *EBF3* was overexpressed at least 60-fold (Figure 3.10). However, I could not detect overexpression of *CITED2* in any samples (Figure 3.10.a). Inspection of Ct (Threshold cycle) values revealed that *CITED2* is a highly expressed gene (Average Ct=18) in fibroblast, which may explain my inability to overexpress it. Taken together, these overexpression studies demonstrate that none of these 4 transcription factors alone are sufficient to reverse the increased reprogramming phenotype of SETD2 inhibition.

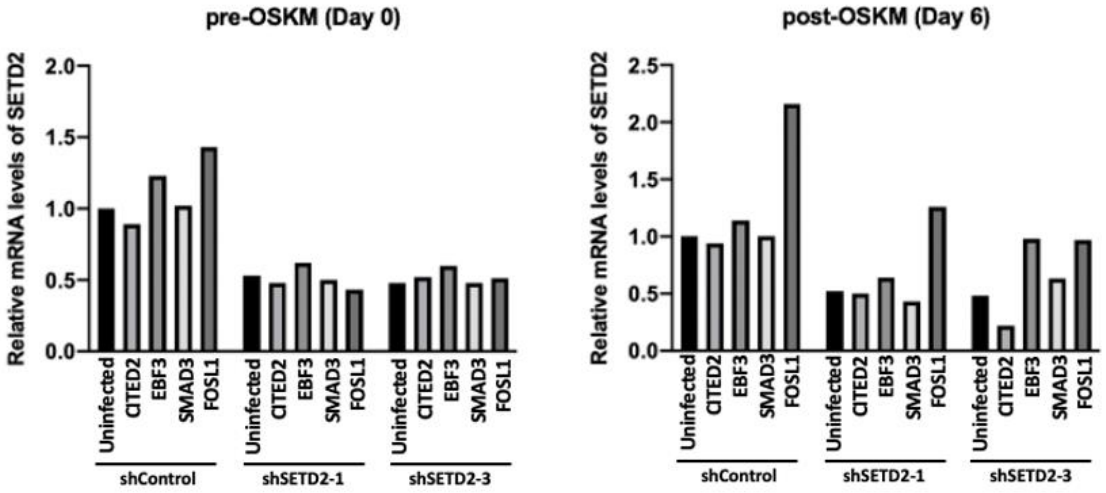


Figure 3.9: Expression levels of SETD2 in SETD2 knockdown samples after *CITED2, EBF3, SMAD3* and *FOSL1* overexpression. Relative mRNA levels of SETD2 were detected by RT-qPCR normalizing to *Beta-Actin* mRNA levels at pre-OSKM (left) and post-OSKM (right) samples. The results were shown as fold change over shControl cells without overexpressing any transcription factors. RT-qPCR was performed as 2 technical replicates and bar graphs indicate means.

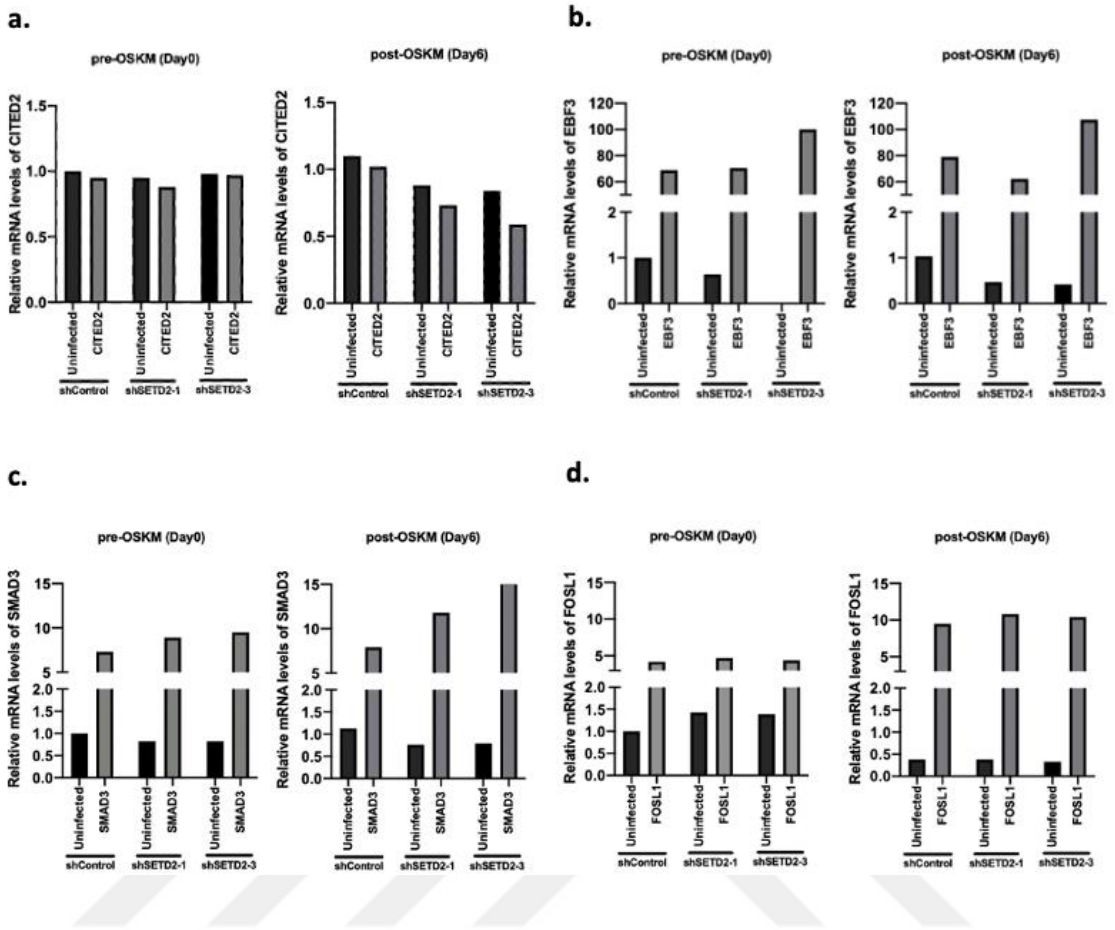
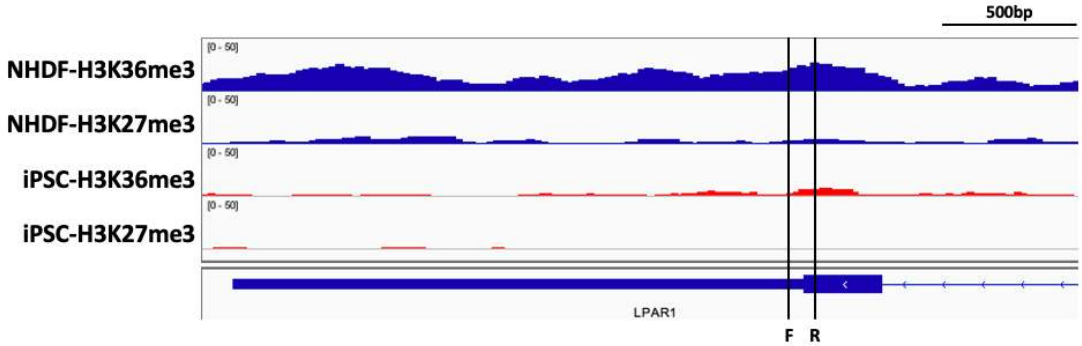


Figure 3.10: The changes in expression levels of candidate transcription factors upon SETD2 knockdown and their overexpression. Relative mRNA levels of *CITED2* (**a.**), *EBF3* (**b.**), *SMAD3* (**c.**) and *FOSL1* (**d.**) were detected by RT-qPCR normalizing to *Beta-Actin* mRNA levels at pre-OSKM and post-OSKM samples. The results were shown as fold change over uninfected shControl-expressed pre-OSKM samples. RT-qPCR was performed as 2 technical replicates and bar graphs indicate their means.

3.4 The changes in the distribution of H3K36me3 and H3K27me3 marks upon SETD2 knockdown

There is a reciprocal relationship between active H3K36me3 and the repressive H3K27me3 marks shown as H3K36me3 prevented deposition of the H3K27me3 by the polycomb repressive complex 2 (PRC2) (Schmitges et al., 2011; Yuan et al., 2011). In addition, these two marks seem to have opposite effects on somatic cell reprogramming. Onder *et al.* showed that inhibition of all major components of the PRC2 complex (Ezh2, Eed and Suz12) catalyzing H3K27me3 decreased reprogramming efficiency (Onder et al., 2012). On the other hand, H3K36me3 loss upon SETD2 depletion enhances reprogramming efficiency. Therefore, I hypothesized that H3K36me3 may act as an anti-silencing mark on fibroblast-specific genes that would otherwise be targeted by PRC2 prior to reprogramming. To test this hypothesis, I performed chromatin immunoprecipitation (ChIP) for H3K36me3 and H3K27me3 marks in shControl and shSETD2-expressed dH1f cells. The enrichment of these marks on upregulated *NANOG* and several downregulated genes upon SETD2 knockdown were analyzed using locus-specific primer sets listed in Table 2.23 (Ozcimen, 2018). In this way, I would like to check whether H3K27me3 is accumulated on the genes which are downregulated by SETD2 knockdown. H3K36me3 and H3K27me3 distributions on the analyzed regions of selected genes with primer sets were visualized with IGV viewer using the publicly available ChIP data obtained from normal human dermal fibroblasts (NHDF) and human iPSC cell lines in Figure 3.11.



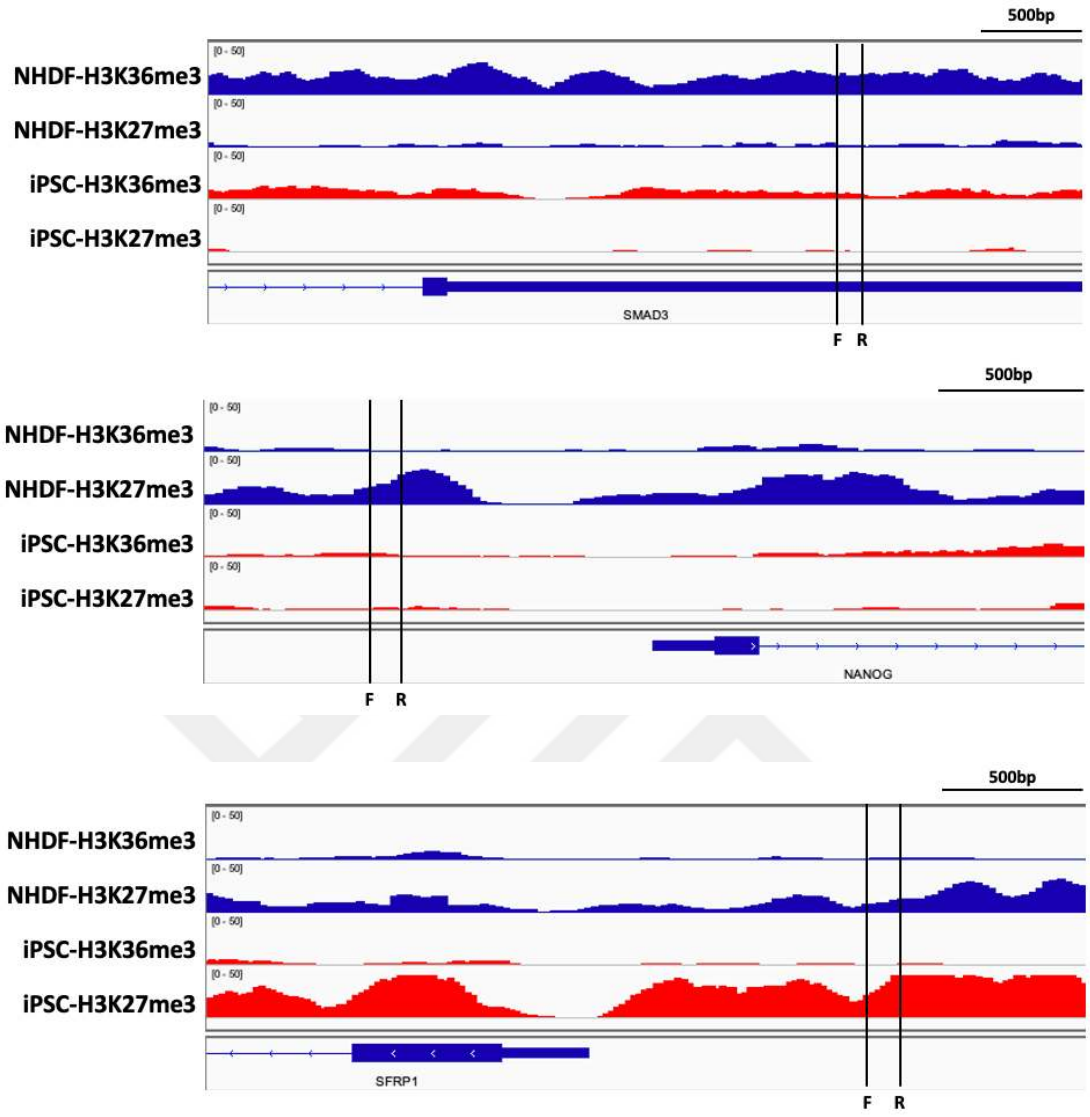


Figure 3.11: Illustration of ChIP-seq data from NHDF and iPSC cell lines for H3K36me3 and H3K27me3 marks on selected genes. Distribution of H3K36me3 and H3K27me3 on *SMAD3*, *LPAR1*, *NANOG* and *SFRP1*. NHDF-H3K36me3 (GSM733733), NHDF-H3K27me3 (GSM733745), iPSC-H3K36me3 (GSM772846) and iPSC-H3K27me3 (GSM621416) data were obtained from Encode Project. ChIP data of NHDF cell line was shown with blue tracks while ChIP data of iPSC was shown with red tracks. The amplification regions of primer sets were pointed out as two vertical black lines ("F" denotes forward primer while "R" denotes reverse primer).

First, SETD2 knockdown dH1f cells were obtained using SETD2 shRNA (shSETD2-3). SETD2 knockdown was confirmed by western blotting of H3K36me3 and significant decrease in global H3K36me3 levels was observed compared to shControl group (Figure 3.12.a). Next, SETD2 knockdown and control dH1f cells were subjected to OSKM-mediated reprogramming to validate the phenotype of SETD2 inhibition in reprogramming. An expected increase in TRA-1-60+ colony numbers was observed upon SETD2 knockdown (Figure 3.12.b).

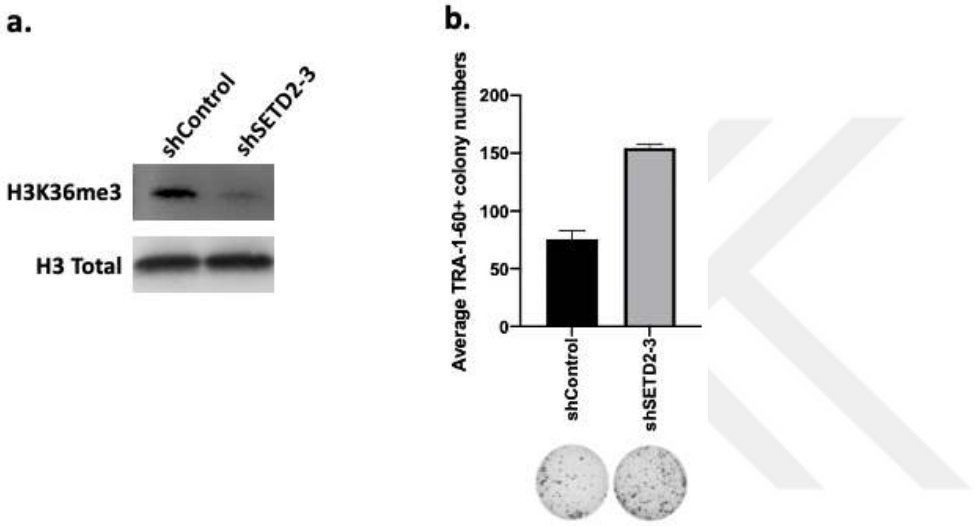


Figure 3.12: Verification of SETD2 knockdown in cells used for ChIP of H3K36me3 and H3K27me3 marks. a. Immunoblot analysis of H3K36me3 mark in shControl and shSETD2-3 expressed cells. Total histone H3 was used as nuclear loading control. **b.** Number of TRA-1-60-positive colonies generated from shControl and shSETD2-3 expressed cells by OSKM induction. Bar graphs indicate means of 3 technical replicates and error bars represent standard deviation. Representative well images of each condition are above the graph.

After validating the SETD2 knockdown and its phenotypic effect, ChIP was performed in shControl and shSETD2-3 samples with histone H3 total, H3K36me3 and H3K27me3 pull-down. Pull-down DNA fragments were obtained by decrosslinking and subjected to qPCR analysis. The enrichment of H3K36me3 and H3K27me3 at promoter regions of *NANOG* and *SFRP1* and 3'ends of *SMAD3* and *LPAR1* was analyzed. *SFRP1*,

SMAD3 and *LPAR1* were chosen from the downregulated genes upon SETD2 knockdown detected by preliminary RNA-seq (Ozcimen, 2018). *NANOG* was chosen as a representative upregulated gene upon SETD2 suppression (Ozcimen, 2018). Although % enrichment of H3K36me3 levels were low in all samples, SETD2 knockdown seemed to cause a decrease in H3K36me3 enrichment at 3'ends of *SMAD3* and *LPAR1* as expected (Figure 3.13). However, no significant change in H3K27me3 enrichment was detected at the promoter region of SFRP1 by SETD2 knockdown (Figure 3.13). On the other hand, SETD2 knockdown appeared to reduce deposition of H3K27me3 at promoter regions of *NANOG* and *SFRP1* (Figure 3.13).

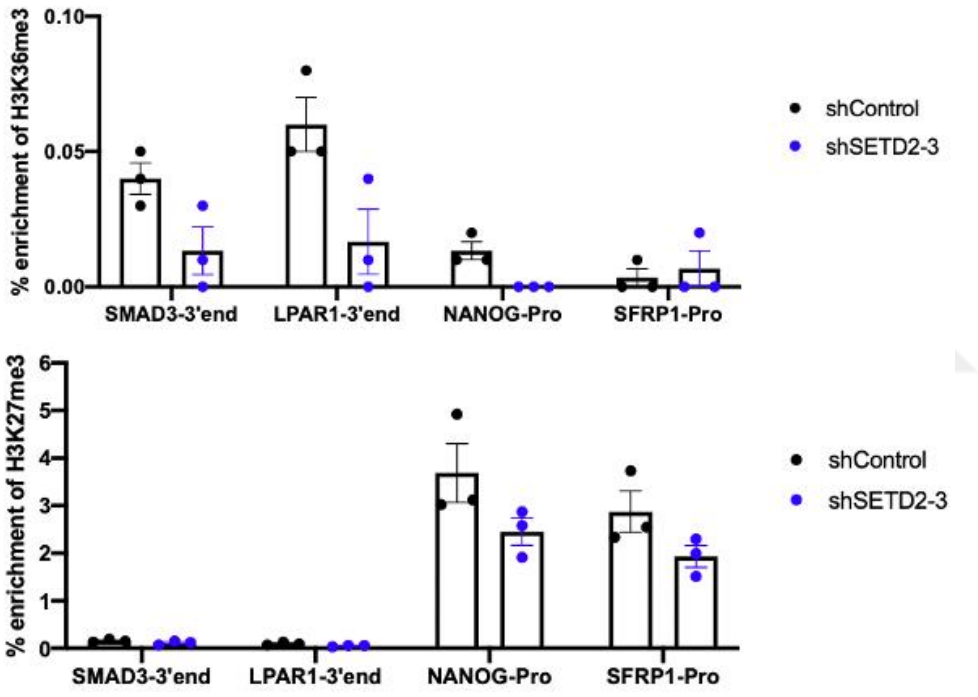


Figure 3.13: Analysis of H3K36me3 and H3K27me3 deposition in control and SETD2 knockdown cells by ChIP-qPCR. H3K36me3 (**a.**) and H3K27me3 (**b.**) % enrichment levels on 3'ends of *SMAD3* and *LPAR1* and promoter regions of *NANOG* and *SFRP1* in control and SETD2 knockdown cells were calculated relative to input and normalized with histone H3 total enrichment. n=3, each biological replicates contained two technical replicates. Bar graphs indicate means and error bars represent standard error of means.

I then designed primer sets to explore enrichment of these two marks at *EBF3* gene which was among the downregulated genes upon SETD2 knockdown and among the transcription factors that I tested as candidate downstream gene of SETD2 mentioned above (Figure 3.8). 5'end of *EBF3* gene were expected to be enriched for H3K27me3 higher than H3K36me3 while at the 3'end of *EBF3* gene H3K36me3 deposition is expected to be higher. The positions of designed primer sets and distribution of H3K36me3 and H3K27me3 on these positions were illustrated with IGV viewer in Figure 3.14.

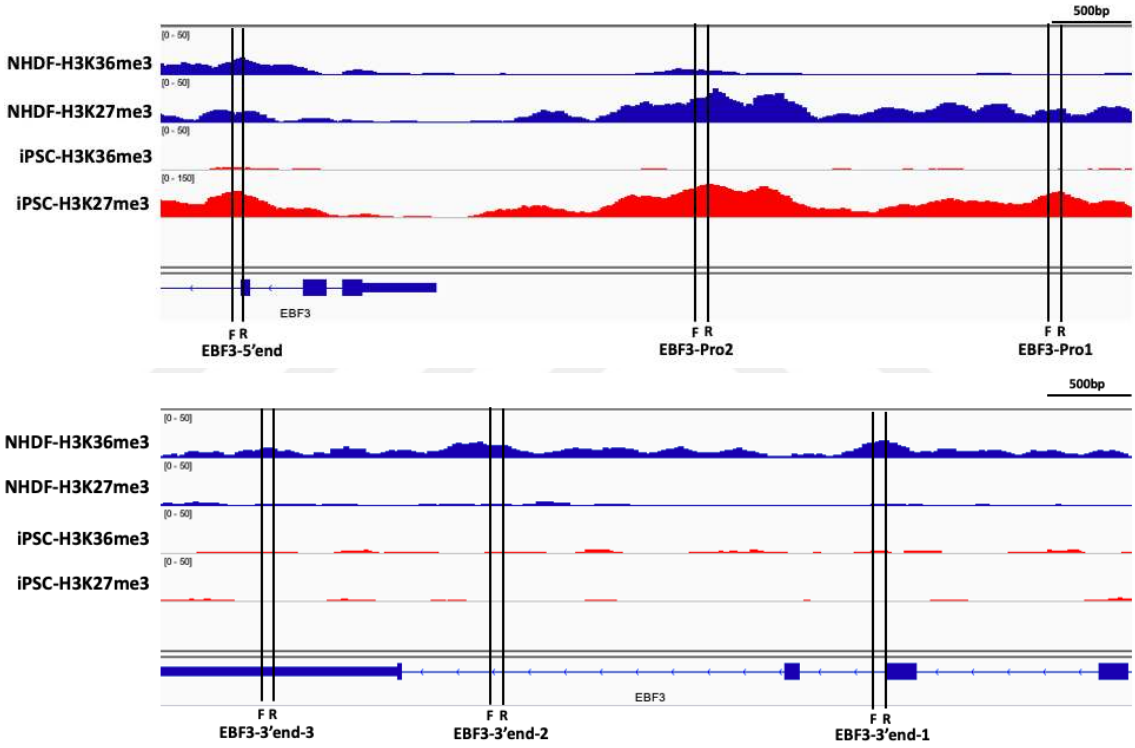


Figure 3.14: Illustration of ChIP-seq data from NHDF and iPSC cell lines for H3K36me3 and H3K27me3 marks on 3'end and 5'end of *EBF3* gene. Distribution of H3K36me3 and H3K27me3 through 5'end (*top*) and 3'end (*bottom*) of *EBF3* gene were demonstrated. ChIP data of NHDF cell line was shown with blue tracks while ChIP data of iPSC was shown with red tracks. The amplification regions of depicted primer sets were pointed out as two black lines ("F" denotes forward primer while "R" denotes reverse primer).

According to ChIP-qPCR results, the enrichment levels of H3K36me3 appeared to be low overall as in the previous analysis. However, relatively higher H3K36me3 enrichment at 3'end was observed compared to promoter and 5'end of *EBF3* gene as expected (Figure 3.15). A decrease in H3K36me3 enrichment at 3'end of EBF3 was seen upon SETD2 knockdown but an increase in H3K27me3 level at 5'end of EBF3 was not detected as hypothesized (Figure 3.27). H3K27me3 enrichment was higher in the promoter region and 5'end of *EBF3* compared to the 3'end. Although a slight decrease of H3K27me3 was detected at the promoter region (EBF3-Pro1 primer set), I cannot conclude that a major change in the level of H3K27me3 enrichment occurred upon SETD2 knockdown (Figure 3.15). Taken together, these ChIP experiments demonstrated that loss of SETD2 affects H3K27me3 enrichment across some of the genes examined, however a genome-wide approach using ChIP-Seq is warranted to fully understand the distribution of this mark.

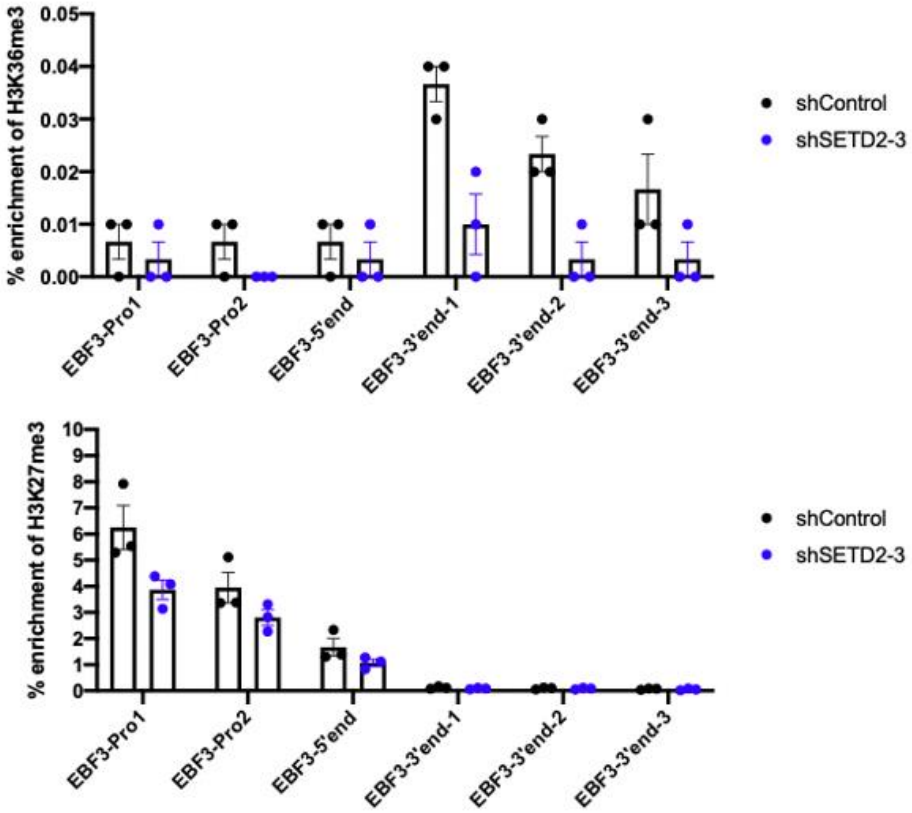


Figure 3.15: Analysis of H3K36me3 and H3K27me3 deposition on 3' and 5' ends of EBF3 gene by ChIP-qPCR. H3K36me3 (**a.**) and H3K27me3 (**b.**) enrichment levels on promoter, 5' and 3'ends of EBF3 gene in control and SETD2 knockdown cells were

calculated relative to input and normalized with total histone H3 enrichment. n=3, each biological replicates contained two technical replicates. Bar graphs indicate means and error bars represent standard error of means.

3.5 Investigating the roles of H3K36me3 readers in reprogramming

SETD2 is involved in different cellular processes by mediating H3K36 trimethylation which is recognized by numerous reader proteins in mammals (Wagner and Carpenter, 2012). Therefore, I hypothesized that SETD2 may serve as a barrier against reprogramming via the functioning of H3K36me3 readers. To test this hypothesis, I designed a small CRISPR/Cas9-based screen targeting all H3K36me3 readers found in mammals to understand their effects on cellular reprogramming. Three independent sgRNAs were designed for each reader protein to generate knock-outs. Using lentiviral vector (pLentiCRISPR_v2) expressing cloned sgRNA and Cas9, H3K36me3 readers were individually suppressed in dH1f cells (Figure 3.16). Then, these cells were subjected to OSKM-mediated reprogramming and results were quantified by TRA-1-60 staining (Figure 3.16).

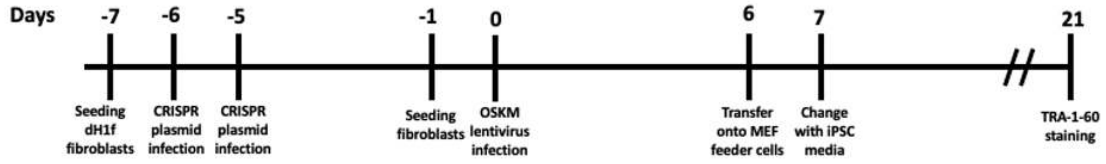


Figure 3.16: Timeline for CRISPR/Cas9-knockout screening of all mammalian H3K36me3 readers in OSKM-reprogramming. CRISPR plasmids containing sgRNAs were expressed in 50,000 dH1f cells by double lentiviral infection. After 9 days from the first infection, cells were subjected to OSKM-mediated reprogramming.

The experiment was performed as 2 technical replicates. Reprogramming data show that genetic suppression of some readers phenocopied SETD2 depletion and enhanced reprogramming efficiency of dH1f cells (Figure 3.17). I defined these readers which are *PSIP1, MRG15, MSH6, MSL3* and *NSD3* as candidate gene hits (Figure 3.17).

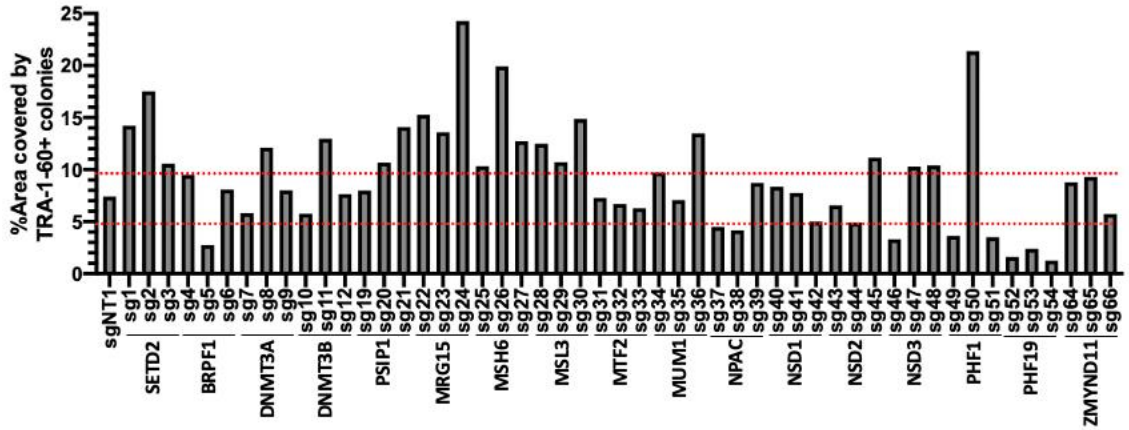


Figure 3.17: CRISPR/Cas9-knockout screening of all mammalian H3K36me3 readers in OSKM-reprogramming. % Area covered by Tra-1-60+ colonies quantified at 21 days after OSKM induction of dH1f cells previously infected with 3 independent sgRNAs targeting each H3K36me3 reader. The dotted red lines indicate 2 standard deviations from the mean % area in sgNT1-infected control wells.

To validate the reprogramming phenotype observed with depletion of the candidate genes, reprogramming experiments were repeated. For validation, the lentiviral production was performed using midi-prepped sgRNA constructs (pLentiCRISPR_v2). Midi-prep plasmid preparation gives higher purity and yield (Joung et al., 2017). dH1f cells were infected with these lentiviruses and subjected to OSKM-mediated reprogramming. This experiment was performed as 3 biological replicates, each composed of 3 technical replicates and quantified by TRA-1-60 staining. Similar increasing phenotype in screen data were obtained with validation experiments by knockout of the hits (Figure 3.18). However, a significant increase in iPSC generation was only observed by two of three sgRNAs targeting *MRG15* (sg22 and sg24) and one of *PSIP1* (sg21) compared to sgNT1-expressed control group (Figure 3.18). Hence, *MRG15* which is a H3K36me3 reader involved in alternative splicing was chosen as the best candidate gene for further analysis among these hits.

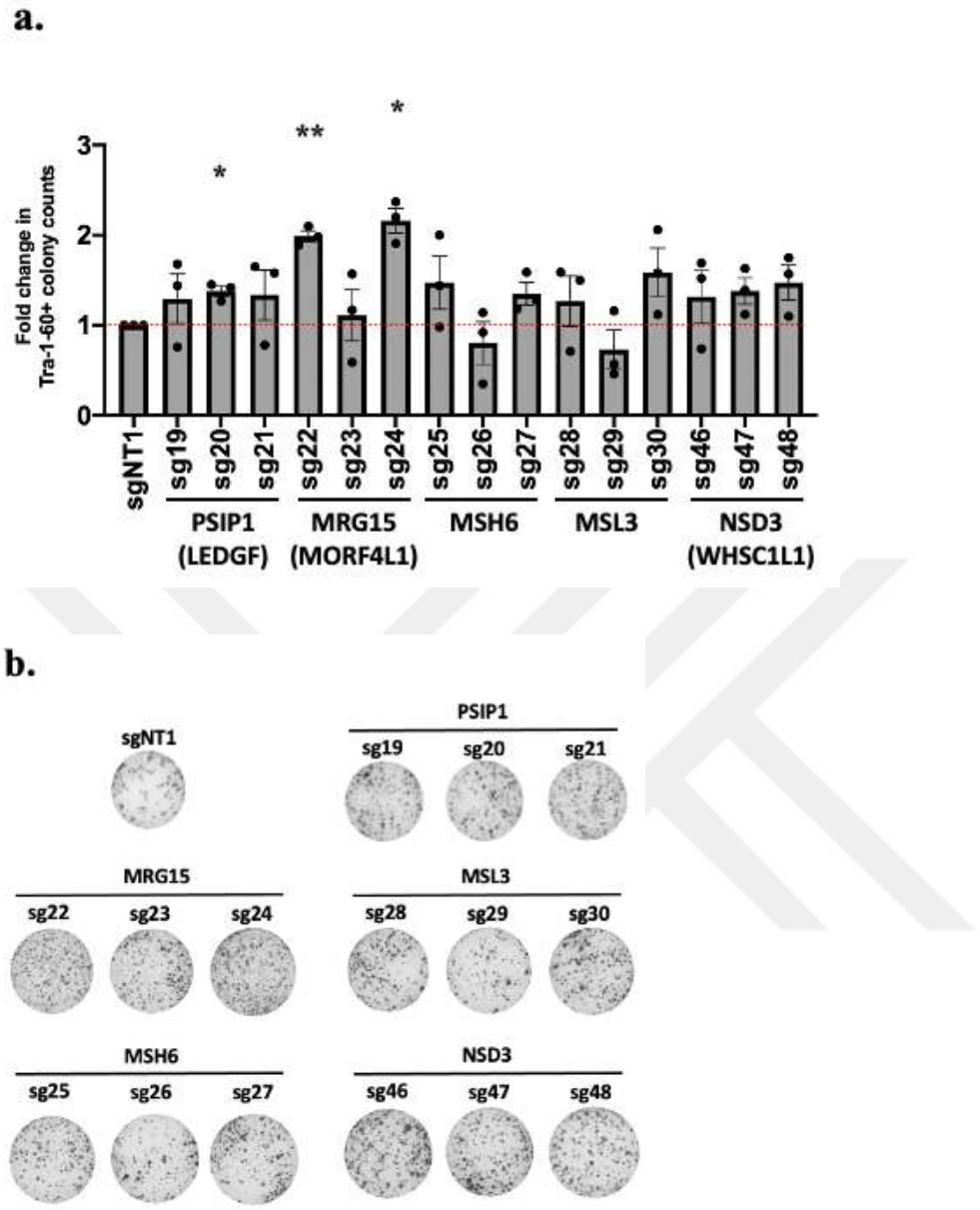


Figure 3.18: Validation of screen hits that increase reprogramming efficiency. a. Fold change in the number of TRA-1-60-positive colonies generated by OSKM induction over sgNT1-expressed control cells. Bar graphs indicate means and error bars represent standard error of means. n=3, each biological replicates contained 3 technical replicates. P-values were calculated by two-tailed t-test. * denotes $p<0.05$, ** denotes $p<0.005$. **b.** Representative well images of each sgRNA.

The relative mRNA levels of *PSIP1, MRG15, MSH6, MSL3* and *NSD3* were examined by RT-qPCR analysis to confirm knock-out of screen hits. RT-qPCR data

demonstrated that all sgRNAs targeting *PSIP1* reduced expression below 40% compared to sgNT1 control (Figure 3.19). While sg22 and sg24 targeting *MRG15* appeared to silence gene expression efficiently, sg23 had no effect (Figure 3.19). For *MSH6*, sg25 and sg26 seemed to decrease mRNA levels more than sg27 (Figure 3.19). All sgRNAs targeting *NSD2* reduced gene expression to almost half similarly (Figure 3.19). sgRNAs targeting *MSL3* and *NSD3* on the other hand seemed to not cause a substantial change in their mRNA levels (Figure 3.19).

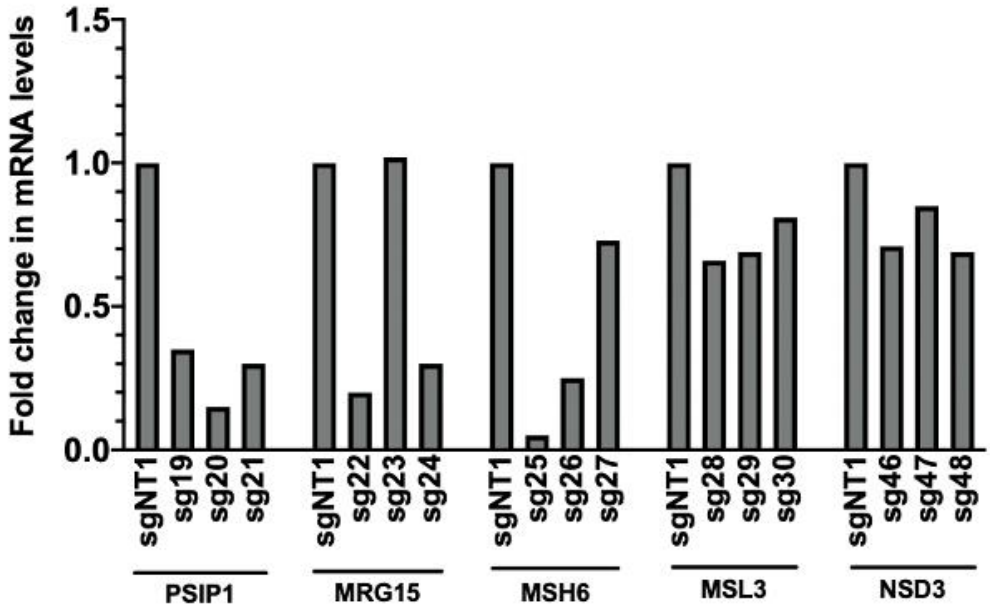


Figure 3.19: Analysis of screen hits knockout at mRNA levels by RT-qPCR. Relative mRNA levels of *PSIP1, MRG15, MSH6, MSL3* and *NSD3* were determined by normalizing to *Beta-Actin* mRNA levels. The results were shown as fold change over sgNT1-infected cells. RT-qPCR was performed as 2 technical replicates and bar graphs indicate their means.

Before analyzing the protein levels of *PSIP1, MRG15, NSD2* and *NSD3* upon their knock-out, I performed a pilot study of their immunoblots in which I checked their levels in whole cell lysate and nuclear fraction isolated from dH1f cells (Figure 3.20). The pilot

study showed that *PSIP1, NSD2* and *NSD3* proteins were better detected in nuclear fraction while *MRG15* protein could be detected in whole cell lysate (Figure 3.20).

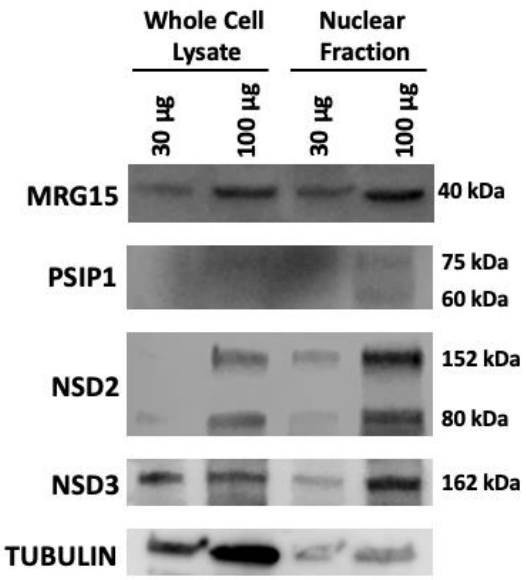


Figure 3.20: Immunoblot of *MRG15, PSIP1, NSD2* and *NSD3* in whole cell lysate and nuclear fraction. Pilot study of *MRG15, PSIP1, NSD2* and *NSD3* immunoblotting in 30 and 100 µg of both whole cell lysate and nuclear fraction isolated from dH1f cells. Tubulin was used as loading control.

I next checked their protein levels in their knockout samples by western blot. Western blot data for *MRG15* showed that sg22 and sg24 led to complete loss of *MRG15* protein, while sg23 seemed to be not functional, consistent with RT-qPCR results (Figure 3.21). Sg19 and sg20 targeting *PSIP1* appeared to efficiently deplete *PSIP1* protein (Figure 3.21). All sgRNAs targeting *NSD2* decreased the protein level efficiently (Figure 3.21). However, sgRNAs targeting *NSD3* seemed to cause some reduction in the protein level similar to a knockdown instead of knockout (Figure 3.21). Taken together, these results show that the CRISPR/Cas9 system seemed to efficiently knockout the majority of target genes as assessed by RT-qPCR and western blot analysis.

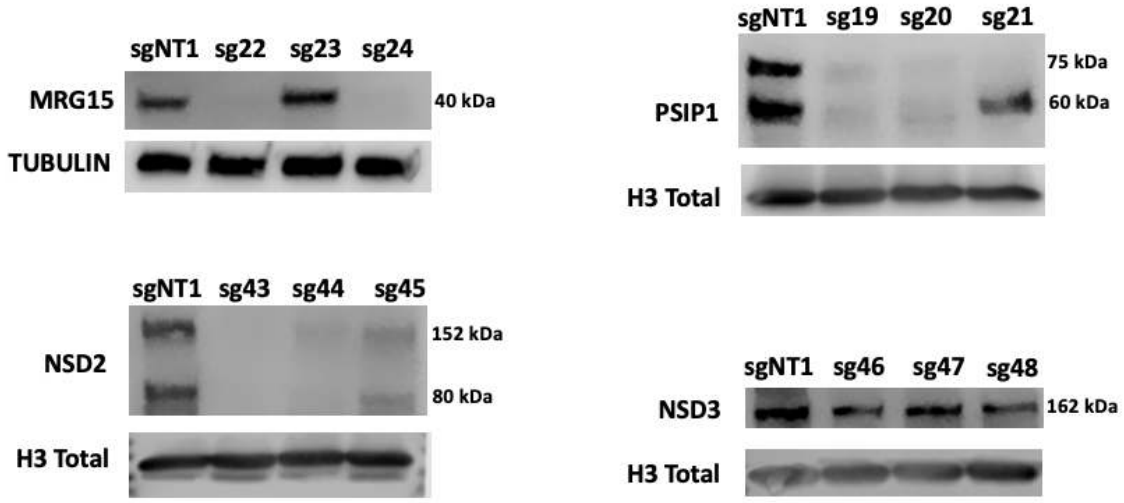


Figure 3.21: Analysis of screen hits knockout at protein levels by western blot. Immunoblot analysis of *MRG15, PSIP1, NSD2* and *NSD3* in dH1f cells where each was targeted by three independent sgRNAs. sgNT1 is control non-targeting cell lysate. Tubulin was used as cytosolic loading control while total histone H3 was used as nuclear loading control.

3.6 Analyzing alternative splicing of the genes significant for reprogramming in SETD2 knockdown cells

During reprogramming, the splicing profile of somatic cells is progressively reversed into the splicing profile of pluripotent cells like ESCs (Ohta et al., 2013). SETD2 is known to modulate alternative splicing events through its direct interaction with RNA Pol II and hnRNP-L (heterogenous nuclear ribonucleoprotein L) and its indirect effect on *MRG15, PSIP1* and *ZMYND11* which are H3K36me3 readers functioning in alternative splicing (Kizer et al., 2005; Bhattacharya et al., 2021; Xu et al., 2018; Bartels et al., 2002). Through the CRISPR/Cas9-knockout screen, I identified *MRG15* and *PSIP1* as barriers of reprogramming.

There are genes whose alternative splicing was shown to be significant for somatic cell reprogramming and their isoforms are differentially expressed in somatic and pluripotent cells (Pradella et al., 2017). Therefore, I hypothesized that SETD2 inhibition may enhance reprogramming by regulating alternative splicing of these key factors. To test my hypothesis, I analyzed the effect of SETD2 knockdown on the alternative splicing

patterns of *FOXP1, MBD2, PRDM14, ZNF207, GRHL1* and *TCF3* whose alternative splicing was reported to be important for reprogramming (Pradella et al., 2017). Firstly, SETD2 was knockdown in dH1f cells using two independent shRNAs (shSETD2-1 and shSETD2-3), and the cells were then reprogrammed by OSKM-induction (Figure 3.22). Before OSKM-transduction, RNA was isolated from both SETD2-knockdown and shControl cells as pre-OSKM samples (Figure 3.22). In addition, RNA isolation was performed at day 6 of reprogramming process as post-OSKM samples (Figure 3.22).

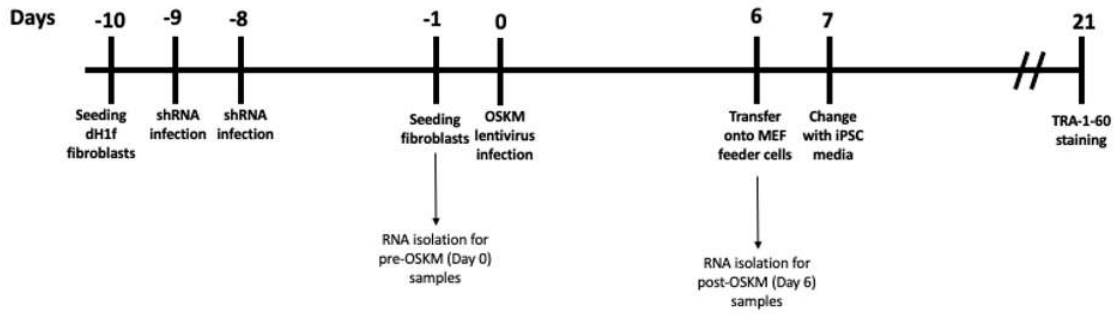


Figure 3.22: Timeline of RNA sample collections for analyzing alternative splicing of the genes in SETD2 knockdown cells. SETD2 was knockdown in dH1f cells by double infection of shRNAs. The day before OSKM-induction, RNA samples were collected from both shControl and shSETD2 expressed cells as preOSKM (Day 0) samples. RNA samples were also collected while transferring those cells onto MEF feeder cells at day 6 of reprogramming as postOSKM (Day 6) samples.

Knockdown of SETD2 was confirmed by RT-qPCR analysis (Figure 3.23.a). Also, H3K36me3 level was checked by western blotting upon SETD2 knockdown. The western blot data showed that global H3K36me3 level was decreased by shSETD2-1 and lost by shSETD2-3 (Figure 3.23.b). Furthermore, the increasing phenotype of SETD2-knockdown in reprogramming was verified by TRA-1-60 staining at day 21 of reprogramming process (Figure 3.23.c). After all these verifications, mRNA levels of different isoforms of some genes in shControl and shSETD2-3 expressed cells were determined by RT-qPCR analysis using the primers listed in Table 2.21. This experiment was carried out as 3 biological replicates per sample, each composed of 2 technical replicates.

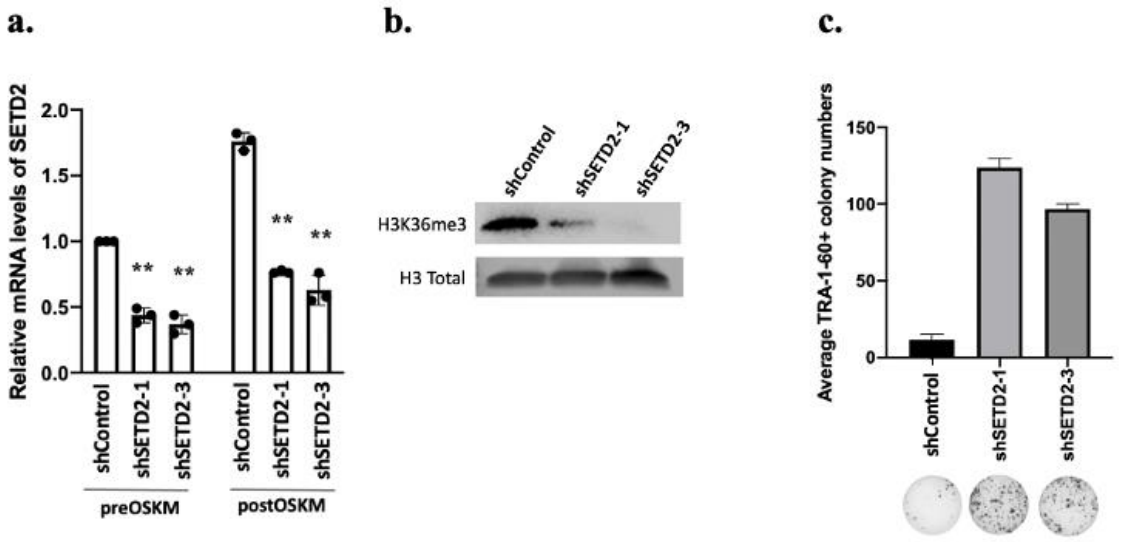


Figure 3.23: Verification of SETD2 knockdown in cells from which RNA samples collected for analyzing alternative splicing of the genes significant for reprogramming. a. RT-qPCR analysis showing mRNA levels of SETD2 in shControl and two independent shSETD2 expressed cells. n=3, each biological replicates contained two technical replicates. P-values were calculated by two-tailed t-test. ** denotes $p<0.005$. **b.** Immunoblot analysis of H3K36me3 mark in shControl and two independent shSETD2 expressed cells. Total histone H3 was used as nuclear loading control. **c.** Number of TRA-1-60-positive colonies generated from shControl and two independent shSETD2 expressed cells by OSKM induction. Bar graphs indicate means of 3 technical replicates and error bars represent standard deviation. Representative well images of each condition are above the graph.

FOXP1 is a transcription factor having a pluripotent-specific isoform, FOXP1-ES, in which Exon 18b is included instead of canonical Exon 18a (Figure 3.24) (Gabut et al., 2011). With the inclusion of Exon 18b, the encoded protein has different DNA binding properties than the somatic cell isoform, and as a result, it is involved in reactivation of *OCT4* and *NANOG*, key pluripotency genes (Gabut et al., 2011). *MBD2* is a methyl-CpG binding protein enriched at promoter regions of *OCT4* and *NANOG* and has two isoforms, MBD2a and MBD2c, expressed in somatic and pluripotent cells, respectively (Figure 3.24) (Lu et al., 2014). While MBD2a overexpression represses *OCT4* and *NANOG* in somatic cells, MBD2c overexpression was shown to enhance somatic cell reprogramming

(Lu et al., 2014). *PRDM14*, a transcriptional regulator of pluripotency, has a skipped exon isoform in differentiated somatic cells (Figure 3.24). Its full-length isoform (PRDM14-ES) was shown to promote cellular reprogramming (Lu et al., 2014). Another key factor, *ZNF207* is regulated by SRSF11 which promotes its shorter isoform (ZNF207B) and was reported to impede reprogramming process (Figure 3.24) (Toh et al., 2016). *GRHL1* also shows exon skipping in preferentially somatic cells while its full-length isoform (GRHL1-FL) increases reprogramming efficiency (Figure 3.24) (Toh et al., 2016). Similar to *FOXP1*, *TCF3,* a Wnt signaling transcription factor, has mutually exclusive exons. Chen *et al.* identified the somatic cell isoform of *TCF3* (TCF3a) which contain Exon 18a that seems to repress *NANOG* expression (Figure 3.24) (Chen et al., 2015b).

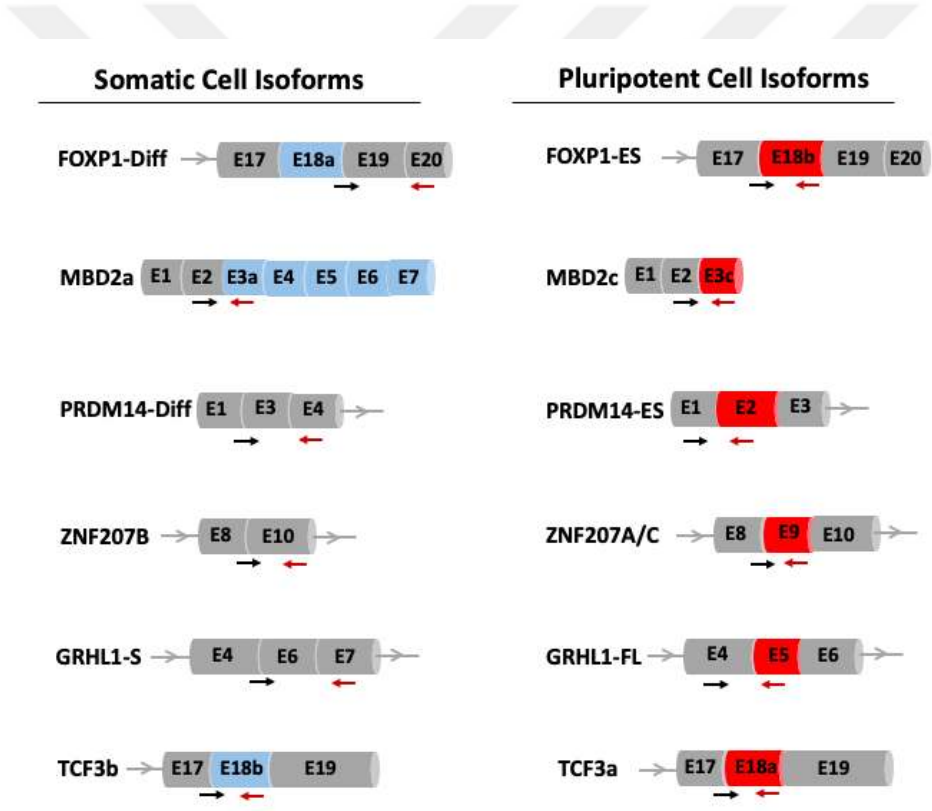
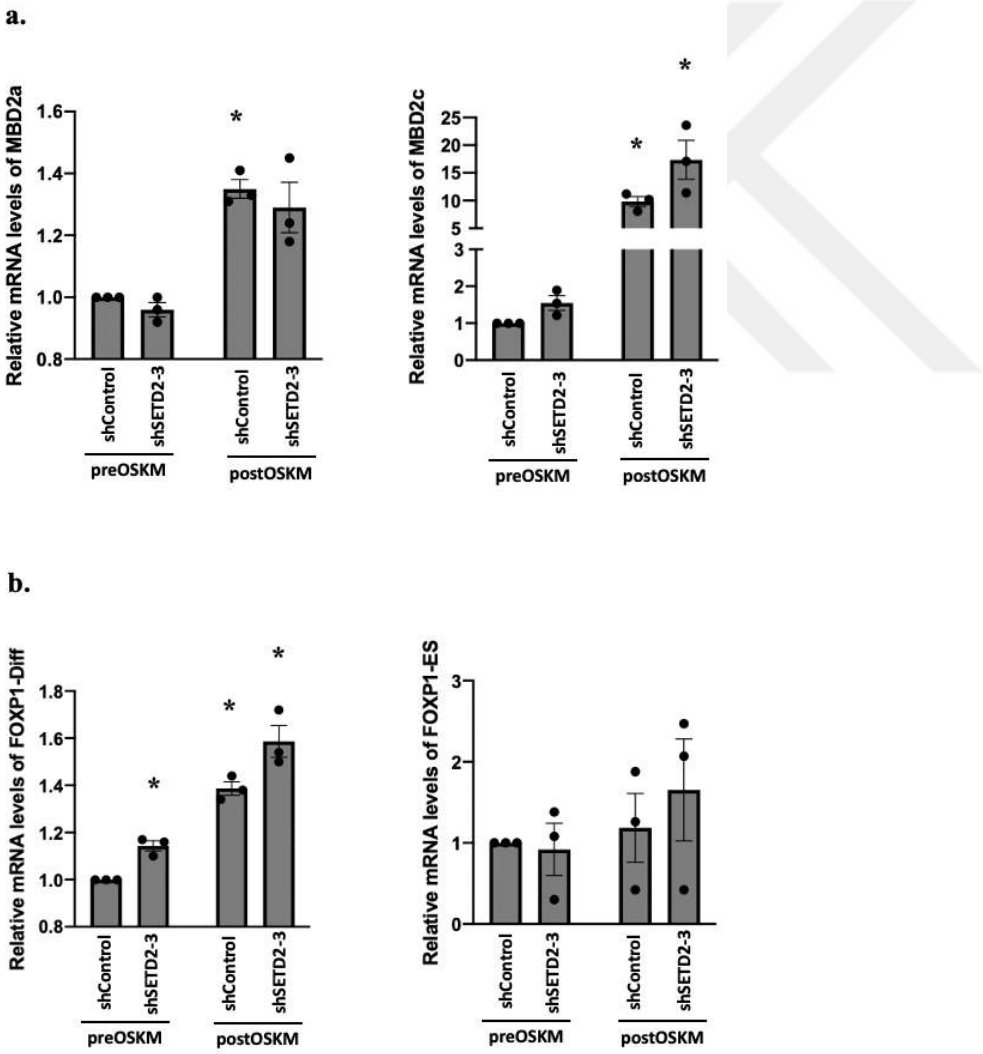


Figure 3.24: Schematic representation for alternative isoforms of the genes significant in reprogramming process. In the left part of the scheme, somatic cell isoforms of the selected genes were shown. Alternative exons found in somatic cells are indicated in blue. Pluripotent cell-specific isoforms were shown in the right part. Alternative exons found in pluripotent cells are indicated in red. The arrows shown below the exons indicate the binding site of the primers used in RT-qPCR analysis for detection of isoforms (Black arrows for forward primers; red arrows for reverse primers).

RT-qPCR analysis of splicing events for these selected genes revealed that the expression of the pluripotent cell-specific isoform of *MBD2*, MBD2c, was highly increased upon SETD2 inhibition. I detected almost 10-fold increase in relative mRNA levels of MBD2c in shControl cells after OSKM-induction (Figure 3.25.a). SETD2 knockdown in post-OSKM samples exhibited a further 7-fold increase in MBD2c expression levels (Figure 3.25.a). These results are consistent with the increasing effect of MBD2c overexpression observed on somatic cell reprogramming (Lu et al., 2014). *FOXP1*, *ZNF207* and *TCF3* did not show drastic changes in expression levels of their different isoforms upon SETD2 knockdown (Figure 3.25).



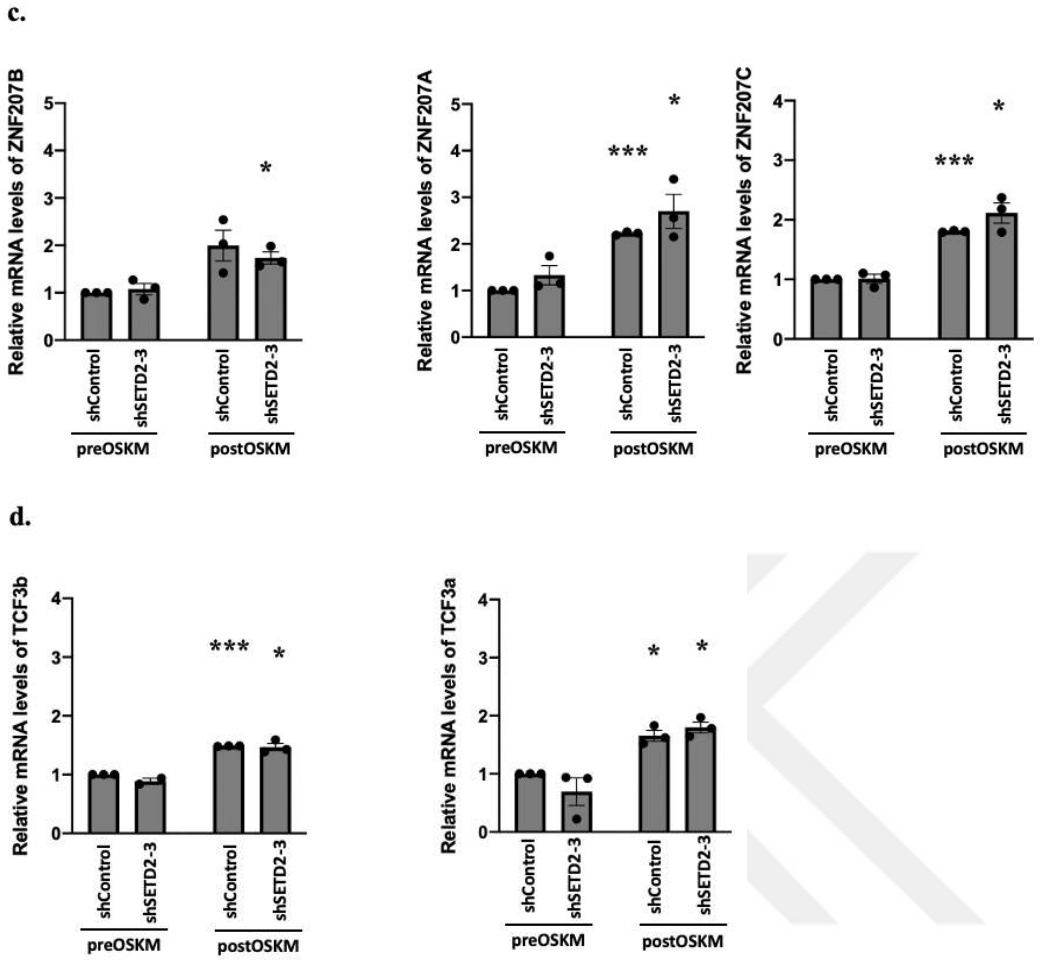


Figure 3.25: RT-qPCR analysis for alternative isoforms of the genes significant in reprogramming process in SETD2 knockdown cells. Relative mRNA levels of different *MBD2* (**a.**), *FOXP1* (**b.**), *ZNF207* (**c.**) and *TCF3* (**d.**) variants were determined by normalizing to *Beta-Actin* mRNA levels. The results were shown as fold change over shControl-expressing preOSKM samples. n=3, each biological replicates contained 2 technical replicates. Bar graphs indicate means and error bars represent standard error of means. P-values were calculated by two-tailed t-test. * denotes $p<0.05$, *** denotes $p<0.0005$.

PRDM14 and *GRHL1* isoforms were not detected by RT-qPCR which indicates that they are not expressed in dH1f cells. *PRDM14* and *GRHL1* appeared to be almost unexpressed in both control and SETD2 knockdown dH1f cells according to their TPM expression levels from the preliminary RNA-seq data (Figure 3.26) (Ozcimen, 2018).

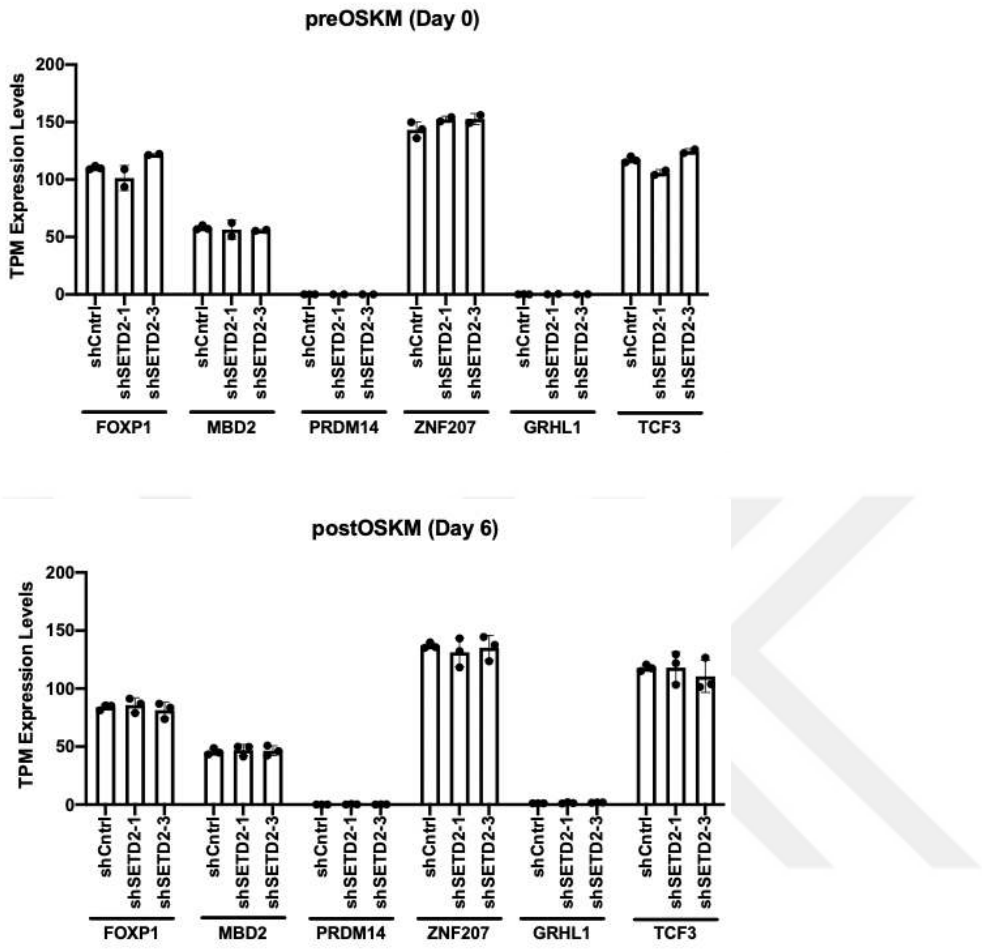


Figure 3.26: Average expression levels of *FOXP1, MBD2, PRDM14, ZNF207, GRHL1* and *TCF3* in transcripts per million (TPM) across control and SETD2 knockdown fibroblasts. *Above.* Data from preOSKM (Day 0) samples. *Below*. Data from postOSKM (Day 6) samples. Bar graphs indicate means and error bars represent standard error of means. n=3, biological replicates. FeatureCounts data of RNA-seq was converted to TPM by Arda Odabaş.

3.7 Analyzing alternative splicing of the genes significant for reprogramming in MRG15-depleted cells

Since I identified *MRG15* as the best candidate gene for the downstream regulator of SETD2 and it is known to be involved in alternative splicing, I also examined how *MRG15* depletion regulates the splicing patterns of these key genes significant for reprogramming. Likewise, *MRG15* was silenced in dH1f cells using two independent

sgRNAs (sg22 and sg24), and the cells were reprogrammed by OSKM lentiviruses. Before OSKM-transduction, RNA was isolated from both *MRG15* knockout and sgNT1 control cells as pre-OSKM samples. In addition, RNA isolation was performed at day 6 of reprogramming process as post-OSKM samples. Knockout of *MRG15* was confirmed by western blotting which demonstrated that *MRG15* level was almost completely lost by both sgRNAs (Figure 3.27.a). Furthermore, the increasing phenotype of *MRG15* knockout in reprogramming was verified by TRA-1-60 staining at day 21 of reprogramming process (Figure 3.27.b).

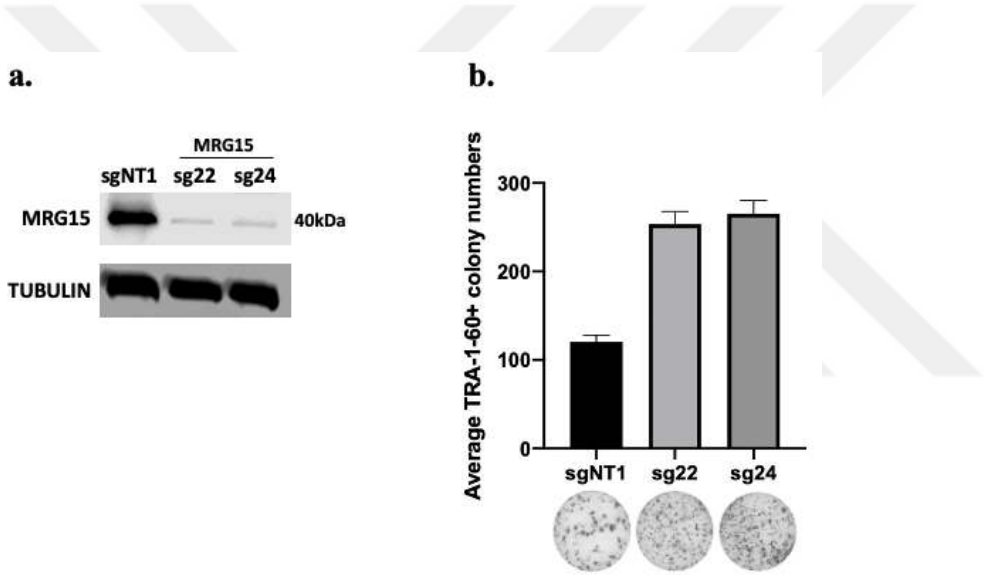
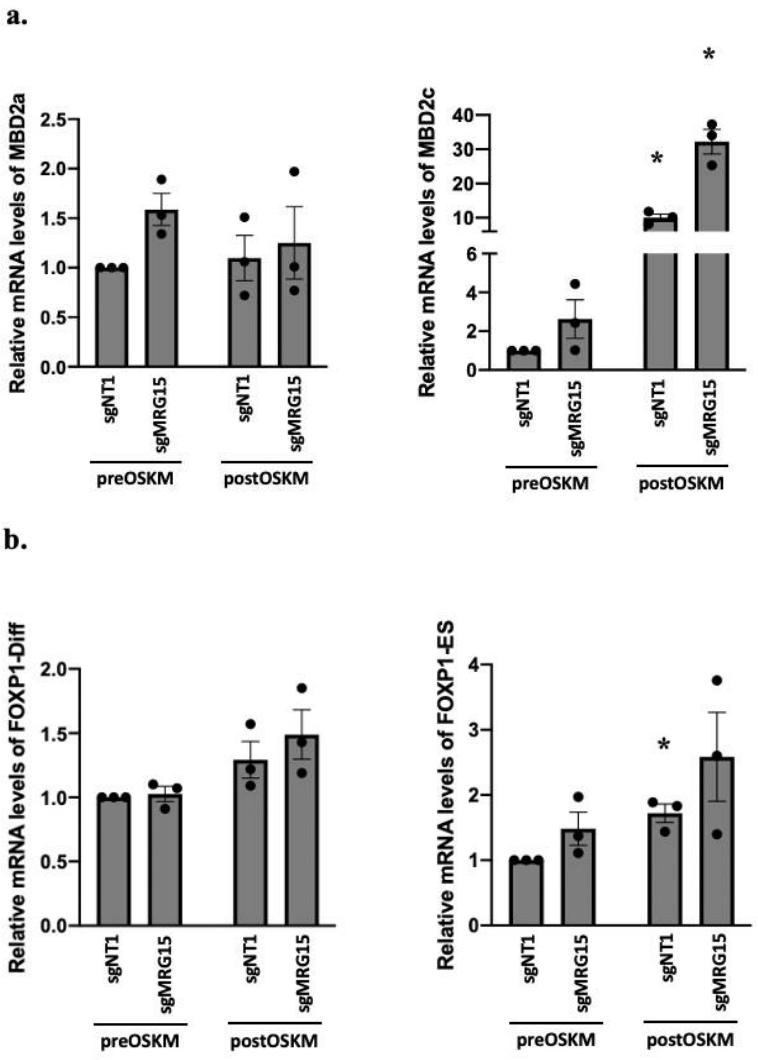


Figure 3.27: Verification of *MRG15* knockout in cells from which RNA samples collected for analyzing alternative splicing of the genes significant in reprogramming. a. Immunoblot analysis of *MRG15* in sgNT1 and two independent sgMRG15 expressed cells. Tubulin was used as loading control. **b.** Number of TRA-1-60-positive colonies generated from sgNT1 and two independent sgMRG15 expressed cells by OSKM induction. Bar graphs indicate means of 3 technical replicates and error bars represent standard deviation. Representative well images of each condition are above the graph.

mRNA levels of different isoforms of genes in sgNT1 and sg22 cells were determined by RT-qPCR analysis. This experiment was carried out as 3 biological

replicates per sample, each composed of 2 technical replicates. Similar results with SETD2-knockdown were observed for *MRG15* knockout. High Ct values were again obtained from *PRDM14* and *GRHL1* genes due to their low expression levels in fibroblasts. *FOXP1, ZNF207* and *TCF3* did not show drastic changes in the expression levels of different isoforms (Figure 3.28). However, the pluripotent-specific isoform of *MBD2* (MBD2c) showed again almost 10-fold increase in expression after OSKM-transduction (Figure 3.28.a). *MRG15* depletion led to further 20-fold higher expression with OSKM-transduction (Figure 3.28.a). Taken together, these alternative splicing analyses indicate that SETD2 inhibition could prevent functioning of *MRG15*, a H3K36me3 reader regulating alternative splicing, which resulted in higher expression of pluripotent-specific isoform of MBD2 gene during reprogramming.



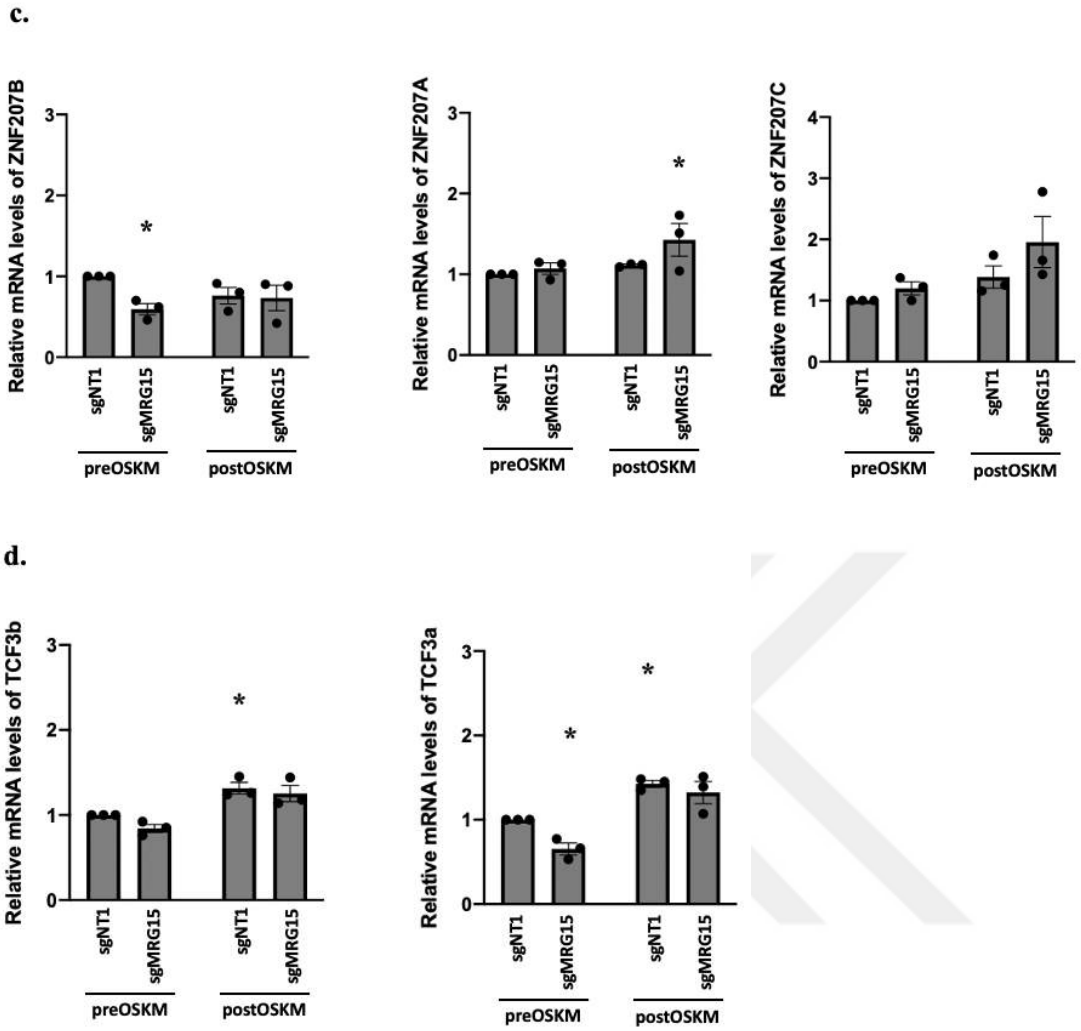


Figure 3.28: RT-qPCR analysis for alternative isoforms of the genes significant in reprogramming process in *MRG15*-depleted cells. Relative mRNA levels of different *FOXP1* (a.), *MBD2* (b.), *ZNF207* (c.) and *TCF3* (d.) variants were determined by normalizing to *Beta-Actin* mRNA levels. The results were shown as fold change over sgNT1-expressed preOSKM samples. n=3, each biological replicates contained 2 technical replicates. Bar graphs indicate means and error bars represent standard error of means. P-values were calculated by two-tailed t-test. * denotes $p<0.05$.

Chapter 4: DISCUSSION

In this thesis, I investigated how H3K36me3 loss enhances reprogramming of somatic cells into pluripotent cells. The main question of this thesis emerged from a previous study where SETD2 inhibition was found to enhance reprogramming efficiency (Ozcimen, 2018). Tri-methylation of H3K36 is the best-defined function of SETD2 and it is the only known enzyme catalyzing H3K36me3 in mammals (McDaniel and Strahl, 2017). Therefore, this study suggested that the H3K36me3 mark may be a roadblock against somatic cell reprogramming. It was a novel finding since the exact role of H3K36 methylation in reprogramming has not been elucidated yet despite several attempts (Esteban et al., 2010; Liang et al., 2012; Wei et al., 2017).

Very low efficiency of somatic cell reprogramming suggests the presence of barriers against reprogramming in somatic cells. Epigenetic modifications and their regulators are important candidates for these barriers. H3K36 tri-methylation is a stable epigenetic modification mainly associated with active transcription. It can be recognized by the number of reader proteins involved in various cellular processes such as alternative splicing, DNA repair, transcription elongation and histone modifications. The roles in transcription and diverse interactions of H3K36me3 make studying its role in reprogramming important.

4.1 The effects of histone H3.3 mutants on somatic cell reprogramming

Before exploring the downstream effects of H3K36me3 loss by SETD2 inhibition, I tested several mutants of histone H3.3 which affect SETD2 activity in reprogramming. Using a histone mutant as a tool is an effective technique to study histone modifications and their roles in cellular processes (Brumbaugh et al., 2019). Histone mutants can prevent modifications at their respective sites which eliminates the need of genetic manipulations. Additionally, expression of these mutants was seen to cause decreased levels of histone marks instead of complete loss which maybe toxic to the cells (Brumbaugh et al., 2019). These aspects of histone mutants make them advantageous over knockdown/knockout approaches while studying the role of histone modifications in reprogramming.

In the preliminary study, the increasing effect of H3.3K36M mutation on reprogramming efficiency was shown (Ozcimen, 2018). It was already known that

H3.3K36M decreases trimethylation of H3K36 (Fontebasso et al., 2013). These findings indicated that decreased H3K36me3 levels without genetic suppression of SETD2 could suffice to enhance reprogramming efficiency. I also tested H3.3K36M mutant in both OSKM- and OS-reprogramming as well as combining with DOT1L inhibition. I observed an additive effect of H3.3K36M and iDOT1L (DOT1L inhibitor) on OSKM-reprogramming, consistent with the previous finding (Ozcimen, 2018). I observed similar effect on OS-reprogramming as well. The results showed that H3.3K36M and DOT1L inhibition could both replace the expression of *KLF4* and *c-MYC*, but when combined, led to higher enhancing effect on OS-reprogramming. From these results, I can conclude that inhibition of H3K36 trimethylation and H3K79 methylation significantly enhances reprogramming efficiency, but they do not affect somatic cell reprogramming through the same pathway. Furthermore, H3.3K36M mutation and DOT1L inhibition eliminate *KLF4* and *c-MYC* in OS-reprogramming of fibroblasts into iPSC.

Apart from the mutation on 36th residue, there are also mutations on other residues of histone H3.3 that were shown to regulate SETD2 activity. Due to their effects on SETD2 activity, I also tested these mutants in OSKM-reprogramming. H3.3G34W/R mutations were reported to impede SETD2 activity and led to redistribution H3K36me3 (Nohr et al., 2017). I observed a moderate increase in reprogramming efficiency with H3.3G34 mutations and they did not cause a significant change in the levels of H3K36me2/3. Hence, H3.3K36M seems to have stronger enhancement on reprogramming than H3.3G34 mutations by directly blocking methylation of H3.3K36 residue. On the other hand, H3.3S31E/A mutations were shown to enhance catalytic activity of SETD2 on LPS-induced genes (Armache et al., 2020). My results showed that H3.3S31E/A mutations led to 1.5-fold increase in H3K36me3 levels and almost 50% decrease in reprogramming efficiency. With these findings, I demonstrated the barrier role of H3K36me3 and SETD2 in reprogramming without genetic suppression of SETD2 as well.

4.2 Testing candidate SETD2-downstream transcription factors in reprogramming

In the first main part of my thesis, I focused on the transcription factors to understand how SETD2 inhibition increases reprogramming efficiency. Transcription factors are important regulators of gene expression and somatic cell reprogramming requires changes in gene expression profile of somatic cells. Hence, SETD2 may regulate reprogramming process, via regulating expression of key transcription factors. In the previous study, differentially expressed genes upon SETD2 knockdown at before and after OSKM-induction in fibroblasts were determined by RNA-sequencing (Ozcimen, 2018). From the 274 genes commonly downregulated by both SETD2 shRNAs, four transcription factors (*CITED2, EBF3, SMAD3* and *FOSL1*) were chosen as putative SETD2-downstream genes. *CITED2* was shown to act as an enhancer of mouse reprogramming but its role in human reprogramming has not been studied yet (Charneca et al., 2017). Ectopic expression of *CITED2* in MEFs prior to OSKM-mediated reprogramming provided generation of iPSC that have less variable expression levels of pluripotency-related factors (Charneca et al., 2017). The barrier roles of *SMAD3* and *FOSL1* in human reprogramming were already reported (Toh et al., 2016; Xing et al., 2020). *SMAD3* was identified from a siRNA screen searching effectors and repressors of human reprogramming and its knockdown led to significant increase in iPSC generation (Toh et al., 2016). *FOSL1* was identified as a key regulator of intermediate phase of human reprogramming through scATAC-seq analysis and its siRNA-mediated depletion enhanced reprogramming efficiency (Xing et al., 2020).

According to the RNA-seq data, *CITED2* and *SMAD3* have higher expression levels than *EBF3* and *FOSL1* in dH1f cells (Ozcimen, 2018). SETD2 knockdown resulted in almost 50% decrease in expression levels of *CITED2* and *SMAD3* while almost 40% decrease in *EBF3* and *FOSL1* levels in post-OSKM samples. I overexpressed these transcription factors separately following SETD2 knockdown to check whether they are functional at the downstream of SETD2 during reprogramming. I detected a decrease in enhanced reprogramming efficiency upon SETD2 knockdown by only *EBF3* overexpression (Figure 3.8). However, it could not completely reverse the phenotype of SETD2 inhibition in human reprogramming. These findings indicate that *EBF3* may be a

downstream transcription factor of SETD2, but it seems that it is not the only route that SETD2 regulates reprogramming.

When I checked the expression levels of *CITED2, EBF3, SMAD3* and *FOSL1* after these treatments by RT-qPCR, *CITED2* did not seem to be overexpressed probably due to its high expression level in dH1f cells. *SMAD3* and *FOSL1* were both almost 10-fold overexpressed in post-OSKM samples. However, an expected decrease in *FOSL1* expression was not detected by SETD2 knockdown before overexpressing *FOSL1*. On the other hand, *FOSL1* overexpression seems to increase SETD2 expression after OSKM induction in both shControl and shSETD2 samples (Figure 3.9). This finding led to me think that *FOSL1* might function upstream rather than downstream of SETD2 while preventing reprogramming.

On the other hand, SETD2 knockdown appeared to sufficiently reduce the expression level of *EBF3* at both pre- and post-OSKM samples and *EBF3* overexpression led to almost 60-80-fold increase in its expression levels. However, so much expression of a transcription factor may make difficult to properly define its role in cellular processes. Hence, I speculated that *EBF3* might fail to fully demonstrate its effect on reprogramming due to its excess overexpression. After *EBF3* expression reached a certain level, it might cause a negative feedback loop that suppress its effect on reprogramming. To overcome this problem, the cDNA cassette of *EBF3* could be cloned into a vector containing tet-inducible promoter. In this way, the expression level of *EBF3* can be adjusted by the amount of doxycycline (Dox) treatment. As an alternative and easier way, the overexpression level of *EBF3* can be reduced by using less amount of virus while transduction.

4.3 Analyzing the distributions of H3K36me3 and H3K27me3 upon SETD2 knockdown

In the second part of my thesis, I analyzed the distribution of H3K36me3 and H3K27me3 marks upon SETD2 knockdown by ChIP-qPCR since they were shown to have a reciprocal relationship (Schmitges et al., 2011). Yuan *et al*. demonstrated that H3K36me2/3 prevents tri-methylation of H3K27 residue by PRC2 complex on the same histone H3 tails (Yuan et al., 2011). I hypothesized that H3K36me3 being an active histone modification and a barrier for reprogramming, may be acting as an anti-silencing

mark on fibroblast-specific genes. H3K27me3 is, on the other hand, a repressive histone mark known to inhibit expression of Hox-family genes important in development (Qin et al., 2013). In addition, PRC2 complex was shown to have the opposite role of SETD2 in reprogramming as shRNA-mediated knockdown of its components decreased reprogramming efficiency (Onder et al., 2012). Therefore, I had hypothesized that these two marks would be deposited in a mutually exclusive manner on SETD2-regulated genes during reprogramming. I checked the levels of H3K36me3 and H3K27me3 marks on *SMAD3, LPAR1* and *SFRP1* which are downregulated by SETD2 knockdown and *NANOG*, a pluripotency marker which is upregulated (Ozcimen, 2018). I expected to observe accumulation of H3K27me3 at promoter regions of repressed genes upon H3K36me3 loss.

Firstly, I visualized H3K36me3 and H3K27me3 peaks on these selected genes by IGV viewer using the publicly available ChIP-seq data of NHDF and iPSC cell lines. The primers that I used in ChIP-qPCR analysis were designed to 3'ends of *LPAR1* and *SMAD3* where H3K36me3 is expected to be abundant. On the other hand, primers were designed to target promoter regions of *NANOG* and *SFRP1* where H3K27me3 is expected to be abundant. According to ChIP-qPCR results, I obtained very low enrichment with H3K36me3-pulldown so the H3K36me3 data was not informative. The amount of chromatin may not have been sufficient for the antibody used in ChIP to pulldown H3K36me3, so higher H3K36me3 enrichment levels could be obtained by using higher amount of chromatin. H3K27me3-pulldown on the other hand appeared to work as 3-4% enrichment was observed on the promoter regions of *NANOG* and *SFRP1*. However, an expected accumulation of H3K27me3 enrichment was not detected at the promoter region of *SFRP1* upon SETD2 knockdown.

I then made a closer analysis of *EBF3* which is also among the downregulated genes by SETD2 knockdown. It was also the most likely candidate SETD2-downstream transcription factor mentioned above. Three different regions at both 3' and 5'ends of *EBF3* were analyzed for H3K36me3 and H3K27me3 enrichment by ChIP-qPCR. However, there was no accumulation of H3K27me3 at the regions through 5'end of *EBF3* upon SETD2 knockdown although there was decrease in H3K36me3 levels at 3'ends. Looking at H3K27me3 enrichment data, there does not seem to be a reciprocal relationship between H3K36me3 and H3K27me3 deposition at *SMAD3, LPAR1, SFRP1* and *EBF3* genes upon SETD2 knockdown. These findings suggest that reprogramming

enhancement upon H3K36me3 loss is not likely to occur via redistribution of repressive H3K27me3 mark. However, comprehensive genome-wide analysis for these two marks should be considered by ChIP-sequencing.

4.4 CRISPR/Cas9 knockout screening of H3K36me3 readers in reprogramming

I performed a screen of all mammalian H3K36me3 readers in human somatic cell reprogramming using CRISPR/Cas9-mediated knockout system (Table 4.1). I aimed to find out through which H3K36me3 readers H3K36me3 loss enhances OSKM-reprogramming. Therefore, I sought the readers which would mimic the phenotype of SETD2 inhibition in reprogramming when knocked-out. I considered a reader as a hit if two out of three sgRNAs targeting it resulted in increased reprogramming efficiency. These hits were *PSIP1 (LEDGF)*, *MRG15 (MORF4L1), MSH6, MSL3* and *NSD3*. The roles of any of these reader proteins in somatic cell reprogramming have not been shown to date. Hence, this is the first time their inhibitory effects on human reprogramming were demonstrated.

Table 4.1: The list of H3K36me3 readers targeted by CRISPR/Cas9 knockout screen with their domains recognizing H3K36me3 mark and main functions.

Reader	Binding Domain for H3K36me3	Main Functions
DNMT3A	PWWP	Intergenic DNA methylation (Weinberg et al., 2019)
DNMT3B	PWWP	Gene body DNA methylation (Baubec et al., 2015)
PHF1	Tudor Domain	Inhibits methyltransferase activity of PRC2 for methylation of H3K27 (Musselman et al., 2013)
PHF19	Tudor Domain	Recruits PRC2 to actively transcribed genes and associates with an H3K36me3 demethylase (NO66) to further silence of these genes (Brien et al., 2012)

MTF2	Tudor Domain	Homologues of PHF1, part of PRC2 complex so has function in H3K27 methylation (Qin et al., 2013)
BRPF1	PWWP	Recruits MOZ/MORF histone acetyltransferase (HAT) complexes and promotes histone acetylation at active chromatin regions (Vezzoli et al., 2010)
MSH6	PWWP	Being part of MutSa complex regulates DNA mismatch repair (Li et al., 2013)
MUM1	PWWP	By direct interaction with 53BP1 tumor suppresor, involves in DNA damage repair (Huen et al., 2010)
PSIP1 (LEDGF)	PWWP	Regulates alternative splicing by recruting SRSF1 (Pradeepa et al., 2012)
ZMYND11	PWWP	Regulates alternative splicing and transcriptional elongation (Bartels et al., 2002; Wen et al., 2014)
MRG15 (MORF4L1)	Chromodomain	Regulates alternative splicing by recruiting PTBP1 and controls H3K4 demethylation (Luco et al., 2010; Xu et al., 2018)
NSD1/2/3	PWWP	Mono- and di-methylation of histone H3 at K36 residue (Vermeulen et al., 2010; Wu et al., 2011)
MSL3	Chromodomain	Enhances transcription by H4K16 acetylation (Moore et al., 2010)
NPAC (GLYR1)	PWWP	Recruits LSD2 which is H3K4 demethylase (Fang et al., 2014)

Although *DNMT3A* and *DNMT3B* were known to act as barrier against reprogramming, their knockout did not fulfill the criteria to be hits in my screen (Gao et al., 2013). *PHF1*, *PH19* and *MTF2* are components of PRC2 complex catalyzing repressive H3K27 methylation marks (Brien et al., 2012). *PHF19* is known to be essential

for differentiation in mouse (Brien et al., 2012). I observed that its knockout by all three sgRNAs decreased reprogramming efficiency substantially, indicating a necessary role for *PHF19* in human reprogramming. The knockout of *PHF1* led to similar phenotype but one of its sgRNAs gave an extremely high signal which may be due to an off-target effect. I did not detect an observable effect by *MTF2* knockout in my screen despite the fact that it was reported to act as an enhancer of mouse somatic cell reprogramming. (Zhang et al., 2011). *BRPF1* has a crucial role for mouse development as it was shown to promote proliferation of MEFs, and my screen showed that its suppression did not affect human somatic cell reprogramming (You et al., 2015). On the other hand, the findings I obtained for *NPAC* were similar to its effect demonstrated in mouse reprogramming as its genetic suppression decreased human reprogramming efficiency (Yu et al., 2021). Apart from these, knockout of *MUM1, NSD1/2* and *ZMYND11* did not lead to a significant phenotype in my screen.

From my screen hits, *MSH6* which has a PWWP domain to recognize H3K36me3 is a part of DNA mismatch repair complex (MutSa) (Li et al., 2013). Binding of *MSH6* to H3K36me3-marked regions was shown to prevent microsatellite instability on those actively transcribed sites (Huang et al., 2018). *MSL3* is a part of MOF-containing MSL complex catalyzing acetylation of H4K16 which enhances transcription (Moore et al., 2010). Furthermore, MSL complex was reported to regulate self-renewal and pluripotency-related gene expression in mice (Chen et al., 2015a). *MSL3* has a chromodomain through which it can recognize both H4K20me1/2 and H3K36me3 (Moore et al., 2010). *NSD3* is a methyltransferase responsible for mono- and di-methylation of H3K36 and it contains a PWWP domain recognizing H3K36me2/3 (Vermeulen et al., 2010). Despite being both a reader and a writer of H3K36 methylation, a positive feedback regulation of H3K36 methylation by *NSD3* has not been elucidated yet. Another point about *NSD3* is its binding ability to both H3K36me2 and H3K36me3 makes it difficult to evaluate the results, because H3K36me3 loss causes an increase in H3K36me2 levels.

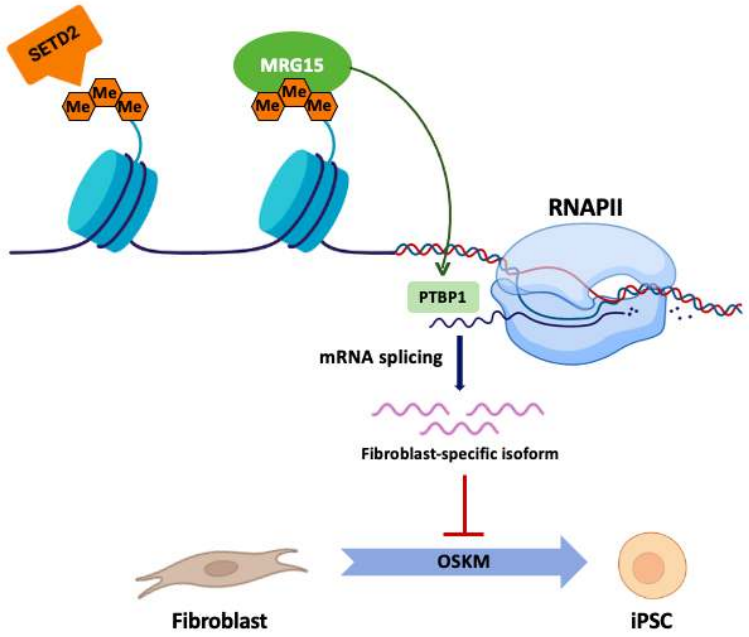


Figure 4.1: Model for the barrier role of SETD2-mediated H3K36me3 in reprogramming of fibroblasts into iPSC.

The remaining two hits are *PSIP1* and *MRG15* which are both known to regulate alternative splicing (Xu et al., 2018). *PSIP1* has a PWWP domain to bind H3K36me3-positive regions and this binding was shown to enhance the inclusion level of the targeted exon by recruiting *SRSF1* (Pradeepa et al., 2012; Xu et al., 2018). In contrast, *MRG15* has a chromodomain to recognize H3K36me3 and it reduces the inclusion level of the targeted exon by recruiting *PTBP1* (Xu et al., 2018). Since these two of my screen hits mainly regulate alternative splicing, I speculated that H3K36me3 shows its effect on somatic cell reprogramming through favoring alternative splicing of somatic cell-specific isoform (Figure 4.1). Additionally, I had obtained the strongest signal with *MRG15* knockout from both the screen and its validation experiments. Therefore, I considered *MRG15* as the best candidate gene affected from H3K36me3 loss during reprogramming and alternative splicing as the most likely downstream mechanism of H3K36me3 loss.

To characterize *MRG15* as the downstream regulator of H3K36me3 loss in reprogramming, some further tests can be performed. Firstly, the effect of SETD2 inhibition and *MRG15* knockout together on somatic cell reprogramming should be tested. If *MRG15* acts as barrier against reprogramming depending on only H3K36me3,

inhibition of both SETD2 and *MRG15* should not show a synergistic effect on the enhancement of reprogramming efficiency. In addition, if *MRG15* is a barrier against somatic cell reprogramming, its overexpression should lead to an observable decrease in reprogramming efficiency. To further associate the barrier role of *MRG15* in reprogramming with its binding property to H3K36me3, the chromodomain of *MRG15* recognizing H3K36me3 could be mutated, and mutant form could be tested in reprogramming. *MRG15* acts as an adapter between H3K36me3 and *PTBP1* by recruiting *PTBP1* to H3K36me3-marked regions and *PTBP1* regulates the inclusion level of targeted exon, so testing *PTBP1* knockout in reprogramming could be useful to associate the effect of *MRG15* on reprogramming with its role in alternative splicing. Therefore, I would expect to obtain similar results with *MRG15* knockout by *PTBP1* knockout in reprogramming of dH1f cells. Beyond these tests, the genome-wide enrichment of *MRG15* in control and SETD2 knockdown cells could be compared by ChIP-sequencing to address on which genes *MRG15* localization changes upon SETD2 knockdown.

MRG15 is a 323 amino acid-long transcription factor composed of two domains (Zhang et al., 2006b). Its chromodomain recognizing H3K36me3 is found at N-terminus while there is a conserved MRG domain at C-terminus and they are linked by 65 amino acid-long flexible region (Bertram and Pereira-Smith, 2001). Through its MRG domain, *MRG15* interacts with various proteins such as *RBP2, PAM14* and *KDM5B*. *PAM14* is a nucleoprotein involved in transcriptional regulation by interacting with *MRG15* (Zhang et al., 2006b). *RBP2* and *KDM5B* are responsible for demethylation of H3K4. *RBP2* was shown to reduce H3K4 methylation at transcribed genes to maintain transcriptional elongation via its association with *MRG15* (Hayakawa et al., 2007). On the other hand, Xie *et al.* showed that *KDM5B* was recruited to H3K36me3-positive gene bodies by interacting with *MRG15* and catalyzed demethylation of H3K4 (Xie et al., 2011). In that study, *MRG15* knockout led to increase in intragenic H3K4me3 levels which prevented transcriptional elongation of *KDM5B*-targeted genes (Xie et al., 2011). It is therefore possible that H3K36me3 loss by SETD2 inhibition may increase intragenic H3K4me3 levels on fibroblast-specific genes. Hence, it may be useful to analyze genome-wide deposition of H3K4me3 in SETD2 knockdown samples by ChIP-sequencing to test this relationship.

4.5 Analyzing alternative splicing of key genes showing differential splicing during reprogramming

For function-related assays, I performed analysis of differential splicing of key factors that were known to regulate somatic cell reprogramming. Pradella *et al.* specified some of these factors, which are *FOXP1, MBD2, PRDM14, GRHL1* and *ZNF207* expressing different isoforms in differentiated and pluripotent cells (Pradella et al., 2017). I also included *TCF3,* a Wnt-signaling transcription factor, whose somatic cell isoform was shown to repress NANOG expression (Chen et al., 2015b). Alternative splicing patterns of *FOXP1, MBD2*, *ZNF207* and *TCF3* upon SETD2 knockdown were analyzed by RT-qPCR in which primers were designed to detect different isoforms expressed in somatic and pluripotent cells. These analyses were also performed in *MRG15* knockout samples. Despite very low expression levels of *PRDM14* and *GRHL1* in dH1f cells, I tried to quantify the expression levels of their pluripotent isoforms, but I did not detect observable changes.

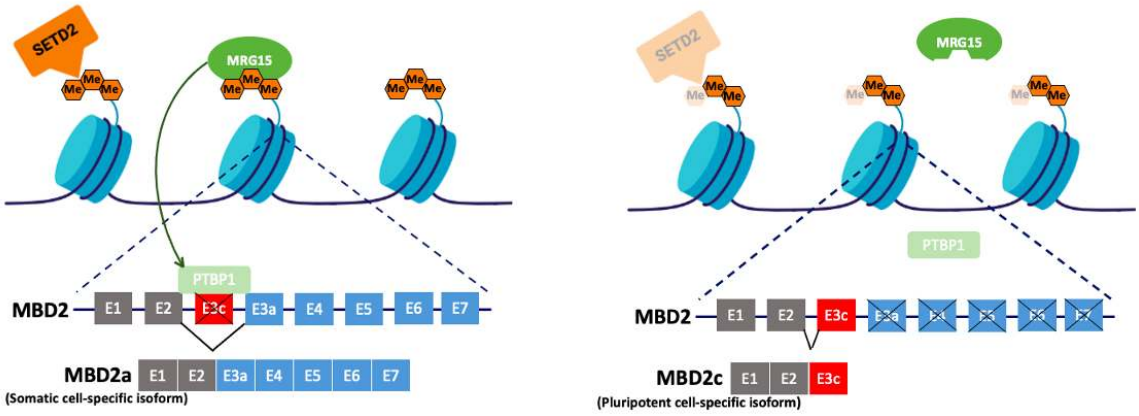


Figure 4.2: Model for the effect of H3K36me3 loss by SETD2 suppression on alternative splicing of MBD2 gene.

Among tested genes, the analysis of *MBD2* isoforms gave meaningful results as its pluripotent-specific isoform, MBD2c, showed higher expression upon SETD2 knockdown and *MRG15* knockout. *MBD2* is a methyl-CpG binding protein which was shown to be functional at promoter regions of pluripotency factors such as *OCT4* and *NANOG* (Lu et al., 2014). Its shorter isoform MBD2c, expressed in pluripotent cells, contains three exons while somatic cell-specific isoform, MBD2a, has seven exons (Pradella et al., 2017). In its shorter form, *MBD2* protein lack C-terminal TRD domain which provides interaction with NuRD complex to repress pluripotency factors (Wood and Zhou, 2016). Moreover, the effect of MBD2c overexpression in human somatic cell reprogramming was reported and it enhanced reprogramming efficiency (Lu et al., 2014). Therefore, my findings suggest that H3K36me3 loss by SETD2 inhibition prevents *MRG15* and increases the inclusion of exon 3c of *MBD2* gene favoring the expression of pluripotent cell-specific isoform, MBD2c (Figure 4.2). To support this notion, it would be nice to knockout two isoforms of *MBD2* gene by CRISPR/Cas9 and check whether MBD2c knockout phenocopies SETD2 inhibition during reprogramming. Additionally, a comprehensive analysis to detect all differential splicing events upon SETD2 knockdown should be performed by long-read RNA-sequencing.

In conclusion, three different hypotheses were tested in this thesis to understand how H3K36me3 loss enhances human somatic cell reprogramming. CRISPR/Cas9-mediated knockout screen revealed some of the H3K36me3 readers *PSIP1, MRG15, MSH6, MSL3* and *NSD3* as barriers to reprogramming. *MRG15,* a transcription factor functioning in alternative splicing, had the strongest effect on reprogramming efficiency. Hence, the presented findings suggest that H3K36me3 loss affects reprogramming efficiency by regulating alternative splicing events via its reader *MRG15*.

BIBLIOGRAPHY

Aguilo, F., & Walsh, M. J. (2017). The N6-methyladenosine RNA modification in pluripotency and reprogramming. *Current opinion in genetics & development*, *46*, 77-82.

Alison, M. R., Poulsom, R., Forbes, S., & Wright, N. A. (2002). An introduction to stem cells. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, *197*(4), 419-423.

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(5), 786.

Amabile, G., & Meissner, A. (2009). Induced pluripotent stem cells: current progress and potential for regenerative medicine. *Trends in molecular medicine*, *15*(2), 59-68.

Arabacı, D. H., Terzioğlu, G., Bayırbaşı, B., & Önder, T. T. (2021). Going up the hill: chromatin-based barriers to epigenetic reprogramming. *The FEBS Journal*, *288*(16), 4798-4811.

Armache, A., Yang, S., de Paz, A. M., Robbins, L. E., Durmaz, C., Cheong, J. Q., ... & Josefowicz, S. Z. (2020). Histone H3. 3 phosphorylation amplifies stimulation-induced transcription. *Nature*, *583*(7818), 852-857.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, *21*(3), 381-395.

Bannister, A. J., Schneider, R., & Kouzarides, T. (2002). Histone methylation: dynamic or static?. *Cell*, *109*(7), 801-806.

Bartels, C., Klatt, C., Lührmann, R., & Fabrizio, P. (2002). The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. *EMBO reports*, *3*(9), 875-880.

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

Bertram, M. J., & Pereira-Smith, O. M. (2001). Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif. *Gene*, *266*(1-2), 111-121.

Bhattacharya, S., Levy, M. J., Zhang, N., Li, H., Florens, L., Washburn, M. P., & Workman, J. L. (2021). The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors through its SHI domain. *Nature communications*, *12*(1), 1-16.

Bohnsack, J. P., & Pandey, S. C. (2021). Histone modifications, DNA methylation, and the epigenetic code of alcohol use disorder. In *International Review of Neurobiology* (Vol. 156, pp. 1-62). Academic Press.

Bongso, A., Fong, C. Y., Ng, S. C., & Ratnam, S. (1994). Fertilization and early embryology: Isolation and culture of inner cell mass cells from human blastocysts. *Human Reproduction*, *9*(11), 2110-2117.

Bradley, A., Evans, M., Kaufman, M. H., & Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, *309*(5965), 255-256.

Brien, G. L., Gambero, G., O'connell, D. J., Jerman, E., Turner, S. A., Egan, C. M., ... & Bracken, A. P. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nature structural & molecular biology*, *19*(12), 1273-1281.

Brumbaugh, J., Kim, I. S., Ji, F., Huebner, A. J., Di Stefano, B., Schwarz, B. A., Charlton, J., Coffey, A., Choi, J., Walsh, R. M., Schindler, J. W., Anselmo, A., Meissner, A., Sadreyev, R. I., Bernstein, B. E., Hock, H., & Hochedlinger, K. (2019). Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nature cell biology*, *21*(11), 1449–1461.

Carvalho, S., Raposo, A. C., Martins, F. B., Grosso, A. R., Sridhara, S. C., Rino, J., ... & de Almeida, S. F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic acids research*, *41*(5), 2881-2893.

Carvalho, S., Vítor, A. C., Sridhara, S. C., Martins, F. B., Raposo, A. C., Desterro, J. M., ... & de Almeida, S. F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *Elife*, *3*, e02482.

Charneca, J., Matias, A. C., Escapa, A. L., Fernandes, C., Alves, A., Santos, J. M., ... & Bragança, J. (2017). Ectopic expression of CITED2 prior to reprogramming, promotes

and homogenises the conversion of somatic cells into induced pluripotent stem cells. *Experimental cell research*, *358*(2), 290-300.

Chen, Q. Y., Costa, M., & Sun, H. (2015a). Structure and function of histone acetyltransferase MOF. *AIMS biophysics*, *2*(4), 555.

Chen, K., Dai, X., & Wu, J. (2015b). Alternative splicing: An important mechanism in stem cell biology. *World journal of stem cells*, *7*(1), 1.

Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure*, *16*(3), 341-350

Cossec, J. C., Theurillat, I., Chica, C., Aguín, S. B., Gaume, X., Andrieux, A., ... & Dejean, A. (2018). SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states. *Cell stem cell*, *23*(5), 742-757.

Creighton, H., & Waddington, C. H. (1958). The Strategy of the Genes. *AIBS Bulletin, 8(2), 49.*

DasGupta, A., Lee, T. L., Li, C., & Saltzman, A. L. (2020). Emerging Roles for Chromo Domain Proteins in Genome Organization and Cell Fate in C. elegans. *Frontiers in Cell and Developmental Biology*, *8*, 1027.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L. K., Beck, H., Sørensen, C. S., ... & Jäättelä, M. (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nature structural & molecular biology*, *19*(8), 803-810.

David, L., & Polo, J. M. (2014). Phases of reprogramming. *Stem cell research*, *12*(3), 754-761.

Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, *51*(6), 987-1000.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *Journal of Biological Chemistry*, *285*(34), 26114-26120.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., ... & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, *34*(2), 184-191.

Dong, C., Nakagawa, R., Oyama, K., Yamamoto, Y., Zhang, W., Dong, A., ... & Min, J. (2020). Structural basis for histone variant H3tK27me3 recognition by PHF1 and PHF19. *Elife*, *9*, e58675.

Dvash, T., Ben-Yosef, D., & Eiges, R. (2006). Human embryonic stem cells as a powerful tool for studying human embryogenesis. *Pediatric research*, *60*(2), 111-117.

Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal*, *27*(2), 406-420.

Esteban, M. A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., ... & Pei, D. (2010). Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell stem cell*, *6*(1), 71-79.

Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *nature*, *292*(5819), 154-156.

Fang, D., Gan, H., Lee, J. H., Han, J., Wang, Z., Riester, S. M., ... & Zhang, Z. (2016). The histone H3. 3K36M mutation reprograms the epigenome of chondroblastomas. *Science*, *352*(6291), 1344-1348.

Fang, R., Chen, F., Dong, Z., Hu, D., Barbera, A. J., Clark, E. A., ... & Shi, Y. G. (2013). LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Molecular cell*, *49*(3), 558-570.

Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I. B., Finkl, K., Schmid, K., ... & Müller, J. (2020). Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *Elife*, *9*, e61964.

Fontebasso, A. M., Schwartzentruber, J., Khuong-Quang, D. A., Liu, X. Y., Sturm, D., Korshunov, A., ... & Majewski, J. (2013). Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta neuropathologica*, *125*(5), 659-669.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., & Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy, Series B*, *85*(8), 348-362.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H. K., ... & Blencowe, B. J. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, *147*(1), 132-146.

Gaidano, G., & Carbone, A. (2000). MUM1: a step ahead toward the understanding of lymphoma histogenesis. *Leukemia*, *14*(4), 563-566.

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., ... & Gao, S. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell stem cell*, *12*(4), 453-469.

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., ... & Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl j Med*, *366*, 883-892.

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, *105*(1), 4-13.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C. L., Trcka, D., & Wrana, J. L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell stem cell*, *11*(6), 769-782.

Greer, E. L., & Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, *13*(5), 343-357.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature*, *389*(6649), 349-352.

Guo, R., Zheng, L., Park, J. W., Lv, R., Chen, H., Jiao, 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.

Gurdon, J. B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Journal of Embryology and Experimental Morphology*

Hayakawa T, Ohtani Y, Hayakawa N, Shinmyozu K, Saito M, Ishikawa F, Nakayama J. RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. *Genes Cells*. 2007 Jun;12(6):811-26.

Holliday, R. (2006). Epigenetics: a historical overview. *Epigenetics*, *1*(2), 76-80.

Hu, M., Sun, X. J., Zhang, Y. L., Kuang, Y., Hu, C. Q., Wu, W. L., ... & Chen, Z. (2010). Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proceedings of the National Academy of Sciences*, *107*(7), 2956-2961.

Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B. S., Sun, M., ... & Chen, J. (2019). Histone H3 trimethylation at lysine 36 guides m 6 A RNA modification co-transcriptionally. *Nature*, *567*(7748), 414-419.

Huang, Y., Gu, L., & Li, G. M. (2018). H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *Journal of Biological Chemistry*, *293*(20), 7811-7823.

Huen, M. S., Huang, J., Leung, J. W., Sy, S. M. H., Leung, K. M., Ching, Y. P., ... & Chen, J. (2010). Regulation of chromatin architecture by the PWWP domain-containing DNA damage-responsive factor EXPAND1/MUM1. *Molecular cell*, *37*(6), 854-864.

Jain, S. U., Khazaei, S., Marchione, D. M., Lundgren, S. M., Wang, X., Weinberg, D. N., ... & Lewis, P. W. (2020). Histone H3. 3 G34 mutations promote aberrant PRC2 activity and drive tumor progression. *Proceedings of the National Academy of Sciences*, *117*(44), 27354-27364.

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., ... & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature protocols*, *12*(4), 828-863.

Khodeer, S. M., Klungland, A., & Dahl, J. A. (2021). ALKBH5 regulates somatic cell reprogramming in a phase specific manner. *bioRxiv*.

Kim, D., Blus, B. J., Chandra, V., Huang, P., Rastinejad, F., & Khorasanizadeh, S. (2010). Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nature structural & molecular biology*, *17*(8), 1027-1029.

Kim, D., Kim, C. H., Moon, J. I., Chung, Y. G., Chang, M. Y., Han, B. S., ... & Kim, K. S. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell stem cell*, *4*(6), 472.

Kim, J. H., Auerbach, J. M., Rodriguez-Gomez, J. A., Velasco, I., Gavin, D., Lumelsky, N., ... & McKay, R. (2002). Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature*, *418*(6893), 50-56.

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., & Strahl, B. D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology*, *25*(8), 3305-3316.

Kobayashi, M., Ohsugi, M., Sasako, T., Awazawa, M., Umehara, T., Iwane, A., ... & Ueki, K. (2018). The RNA methyltransferase complex of WTAP, METTL3, and METTL14 regulates mitotic clonal expansion in adipogenesis. *Molecular and cellular biology*, *38*(16), e00116-18.

Koche, R. P., Smith, Z. D., Adli, M., Gu, H., Ku, M., Gnirke, A., ... & Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell stem cell*, *8*(1), 96-105.

Kolundzic, E., Ofenbauer, A., Bulut, S. I., Uyar, B., Baytek, G., Sommermeier, A., ... & Tursun, B. (2018). FACT sets a barrier for cell fate reprogramming in Caenorhabditis elegans and human cells. *Developmental cell*, *46*(5), 611-626.

Lee, Y., & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annual review of biochemistry*, *84*, 291-323.

Lewis, P. W., Müller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., ... & Allis, C. D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*, *340*(6134), 857-861.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., & Li, G. M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell*, *153*(3), 590-600.

Li, J., Duns, G., Westers, H., Sijmons, R., van den Berg, A., & Kok, K. (2016). SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget*, *7*(31), 50719.

Li, Y., Trojer, P., Xu, C. F., Cheung, P., Kuo, A., Drury, W. J., ... & Reinberg, D. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *Journal of Biological Chemistry*, *284*(49), 34283-34295.

Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., ... & Deng, H. (2011). Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell research*, *21*(1), 196-204.

Liang, G., He, J., & Zhang, Y. (2012). Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nature cell biology*, *14*(5), 457-466.

Liu, J., Gao, M., He, J., Wu, K., Lin, S., Jin, L., ... & Chen, J. (2021). The RNA m 6 A reader YTHDC1 silences retrotransposons and guards ES cell identity. *Nature*, *591*(7849), 322-326.

Lu, C., Jain, S. U., Hoelper, D., Bechet, D., Molden, R. C., Ran, L., ... & Lewis, P. W. (2016). Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science*, *352*(6287), 844-849.

Lu, P. J., Zhou, X. Z., Shen, M., & Lu, K. P. (1999). Function of WW domains as phosphoserine-or phosphothreonine-binding modules. *Science*, *283*(5406), 1325-1328.

Lu, X., Göke, J., Sachs, F., Jacques, P. É., Liang, H., Feng, B., ... & Ng, H. H. (2013). SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nature cell biology*, *15*(10), 1141-1152.

Lu, Y., Loh, Y. H., Li, H., Cesana, M., Ficarro, S. B., Parikh, J. R., ... & Marto, J. A. (2014). Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. *Cell stem cell*, *15*(1), 92-101.

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., & Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science*, *327*(5968), 996-1000.

Luger, K., & Richmond, T. J. (1998). The histone tails of the nucleosome. *Current opinion in genetics & development*, *8*(2), 140-146.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251-260.

Lutsik, P., Baude, A., Mancarella, D., Öz, S., Kühn, A., Toth, R., ... & Plass, C. (2020). Globally altered epigenetic landscape and delayed osteogenic differentiation in H3. 3-G34W-mutant giant cell tumor of bone. *Nature communications*, *11*(1), 1-16.

Ma, D. K., Chiang, C. H. J., Ponnusamy, K., Ming, G. L., & Song, H. (2008). G9a and Jhdm2a regulate embryonic stem cell fusion-induced reprogramming of adult neural stem cells. *Stem cells*, *26*(8), 2131-2141.

Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., & Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature*, *382*(6592), 646-649.

Mali, P., Chou, B. K., Yen, J., Ye, Z., Zou, J., Dowey, S., ... & Cheng, L. (2010). Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem cells*, *28*(4), 713-720.

Mansour, A. A., Gafni, O., Weinberger, L., Zviran, A., Ayyash, M., Rais, Y., ... & Hanna, J. H. (2012). The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature*, *488*(7411), 409-413.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, *78*(12), 7634-7638.

McDaniel, S. L., & Strahl, B. D. (2017). Shaping the cellular landscape with Set2/SETD2 methylation. *Cellular and Molecular Life Sciences*, *74*(18), 3317-3334.

Melamed, P., Yosefzon, Y., David, C., Tsukerman, A., & Pnueli, L. (2018). Tet enzymes, variants, and differential effects on function. *Frontiers in cell and developmental biology*, *6*, 22.

Mikkelsen, T. S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., ... & Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature*, *454*(7200), 49-55.

Miles, D. C., de Vries, N. A., Gisler, S., Lieftink, C., Akhtar, W., Gogola, E., ... & van Lohuizen, M. (2017). TRIM28 is an epigenetic barrier to induced pluripotent stem cell reprogramming. *Stem Cells*, *35*(1), 147-157.

Moore, S. A., Ferhatoglu, Y., Jia, Y., Al-Jiab, R. A., & Scott, M. J. (2010). Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono-or dimethyllysine 20 on histone H4. *Journal of Biological Chemistry*, *285*(52), 40879-40890.

Musselman, C. A., Gibson, M. D., Hartwick, E. W., North, J. A., Gatchalian, J., Poirier, M. G., & Kutateladze, T. G. (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nature communications*, *4*(1), 1-9.

Nohr, E., Lee, L. H., Cates, J. M., Perizzolo, M., & Itani, D. (2017). Diagnostic value of histone 3 mutations in osteoclast-rich bone tumors. *Human pathology*, *68*, 119-127.

Ohta, S., Nishida, E., Yamanaka, S., & Yamamoto, T. (2013). Global splicing pattern reversion during somatic cell reprogramming. *Cell reports*, *5*(2), 357-366.

Okita, K., Ichisaka, T., & Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *nature*, *448*(7151), 313-317.

Onder, T. T., Kara, N., Cherry, A., Sinha, A. U., Zhu, N., Bernt, K. M., ... & Daley, G. Q. (2012). Chromatin-modifying enzymes as modulators of reprogramming. *Nature*, *483*(7391), 598-602.

Oshita, F., Kasajima, R., & Miyagi, Y. (2016). Multiplex genomic test of mutation and fusion genes in small biopsy specimen of lung cancer. *Journal of experimental therapeutics & oncology*, *11*(3).

Ozcimen, B. (2018). The role of histone H3 lysine 36 methylation in reprogramming of fibroblasts and on induced pluripotent stem cell generation. (Publication No. 520753) [Doctoral dissertation, Koc University]. Council of Higher Education Thesis Center.

Paik, D. T., Chandy, M., & Wu, J. C. (2020). Patient and disease–specific induced pluripotent stem cells for discovery of personalized cardiovascular drugs and therapeutics. *Pharmacological reviews*, *72*(1), 320-342.

Park, I. H., Zhao, R., West, J. A., Yabuuchi, A., Huo, H., Ince, T. A., ... & Daley, G. Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. *Nature*, *451*(7175), 141-146.

Pfister, S. X., Ahrabi, S., Zalmas, L. P., Sarkar, S., Aymard, F., Bachrati, C. Z., ... & Humphrey, T. C. (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell reports*, *7*(6), 2006-2018.

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), e1002717.

Pradella, D., Naro, C., Sette, C., & Ghigna, C. (2017). EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. *Molecular cancer*, *16*(1), 1-19.

Qin, S., Guo, Y., Xu, C., Bian, C., Fu, M., Gong, S., & Min, J. (2013). Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3. *Biochemical and biophysical research communications*, *430*(2), 547-553.

Ratajczak, M. Z., Zuba-Surma, E., Kucia, M., Poniewierska, A., Suszynska, M., & Ratajczak, J. (2012). Pluripotent and multipotent stem cells in adult tissues. *Advances in medical sciences*, *57*(1), 1-17.

Razin, A. (1998). CpG methylation, chromatin structure and gene silencing—a three-way connection. *The EMBO journal*, *17*(17), 4905-4908.

Rona, G. B., Eleutherio, E. C., & Pinheiro, A. S. (2016). PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophysical reviews*, *8*(1), 63-74.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H. K., Beyer, T. A., Datti, A., ... & Wrana, J. L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell stem cell*, *7*(1), 64-77.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual* (No. Ed. 2). Cold spring harbor laboratory press.

Sanjana, N. E., Shalem, O., & Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nature methods*, *11*(8), 783-784.

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.

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., ... & Thomä, N. H. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Molecular cell*, *42*(3), 330-341.

Schneuwly, S., Klemenz, R., & Gehring, W. J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. *Nature*, *325*(6107), 816-818.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., ... & Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, *343*(6166), 84-87.

Shenghui, H. E., Nakada, D., & Morrison, S. J. (2009). Mechanisms of stem cell self-renewal. *Annual Review of Cell and Developmental*, *25*, 377-406.

Shi, Y., Desponts, C., Do, J. T., Hahm, H. S., Schöler, H. R., & Ding, S. (2008). Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell stem cell*, *3*(5), 568-574.

Shi, Y., Inoue, H., Wu, J. C., & Yamanaka, S. (2017). Induced pluripotent stem cell technology: a decade of progress. *Nature reviews Drug discovery*, *16*(2), 115-130.

Siller, R., Greenhough, S., Park, I. H., & J Sullivan, G. (2013). Modelling human disease with pluripotent stem cells. *Current gene therapy*, *13*(2), 99-110.

Skucha, A., Ebner, J., & Grebien, F. (2019). Roles of SETD2 in leukemia—transcription, DNA-damage, and beyond. *International journal of molecular sciences*, *20*(5), 1029.

Sommer, C. A., Stadtfeld, M., Murphy, G. J., Hochedlinger, K., Kotton, D. N., & Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem cells*, *27*(3), 543-549.

Spemann, H. (1938). Embryonic Development and Induction, 1988 (reprinted) ed.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., ... & Plath, K. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. *Nature cell biology*, *15*(7), 872-882.

Stabell, M., Larsson, J., Aalen, R. B., & Lambertsson, A. (2007). Drosophila dSet2 functions in H3-K36 methylation and is required for development. *Biochemical and biophysical research communications*, *359*(3), 784-789.

Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., & Hochedlinger, K. (2008). Induced pluripotent stem cells generated without viral integration. *Science*, *322*(5903), 945-949.

Steger, D. J., Lefterova, M. I., Ying, L., Stonestrom, A. J., Schupp, M., Zhuo, D., ... & Vakoc, C. R. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Molecular and cellular biology*, *28*(8), 2825-2839.

Stevens, L. C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre-and postimplantation mouse embryos. *Developmental biology*, *21*(3), 364-382.

Stevens Jr, L. C., & Little, C. C. (1954). Spontaneous testicular teratomas in an inbred strain of mice. *Proceedings of the National Academy of Sciences of the United States of America*, *40*(11), 1080.

Sweeney, S. M., Cerami, E., Baras, A., Pugh, T. J., Schultz, N., Stricker, T., Lindsay, J., Del Vecchio Fitz, C., Kumari, P., Micheel, C., Shaw, K., Gao, J., Moore, N., Stricker, T., Kandoth, C., Reardon, B., Lepisto, E., Gardos, S., Dang, K., … Shaw, K. (2017). AACR project genie: Powering precision medicine through an international consortium. *Cancer Discovery, 7(8), 818-831.*

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., & Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Current Biology*, *11*(19), 1553-1558.

Taei, A., Rasooli, P., Braun, T., Hassani, S. N., & Baharvand, H. (2020). Signal regulators of human naive pluripotency. *Experimental cell research*, *389*(2), 111924.

Takahashi, K. (2014). Cellular reprogramming. *Cold Spring Harbor perspectives in biology*, *6*(2), a018606.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, *126*(4), 663-676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, *131*(5), 861-872.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., & Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. *Proceedings of the National Academy of Sciences*, *110*(30), 12172-12179.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *science*, *282*(5391), 1145-1147.

Toh, C. X. D., Chan, J. W., Chong, Z. S., Wang, H. F., Guo, H. C., Satapathy, S., ... & Loh, Y. H. (2016). RNAi reveals phase-specific global regulators of human somatic cell reprogramming. *Cell reports*, *15*(12), 2597-2607.

Ullah, M., Pelletier, N., Xiao, L., Zhao, S. P., Wang, K., Degerny, C., ... & Yang, X. J. (2008). Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Molecular and cellular biology*, *28*(22), 6828-6843.

van den Berg, E. (2013). Renal Cell Carcinoma. In *Brenner's Encyclopedia of Genetics: Second Edition*.

van Nuland, R., van Schaik, F. M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., ... & van Ingen, H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics & chromatin*, *6*(1), 1-12.

Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., ... & Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, *142*(6), 967-980.

Vezzoli, A., Bonadies, N., Allen, M. D., Freund, S. M., Santiveri, C. M., Kvinlaug, B. T., ... & Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature structural & molecular biology*, *17*(5), 617-619.

Wade, P. A. (2001). Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. *Human molecular genetics*, *10*(7), 693-698.

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.

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., ... & Pei, D. (2011). The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell stem cell*, *9*(6), 575-587.

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., ... & Rossi, D. J. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell stem cell*, *7*(5), 618-630.

Waterland, R. A. (2006). Epigenetic mechanisms and gastrointestinal development. *The Journal of pediatrics*, *149*(5), S137-S142.

Weber, M., & Schübeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Current opinion in cell biology*, *19*(3), 273-280.

Wei, J., Antony, J., Meng, F., MacLean, P., Rhind, R., Laible, G., & Oback, B. (2017). KDM4B-mediated reduction of H3K9me3 and H3K36me3 levels improves somatic cell reprogramming into pluripotency. *Scientific reports*, *7*(1), 1-14.

Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., ... & Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*, *573*(7773), 281-286.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., ... & Shi, X. (2014). ZMYND11 links histone H3. 3K36me3 to transcription elongation and tumour suppression. *Nature*, *508*(7495), 263-268.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., & Campbell, K. H. (1997). Viable offspring derived from fetal and adult mammalian cells. *nature*, *385*(6619), 810-813.

Wood, K. H., & Zhou, Z. (2016). Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. *Frontiers in genetics*, *7*, 93.

Wozniak, G. G., & Strahl, B. D. (2014). Hitting the 'mark': interpreting lysine methylation in the context of active transcription. *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1839*(12), 1353-1361.

Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M. F., Xu, C., ... & Min, J. (2011). Structural and histone binding ability characterizations of human PWWP domains. *PloS one*, *6*(6), e18919.

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. *The EMBO journal*, *30*(8), 1473–1484.

Xie, P., Tian, C., An, L., Nie, J., Lu, K., Xing, G., ... & He, F. (2008). Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. *Cellular signalling*, *20*(9), 1671-1678.

Xing, Q. R., El Farran, C. A., Gautam, P., Chuah, Y. S., Warrier, T., Toh, C. X. D., ... & Loh, Y. H. (2020). Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Science advances*, *6*(37), eaba1190.

Xu, W., Li, J., He, C., Wen, J., Ma, H., Rong, B., ... & Shen, H. (2021). METTL3 regulates heterochromatin in mouse embryonic stem cells. *Nature*, *591*(7849), 317-321.

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.

Yamanaka, S. (2012). Induced pluripotent stem cells: past, present, and future. *Cell stem cell*, *10*(6), 678-684.

Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., ... & Kouzarides, T. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*, *593*(7860), 597-601.

You, L., Yan, K., Zou, J., Zhao, H., Bertos, N. R., Park, M., ... & Yang, X. J. (2015). The chromatin regulator Brpf1 regulates embryo development and cell proliferation. *Journal of Biological Chemistry*, *290*(18), 11349-11364.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. *Science*, *324*(5928), 797-801.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., ... & Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *science*, *318*(5858), 1917-1920.

Yu, S., Li, J., Ji, G., Ng, Z. L., Siew, J., Lo, W. N., ... & Wu, Q. (2021). Npac Is a Co-factor of Histone H3K36me3 and Regulates Transcriptional Elongation in Mouse Embryonic Stem Cells. *Genomics, Proteomics & Bioinformatics*.

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.

Yun, M., Wu, J., Workman, J. L., & Li, B. (2011). Readers of histone modifications. *Cell research*, *21*(4), 564-578.

Zaccara, S., Ries, R. J., & Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nature reviews Molecular cell biology*, *20*(10), 608-624.

Zaghi, M., Broccoli, V., & Sessa, A. (2020). H3K36 methylation in neural development and associated diseases. *Frontiers in genetics*, *10*, 1291.

Zaidan, N. Z., & Sridharan, R. (2020). HP1γ regulates H3K36 methylation and pluripotency in embryonic stem cells. *Nucleic acids research*, *48*(22), 12660-12674.

Zhang, J., Han, C., Song, K., Chen, W., Ungerleider, N., Yao, L., ... & Wu, T. (2020). The long-noncoding RNA MALAT1 regulates TGF-β/Smad signaling through formation of a lncRNA-protein complex with Smads, SETD2 and PPM1A in hepatic cells. *PloS one*, *15*(1), e0228160.

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., ... & Ding, J. (2006a). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic acids research*, *34*(22), 6621-6628.

Zhang, P., Zhao, J., Wang, B., Du, J., Lu, Y., Chen, J., & Ding, J. (2006b). The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. *Protein science : a publication of the Protein Society*, *15*(10), 2423–2434.

Zhang, Y. L., Sun, J. W., Xie, Y. Y., Zhou, Y., Liu, P., Song, J. C., ... & Huang, Q. H. (2018). Setd2 deficiency impairs hematopoietic stem cell self-renewal and causes malignant transformation. *Cell research*, *28*(4), 476-490.

Zhang, Y., Shan, C. M., Wang, J., Bao, K., Tong, L., & Jia, S. (2017). Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Scientific reports*, *7*(1), 1-9.

Zhang, Y., Xie, S., Zhou, Y., Xie, Y., Liu, P., Sun, M., ... & Chen, S. (2014). H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. *Cell reports*, *8*(6), 1989-2002.

Zhang, Z., Jones, A., Sun, C. W., Li, C., Chang, C. W., Joo, H. Y., ... & Wang, H. (2011). PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprograming. *Stem cells*, *29*(2), 229-240.

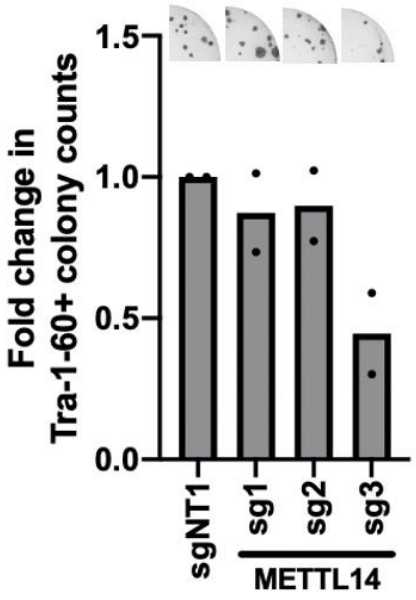
Zhao, W., Ji, X., Zhang, F., Li, L., & Ma, L. (2012). Embryonic stem cell markers. *Molecules*, *17*(6), 6196-6236.

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., ... & Ding, S. (2010). Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell stem cell*, *7*(6).

Zhu, X., He, F., Zeng, H., Ling, S., Chen, A., Wang, Y., ... & Wang, Q. F. (2014). Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nature genetics*, *46*(3), 287-293.

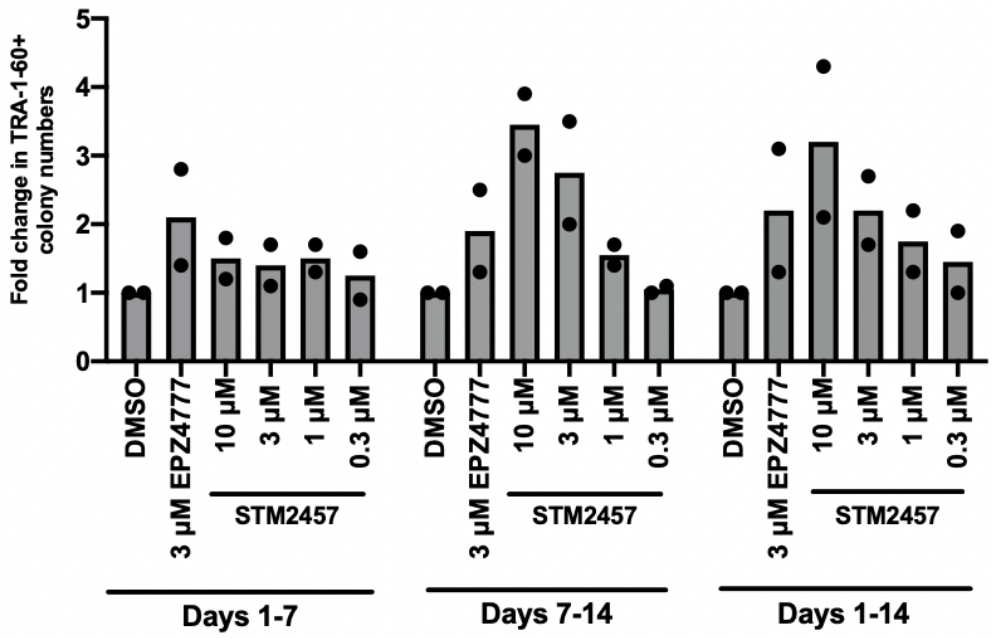
APPENDIX A: INVESTIGATING THE ROLE OF N^6-METHYLADENOSINE (M^6A) IN HUMAN SOMATIC CELL REPROGRAMMING

N^6-methyladenosine (m^6A) is the most common mRNA modification in eukaryotes and significant for gene regulation and involved in many cellular processes (Zaccara et al., 2019). The role of m^6A modification in somatic cell reprogramming is not fully defined despite several studies performed in mouse (Aguilo and Walsh, 2017; Khodeer et al., 2021). It is known that this modification is dynamically regulated by its writer, reader and erasers but the underlying mechanism of its deposition on the transcriptome has not been shown precisely as well. One possible mechanism was demonstrated by Huang *et al.* as m^6A modification was globally controlled by H3K36me3 (Huang et al., 2019). m^6A methyltransferase complex (MTC) is composed of METTL3 having catalytic activity, METTL4 having RNA binding property and WTAP the regulatory subunit (Kobayashi et al., 2018). In the study of Huang *et al.,* H3K36me3 was shown to be recognized and bound by METTL14 and MTC catalyzes m^6A modification co-transcriptionally (Huang et al., 2019). In addition, SETD2 inhibition led to reduced m^6A level transcriptome-wide and on mRNAs of pluripotency-related genes *OCT4, SOX2, KLF4* and *NANOG* resulting in enhanced stemness of mouse ESCs (Huang et al., 2019). Therefore, I hypothesized that H3K36me3 loss may enhance human somatic cell reprogramming via regulation of m^6A modification. To test this hypothesis, I knocked-out METTL14 in dH1f cells using three independent sgRNAs and detected its effect on OSKM-mediated reprogramming. However, I could not observe a significant change with its two out of three sgRNAs while the sg3 led to 50% decrease in reprogramming efficiency (Appendix A – Figure 1).



Appendix A - Figure 1: The effect of METTL14 knockout on human somatic cell reprogramming. Fold change in the number of TRA-1-60-positive colonies generated by OSKM induction over sgNT1-expressed control cells Bar graphs indicate means of biological replicates. n=3, each contained 3 technical replicates.

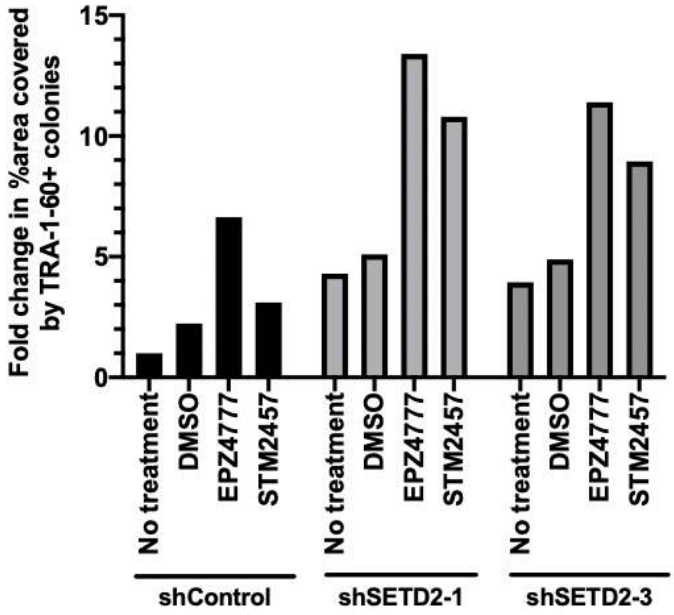
I next tested a catalytic inhibitor of METTL3 (STM2457) in reprogramming which was recently identified and shown to inhibit MTC to catalyze m^6A modification (Yankova et al., 2021). Yankova *et al*. tested STM2457 in mouse models of acute myeloid leukemia and suggested it as a potential therapeutic strategy (Yankova et al., 2021). I performed time course experiments with STM2457 using four different concentrations (10 μM, 3 μM, 1 μM and 0.3 μM) in OSKM-mediated reprogramming of dH1f cells. In these experiments, EPZ4777 (DOT1L inhibitor) which is known to enhance reprogramming efficiency at least 2-fold was used as positive control (Onder et al., 2012). Inhibition of METTL3 resulted in increased reprogramming efficiency and it seems that it showed its effect on late reprogramming (Appendix A – Figure 2). I detected the highest increase as 4-fold change in reprogramming efficiency with 10 μM STM2457 treatment during days 7-14 of reprogramming (Appendix A – Figure 2). These findings indicate that METTL3 is a roadblock against human somatic cell reprogramming and not effective at early reprogramming.



Appendix A - Figure 2: Time course experiment of METTL3 inhibitor, STM2457, in human somatic cell reprogramming. Fold change in the number of TRA-1-60-positive colonies generated by OSKM induction over DMSO treated control cells. STM2457 was used in decreasing concentrations from 10 to 0.3 µM. The cells were also treated with 3 µM EPZ4777 (DOT1L inhibitor) at each time window. Bar graphs indicate means of biological replicates. n=2, each contained 3 technical replicates.

After I detected the increasing effect of STM2457 in cellular reprogramming, I also tested it in combination with SETD2 knockdown to check whether they show an additive effect. The additive effect of SETD2 knockdown and DOT1L inhibition on reprogramming efficiency was demonstrated before (Ozcimen, 2018). Therefore, 3 µM EPZ4777 treatment at first seven days of reprogramming was used as positive control in this experiment as well. SETD2 was knocked-down by two independent shRNAs in dH1f cells, and then the cells were subjected to OSKM-mediated reprogramming. During reprogramming, the cells were treated with 3 µM STM2457 between days 7 and 14. METTL3 inhibition in shControl cells caused 3-fold increase and SETD2 knockdown led to almost 5-fold increase in reprogramming efficiency compared to untreated shControl-expressed cells (Appendix A – Figure 3). When they combined, 10-fold increase was observed so they appeared to have a synergistic effect rather than an additive effect on

OSKM-reprogramming of dH1f cells (Appendix A – Figure 3). Taken together these findings indicate that m^6A loss by METTL3 inhibition enhances human somatic cell reprogramming and this effect appeared to be not related with H3K36me3.



Appendix A - Figure 3: The effect of STM2457 treatment in combination with SETD2 knockdown on reprogramming efficiency. Fold change in % area covered by TRA-1-60-positive colonies generated by OSKM induction over untreated shControl-expressed cells. 3 μM STM2457 was used during days 7-14 of reprogramming. The cells were also treated with 3 μM EPZ4777 (DOT1L inhibitor) during days 1-7. Bar graphs indicate means of three technical replicates.