IDENTIFICATION OF TRANSCRIPTION FACTORS THAT INTERACT WITH ARID4B PROTEIN IN MOUSE EMBRYONIC STEM CELLS

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

BY

EZGI GÜL KESKIN

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

SEPTEMBER 2019

Approval of the thesis:

IDENTIFICATION OF TRANSCRIPTION FACTORS THAT INTERACT WITH ARID4B PROTEIN IN MOUSE EMBRYONIC STEM CELLS

submitted by **EZGI GÜL KESKIN** in partial fulfillment of the requirements for the degree of **Master of Science in Molecular Biology and Genetics Department, Middle East Technical University** by,

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

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

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

Examining Committee Members:

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

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

Assoc. Prof. Dr. Erkan Kiriş
Biology, METU _____

Date: 05.09.2019

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

Name, Surname: Ezgi Gül Keskin

Signature:

ABSTRACT

IDENTIFICATION OF TRANSCRIPTION FACTORS THAT INTERACT WITH ARID4B PROTEIN IN MOUSE EMBRYONIC STEM CELLS

Keskin, Ezgi Gül
Master of Science, Molecular Biology and Genetics
Supervisor: Assist. Prof. Dr. Nihal Terzi Çizmecioğlu

September 2019, 58 pages

Embryonic development is a process that is highly regulated at the level of gene expression. Throughout this process, the zygote divides and differentiates to form all cell types in the adult body. In the early stages of development, pluripotent embryonic stem cells (ESC) can be isolated and used to study cell fate mechanisms during development in vitro. The pluripotency of ESCs is governed by certain transcription factors (TF) (Oct4, Nanog, etc.). These TFs construct a transcription network specific to ESC and enable the high and stable expression of pluripotency genes. Chromatin factors work alongside TFs to maintain pluripotency.

In our previous studies, we have identified Arid4b protein in the Sin3a corepressor complex is critical for proper ESC differentiation into mesoderm and endoderm lineages. Arid4b might direct the Sin3a complex to the ESC differentiation related genes with the aid of its histone modification and DNA binding domains. Histone deacetylases in the complex is critical for the suppression of gene expression.

To better understand the role of Arid4b protein in ESC differentiation, we examined the histone 3 lysine 27 acetylation (H3K27Ac) genome-wide distribution in wildtype and arid4bΔ mesoderm or endoderm differentiated cells. H3K27Ac is a marker of active enhancer regions as well as active promotor regions. We found that

the higher level of H3K27Ac in the arid4bΔ cells when compared to wildtype cells. Therefore, Sin3a complex may be required for the suppression of these genes during ESC differentiation.

Upon bioinformatics analyses, we identified TF consensus binding sequences enriched in the regions with increased H3K27Ac levels in arid4bΔ cells. We analyzed the mRNA expression level of 10 such TFs in WT and arid4bΔ ESCs. We further validated the protein expression of selected TFs. Of these, we showed that Tfap2c physically interacts with Arid4b and Sin3a in ESCs through endogenous coimmunoprecipitation experiments.

Keywords: Embryonic Stem Cell, Transcription Factor, Chromatin

ÖZ

FARE EMBRİYONİK KÖK HÜCRELERDE ARID4B PROTEİNİYLE ETKİLEŞEN TRANSKRİPSİYON FAKTÖRLERİNİN BULUNMASI

Keskin, Ezgi Gül
Yüksek Lisans, Moleküler Biyoloji ve Genetik
Tez Danışmanı: Dr. Öğr. Üyesi Nihal Terzi Çizmecioğlu

Eylül 2019, 58 sayfa

Embriyonik gelişim, gen ifadesinin kontrolü seviyesinde, son derece planlı ve düzenli işleyen bir süreçtir. Bu süreç boyunca, zigot bölünür ve erişkin vücudundaki bütün hücre tiplerini oluşturacak şekilde farklılaşır. Gelişimin erken safhalarında pluripotent özellik gösteren embriyonik kök hücreler (EKH) izole edilip laboratuvar ortamında gelişim sırasındaki hücre kader mekanizmaları üzerinde çalışmak için kullanılabilir. EKH'lerin pluripotent karakteri belli transkripsiyon faktörleri (TF) (Oct4, Sox2 vb.) tarafından yönetilir. EKH'lere özgü bir transkripsiyon ağı kuran bu TF'ler pluripotentlik için gerekli genlerin yüksek ve dengeli bir miktarda ifade edilmesini sağlar. TF'lerin bu görevlerinde kromatin faktörleri de görev alır.

Daha önceki çalışmalarımızda EKH'lerin mezoderm ve endoderm tabakalarına farklılaşmasını sağlayan kromatin faktörlerini belirledik. Bu faktörlerden biri, Sin3a korepresör kompleksinin içinde bulunan Arid4b proteinidir. Arid4b proteininin histon modifikasyon ve DNA bağlanma bölgeleri yardımıyla Sin3a kompleksini, EKH farklılaşması için gerekli olan genlere yönlendirdiğini düşünüyoruz. Kompleks içerisinde olan histon deasetilazlar sayesinde bu kompleks, gen ifadesini baskılayıcı bir görev üstlenebilmektedir. Arid4b proteininin EKH farklılaşmasındaki görevini daha detaylı anlayabilmek için, normal ve arid4bΔ EKH'leri mezoderme ve endoderme farklılaştırıp genomdaki histon 3 lizin 27

asetilasyon (H3K27Ac) dağılımını inceledik. H3K27Ac aktif arttırıcı bölgelerinde; aktif promotor bölgelerinde görülür. arid4bΔ hücrelerindeki H3K27Ac seviyesinin normal hücrelere oranla daha yüksek olduğunu belirledik.

Arid4b ve içinde bulunduğu Sin3a kompleksi, bu genlerin EKH farklılaşması sırasında baskılanması görevinde bulunuyor olabilir. Bu görevi yerine getirirken, Arid4b'nin transkripsiyon faktörleri ile çalıştığını düşünüyoruz. H3K27Ac seviyesinde artış görünen bölgelerde konsensüs bağlanma sekansı sık bulunan bir TF olan Tfap2c proteinin normal ve arid4bΔ EKH'lerde mRNA ve protein seviyesinde ifade edildiğini teyit ettik. Yaptığımız eşimmünoçökeltme deneylerine göre, Arid4b ve Sin3a proteinin Tfap2c ile fiziksel etkileşimi olduğunu gösterdik. Sonuç olarak, Arid4b'nin Tfap2c proteini ile fiziksel etkileşim sayesinde genlerin baskılanmasını sağladığını düşünüyoruz. Bu projede, EKH'lerin mezoderm ve endoderme farklılaşma süreçlerini kontrol eden Sin3a korepresör kompleksi ve Arid4b proteinin bir transkripsiyon faktörü olan Tfap2c ile fiziksel etkileşimini belirledik.

Anahtar Kelimeler: Embriyonik Kök Hücre, Transkripsiyon Faktörü, Kromatin

This thesis is dedicated to my entire family who have been my source of inspiration and gave me strength when I thought of giving up.

ACKNOWLEDGEMENTS

I would like to thank to my supervisor Assist Prof Dr. Nihal Terzi Çizmecioğlu for her support and advice throughout my graduate study. I have learned a lot about scientific approach, critical thinking and planning future steps. I would like to thank her for patience even with a very tiny problem. I will always be grateful for her counseling that encourages me to be a well-rounded scientist.
I would like to thank Prof. Dr. Stuart H. Orkin for providing mouse embryonic stem cells and cell strains to our research.
I would like to thank all my jury members; Assoc. Prof. Dr. Işık Yuluğ and Assoc. Prof. Dr. Erkan Kiriş.
I would like to thank Prof. Dr. Mesut Muyan for his support. He always motivated me to keep going and I have learned a lot about scientific approach, nature and art from him. I will always see him as role model.
I would like to thank Gözde Güven as my lab mate. We have overcome many problems together since EpiStem lab was established.
I am deeply grateful and thankful to former and present Lab B59 family; Esin G. Seza, Melis Çolakoğlu, Sinem Ulusan, Burak Kızıl, Aydan Torun, İsmail Güderer, Çağdaş Ermiş, Hoşnaz Tuğral for their support and help in most challenging problem.
I would like to thank Pelin Yaşar, Gamze Ayaz, Gizem Kars, Gizem Turan, Kerim Yavuz, Çağla Olgun, Negin Razizadeh for their valuable help to solve any kind of problems and for allowing us to use their equipment and materials.
I would like to thank Ayça H. Çırçır, Murat Erdem, Harun Cingöz, İbrahim Özgül and other ER lab members for help allowing us to use their equipment and materials.
I would like to thank Asena Şanlı and İ.Cihan Ayanoğlu and other Mayda Gürsel lab members for exchange information and help allowing us to use their equipment and materials.
I would like to thank Dr. Semir Beyaz for contributions to my adventure of science. He guided me when I sank into pessimism about my career and inspired me with his motivation in science.
I would like to thank my beloved friend Gökçe Senger for friendship from IYTE to METU.
Many thanks to my love Yiğit Tunçtürk. He always listened my complaints and problems, helped me to solve them. I always get strength from him.
I would like to thank my beloved family for stay by me every condition, their endless love and supports. I cannot thank you enough to them. My father, Mustafa Zeki Keskin, my mother, Meral Keskin and my sister Özge Sevin Keskin always lift me up every time I fall. My only and the greatest luck in my life is my family and I am grateful to everything they did.
Finally, I would like to thank TUBITAK for supporting through 117Z923.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

Arid4b	AT-Rich Interaction Domain 4B
Bach1	BTB Domain and CNC Homolog 1
Ddit3	DNA Damage Inducible Transcript 3
DNMT	DNA Methyltransferases
GTF	General Transcription Factors
HAT	Histone Acetyltransferase
HDAC	Histone deacetylases
HMT	Histone Methyltransferase
H3K4eme3	H3 tri-methylation on Lysine 4
H3K27Ac	H3 mono-acetylation on Lysine 27
H3K27me3	H3 tri-methylation on Lysine 27
ICM	Inner Cell Mass
IP	Immunoprecipitation
LIF	Leukemia Inhibitory Factor
mESC	Mouse Embriyonic Stem Cells
Oct4	Octamer-binding transcription factor 4
PcG	Polycomb Group Protein
PGC	Primordial Germ Cell
Pol	RNA Polymerase

PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
Prrx2	Paired Related Homeobox 2
PTM	Post-translational Modification
qRT-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
RNAPII	RNA Polymerase II
Sin3a	SIN3 Transcription Regulator Family Member A
SMC	Structural Maintenance of Chromosome Protein
Sox2	Sex Determining Region Y-Box 2
Stat3	Signal Transducer and Activator of Transcription 3
TBP	TATA-box Binding Protein
TF	Transcription Factor
Tfap2c	Transcription factor AP-2 gamma
Znf354c	Zinc Finger Protein 354C

CHAPTER 1

INTRODUCTION

1.1. Transcription

Gene expression is a tightly regulated process. During this process, the instructions in the DNA are turned into RNA and proteins. The first step of gene expression is transcription where DNA is copied into RNA by an enzyme called RNA polymerase. RNA polymerases are very conserved from bacteria to eukaryotes. Bacteria possess one polymerase but eukaryotes possess at least three different polymerases that specialize in the synthesis of distinct classes of RNA molecules (Pol I, Pol II, Pol III; in plants Pol IV and Pol V, additionally) (Vannini & Cramer, 2012).

Transcription can be studied in three different stages: initiation, elongation, termination. RNA Polymerase II (RNAPII) work with initiation factors to perform efficient and promoter-specific initiation in eukaryotes. The initiation factors are named as general transcription factors (GTF) which aid RNAPII in recruitment on the appropriate upstream region of the gene, called a promoter. These factors help RNAPII to reach template strand by melting DNA by hydrolysis of ATP and to stabilize its binding on promoters to avoid fall off until elongation phase. Also, GTFs have role in promoter escape and help to enter on elongation phase (Krishnamurthy & Hampsey, 2009). In eukaryotes, DNA is compacted with specific proteins called histones to form nucleosome. Therefore, chromatin modifying enzyme and chromatin remodeler complex are required to provide open chromatin structure to RNAPII for efficient initiation. In addition to that, Mediator complex and transcriptional activator proteins are needed for efficient initiation. Mediator regulates the cytoplasmic domain kinase in TFIIH, one of the GTF, thereby helps initiation. Activators are the transcription regulatory proteins which assist the recruitment of RNAPII to the

promoter and stabilize its binding to promoter (Kornberg, 2007; Krishnamurthy & Hampsey, 2009).

Reviewed in Krishnamurthy, all GTFs and RNAPII bind together to the promoter and get ready for initiation. This complex is called preinitiation complex (PIC). Pol II promoters might include a TATA box. TFIID which is a GTF has subunits like TATA box binding protein (TBP) that recognizes TATA box and TATA box associated protein (TAF). TAFs also recognizes the core promoter elements like downstream core element (DCE), downstream promoter element (DPE), initiator (Inr)(Krishnamurthy & Hampsey, 2009).

Krishnamurthy and Hampsey also reviewed that once TFIID recognizes and binds to the promoter, RNAPII and the other GTFs are recruited to the promoter site. GTFs recruited to promoter are TFIIA, TFIIB, TFIIF. Then, TFIIE and TFIIH are recruited to the promoter. After formation of preinitiation complex, promoter is melted as a result of hydrolysis of ATP mediated by TFIIH in eukaryotes. Besides, TFIIH has kinase activity which is important promoter escape. Carboxy-terminal domain (CTD) tail of RNA Pol II (RNAPII) include repeats of heptapeptide sequence (Tyr-Ser-Pro-Thr-Ser-Pro-Ser). Phosphorylation of CTD tail of RNAPII is critical for promoter escape and embarking on elongation. The tail is unphosphorylated in preinitiation stage, however, phosphorylation of serine 5 by TFIIH causes dissociation of mediator from RNAPII thereby promoter escape happens. At the same time, phosphorylation of serine 5 is a signal for 5' capping, as well.(Krishnamurthy & Hampsey, 2009).

After successful promoter escape of the RNA Pol II, the elongation phase begins. In this step, general transcription factors that form preinitiation complex, are changed by different set of transcription factors called elongation factors. Additionally, RNA processing machinery (for 5' capping and polyadenylation) are recruited to CTD tail of RNAPII. The phosphorylation of serine 2 in CTD of RNAPII leads to start of elongation. This phosphorylation is performed by P-TEFb which is recruited to Pol II by activators and binds to Pol II. P-TEFb can also phosphorylate SPT5 to activate it.

Another elongation factor that is recruited by P-TEFb is TAT-SF1. The length of time that RNAPII stops is limited by some of elongation factors to ensure continuity of elongation. ELL family which is a class of elongation factors and TFIIIS limit the length of time(Maniatis & Reed, 2002; Zhou, Li, & Price, 2012). In addition, DNA is folded by specialized proteins called histones. Therefore, RNAPII must overcome histones during elongation. To do this, there are proteins which are specified to disassemble and reassemble histones during this process. In human, FACT (facilitates chromatin transcription) complex deals with histones. FACT is formed by two highly conserved proteins, Spt16 and SSRP1. Spt16 binds to H2A:H2B dimer whereas SSRP1 binds to H3:H4 tetramer. During elongation, FACT dismantles H2A:H2B dimer and this allows RNAPII to proceed easily. Then, H2A:H2B dimer is restituted immediately after RNAPII by help of histone chaperone activity of FACT. This way, the integrity of chromatin is protected during continuous cycles of transcription (Cao et al., 2003; Orphanides, LeRoy, Chang, Luse, & Reinberg, 1998)

When RNAPII reaches poly-A signal sequence in DNA, cleavage and polyadenylation is stimulated. The polyadenylation triggers the termination of transcription. When poly-A signal sequence is transcribed, polyadenylation and cleavage factor pass from CTD tail of RNAPII to the RNA. Then, mRNA is cleaved, and a poly-A stretch is added to its 3' end (Zhou et al., 2012). The remaining RNA piece which is still linked to RNAPII is degraded by ribonucleases. Finally, transcription is terminated. Although the exact mechanism of transcription termination is still under investigation, there are two possible models that fit current data in the field. One of them is the torpedo model detailed above with the help of RNases. The other one is allosteric model. In this model, termination is performed because of conformational change in the RNAPII that decreases the processivity of RNAPII. Polyadenylation might trigger the conformational change. For instance, passing of polyadenylation and cleavage factor to form tail of RNAPII to RNA might be the reason (Rosonina, Kaneko, & Manley, 2006).

1.2. Chromatin

The nucleosome is the smallest unit of chromatin. It is formed by histone proteins wrapped by 147 bp of DNA. The histones are positively charged small proteins and there are four types of histones in nucleosome as H2A, H2B, H3 and H4. H2A and H2B form a dimer whereas H3 and H4 constitute a tetramer, they are called as canonical core histones. Another important aspect of histones is that they have amino-terminal tail which is unstructured and available to proteases. There are many serine, tyrosine and arginine residues that undergo different types of post-translational modifications (PTM). These modifications change function or the chemical nature of the nucleosome (K. K. Lee & Workman, 2007). H1 is another type of histone that is found between two nucleosomes as linker. While H1 introduces high-order chromatin structure, it provides approx. 40-fold compaction. It means that further compaction is still required to fit DNA into nucleus. Moreover, this compaction affects the accessibility of the DNA by DNA-dependent enzymes. (Lessard & Crabtree, 2010). In high compaction level, there are two different chromosomal regions as heterochromatin and euchromatin. Heterochromatin is the region where DNA is densely packaged however, euchromatin has less compaction level. In this sense, heterochromatin regions are less accessible and therefore gene expression on that region is generally low. The heterochromatin and euchromatin regions are regulated dynamically to provide control of gene expression (Li & Reinberg, 2011; Sarma & Reinberg, 2005). Methylation of DNA, PTMs of histones and its tail, chromatin modifying enzyme, and chromatin remodeler have role in the regulation of chromatin structure (Li & Reinberg, 2011). For further compaction, large loops of nucleosomes are formed by 30-nm fibers. It makes DNA more compact for fitting into nucleus. The class of proteins that have role in generating loops are topoisomerase II (Topo II) and structural maintenance of chromosome (SMC) proteins. During transcription, supercoiling is fixed by topoisomerases (Naughton et al., 2013).

SMC proteins are encoded by at least six genes in eukaryotes and work with other protein in chromosome condensation, recombination, repair of DNA and epigenetic

repression of gene expression (Harvey, Krien, & O'Connell, 2002). Also, they have role in condensation of sister chromatin and holding of sister chromatin together after chromosome duplication (W.-H. Tan, Gilmore, & Baris, 2013).

Besides, there are also histone variants such as H3.3, H2AZ, H2AX in eukaryotes. They get incorporated into nucleosome instead of canonical core histones and provide distinct functions in diverse cellular processes. They regulate the structure of chromatin (Sarma & Reinberg, 2005). For example, H2A has four variants, one of which is H2AX. When DNA is damaged, a serine residue of H2AX which localized to the double stand break is phosphorylated. DNA repair machinery recognizes the phosphorylated H2AX thus, the needed proteins are recruited to site of damage (Sarma & Reinberg, 2005).

As mentioned before, histone tails undergo plenty of PTMs. These modifications involve acetylation, methylation, phosphorylation, ubiquitination and sumoylation (K. K. Lee & Workman, 2007). Recently, new histone modifications such as propionylation, butyrylation, crotonylation, glutarylation were identified. (Yue Chen et al., 2007; M. Tan et al., 2011, 2014). Histone modifications affect chromatin compaction, gene expression, genome organization, cell division and repair. Combination of modifications indicates whether a gene is repressed or active (Lee & Workman, 2007). Also, modification on different amino acid on tail of histone influences the gene expression differently. The gene might be activated or repressed based on the modification type and amino acid position on the tail. The multiple modifications that function in combination or in a sequential manner on one histone tails or multiple tails give rise to unique downstream functions. This is known as histone code (Strahl & Allis, 2000).

1.3. Early Development and Embryonic Stem Cells

The early development in mouse embryo is a process that involves fertilization of egg to gastrulation embryo. After fertilization of the egg, it undergoes series of cleavage divisions and forms 2 cells to 8 cells. Then, the cluster of cells forms a 3D ball like

structure, this stage is called morula stage. After morula formation (8-cell stage which is called early morula), cells undergo compaction leading to increase in cell-cell junction and cells become stack. Then, cells continue to grow and morula undergoes cleavage to generate blastocyst which is a hollow sphere-like structure. After symmetric and asymmetric division, a fluid-filled cavity which is called blastocoel is formed. The inner cells gather on a region, this is called inner cell mass (ICM). The outer cells form the trophoblast region. The formation of ICM and trophoblast cells is the first cell fate decision in the early development. After implementation, epiblast and primitive endoderm are formed from ICM. The formation of epiblast and primitive endoderm is known as second cell fate decision. Embryonic stem cells (ESC) are derived from ICM (Takaoka & Hamada, 2012). ESCs have self-renewal ability which means that they can divide indefinitely. Interestingly, they have potential to produce any type of cell in a body which is called pluripotency. ESCs are able to differentiate to three primitive germ layers, endoderm, mesoderm and ectoderm (Gaspar-Maia, Alajem, Meshorer, & Ramalho-Santos, 2011). ESCs are suitable model organism for the study of development and disease.

1.4. Transcription Factors in Embryonic Stem Cell Pluripotency

There are core pluripotency factors to maintain pluripotency in embryonic stem cells These are Oct4, Nanog Sox2, Klf4 and c-Myc. Additionally, Sall4, Rex1, Dax1 and Tcl1 have important functions in ESCs (Orkin & Hochedlinger, 2011).

Oct4 is a key transcription factor that is expressed from *Pou5f1* gene. Its expression is downregulated during differentiation. One of the class of the POU (Pit/Oct/Unc) family of homeodomain proteins is octamer transcription factors. They prefer binding to an 8-basespair DNA sequence. A class of evolutionary conserved protein is called homeodomain proteins which have essential function in development. Oct4 belongs to this class. Oct4 is highly expressed in ICM and also in epiblast and primordial germ cells (PGCs) (Pesce & Schöler, 2004). After gastrulation, Oct4 expression is limited to germ cells. Oct4 knockout embryos die at the preimplantation stage with problems

of pluripotency loss and abnormal expression of trophoblast markers at ICM. (Nichols et al., 1998). A study also showed that repression of Oct4 in ESCs leads to loss of self-renewal and differentiation into trophoblast. On the other side, overexpression of it supports the differentiation into multiple cell types (Niwa, Miyazaki, & Smith, 2000).

The other core factor is Sox2 which is SRY (Sex determining Region Y) -related TFs. It has high mobility group (HMG) DNA domain and expressed in other cell types like neural progenitor cells (Heng & Ng, 2010). Deletion of Sox2 in ESCs causes loss of self-renewal however it was shown that the ectopic expression of Oct4 rescues the self-renewal ability of Sox2 $^{-/-}$ ESCs. Therefore, it is claimed that Sox2 maintains ESC stability by keeping Oct4 expression at certain level (Masui et al., 2007).

Nanog is specifically expressed in ICM and PGCs. Like Oct4, it is a member homeodomain TFs and forms a homodimer with itself or a hetero-dimer with Smad1 to maintain pluripotency and self-renewal (Mullin et al., 2008). In ESCs, it is also critical to maintain pluripotency independent of the LIF-Stat3 pathway (Mitsui et al., 2003). Like Oct4, expression of Nanog is reduced during differentiation and deletion of Nanog causes differentiation of ESCs into extraembryonic endoderm lineage (Ivanova et al., 2006; Mitsui et al., 2003).

1.5. Chromatin and Embryonic Stem Cells

The chromatin contains distinct region as heterochromatin and euchromatin. The euchromatin region is more accessible to the DNA-related enzymes such as RNAPII, therefore gene expression in euchromatin is generally higher. The chromatin structure in pluripotent stem cell is more open compared to chromatin structure of somatic cells (Gaspar-Maia et al., 2011). The electron microscope image of chromatin in ESC proved that euchromatin was more widespread in ESCs than in differentiated cells (Efroni et al., 2008). Another important finding is that ICM of mouse blastocyst at day 3.5 have a similar open chromatin conformation with ESCs (Ahmed et al., 2010). H3 tri-methylation on Lysine 9 (H3K9me3) is enriched in heterochromatic regions because this modification is a repressive signal for transcription by recruiting silencing

factors. Western blot and immunofluorescence experiments of such modification support the idea that ESCs possess less heterochromatin than differentiated cells (Meshorer et al., 2006). Conversely, acetylation which is a key marker of open chromatin is decreased during ESC differentiation (Krejčí et al., 2009).

In ESCs, histone PTMs might provide regulation of chromatin state to affect pluripotency. There are chromatin modifying enzymes and chromatin remodelers that affects the chromatin states. Mostly known modifying enzymes are histone methyltransferase (HMTs), histone demethylases (HDMs), histone acetyltransferases (HATs), histone deacetylases (HDACs). SWI/SNF, CHD, ISWI, INO80 are the family of the multi-subunit chromatin-remodeling proteins (Gaspar-Maia et al., 2011).

The open chromatin of ESCs has abundant levels of epigenetic modification that correlate with active transcription such as, H3K4me3 or H3K27ac. H3K4me3 is associated with 5' end of active genes. The presence of this modification on 5' region of genes positively correlates with active RNAPII occupancy, transcription rate. These modifications are indicative of active or bivalent promoter (H3K4me3) or active enhancers (H3K27ac) (Azuara et al., 2006; Creyghton et al., 2010). There is a fine-tuning that provides repression of genes responsible for differentiation while activation of pluripotency genes in ESCs. The differentiation genes are repressed however they are poised for activation when there are activating marker and repressive marker on the promoter of interested gene. When differentiation is stimulated, the repressive mark, H3K27me3 on bivalent promoter of lineage-specific genes is lost and it causes rapid activation of lineage-specific genes (Bernhart et al., 2016). H3K4me3 and H3K27me3 are bivalency marker indicates activation of transcription and repression of it, respectively. The polycomb group (PcG) proteins regulated the repressive H3K27me3. PcG proteins involves two complex, polycomb repressive complex 2 (PRC2) and polycomb repressive complex 1 (PRC1). PRC2 methylate 27^{th} lysine residue on H3 (H3K27me3) which is recognized by PRC1. It was shown that PcG proteins are enriched on the repressive development regulatory genes in human and mouse ESCs(Boyer et al., 2006). Additionally, the co-occupancy of PcG protein

target genes with pluripotency key factors (Oct4, Sox2, Nanog) was observed (T. I. Lee et al., 2006).

1.6. Sin3 Corepressor Complex

In mammals, there are three isoforms of Sin3 encoded by two different genes, SIN3A and SIN3B (Grzenda, Lomberk, Zhang, & Urrutia, 2009). Sin3a is a scaffold protein of multi-protein Sin3a complex and its sequence is very similar by 98 percent in human and mouse. In mammals, it includes four paired amphipathic helix (PAH) domains and HDAC interaction domain (HID). Sequence-specific transcriptional regulators are recognized by PAH domains (PAH 1-4). The human protein MAD/MAX, mouse proteins Mxd1, are examples of transcriptional regulators (Kadamb, Mittal, Bansal, Batra, & Saluja, 2013). These regulators help to recruit Sin3a proteins to the specific loci because Sin3a complex doesn't have an intrinsic DNA binding activity (Grzenda et al., 2009). Sin3a complex includes Hdac1 and Hdac2 which regulate chromatin structure by deacetylase activity and gene expression in eukaryotes (Kelly & Cowley, 2013). Other proteins are also involved in the core complex, RbAp4, RbAp7, SAP30, SAP18, and SDS3, CpG methylated binding protein (MeCP2) (Grzenda et al., 2009; Kadamb et al., 2013). Sin3a is also associated with different enzymatic activities, containing O-linked N-acetylglucosamine transferase (OGT8), helicases of the Swi/Snf chromatin remodeling complex. Additionally, it interacts with Ten-eleven translocation 1 (Tet1) which is a 5-methylcytosine (5mC) dioxygenase (Sif, Saurin, Imbalzano, & Kingston, 2001; Williams et al., 2011; Yang, Zhang, & Kudlow, 2002). Chromatin structure and gene expression is also regulated by DNA methylation through DNA methyltransferases such Dnmt1. Tet1 proteins can initiate DNA demethylation. The co-occupancy of Sin3a and Tet1 protein on many same target genes in ESCs was shown (Williams et al., 2011). Additionally, highly conserved Tet1-Sin3a interaction domain (Tet1-SID) was identified, Tet1-SID interacts with PAH1 domain of Sin3a (Chandru, Bate, Vuister, & Cowley, 2018). The demethylation activity of Sin3a/Hdac1 through Tet1 proteins might be important for ESCs.

Sin3a is important for early development and efficient self-renewal and somatic cell reprogramming. The mechanism behind the regulation of transcription in pluripotency and reprograming is not clear. It was shown that Sin3a/Hdac1 complex and core pluripotency protein Nanog physically interacts with each other in ESCs (Liang et al., 2008). Another recent study showed that Sin3a co-activator function is enhanced when it worked with Nanog. This is important for stem cell maintenance. The same study also suggested that synergistic function of Nanog and Sin3a are required for activation of pluripotency genes and repression reprogramming barrier genes for efficient reprogramming (Saunders et al., 2017). In our studies, we focused on Hdac1 and Arid4b proteins that are the members of Sin3a corepressor complex.

The human AT-rich interaction domain (ARID) family involves 15 members. All members of ARID family include a DNA-binding domain. The ARID family might have role in cell growth, differentiation and development and human cancer. The one of the subfamilies of ARID is ARID4 which includes two members, ARID4A and ARID4B. Their gene location on chromosome are 14q23.1 and 1q42.1-q43, respectively. The sequence of Arid4a and Arid4b proteins are 40%-50% identical. They possess an ARID domain, a Tudor domain and N-Terminal retinoblastoma-binding protein-1 family domain. Unlike ARID4B, ARID4A includes retinoblastoma (RB) binding motif. ARID4B protein is member of mSin3a/Hdac complex. The *ARID4B* gene is also called *breast cancer associated antigen 1* (*BRCAA1*) *or Sin3-Associated Polypeptide P180* (*SAP180) or retinoblastoma binding protein 1 like protein 1 (RBP1L1).* Its expression is high in human colon, pancreatic, lung, breast, pancreatic and ovarian cancers and in normal testis. However, gene expression of it is limited in other normal tissues (Lin et al., 2014). In development of metastatic breast cancer, a causative role of ARID4B is supported by recent studies (Winter, Lukes, Walker, Welch, & Hunter, 2012). ARID domain provides DNA binding whereas the chromodomain and tudor domain play role in protein-protein interaction. Therefore, it is suggested that ARID4B recruit mSin3a/Hdac complex to DNA and stabilizes the complex by multiple interaction which are not dependent on sequence or recruits

additional factors to support repressor function of the complex(Fleischer, Yun, & Ayer, 2003).

The lysine residue undergo acetylation which is well-known PTMs and it is indicated that this modification has regulatory roles in many cellular processes(Choudhary et al., 2009). Acetylation is catalyzed by histone acetyltransferases (HAT). Histone deacetylases (HDAC) remove acetyl group. The N-terminal tails of the core histones H3 and H4 contain high amount of lysine residues that is charged positively, this allows protein-protein and DNA-protein interaction however, it limits the accessibility to the DNA. The acetylation of lysine residue in H3 (Lys^{9}, Lys^{14}, Lys^{18} and Lys^{23}) and H4 (Lys^{5}, Lys^{8}, Lys^{12} and Lys^{16}) neutralize the electrostatic effects and the association of H1 histone that compacts DNA is inhibited by acetylation. Acetylated lysine of H3 and H4 is recognized by bromodomain-containing proteins such as, P/CAF (p300/CREB (cAMP-response element-binding protein)-binding protein-associated factor) and some HATs themselves. Thus, open chromatin structure and gene activation is stimulated(Kelly & Cowley, 2013).

In mammals, there are lots of HDACs and HATs, however not all of them have active role in nucleus. Regulatory role on gene expression is performed by the class-1 HDACs, 1,2,3, and 8. They form the catalytic core of Sin3, NuRD (nucleosome remodeling and deacetylation) CoREST (co-repressor for element-1-silencing transcription factor) and SMRT (silencing mediator of retinoid and thyroid receptors) /NCoR (nuclear receptor co-repressor) co-repressor complex. HDAC1/2 don't have DNA binding ability therefore they need interacting partners (Kelly & Cowley, 2013).

HDAC1 is important for development in mouse (Lagger et al., 2002). Conditional deletion of *Hdac1* not *Hdac2* in ESCs leads to precocious differentiation. This finding supports the idea that HDAC1 plays a role in early development and cell cycle regulation (Dovey, Foster, & Cowley, 2010). Although HDACs are mostly known as histone deacetylases, they also catalyze the deacetylation non-histone proteins such as

p53, E2F1, GATA4 and NF-KB (Ying Chen et al., 2011; LeBoeuf et al., 2010; Martínez-Balbás, Bauer, Nielsen, Brehm, & Kouzarides, 2000; Trivedi et al., 2010).

1.7. Preliminary Data

In the previous study, a pooled shRNA screen was performed in mESCs to find the epigenetic factors that control mesoderm and endoderm differentiation (In preparation). It is found that almost 30 chromatin related factors have critical role in efficient endoderm and mesoderm differentiation. One of these factors is Arid4b. Therefore, the study has been focused on the Sin3a corepressor complex and its components, Hdac1, Hdac2 and Arid4b (Sap180).

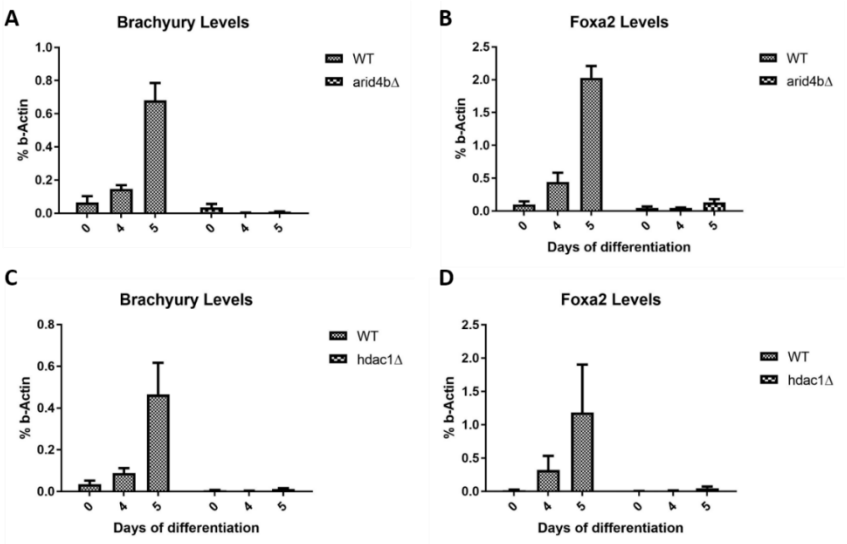


Figure 1.1. The gene expression level of Brachyury and Foxa2 in wild type (WT), arid4bΔ and hdac1Δ cells that is committed to endoderm. Brachyury is a meso-endoderm marker and Foxa2 is endoderm markers. **A, B**, arid4bΔ cell does not express Brachyury and Foxa2 genes. **C, D**, the level of Brachyury and Foxa2 gene expression in hdac1Δ cells is significantly reduced during endoderm differentiation. Data normalized to b-actin and represented as %b-actin. Data are shown as mean ± SEM of three biological replicates.

When Arid4b is knocked down by shRNA and deleted by CRISPR-Cas9 technology, it was observed that mouse embryonic stem cells cannot differentiate into mesoderm

or endoderm (Figure 1.1). Brachyury is a meso-endoderm marker and Foxa2 is an endoderm marker. When cells are committed to endoderm, brachyury gene is expressed. In the forthcoming days, Foxa2 genes is expressed and it shows that mESCs differentiate into endoderm germ layer (Figure 1.1).

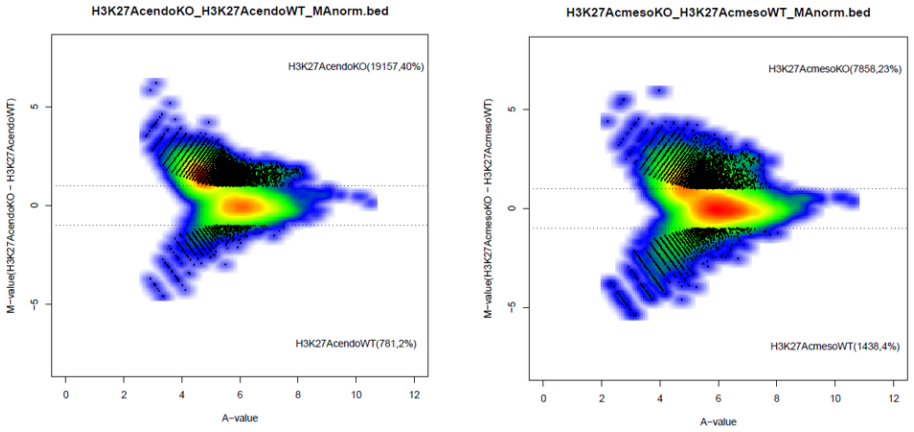


Figure 1.2. The overlap of H3K27Ac peaks between normal (wildtype, WT) and arid4bΔ (KO) cells which committed to endoderm and mesoderm. The overlap of H3K27Ac peaks between WT and arid4bΔ cells which committed to endoderm (left) and mesoderm (right). was analyzed by MA-norm (Shao, Zhang, Yuan, Orkin, & Waxman, 2012). The position of peaks of H3K27Ac ChIP-Seq in WT and KO cells were compared. X-axis presents peak density for each position on genome, Y-axis indicates the differences of amount of H3K27Ac signal between WT and KO for each position on genome. The dots parallel to x-axis and the rectangle including zero point on y-axis shows the overlap peaks. The dots on upper part of the rectangle shows the region on genome where H3K27Ac is higher in KO cells. The dots on lower part of the rectangle shows the region on genome where H3K27Ac is higher in WT cells. The frequency of dots is showed by color, blue means more sparse, black refers frequent. In both graph (left and right ones), overlap of peaks between KO and WT was observed (the region in the triangle). The area of black regions is more above the triangle compared the lower part. It means the level of H3K27Ac is higher in KO cells than that of in WT.

This finding shows the absence of Sin3a corepressor complex leads to a dramatic defect in ESC differentiation process. To understand the mechanism of differentiation defect due to the loss of Arid4b, whole genome chromatin precipitation and sequencing (ChIP- Seq) was performed for H3K4me3, H3K27me3 and H3K27Ac. In arid4bΔ endoderm directed cells, the H3K27Ac modification which is commonly found on the active enhancers and promoters as well as bivalent promoters are higher

compared to normal (wildtype, WT) endoderm differentiated cells (Figure 1.2). The genes that have an increased level of H3K27Ac are transcriptionally upregulated in arid4bΔ cells. This suggests that the H3K27Ac modification on genome must be kept at a certain level by the Sin3a corepressor complex during ESC differentiation.

Sin3a corepressor complex cannot recognize DNA sequence. Also, it is known that transcription factors can function properly when the chromatin environment is favorable. Besides, it is showed that the chromatin factors can work with transcription factors. In this sense, Sin3a corepressor complex may prepare the suitable chromatin environment for TFs and Sin3a corepressor complex might be recruited to targeted DNA region by transcription factors. In a summary, the function of TFs might depend on the function of the Sin3a corepressor complex or vice versa. Therefore, we focus on the transcription factors whose consensus binding sequences are enriched on the region where the level H3K27Ac is higher compared to the other region on the genome (Table 1.1). Based on the results of bioinformatic analysis (HOMER, motif analysis), we identified consensus binding sequences of transcription factors (TF) that are enriched on the region where the level H3K27Ac is higher compared to the other region on the genome (Table 1.1). The Sin3a corepressor complex that contain Arid4b leads to repression of some genes by deacetylase activity of Hdac1 during differentiation. Therefore, the effect of Arid4b on only these genes may arise from working with specific TF to these genes. Our hypothesis was that Arid4b might interacts with one or several TFs on the list to direct the differentiation process. To test this hypothesis, we chose the TFs that have active (H3K4me3) or bivalent marker (H3K4me3 ve H3K27me3). Then, the expression of them in the mESCs were searched in the BioGPS database (Figure 3.1). TFs that we anticipate the expression in mESC based of BioGPS were chosen to further validation with qRT-PCR (Figure 3.2). Accordingly, the presence of protein of selected TFs were checked with western blot (Figure 3.3). Then, we checked the interaction between Arid4b and TFs that expressed as both RNA and protein in the mESC. We found that Arid4b protein physically interacts with Tfap2c protein (Figure 3.5 and 3.6).

Table 1.1. *Candidate Transcription Factors which work with Arid4b*

Transcription Factor	*Enrichment Value*	Transcription Factor	*Enrichment Value*
Pdx1	300	Sox3	29
Bach1:MafK	300	Sry	25
Prrx2	159	Bhlhe40	25
Znf354c	147	Nkx3-1	24
Mzf1_1-4	99	Creb1	21
Pou2f2	73	Tfap2a	20
Tfap2c	66	En1	17
Ddit3:Cebpa	57	Sox2	17
Stat6	53	Tp53	16
Nobox	51	Nhlh1	14
Atoh1	51	Foxd1	13
TBP	39	Brca1	13
Hinfp	35	MafD	11
Stat2:Stat1	35		

1.8. The Aim of Study

ESCs have huge potential to form any type of cell in a body. Our previous data showed that mESCs cannot differentiate into mesoderm or endoderm when we deleted the arid4b genes. This finding suggested that arid4b gene is critical for commitment of mESCs into germ layer such as endoderm and mesoderm. Also, the result of ChIP showed that the H3K27Ac modification in arid4bΔ endoderm directed cells are higher compared to normal (wildtype, WT) endoderm differentiated cells (Figure 1.2).

We know that Sin3a corepressor complex that contain Arid4b protein cannot recognize DNA sequence. Also, TFs can function properly when the chromatin environment is favorable. Besides, it is showed that the chromatin factors can work with transcription factors. Sin3a corepressor complex may prepare the suitable chromatin environment for TFs and Sin3a corepressor complex might be recruited to targeted DNA region by transcription factors. In a summary, the function of TFs might depend on the function of the Sin3a corepressor complex or vice versa. Therefore, transcription factors (TF) that are enriched on the region where the level H3K27Ac is higher compared to the other region on the genome was identified. Our hypothesis was that Arid4b might interacts with one or several TFs to direct the differentiation process. The aim of this thesis to understand the mechanism of mESC differentiation

into germ layers and explore the epigenetic regulation behind the mechanism of cell fate by identification of TFs that are physically interact with Arid4b protein. If the cell fate decision and differentiation process are understood clearly, ESC might be directed to generate progenitor cells for tissue recovery. This will eventually contribute the regenerative medicine which would be solution for diseases such as neural and pancreatic diseases.

CHAPTER 2

MATERIALS AND METHODS

2.1. Cell Culture

CJ9 and mESCs that arid4b genes deleted by CRISPR-Cas9 system (arid4bΔ) were obtained from laboratory of Prof. Stuart Orkin (Boston Children Hospital). CJ9 and arid4bΔ mESCs were expanded in the high glucose DMEM (Gibco, Thermo Fisher Scientific), including 15% FCS (Gibco, Thermo Fisher Scientific), 1% Pen-strep (Gibco,Thermo Fisher Scientific), 1% Nucleoside Mix (80 mg Adenine, 24 mg thymidine, 85 mg Guanosine 73 mg Uracil ve 73 mg cytosine), 1% Glutamax (200mM, Gibco, Thermo Fisher Scientific), NEAA (nonessential aminoacids) (Gibco,Thermo Fisher Scientific), 10^{-4} M beta-mercaptoethanol on the irradiated mouse embryonic fibroblast (MEF) and cultured at 37ºC at 5% CO2. When cells were 70-80% confluent, cells were split by using 0.25% Trypsin-EDTA (Sigma-Aldrich). These cells that were growth on MEFs were used qRT-PCR experiment. Because of supply difficulty of MEFs, growth medium was changed and we started to expand cells onto gelatinized tissue culture (TC) plates without MEFs. In this system, mESCs were expand in high glucose DMEM including , 4% FBS (Gibco, Thermo Fisher Scientific), 1% Pen-strep (Gibco,Thermo Fisher Scientific), 1% Glutamax (200mM, Gibco, Thermo Fisher Scientific), 0.65 µl / 50 ml 1-Thioglycerol (MTG), 3 µM CHIR99021 (Selleckchem), 1µM PD0325901 (Selleckchem) and 1 % LIF (Millipore). Cells were cultured at 37ºC at 5% CO2. The growth medium is called 2i4 medium. When cells were 70-80% confluent, cells were split by using 1X TrypLE Express enzyme (Sigma Aldrich). Whole cell extracts for western blot and nuclear extracts for co-immunoprecipitation were prepared from these cells. You can see detailed medium recipes in the appendix part (Table A.1, Table A.2, Table A.3).

2.2. RNA Expression Studies

2.2.1. Sample Preparation For RNA Isolation

CJ9 (normal, wildtype) and arid4bΔ mESCs were expanded on the 10 cm gelatinized tissue culture plates. Before, Trypsin-EDTA (0.25%) enzyme was used to detach colonies, attached colonies were washed with DMEM (high glucose) to get rid of dead cells. To quench enzymatic activity of Trypsin-EDTA (0.25%), DMEM was used and then, collected cells were centrifuged at 300 g at +4 ^{o}C for 5 minutes. To get rid of fibroblast, cells were put into 10 cm gelatinized TC dish and incubated at 37 ^{o}C short term. After collecting cells, they were centrifuged at 300 g at +4 ^{o}C for 5 minutes. Next, pellets were resuspended in 1 ml DMEM to count. After count, collected cells were centrifuged at 300 g at +4 ^{o}C for 5 minutes. Then, supernatant was removed and pellet were washed twice with cold PBS. Finally, pellet was resuspended in TRIzol (Thermo Fisher) and stored at – 20 ^{o}C.

2.2.2. RNA Isolation and cDNA Synthesis

For RNA isolation, pellets of normal (wildtype) and arid4bΔ mESCs had been resuspended in TRIzol to get total cell lysate. To obtain total RNA, Qiagen RNeasy Plus Micro Kit (Catalog no: 74034) was used based on the kit's instructions. With Biorad iScript cDNA sythesis kit (Catalog no: 1708890), complementary DNAs (cDNA) were synthesized from total RNA by using oligo dT and random hexamers. cDNAs were stored at -20 ^{o}C

2.2.3. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)

Melting curve for all primers that were used in qRT-PCR was done to check specificity of the primer. The volume of a PCR reaction is 10 µl including forward and reverse primers (each final concentration is 10 µM) and 5 µl of 2X SsoAdvanced universal SYBER Green (Bio-Rad). qRT-PCR was performed in 96 well qPCR plates by using Bio-Rad CFX Connect (Bio-Rad) machines. cDNA dilutions corresponding to18-32

Cq values were chosen so that Cq (quantification cycle) values were on the log-linear phase of amplification. The Cq values that were out of the range were eliminated. To determine the level of amplification, Cq values were normalized to Cq values of b-actin and the fold change is represented in percentage.

The qPCR primers of candidate TFs were obtained from previously used studies. The primer sequence and PubMed ID of the articles which includes primers are in the appendix part (Table B.1).

2.3. Protein Studies

2.3.1. Sample Preparation for Whole Cell and Nuclear Extraction

CJ9 (normal, wildtype) and arid4bΔ mESCs that grew in 2i4 medium were used for further experiments. For western blot results, whole cell lysate were prepared from $1x10^6$ cells. After last centrifugation step as written in 2.1, supernatant was removed and pellet were washed twice with cold PBS. Then, pellet was undergone snap freeze (99% ethanol and dry ice) at stored at -80 ^oC.

2.3.2. Western Blot

Different concentrations of whole cell lysate were tried in western blot experiments and the concentration corresponding to $1x10^6$ cell lysate were suitable for the best upon signal to noise ratio. Each replica including set of wild type and arid4bΔ mESCs was expand at different times. Three biological replicates were done, each whole cell lysate was obtained from $1x10^6$ cell. After thawing cell lysate on the ice, proper amount of 2x Laemmli buffer (Biorad) including β-mercaptoethanol (BME) corresponding to cell number was used to resuspend cell pellet. To denature proteins, cell lysates were incubated at the 95 ^oC for 10 minutes, then put onto ice to cool. After that, sample were centrifuged at maximum speed for 1 minute to get rid of cell debris. Then, cell lysates were load into 12 % polyacrylamide gel. After electrophoresis (PAGE), proteins were transferred to nitrocellulose membrane with Biorad Trans-blot Turbo Transfer System. For blocking the membrane, TBST containing 5% skimmed

milk were used and membrane were incubated in blocking solution for 1 hour at room temperature. Incubation of primary antibodies were done 2 days at cold room (+4 ^{o}C), the antibody names and dilutions are written in Table C.1 in the appendix part. The antibodies that we used yielded weak signal therefore two days incubation worked for these antibodies. After primer antibodies, membranes were washed with TBST three times, 10 minutes for each. Then, secondary antibodies that is linked to horseradish peroxidase (HRP) were put on the membranes and membranes were incubated one hour at RT. The antibody name and dilution are written Table C.1 in the appendix part. After secondary antibodies, membranes were washed with TBST three times, 10 minutes for each. For visualization, Biorad Clarity Western ECL was used with Biorad ChemiDoc MP system. Biorad Clarity Max Western ECL which is an amplified form of ECL was used for antibodies that has weak power signal.

The samples of Arid4b-coIP and Tfap2c reciprocal co-IP was incubated at 95 ^{o}C for 5 minutes to separate proteins on the beads and denature proteins. After reaching to RT, samples were centrifuged at 6000 rpm for 5 minutes. The other steps of western blot were the same as written above. However, the heavy chain of IgG gave signal at 50 kDa where it overlaps with Tfap2c signal. Therefore, the membrane that include Tfap2c membrane was incubated with conformational specific secondary antibody for 1 hour at RT. Then, it was incubated with tertiary antibody for one hour at RT because the conformational specific antibody was not conjugated to HRP (Table C.2). To visualize the membranes for visualization, Biorad Clarity Western ECL was used with Biorad ChemiDoc MP system. Biorad Clarity Max Western ECL which is an amplified form of ECL was used for antibodies that has weak power signal. Also, the Tfap2c protein and Sin3a in the reciprocal co-IP was visualized with Thermo SuperSignal™ West Dura Extended Duration Substrate which was gave best results for them. Different concentration of nuclear lysate was used for co-IP experiment, 400 μg nuclear extract and 20 μg input were the most suitable amount that came up with the best result of western blot of both Arid4b and Tfap2c co-IP. You can find out

detailed information about antibody name and dilution in Table C.2, C.3 in the appendix part.

2.4. Coimmunoprecipitation

2.4.1. Arid4b Coimmunoprecipitation

The replicas that were expanded at different times. The nuclear extracts were obtained from wildtype and arid4bΔ mESCs with Universal Magnetic CoIP Kit (Active Motif, catalog no: 54002). The protocol suggested by kit was followed. As a result of trying different concentration of nuclear lysate, we decided that 20 µg input was enough. Normal rabbit IgG was precipitated in the nuclear extract obtained from wildtype and arid4bΔ mESCs to check the non-specific precipitation. As another control, Arid4b protein was precipitated in the nuclear extract obtained from arid4bΔ mESCs. Because, any transcription factors that interact with Arid4b are not precipitated or very small amount of it are precipitated in the nuclear extract obtained from arid4bΔ mESCs.

In the immunoprecipitation (IP) experiment, 400 µg of CJ9 (wildtype) nuclear lysate were used for normal rabbit IgG- IP and Arid4b-IP. Also, 400 µg of arid4bΔ nuclear lysate were used for Arid4b-IP. To decrease the non-specific binding during co-IP, equal amount of the magnetic beads conjugated Protein G were incubated with nuclear extracts obtained from wildtype and arid4bΔ mESCs for 2 hours on the rotator at +4 ^{o}C (pre-clearing). Precleared lysates were transferred to new eppendorf and 20 µg input from each cell was took, the inputs were mixed with 2x Laemmli containing BME (Biorad) and stored at -80 ^{o}C. Then, their final volumes which were prepared as kit suggested were completed to 500 µl with complete co-IP wash buffer. Following that, arid4b (Bethyl) and normal rabbit IgG (Millipore) antibodies as 5 µg antibody in 500 µl were added to related nuclear lysates. The information about antibodies is written in Table C.3 in the appendix part. After adding antibodies, nuclear lysates were incubated 4 hours on the rotator at +4 ^{o}C. At the end of 4 hours, equal amount of magnetic beads conjugated to Protein G were added to each immunoprecipitation

sample and incubated one hour on rotator +4 ^{o}C. After that, beads were washed with complete co-IP wash buffer two times for 4 minutes on the rotator at +4 ^{o}C and washed with NP-40 three times for 4 minutes on the rotator at +4 ^{o}C. The washed beads were resuspended in 2x Laemmli Buffer containing BME (Biorad) and stored at – 80 ^{o}C for further western blot.

2.4.2. Tfap2c Coimmunoprecipitation

The replicas that were expanded at different times. The nuclear extracts were obtained from wildtype mESCs with Universal Magnetic CoIP Kit (Active Motif). The protocol suggested by kit was followed. We decided that 20 µg input was enough. Normal mouse IgG was precipitated in the nuclear extract obtained from wildtype mESCs to check the non-specific precipitation.

In the reciprocal IP experiment, 350 µg of CJ9 (wildtype) nuclear lysate were used for normal mouse IgG- IP and Tfap2c-IP. To decrease the non-specific binding during co-IP, equal amount of the magnetic beads conjugated Protein G were incubated with nuclear extracts obtained from wildtype for 2 hours on the rotator at +4 ^{o}C (pre-clearing). Precleared lysates were transferred to new eppendorf and 20 µg input from each cell was took, the inputs were mixed with 2x Laemmli containing BME (Biorad) and stored at -80 ^{o}C. Then, their final volumes which were prepared as kit suggested were completed to 500 µl with complete co-IP wash buffer. Following that, Tfap2c (Santa Cruz) and normal mouse IgG (Santa Cruz) antibodies as 2 µg antibody in 500 µl were added to related nuclear lysates. The information about antibodies is written in the Table C.3 in the appendix part. After adding antibodies, nuclear lysates were incubated 4 hours on the rotator at +4 ^{o}C. At the end of 4 hours, equal amount of magnetic beads conjugated to Protein G were added to each immunoprecipitation sample and incubated one hour on rotator +4 ^{o}C. After that, beads were washed with complete co-IP wash buffer two times and with NP-40 three times for 4 minutes on the rotator at +4 ^{o}C. The washed beads were resuspended in 2x Laemmli Buffer containing BME (Biorad) and stored at – 80 ^{o}C for further western blot.

CHAPTER 3

RESULTS

3.1. Expression of Candidate Transcription Factors in mESCs

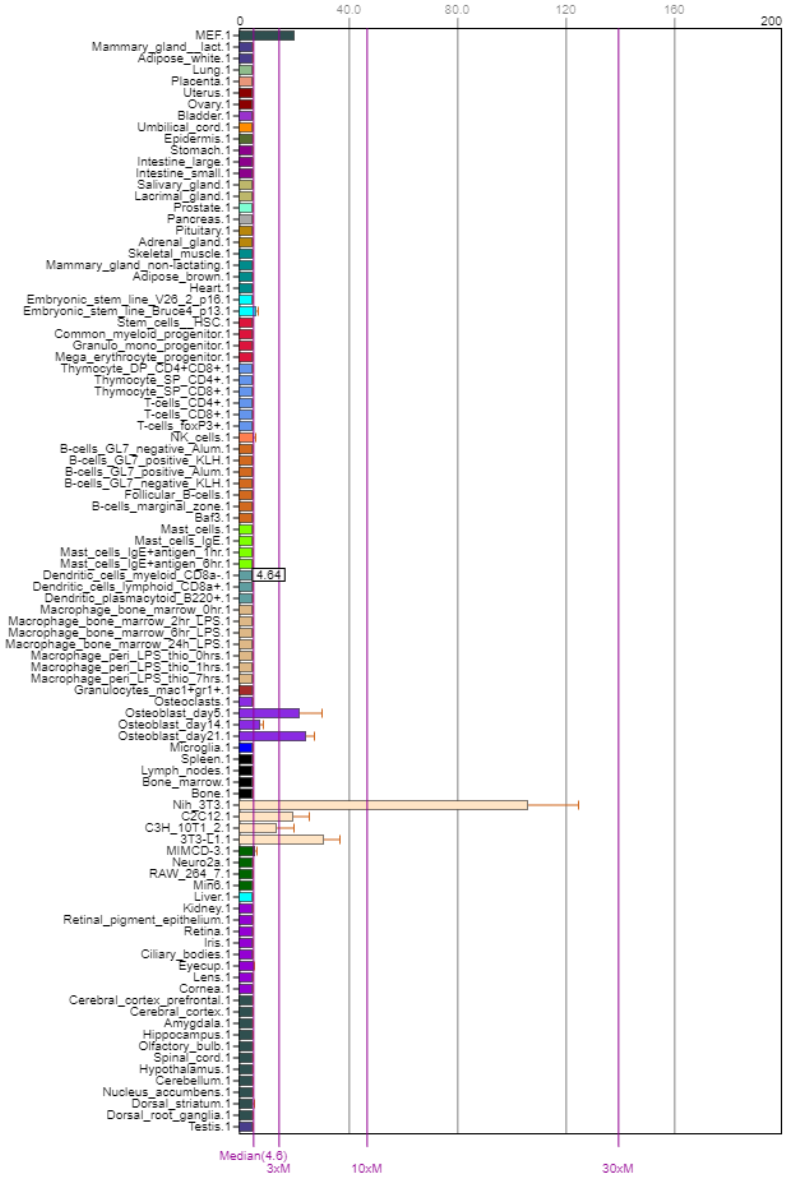
3.1.1 Bioinformatic Studies

ChIP studies during endoderm or mesoderm differentiation cells showed a significantly higher global H3K27Ac signal in arid4bdel cells. (Figure 1.2). For endoderm and mesoderm, consensus binding sequences of 27 transcription factors were enriched in the region that contain higher level of H3K27Ac in proportion to other regions (Table 1.1). To find out candidate transcription factors whose physical or genetic interaction with Arid4b and / or other members of Sin3a corepressor complex has been previously demonstrated in other cell types or other model organisms, we reviewed the literature. Each TF in Table 1.1 was searched in the PubMed database (https://www.ncbi.nlm.nih.gov/pubmed/). For each TFs, 1. the physical or functional relation with Sin3a or/and Arid4b, 2. functional role with ESCs, 3. a role in cell fate mechanisms were checked. TFs which fit all three parameters were selected for further inquiry. Sry, Bhlhe40, Nkx3-1, Creb1, Tfap2a, En1, Sox2, Tp53, Nhlh1, Foxd1, Brca1, Mafb were eliminated.

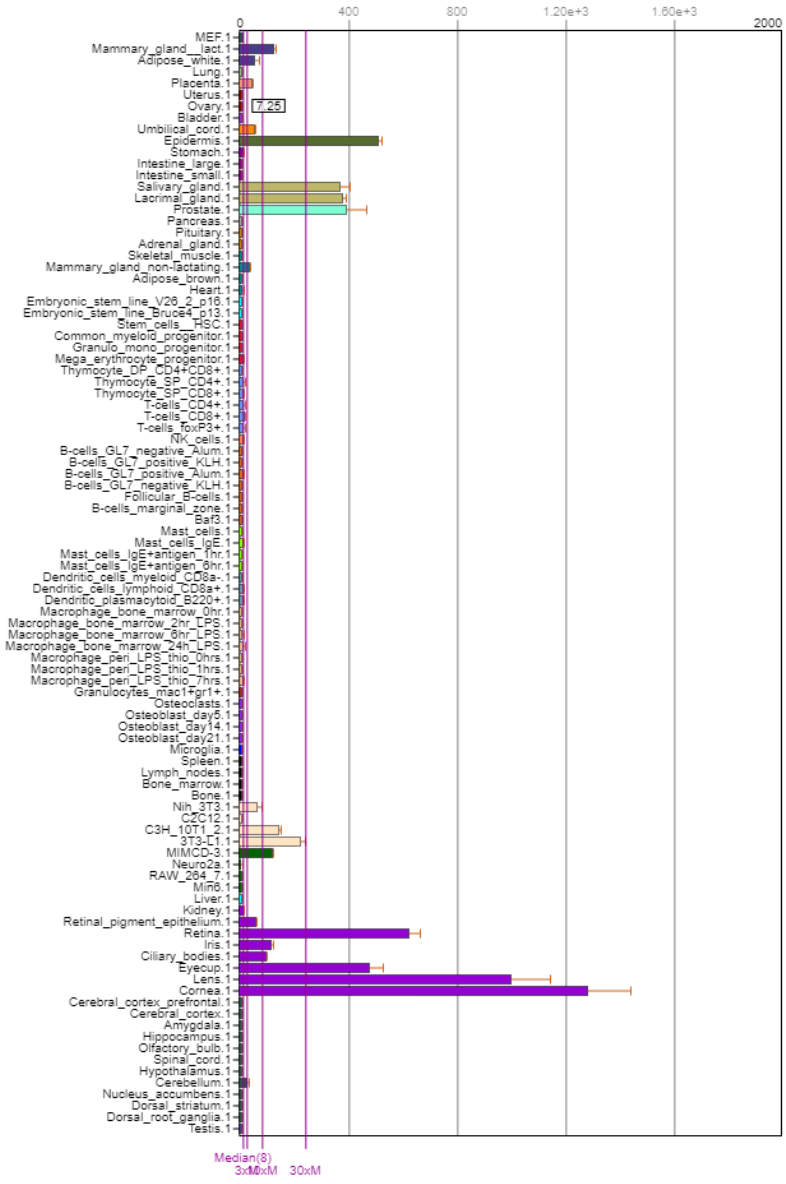
To identify the expression profile of selected TFs across various cell types and during mouse embryonic development, two databases were used: BioGPS (http://biogps.org) and LifeMap (https://discovery.lifemapsc.com). Those TFs that show an expression in ESCs through BioGPS were further tested in LifeMap database. Based on the BioGPS data in Figure 3.1, En1 and Tfap2a are not expressed in mESCs. Tfap2c and Bach1 are expressed in mESCs and different tissues. TFs that are not expressed in mESCs based on databases were eliminated.

The residue (lysine, serine etc.) and level of chromatin modifications can be informative about whether genes are expressed or not. Therefore, the levels of H3K4me3 and H3K27me3 modifications in gene loci of candidate TFs were checked through visualization of our previous ChIP-seq data in Integrative Genomics Viewer (IGV) (https://software.broadinstitute.org/software/igv/) IGV is as viewer tool for large, integrated genomic datasets. H3K4me3 is specific modification found on active enhancer and bivalent promoter and associated with gene activation. Unlikely, H3K27me3 is a modification on the bivalent promoter and associated with gene repression. H3K4me3 and H3K27me3 ChIP-seq data were analyzed with IGV, TFs were divided into classes based on the modification they include. Then, each TFs among the its classes was put in order regarding to the peak height and peak width on the region encoded TFs of interest. The analysis results are shown in Table 3.1. In the table, dark grey presents high level of modification of interest on the region encoded candidate TFs; and light grey refers to low level of modification on that region. No shading means this modification is not observed at a meaningful level. We presumed that the TFs that do not harbor H3K4me3 in their promoter region are not expressed. Therefore, we eliminated TFs in column 3 and 4 of Table 3.1. Bivalent genes are predicted to be expressed in low levels. Therefore, TFs which don't have modification, such as En1, Nobox were eliminated. Additionally, TFs which are bivalent but not expressed in ESC based on LifeMap and BioGPS were also eliminated (Figure 3.1b). As a conclusion, TFs that had the highest level of modification colored with dark and light gray was selected from column 1 and 2 of Table 3.1. 10 TFs which are bivalent and includes only H3K4me3 modification were selected to check gene expression at transcription level by RT-qPCR.

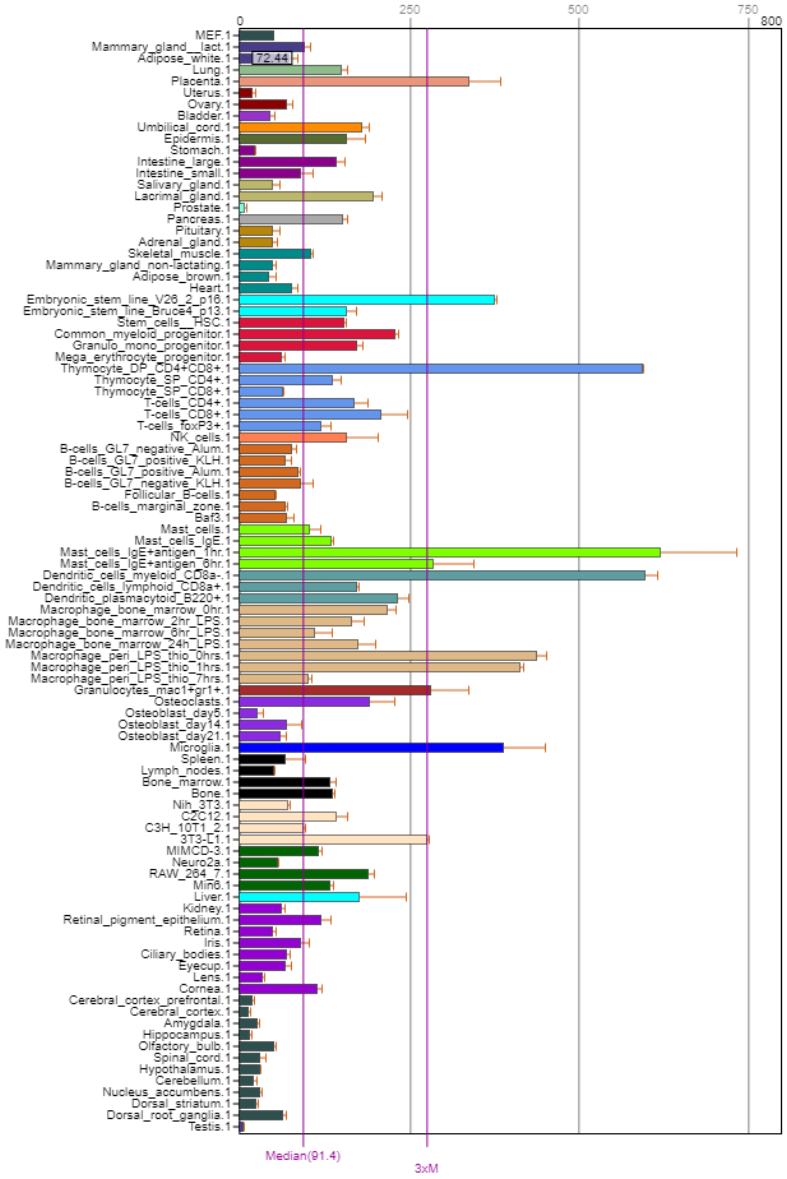
a.


0
40.0
80.0
120
160
200
MEF.1
Mammary_gland__lact.1
Adipose_white.1
Lung.1
Placenta.1
Uterus.1
Ovary.1
Bladder.1
Umbilical_cord.1
Epidermis.1
Stomach.1
Intestine_large.1
Intestine_small.1
Salivary_gland.1
Lacrimal_gland.1
Prostate.1
Pancreas.1
Pituitary.1
Adrenal_gland.1
Skeletal_muscle.1
Mammary_gland_non-lactating.1
Adipose_brown.1
Heart.1
Embryonic_stem_line_V26_2_p16.1
Embryonic_stem_line_Bruce4_p13.1
Stem_cells__HSC.1
Common_myeloid_progenitor.1
Granulo_mono_progenitor.1
Mega_erythrocyte_progenitor.1
Thymocyte_DP_CD4+CD8+.1
Thymocyte_SP_CD4+.1
Thymocyte_SP_CD8+.1
T-cells_CD4+.1
T-cells_CD8+.1
T-cells_foxP3+.1
NK_cells.1
B-cells_GL7_negative_Alum.1
B-cells_GL7_positive_KLH.1
B-cells_GL7_positive_Alum.1
B-cells_GL7_negative_KLH.1
Follicular_B-cells.1
B-cells_marginal_zone.1
Baf3.1
Mast_cells.1
Mast_cells_IgE.1
Mast_cells_IgE+antigen_1hr.1
Mast_cells_IgE+antigen_6hr.1
Dendritic_cells_myeloid_CD8a-.1
4.64
Dendritic_cells_lymphoid_CD8a+.1
Dendritic_plasmacytoid_B220+.1
Macrophage_bone_marrow_0hr.1
Macrophage_bone_marrow_2hr_LPS.1
Macrophage_bone_marrow_6hr_LPS.1
Macrophage_bone_marrow_24h_LPS.1
Macrophage_peri_LPS_thio_0hrs.1
Macrophage_peri_LPS_thio_1hrs.1
Macrophage_peri_LPS_thio_7hrs.1
Granulocytes_mac1+gr1+.1
Osteoclasts.1
Osteoblast_day5.1
Osteoblast_day14.1
Osteoblast_day21.1
Microglia.1
Spleen.1
Lymph_nodes.1
Bone_marrow.1
Bone.1
Nih_3T3.1
C2C12.1
C3H_10T1_2.1
3T3-L1.1
MIMCD-3.1
Neuro2a.1
RAW_264_7.1
Min6.1
Liver.1
Kidney.1
Retinal_pigment_epithelium.1
Retina.1
Iris.1
Ciliary_bodies.1
Eyecup.1
Lens.1
Cornea.1
Cerebral_cortex_prefrontal.1
Cerebral_cortex.1
Amygdala.1
Hippocampus.1
Olfactory_bulb.1
Spinal_cord.1
Hypothalamus.1
Cerebellum.1
Nucleus_accumbens.1
Dorsal_striatum.1
Dorsal_root_ganglia.1
Testis.1
Median(4.6)
3xM
10xM
30xM

b.


0
400
800
1.20e+3
1.60e+3
2000
7.25
MEF.1
Mammary_gland__lact.1
Adipose_white.1
Lung.1
Placenta.1
Uterus.1
Ovary.1
Bladder.1
Umbilical_cord.1
Epidermis.1
Stomach.1
Intestine_large.1
Intestine_small.1
Salivary_gland.1
Lacrimal_gland.1
Prostate.1
Pancreas.1
Pituitary.1
Adrenal_gland.1
Skeletal_muscle.1
Mammary_gland_non-lactating.1
Adipose_brown.1
Heart.1
Embryonic_stem_line_V26_2_p16.1
Embryonic_stem_line_Bruce4_p13.1
Stem_cells__HSC.1
Common_myeloid_progenitor.1
Granulo_mono_progenitor.1
Mega_erythrocyte_progenitor.1
Thymocyte_DP_CD4+CD8+.1
Thymocyte_SP_CD4+.1
Thymocyte_SP_CD8+.1
T-cells_CD4+.1
T-cells_CD8+.1
T-cells_foxP3+.1
NK_cells.1
B-cells_GL7_negative_Alum.1
B-cells_GL7_positive_KLH.1
B-cells_GL7_positive_Alum.1
B-cells_GL7_negative_KLH.1
Follicular_B-cells.1
B-cells_marginal_zone.1
Baf3.1
Mast_cells.1
Mast_cells_IgE.1
Mast_cells_IgE+antigen_1hr.1
Mast_cells_IgE+antigen_6hr.1
Dendritic_cells_myeloid_CD8a-.1
Dendritic_cells_lymphoid_CD8a+.1
Dendritic_plasmacytoid_B220+.1
Macrophage_bone_marrow_0hr.1
Macrophage_bone_marrow_2hr_LPS.1
Macrophage_bone_marrow_6hr_LPS.1
Macrophage_bone_marrow_24h_LPS.1
Macrophage_peri_LPS_thio_0hrs.1
Macrophage_peri_LPS_thio_1hrs.1
Macrophage_peri_LPS_thio_7hrs.1
Granulocytes_mac1+gr1+.1
Osteoclasts.1
Osteoblast_day5.1
Osteoblast_day14.1
Osteoblast_day21.1
Microglia.1
Spleen.1
Lymph_nodes.1
Bone_marrow.1
Bone.1
Nih_3T3.1
C2C12.1
C3H_10T1_2.1
3T3-L1.1
MIMCD-3.1
Neuro2a.1
RAW_264_7.1
Min6.1
Liver.1
Kidney.1
Retinal_pigment_epithelium.1
Retina.1
Iris.1
Ciliary_bodies.1
Eyecup.1
Lens.1
Cornea.1
Cerebral_cortex_prefrontal.1
Cerebral_cortex.1
Amygdala.1
Hippocampus.1
Olfactory_bulb.1
Spinal_cord.1
Hypothalamus.1
Cerebellum.1
Nucleus_accumbens.1
Dorsal_striatum.1
Dorsal_root_ganglia.1
Testis.1
Median(8)
3xM
10xM
30xM

c.


0
250
500
750
800
72.44
MEF.1
Mammary_gland__lact.1
Adipose_white.1
Lung.1
Placenta.1
Uterus.1
Ovary.1
Bladder.1
Umbilical_cord.1
Epidermis.1
Stomach.1
Intestine_large.1
Intestine_small.1
Salivary_gland.1
Lacrimal_gland.1
Prostate.1
Pancreas.1
Pituitary.1
Adrenal_gland.1
Skeletal_muscle.1
Mammary_gland_non-lactating.1
Adipose_brown.1
Heart.1
Embryonic_stem_line_V26_2_p16.1
Embryonic_stem_line_Bruce4_p13.1
Stem_cells__HSC.1
Common_myeloid_progenitor.1
Granulo_mono_progenitor.1
Mega_erythrocyte_progenitor.1
Thymocyte_DP_CD4+CD8+.1
Thymocyte_SP_CD4+.1
Thymocyte_SP_CD8+.1
T-cells_CD4+.1
T-cells_CD8+.1
T-cells_foxP3+.1
NK_cells.1
B-cells_GL7_negative_Alum.1
B-cells_GL7_positive_KLH.1
B-cells_GL7_positive_Alum.1
B-cells_GL7_negative_KLH.1
Follicular_B-cells.1
B-cells_marginal_zone.1
Baf3.1
Mast_cells.1
Mast_cells_IgE.1
Mast_cells_IgE+antigen_1hr.1
Mast_cells_IgE+antigen_6hr.1
Dendritic_cells_myeloid_CD8a-.1
Dendritic_cells_lymphoid_CD8a+.1
Dendritic_plasmacytoid_B220+.1
Macrophage_bone_marrow_0hr.1
Macrophage_bone_marrow_2hr_LPS.1
Macrophage_bone_marrow_6hr_LPS.1
Macrophage_bone_marrow_24h_LPS.1
Macrophage_peri_LPS_thio_0hrs.1
Macrophage_peri_LPS_thio_1hrs.1
Macrophage_peri_LPS_thio_7hrs.1
Granulocytes_mac1+gr1+.1
Osteoclasts.1
Osteoblast_day5.1
Osteoblast_day14.1
Osteoblast_day21.1
Microglia.1
Spleen.1
Lymph_nodes.1
Bone_marrow.1
Bone.1
Nih_3T3.1
C2C12.1
C3H_10T1_2.1
3T3-L1.1
MIMCD-3.1
Neuro2a.1
RAW_264_7.1
Min6.1
Liver.1
Kidney.1
Retinal_pigment_epithelium.1
Retina.1
Iris.1
Ciliary_bodies.1
Eyecup.1
Lens.1
Cornea.1
Cerebral_cortex_prefrontal.1
Cerebral_cortex.1
Amygdala.1
Hippocampus.1
Olfactory_bulb.1
Spinal_cord.1
Hypothalamus.1
Cerebellum.1
Nucleus_accumbens.1
Dorsal_striatum.1
Dorsal_root_ganglia.1
Testis.1
Median(91.4)
3xM

d.

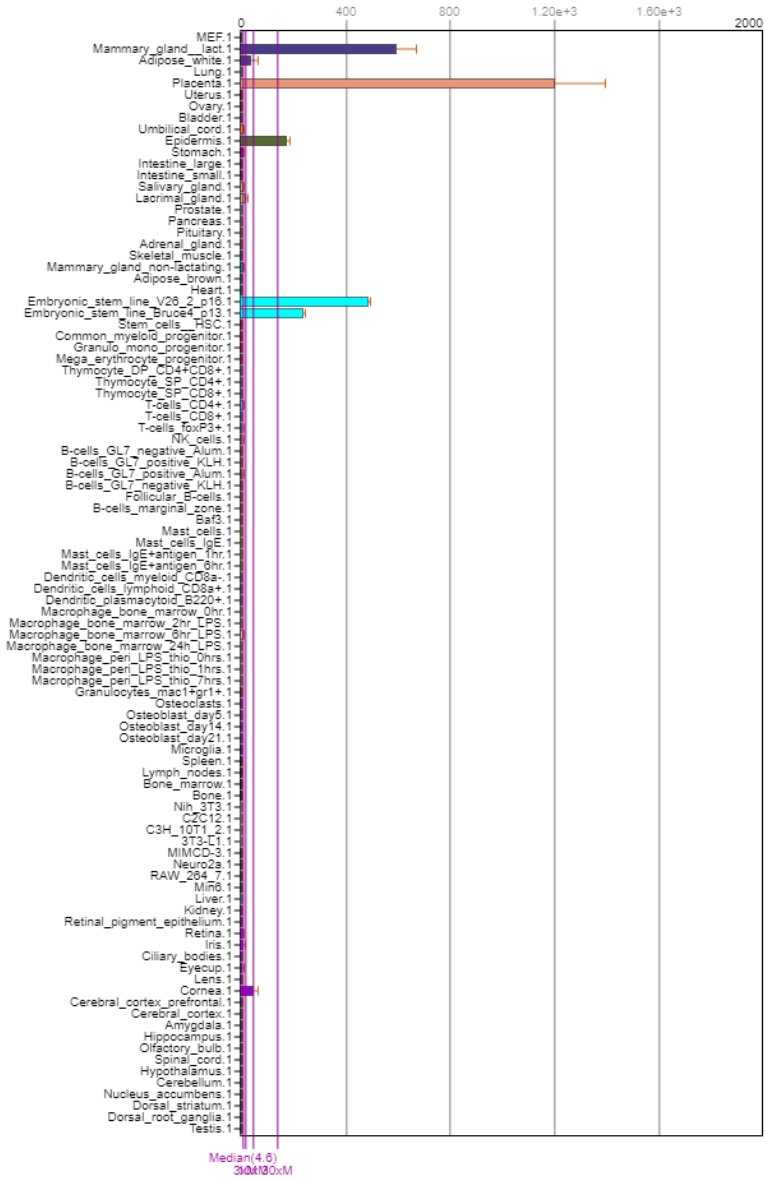


Figure 3.1. Gene expression levels of candidate TFs at transcription level based on BioGPS database. In the figure the expression level of gene of TFs is presented as bar graph. The y axis shows the list of different cell types. All expression data were obtained by using "GeneAtlas MOE430, gcrma" dataset. **a,** En1 **b,** Tfap2a, **c,** Bach1, **d,** Tfap2c

Table 3.1. *List of Modification Found in Transcription Factors as a Result of ChIP-Seq.*

Bivalent (H3K4me3 ve H3K27me3)	H3K4me3	H3K27me3	No Marker
Pdx1	Bach1	En1	Pou2f2
Prrx2	Mafk		Nobox
Tfap2c	Znf354c (Zfp354c)		Sry
Atoh1	Mzf1		Nhlh1
Bhlhe40	Ddit3		Ehf
Tfap2a	Stat6		Sox5
Foxd1	Hinfp		
Mafb	Stat1		
	Stat2		
	Sox3		
	Nkx3-1		
	Creb1		
	Sox2		
	Brca1		
	Myc		
	Arnt		

3.1.2 RNA Level

Chromatin immunoprecipitation and sequencing (ChIP-seq) experiment showed that level of H3K27Ac in arid4b gene deleted mESCs is higher than wildtype mESCs (Figure 1.2). Transcription factors whose consensus binding sequences are enriched in the H3K27Ac upregulated regions were selected through predictions of expression in mESCs (detailed in section 1.7).

To validate the gene expression in RNA level of candidate TFs in wildtype mESCs, we performed qRT-PCR (Figure 3.2). Nanog and Oct4 are key pluripotency regulators and they are highly expressed in mESCs. Therefore, they were used as the positive controls in this experiment. Sox1 is a pioneer transcription factor expressed in neuro-

ectoderm lineage, therefore it is included as a negative control for mESCs. Results of qRT-PCR was normalized to the expression level of β-actin. Expressions of Atoh1 and Mzf1 transcription factors were not observed in mESCs similar to the negative control. Pdx1, MafK, Stat6 were expressed at a low level, therefore, they are not analyzed further. Prrx2, Bach1, Znf354c, Ddit3, Tfap2c expression was observed at variable levels in mESCs (Figure 3.2).

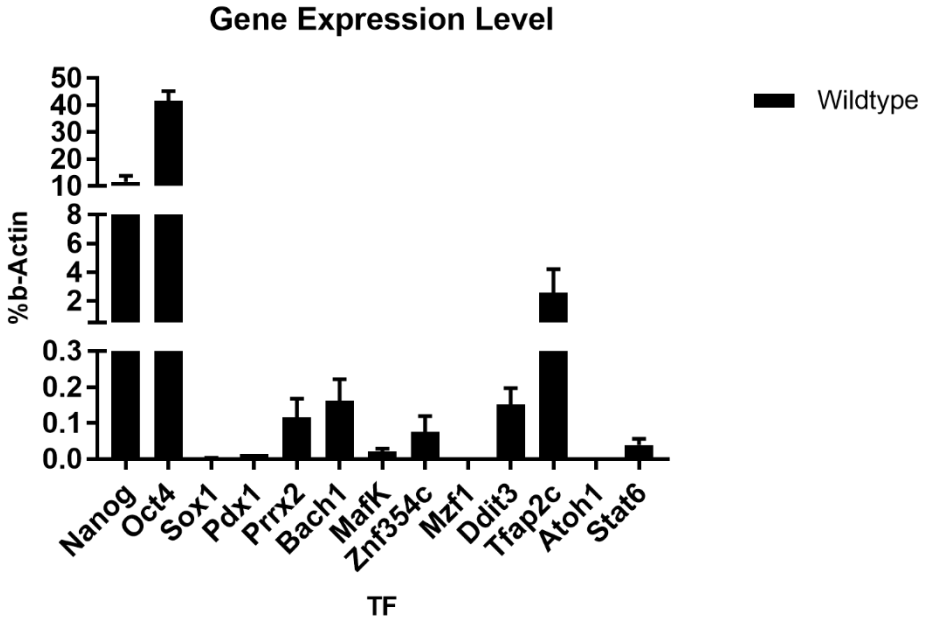


Figure 3.2. Gene Expression of Candidate Transcription Factors in Wildtype ESCs. Gene expression in mRNA level was measured by qPCR. Expression was normalized to b-actin expression level. Nanog and Oct4 are the pluripotent markers used as positive control. Sox1 is the earliest neuro-ectoderm marker used as negative control. CJ9 mouse embryonic stem cells is wildtype cells. Data are shown as mean ± SEM of three biological replicates. (TF, transcription factor).

Expression level of candidate transcription factors in arid4bΔ mESCs were tested by qRT-PCR in the same manner as wildtype cells. Similar to wildtype mESCs, expression of Pdx1, Mzf1 and Atoh1 transcription factors were not observed. MafK and Stat6 were found to be expressed at low level in arid4bΔ mESCs, therefore, they

were eliminated at this step. Consistent with WT mESCs, Prrx2, Bach1, Znf354c, Ddit3, Tfap2c were expressed in arid4bΔ mESCs (Figure 3.3). Therefore, Prrx2, Bach1, Znf354c, Ddit3, Tfap2c were selected for further experiments and tested for protein expression using western blot analysis.

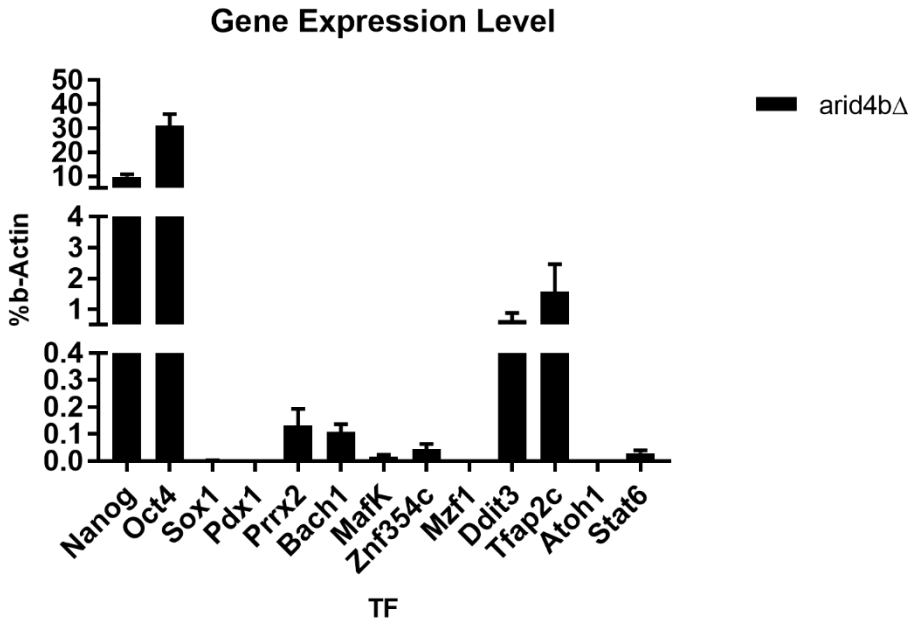


Figure 3.3. Gene Expression of Candidate Transcription Factors in arid4bΔ mESCs. Gene expression in transcription level was measured by qPCR. Expression was normalized to b-actin expression level. Nanog and Oct4 are the pluripotent markers used as positive control. Sox1 is the earliest neuro-ectoderm marker used as negative control. Data are shown as mean ± SEM of three biological replicates. (TF, transcription factor).

3.1.3 Protein Level

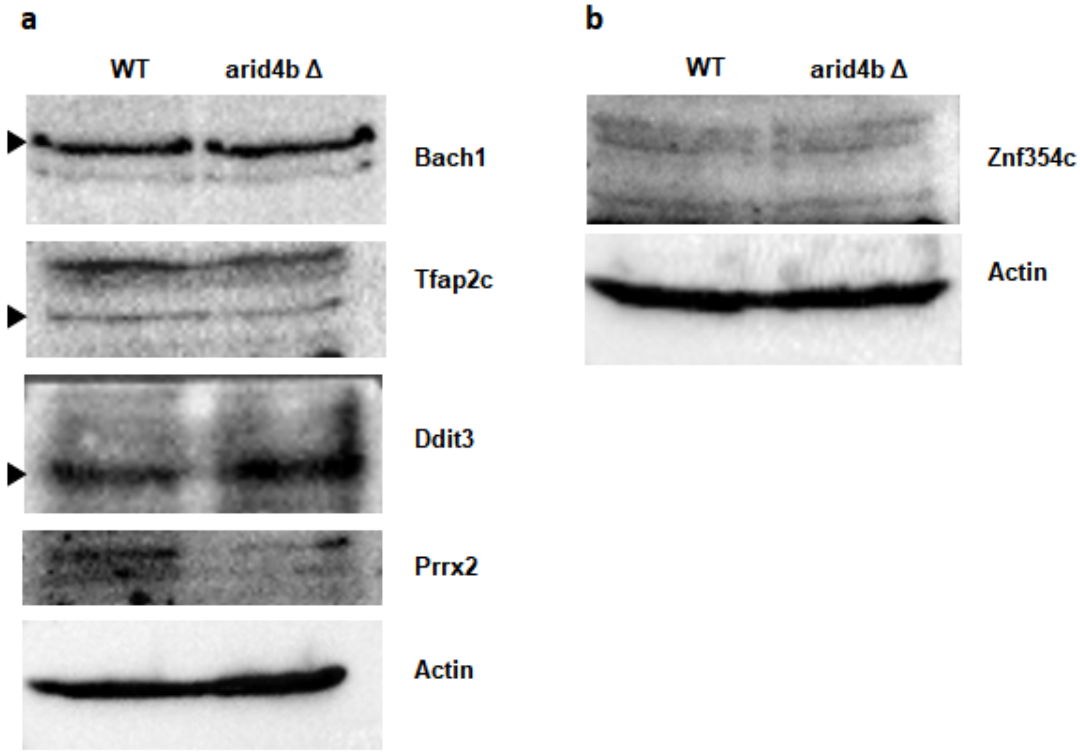


Figure 3.4. Gene Expression of Candidate Transcription Factors in Wildtype and Arid4bΔ mESCs. Gene expression in protein level was measured by western blot. Actin is loading control. Black arrowheads show the predicted size of target protein. a, b; The western blot against Bach1, Tfap2c, Ddit3, Prrx2 proteins, Znf354c proteins in wildtype and arid4bΔ mESCs, respectively. Three biological replicates were done for each candidate transcription factors.

Protein levels of selected transcription factors, Bach1, Tfap2c, Ddit3, Prrx2, Znf354c were checked in wildtype and arid4bΔ mESCs with western blot. B-actin was used as loading controls (Figure 3.4). Bach1 and Tfap2c were expressed in wildtype and arid4bΔ mESCs. As can be seen in Figure 3.4a, molecular weight of Bach1 (82-92 kDA) is consistent with the dark band labeled with black arrowhead. Bach1 protein is upper band labeled with black arrowhead. It is known that there are three different transcript variants, however only one of them is translated into protein (Uniprot). According to co-immunoprecipitation results, only upper band labelled with black arrowhead was observed in nuclear extract (Figure 3.5, fifth panel).

Three different bands were observed on the Tfap2c membrane (Figure 3.4a). There are three isoforms of Tfap2c protein (Uniprot). However, lower band of Tfap2c was observed in nuclear extract (Figure 3.5, fourth panel). Therefore, we predicted that the low MW variant might be the functional one.

Prrx2 protein was detected in both wildtype and knock-out mESCs. Prrx2 protein level was found to be lower in arid4bΔ cells. Ddit3 protein expression was also validated by western blot (Figure 3.4a).

Although we tried different concentrations of whole cell lysate, we always got weak signals for Znf354c. This might be because of low expression of Znf354c in the cells. Another reason might be low antigen-antibody binding affinity (Figure 3.4b).

Based on our presented data, Bach1, Tfap2c, Ddit3 and Prrx2 were chosen to be tested for physical interaction between Arid4b or/and Sin3a proteins and these TFs.

3.2. Determination of Physical Interaction Between Arid4b and Candidate TFs Using Arid4b as Bait

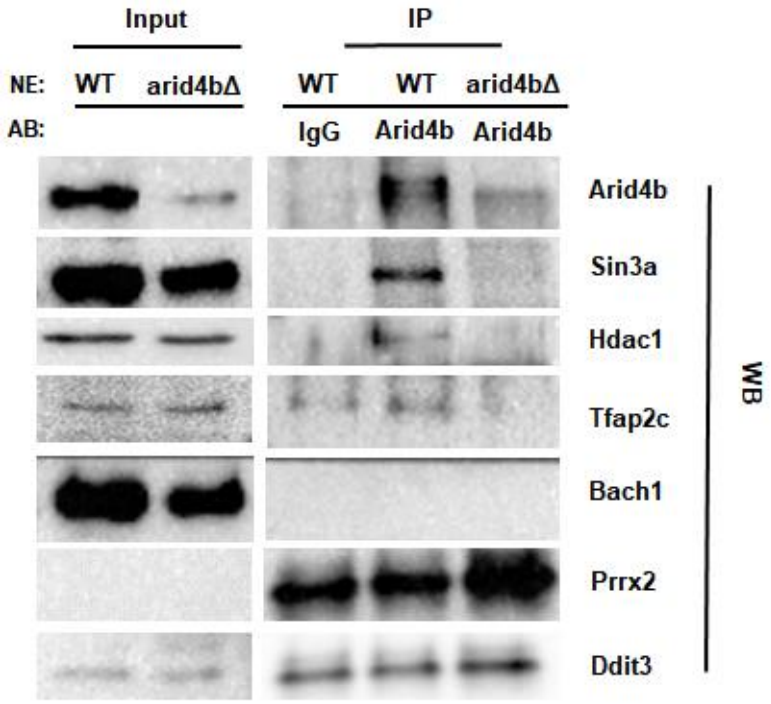


Figure 3.5. Arid4b and Normal Rabbit IgG Co-immunoprecipitation Results in Wildtype (WT) and arid4bΔ (KO) mESCs. 20 ug of input which is nuclear lysate from WT and KO mESCs shown on left side. 400 ug nuclear lysate was used for each immunoprecipitation. To visualized Tfap2c, Prrx2, Ddit3 proteins for input and Hdac1, Bach1, Prrx2, Ddit3 IP samples, Biorad Clarity Max ECL substrate was used. The others were visualized with Biorad Clarity ECL. Signals from inputs and IP samples are on a line at the same exposure time (IP, immunoprecipitation; WB, western blot; NE nuclear extract; AB, antibody used for immunoprecipitation)

Bach1, Tfap2c, Ddit3 and Prrx2 TFs were chosen to check whether there is physical interaction with Arid4b protein. The signal of antibody of Znf354c was weak therefore we eliminated it. Inputs show expression of interested TFs in the nuclear extracts from wildtype and arid4bΔ mESCs. The molecular weight of precipitated proteins is consistent with protein in inputs as can be seen in Figure 3.5.

Arid4b protein was successfully precipitated. Nuclear lysate from arid4bΔ mESC and rabbit IgG immunoprecipitation using wildtype nuclear extract were used as negative control. Low amount of Arid4b protein was precipitated in arid4b gene-deleted cells. This was expected, because arid4b genes were not able to be completely deleted by CRISPR-Cas9 technology. There was less expression of Arid4b in arid4bΔ mESC. It is predicted that transcription factors which have a specific physical interaction with Arid4b should be immunoprecipitated in low/no amount using arid4bΔ nuclear extract or via rabbit IgG using wildtype nuclear extract. As shown in Figure 3.5 upper panel, Arid4b immunoprecipitation was done successfully. The protein level of Arid4b was significantly lower in nuclear lysate of arid4b gene deleted cells than wildtype cells, as expected. The IgG lane is clear therefore it is concluded that Arid4b protein precipitated specifically.

As previously explained, Arid4b protein and Hdac1 are members of Sin3a corepressor complex. Therefore, the interaction of Arid4b protein with Sin3a and Hdac1 was also checked as a positive control. As shown in Figure 3.5, Sin3a and Hdac1 were precipitated with Arid4b protein (Second and third panel).

It was observed that Tfap2c protein was expressed and physically interacted with Arid4b protein in the nucleus. Tfap2c protein was also expressed in nucleus of arid4bΔ mESCs however, it was not precipitated with Arid4b. Additionally, it gave less signal with normal rabbit IgG antibodies compared to Arid4b antibody. These findings support a specific physical interaction between Tfap2c protein and Arid4b protein.

Bach1 transcription factor was present in nucleus, however, no interaction with Arid4b protein was detected. Prrx2 protein was detected in whole cell lysate of wildtype and arid4bΔ (Figure 3.4a). However, it was not detected in nuclear extract input. Prrx2 protein was precipitated with Arid4b in wildtype nuclear lysate but the equal amount of Prrx2 protein was precipitated with normal rabbit IgG. Besides, slightly more Prrx2 protein was precipitated with Arid4b in nuclear lysate of arid4bΔ mESCs. Therefore, we concluded that Prrx2 immunoprecipitation we observed is non-specific.

Lastly, Ddit3 transcription factor was observed in inputs and IP samples. However, equal amount of Ddit3 was precipitated with normal rabbit IgG and Arid4b in wildtype and arid4bΔ nuclear extract. Therefore, interaction between Arid4b and Ddit3 protein was not specific. As a result, Bach1, Prrx2 and Ddit3 transcription factors were eliminated. Tfap2c was selected for reciprocal immunoprecipitation.

3.3 Reciprocal Co-Immunoprecipitation

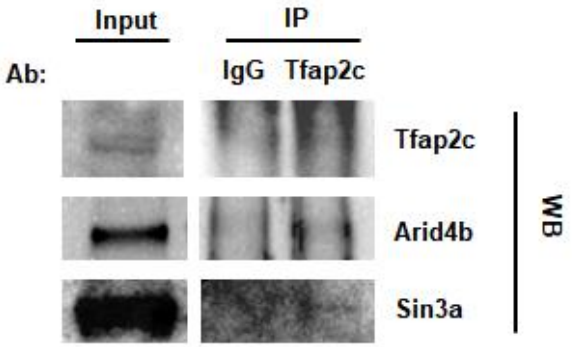


Figure 3.6. Tfap2c and normal mouse IgG co-immunoprecipitation results in wildtype (WT) mESCs. 20 ug of input which is nuclear lysate from WT mESCs shown on left side. Light exposure for input of western blot of Sin3a and Tfap2c proteins, dark exposure for IP samples were chosen. Signals from inputs and IP samples are on a line at the same exposure time (IP, immunoprecipitation; WB, western blot, Ab: antibodies for immunoprecipitation)

Reciprocal co-immunoprecipitation was performed to validate the physical interaction observed between Arid4b and Tfap2c proteins. Tfap2c and normal mouse IgG proteins were endogenously immunoprecipitated from wildtype mESC nuclear extract. To determine the success of immunoprecipitation, the presence of Tfap2c protein in nuclear lysate was checked. In upper panel of Figure 3.6, Tfap2c protein was precipitated in wildtype nuclear lysate and there was no interaction between IgG and Tfap2c proteins. This finding supports that Tfap2c immunoprecipitation is specific to

Tfap2c protein. It was shown that Arid4b protein is present in IP samples by western blot. Low amount of Arid4b protein was immunoprecipitated with Tfap2c protein. Similarly, results of western blot against to Sin3a showed that Sin3a protein was immunoprecipitated with Tfap2c. However, low amount of immunoprecipitated protein would indicate that the interaction between Arid4b and Tfap2c proteins is weak (substoichiometric) and transient. Consequently, we predict that there is a weak physical interaction between Tfap2c and Arid4b protein based on our results of Arid4b and Tfap2c immunoprecipitation experiments.

CHAPTER 4

DISCUSSION

Embryonic development is a well-coordinated process in terms of gene expression. Pluripotency state of embryonic stem cell is maintained by the key transcription factors (Oct4, Sox2, Nanog, Klf4, Myc etc.). These TFs establish transcription networks specific to ESCs and induce expression of genes which are important for pluripotency. Also, TFs works with chromatin factors to maintain pluripotency (Orkin & Hochedlinger, 2011).

We identified chromatin factors that have a role in differentiation of mESC towards mesoderm and endoderm. One of them is Arid4b protein which is a member of Sin3a corepressor complex. Arid4b has a chromodomain and a tudor domain which might recognize histone modifications. Besides, it contains an ARID region which might aid in non-sequence specific DNA binding. Arid4b protein is a member of the Sin3a corepressor complex along with Hdac1. It is thought that through weak interactions with chromatin, Arid4b targets Hdac1 to correct gene regions and enhances its activity there. To understand the role of Arid4b protein in differentiation, normal and arid4bΔ mESCs were differentiated into endoderm and mesoderm. In differentiated cells, genome wide distribution of H3K27Ac was examined by ChIP-Seq. It is known that H3K27Ac histone modification is present on active enhancers or promoter regions. We identified that the level of H3K27Ac was higher in arid4bΔ cells compared to normal cells (Figure 1.2). This might indicate that some genes might get aberrantly active in the absence of Arid4b. Therefore, Arid4b protein and the Sin3a corepressor complex may be required for repression of these genes for successful differentiation of mESCs towards endoderm or mesoderm.

We hypothesized that Arid4b may work with TFs to regulate differentiation process. To find candidate TFs, motif analysis was done using H3K27Ac ChIP-Seq peaks. We obtained the consensus binding sequences for TFs that were statistically enriched in regions where H3K27Ac signal was upregulated in arid4bΔ cells. There were 28 transcription factors, we reduced the number of TFs to 10 by using bioinformatic analysis (Table 1.1, Table 3.1, Figure 3.1). One of the candidate TFs was Tfap2c. In wildtype and arid4bΔ mESCs, the gene expression of Tfap2c at transcriptional level was checked by qRT-PCR (Figure 3.2, 3.3). Tfap2c expression was validated on the protein level. Similar expression level was observed in wildtype and arid4bΔ cells (Figure 3.4). We investigated whether Arid4b and Sin3a has a physical interaction with Tfap2c protein by using coimmunoprecipitation. The Arid4b was precipitated with Sin3a and Hdac1 which is a member of Sin3a together. However, Bach1 expression in nucleus in both WT and arid4bΔ mESC which was shown in input lane in Figure 3.5. However, Bach1 did not precipitated with Arid4b or IgG protein. We were reciprocally precipitated, Bach. Then, we checked co-precipitation of it with the Arid4b, however we did not observe Arid4b precipitation data not shown here. we concluded that Bach1 and Arid4b might not have direct physical interaction. Prrx2 signal was not detected in the nuclear extract which can be seen in input lane in Figure 3.5, we got signal for Prrx2 from immunoprecipitation. The reason might be that precipitation of a protein with antibody from nuclear extract leads to amplify signal for the protein. Therefore, we got higher signal for Prrx2 in immunoprecipitated samples although any signal for Prrx2 were not detected in inputs. It seemed Prrx2 was precipited with Arid4b in wildtype cells. However, almost equal amount of it was also precipitated with IgG. Thus, Prrx2 were nonspecifically precipitated with Arid4b. The molecular weight of Prrx2 is expected 27 kDa which is close to light chain of IgG (25 kDa). When we superposed the image for Prrx2 or Ddit3 (29 kDa) with the image of IgG light chain, the IgG light chain and Prrx2 band or Ddit3 band did not overlap. Also, we used to conformational specific antibody to get rid of light and heavy chain (50 kDa) of IgG. As a conclusion, the signal for Prrx2 and Ddit3 were not signal for IgG light chain. Ddit3 were present in both nucleus of wildtype and arid4bΔ mESC as

shown in Figure 3.5. However, Ddit3 were equally precipitated with IgG and Arid4b in nuclear extract in wildtype. Also, Ddit3 was precipitated with Arid4b in nuclear extract from arid4bΔ mESC, unexpectedly. Therefore, we concluded that Ddit3 did not precipitated with Arid4b, specifically. Overall, we showed the physical interaction between Arid4b and Tfap2c (Figure 3.5).

In addition, the results of endogenous reciprocal coimmunoprecipitation also confirmed the physical interaction between Tfap2c and Arid4b proteins. We detected a weak signal for immunoprecipitated Tfap2c (Figure 3.6). This might be due to a lower concentration of Tfap2c in the nucleus however we observed its expression in the nuclear extracts of wildtype and arid4bΔ. Therefore, this is slight possibility. The other reason might be due to the substoichiometric interaction between Arid4b and Tfap2c. Therefore, a low percentage of the total Arid4b protein pool in the cell might be interacting with Tfap2c at a given time. This result supports a transient, dynamic and weak physical interaction between these two proteins. Protein-protein crosslinking can be utilized in order to stabilize the interaction between Arid4b and Tfap2c.

Cell fate is determined by combinatorial actions of transcription factors. the combined function of Blimp1, Prdm14 and Tfap2c initiates primordial germ cell (PGC). PGCs are originated from pluripotent post-implantation epiblast cells and committed to somatic fate. Blimp1, Prdm14 and Tfap2c support the suppression of mesodermal genes in PGCs. It is suggested that *Tfap2c* is a direct target of Blimp1 and induced in P19EC cells, in vitro. P19EC cells are precursor of PGC and share many features of post implantation epiblast. The induction of *Tfap2c* is probably maintained by Prdm14 in PGCs (Magnúsdóttir et al., 2013). Additionally, it is suggested that Tfap2c acts downstream of Nanog in PGC specification, in vitro (Murakami et al., 2016). These findings support that Tfap2c may be important factors in cell fate determination.

Another study on human embryonic stem cells (hESC) suggests that Tfap2c helps the opening of enhancer of OCT4. In this way, it supports the maintenance of pluripotency

and suppresses neuroectoderm differentiation(Pastor et al., 2018). The neuroectoderm and meso-endoderm germ layer are different from each other. The repression of genes related to neuroectoderm may leads to activation of the genes related to meso-endoderm. Although pluripotency stages of mESC are different from hESC, we suggested that Arid4b may suppress the genes by help of physical interaction with Tfap2c.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

In this study, the level of H3K27Ac modification on promoters and enhancers of arid4bΔ cells committed into mesoderm or endoderm was found higher compared to wild type cells (Figure 1.2). TFs whose consensus binding sequence was enriched in the regions where H3K27Ac was high were identified by HOMER Analysis (Table 1.1). One of the candidate TFs was Tfap2c. Its expression at transcription level was confirmed in wildtype and arid4bΔ mESC by qRT-PCR (Figure 3.2 and 3.3) Additionally, we showed that Tfap2c protein was present in both wildtype cells and arid4bΔ mESC (Figure 3.4). We demonstrated that Arid4b protein is physically interacted with Tfap2c (Figure 3.5). Reciprocal immunoprecipitation was confirmed this physical interaction (Figure 3.6). We concluded that there was physical interaction between Tfap2c and Arid4b protein This interaction was transient, dynamics or weak known as substoichiometric interactions. Further studies written below might be performed to support our data.

Detection of weak interaction by co-immunoprecipitation was difficult. The interaction between arid4b and Tfap2c was weak therefore we increased the amount of nuclear extract to get consistent results from coimmunoprecipitation experiments. Therefore, we grew huge number of cells. This difficulty may overcome with a method known as cross-linking method to enhance the strength of the interaction. UV causes crosslinking and the protein-protein interaction becomes stable(Yakovlev, 2009). The interaction between Tfap2c and Arid4b may be showed by crosslinking method as further study.

As a next step, the role of Tfap2c in ESC differentiation should be further investigated. Another approach might be knocking down of Tfap2c. In this method, we can check

the effect of Tfap2c protein in differentiation. Additionally, *Tfap2c* knock-out mESC and *Arid4b* knock-out (KO) mESC would be induced to differentiate into endoderm or mesoderm and then the phenotype caused by the absence of Tfap2c and Arid4b in KO cell may be studied. For instance, the genes whose expression changed might be analyzed and the overlapped genes whose expression changed in *Tfap2c* KO and *Arid4b* KO cells might be identified to find out any correlation between two genotypes. Also, Tfap2c ChIP-seq could be performed to find out Tfap2c target genes whose have Tfap2c binding sequence. Then, it might be check that whether the regions where Tfap2c binds correlate with the regions where the level of H3K27Ac is higher in arid4bΔ cells. To identify the other protein that physically interact with Tfap2c protein, proteomic approach like mass spectrometry might be used. The other proteins which may physically interact with Arid4b and Tfap2c give an idea about the role of Tfap2c in differentiation. Further approach might be co-immunoprecipitation of Hdac1 and Sin3a protein in wild type mESC to check the protein-protein interaction with Tfap2c

REFERENCES

Adıyaman, Ö. And Chorowicz, J. 2002. Late Cenozoic tectonics and volcanics in the northwestern corner of the Arabian plate: a consequence of the strike-slip Dead Sea Fault Zone and lateral escape of Anatolia, Journal of Volcanology and Geothermal Research, 117. 327-345.

Allmendinger, R. W.1999. Introduction to Structural Geology. Lecture notes. 279 p.

Allmendinger, R. W., Cardozo, N. C., and Fisher, D., 2013. Structural Geology Algorithms: Vectors & Tensors: Cambridge, England, Cambridge University Press, 289 p.

Alpar, B. and Yaltırak, C. 2002. Characteristic features of the North Anatolian Fault in the eastern Marmara region and its tectonic evolution. Marine Geology, 190, 329-350.

Alpaslan, M. 2007. Early to Middle Miocene Intracontinental basaltic volcanism in the northern margin of the Arabian Plate, SE Anatolia, Turkey: geochemistry and petrogenesis. Geological Magazine. 144 (5), 867-882.

Amrouch, K., Lacombe, O., Bellahsen, N., Daniel, J. M. and Callot, J. P. 2010.Stress and strain patterns, kinematics and deformation mechanisms in a basement cored anticline: Sheep Mountain Anticline, Wyoming. Tectonics, 29, TC1005, doi:10.1029/2009TC002525.

Angelier, J. (1990), Inversion of field data in fault tectonics to obtain the regional stress. III: A new rapid direct inversion method by analytical means, Geophys. J. Int., 103, 363-376, doi:10.1111/ j.1365-246X.1990.tb01777.x.

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., & Bazett-Jones, D. P. (2010). Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. *PLoS ONE*, *5*(5), e10531. https://doi.org/10.1371/journal.pone.0010531

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

Bernhart, S. H., Kretzmer, H., Holdt, L. M., Jühling, F., Ammerpohl, O., Bergmann, A. K., … Hoffmann, S. (2016). Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Scientific Reports*, *6*. https://doi.org/10.1038/srep37393

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., … Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in

murine embryonic stem cells. *Nature*, *441*(7091), 349–353. https://doi.org/10.1038/nature04733

Cao, S., Bendall, H., Hicks, G. G., Nashabi, A., Sakano, H., Shinkai, Y., … Ruley, H. E. (2003). The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. *Molecular and Cellular Biology*, *23*(15), 5301–5307. https://doi.org/10.1128/mcb.23.15.5301-5307.2003

Chandru, A., Bate, N., Vuister, G. W., & Cowley, S. M. (2018). Sin3A recruits Tet1 to the PAH1 domain via a highly conserved Sin3-Interaction Domain. *Scientific Reports*, *8*(1), 14689. https://doi.org/10.1038/s41598-018-32942-w

Chen, Ying, Wang, H., Yoon, S. O., Xu, X., Hottiger, M. O., Svaren, J., … Lu, Q. R. (2011). HDAC-mediated deacetylation of NF-κB is critical for Schwann cell myelination. *Nature Neuroscience*, *14*(4), 437–441. https://doi.org/10.1038/nn.2780

Chen, Yue, Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., … Zhao, Y. (2007). Lysine Propionylation and Butyrylation Are Novel Post-translational Modifications in Histones. *Molecular & Cellular Proteomics*, *6*(5), 812–819. https://doi.org/10.1074/mcp.M700021-MCP200

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., … Mann, M. (2009). Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science*, *325*(5942), 834–840. https://doi.org/10.1126/science.1175371

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Dovey, O. M., Foster, C. T., & Cowley, S. M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proceedings of the National Academy of Sciences*, *107*(18), 8242–8247. https://doi.org/10.1073/pnas.1000478107

Dzierzak, E. (2005). The emergence of definitive hematopoietic stem cells in the mammal. *Current Opinion in Hematology*, *12*(3), 197–202. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15867575

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., … Meshorer, E. (2008). Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*, *2*(5), 437–447. https://doi.org/10.1016/J.STEM.2008.03.021

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., … Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early

embryogenesis. *Nature Cell Biology*, *8*(2), 188–194. https://doi.org/10.1038/ncb1353

Fleischer, T. C., Yun, U. J., & Ayer, D. E. (2003). Identification and Characterization of Three New Components of the mSin3A Corepressor Complex Identification and Characterization of Three New Components of the mSin3A Corepressor Complex, *23*(10), 3456–3467. https://doi.org/10.1128/MCB.23.10.3456

Gaspar-Maia, A., Alajem, A., Meshorer, E., & Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3036

Grzenda, A., Lomberk, G., Zhang, J.-S., & Urrutia, R. (2009). Sin3: Master scaffold and transcriptional corepressor. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1789*(6–8), 443–450. https://doi.org/10.1016/J.BBAGRM.2009.05.007

Harvey, S. H., Krien, M. J. E., & O'Connell, M. J. (2002). Structural maintenance of chromosomes (SMC) proteins, a family of conserved ATPases. *Genome Biology*, *3*(2), REVIEWS3003. https://doi.org/10.1186/gb-2002-3-2-reviews3003

Heng, J.-C. D., & Ng, H.-H. (2010). Transcriptional Regulation in Embryonic Stem Cells. In E. Meshorer & K. Plath (Eds.), *The Cell Biology of Stem Cells* (pp. 76–91). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-7037-4_6

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., … Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature*, *442*(7102), 533–538. https://doi.org/10.1038/nature04915

Kadamb, R., Mittal, S., Bansal, N., Batra, H., & Saluja, D. (2013). Sin3: Insight into its transcription regulatory functions. *European Journal of Cell Biology*, *92*(8–9), 237–246. https://doi.org/10.1016/j.ejcb.2013.09.001

Kelly, R. D. W., & Cowley, S. M. (2013). The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochemical Society Transactions*, *41*(3), 741–749. https://doi.org/10.1042/BST20130010

Kornberg, R. D. (2007). The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences*, *104*(32), 12955–12961. https://doi.org/10.1073/pnas.0704138104

Krejčí, J., Uhlířová, R., Galiová, G., Kozubek, S., Šmigová, J., & Bártová, E. (2009). Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. *Journal of Cellular Physiology*, *219*(3), 677–687. https://doi.org/10.1002/jcp.21714

Kriegstein, A., & Alvarez-Buylla, A. (2009). The Glial Nature of Embryonic and Adult Neural Stem Cells. *Annual Review of Neuroscience*, *32*(1), 149–184.

https://doi.org/10.1146/annurev.neuro.051508.135600

Krishnamurthy, S., & Hampsey, M. (2009). Eukaryotic transcription initiation. *Current Biology*, *19*(4), R153–R156. https://doi.org/10.1016/j.cub.2008.11.052

Lagger, G., O'Carroll, D., Rembold, M., Khier, H., Tischler, J., Weitzer, G., … Seiser, C. (2002). Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *The EMBO Journal*, *21*(11), 2672–2681. https://doi.org/10.1093/emboj/21.11.2672

LeBoeuf, M., Terrell, A., Trivedi, S., Sinha, S., Epstein, J. A., Olson, E. N., … Millar, S. E. (2010). Hdac1 and Hdac2 Act Redundantly to Control p63 and p53 Functions in Epidermal Progenitor Cells. *Developmental Cell*, *19*(6), 807–818. https://doi.org/10.1016/j.devcel.2010.10.015

Lee, K. K., & Workman, J. L. (2007). Histone acetyltransferase complexes: One size doesn't fit all. *Nature Reviews Molecular Cell Biology*, *8*(4), 284–295. https://doi.org/10.1038/nrm2145

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., … Young, R. A. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell*, *125*(2), 301–313. https://doi.org/10.1016/j.cell.2006.02.043

Lessard, J. A., & Crabtree, G. R. (2010). Chromatin Regulatory Mechanisms in Pluripotency. *Annual Review of Cell and Developmental Biology*, *26*(1), 503–532. https://doi.org/10.1146/annurev-cellbio-051809-102012

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

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S. Y., … Songyang, Z. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nature Cell Biology*, *10*(6), 731–739. https://doi.org/10.1038/ncb1736

Lin, C., Song, W., Bi, X., Zhao, J., Huang, Z., Li, Z., … Zhao, H. (2014). Recent advances in the ARID family: focusing on roles in human cancer. *OncoTargets and Therapy*, *7*, 315–324. https://doi.org/10.2147/OTT.S57023

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., … Surani, M. A. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature Cell Biology*, *15*(8), 905–915. https://doi.org/10.1038/ncb2798

Maniatis, T., & Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature*, *416*(6880), 499–506. https://doi.org/10.1038/416499a

Martínez-Balbás, M. A., Bauer, U.-M., Nielsen, S. J., Brehm, A., & Kouzarides, T. (2000). Regulation of E2F1 activity by acetylation. *The EMBO Journal*, *19*(4), 662–671. https://doi.org/10.1093/emboj/19.4.662

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., … Niwa, H. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology*, *9*(6), 625–635. https://doi.org/10.1038/ncb1589

Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T., & Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell*, *10*(1), 105–116. https://doi.org/10.1016/j.devcel.2005.10.017

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., … Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, *113*(5), 631–642. https://doi.org/10.1016/s0092-8674(03)00393-3

Mullin, N. P., Yates, A., Rowe, A. J., Nijmeijer, B., Colby, D., Barlow, P. N., … Chambers, I. (2008). The pluripotency rheostat Nanog functions as a dimer. *Biochemical Journal*, *411*(2), 227–231. https://doi.org/10.1042/BJ20080134

Murakami, K., Günesdogan, U., Zylicz, J. J., Tang, W. W. C., Sengupta, R., Kobayashi, T., … Surani, M. A. (2016). NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature*, *529*(7586), 403–407. https://doi.org/10.1038/nature16480

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J. G., Mati, I. K., Eijk, P. P., … Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature Structural & Molecular Biology*, *20*(3), 387. https://doi.org/10.1038/NSMB.2509

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., … Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, *95*(3), 379–391. https://doi.org/10.1016/s0092-8674(00)81769-9

Niwa, H., Miyazaki, J., & Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, *24*(4), 372–376. https://doi.org/10.1038/74199

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

Orphanides, G., LeRoy, G., Chang, C.-H., Luse, D. S., & Reinberg, D. (1998). FACT, a Factor that Facilitates Transcript Elongation through Nucleosomes. *Cell*, *92*(1), 105–116. https://doi.org/10.1016/S0092-8674(00)80903-4

Pastor, W. A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T. J., … Clark, A. T. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. Nature Cell Biology, 20(May). https://doi.org/10.1038/s41556-018-0089-0

Pesce, M., & Schöler, H. R. (2004). Oct-4 : Gatekeeper in the Beginnings of Mammalian Development . *Stem Cells*, *19*(4), 271–278. https://doi.org/10.1634/stemcells.19-4-271

Rosonina, E., Kaneko, S., & Manley, J. L. (2006). Terminating the transcript: breaking up is hard to do. *Genes & Development*, *20*(9), 1050–1056. https://doi.org/10.1101/gad.1431606

Sarma, K., & Reinberg, D. (2005). Histone variants meet their match. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm1567

Saunders, A., Huang, X., Fidalgo, M., Reimer, M. H., Faiola, F., Ding, J., … Wang, J. (2017). The SIN3A/HDAC Corepressor Complex Functionally Cooperates with NANOG to Promote Pluripotency. *Cell Reports*, *18*(7), 1713–1726. https://doi.org/10.1016/j.celrep.2017.01.055

Shao, Z., Zhang, Y., Yuan, G.-C., Orkin, S. H., & Waxman, D. J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biology*, *13*(3), R16. https://doi.org/10.1186/gb-2012-13-3-r16

Sif, S., Saurin, A. J., Imbalzano, A. N., & Kingston, R. E. (2001). Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes & Development*, *15*(5), 603–618. https://doi.org/10.1101/gad.872801

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

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

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., … Zhao, Y. (2011). Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell*, *146*(6), 1016–1028. https://doi.org/10.1016/j.cell.2011.08.008

Tan, M., Peng, C., Anderson, K. A., Chhoy, P., Xie, Z., Dai, L., … Zhao, Y. (2014). Lysine Glutarylation Is a Protein Posttranslational Modification Regulated by SIRT5. *Cell Metabolism*, *19*(4), 605–617.

https://doi.org/10.1016/j.cmet.2014.03.014

Tan, W.-H., Gilmore, E. C., & Baris, H. N. (2013). Human Developmental Genetics. *Emery and Rimoin's Principles and Practice of Medical Genetics*, 1–63. https://doi.org/10.1016/B978-0-12-383834-6.00018-5

Trivedi, C. M., Zhu, W., Wang, Q., Jia, C., Kee, H. J., Li, L., … Epstein, J. A. (2010). Hopx and Hdac2 Interact to Modulate Gata4 Acetylation and Embryonic Cardiac Myocyte Proliferation. *Developmental Cell*, *19*(3), 450–459. https://doi.org/10.1016/j.devcel.2010.08.012

Vannini, A., & Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular Cell*, *45*(4), 439–446. https://doi.org/10.1016/j.molcel.2012.01.023

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J., & Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, *473*(7347), 343–348. https://doi.org/10.1038/nature10066

Winter, S. F., Lukes, L., Walker, R. C., Welch, D. R., & Hunter, K. W. (2012). Allelic variation and differential expression of the mSIN3A histone deacetylase complex gene Arid4b promote mammary tumor growth and metastasis. *PLoS Genetics*, *8*(5), e1002735. https://doi.org/10.1371/journal.pgen.1002735

Yakovlev, A. A. (2009). Crosslinkers and their utilization for studies of intermolecular interactions. Neurochemical Journal, 3(2), 139–144. https://doi.org/10.1134/S181971240902010X

Yang, X., Zhang, F., & Kudlow, J. E. (2002). Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell*, *110*(1), 69–80. https://doi.org/10.1016/s0092-8674(02)00810-3

Zhou, Q., Li, T., & Price, D. H. (2012). RNA Polymerase II Elongation Control. *Annual Review of Biochemistry*, *81*(1), 119–143. https://doi.org/10.1146/annurev-biochem-052610-095910

APPENDICES

A. Medium Recipes Used in Cell Culture

Table A.1. *The recipe of Embryonic stem cell medium (ESC).*

Reagent	***Brand/ Cat. No***	***Final concentration/ percentage***	***Volume for needed for 100 ml medium***
DMEM	Gibco / 41966029	80%	100 ml
FBS	Gibco/10500064	15 %	19 ml
Nucleoside Mix*	Written in Table 0.2	1 %	1.25 ml
GLUTAMAX	Thermo Fisher / 35050038	1 %	1.25 ml
PEN/STREP	Thermo Fisher / 15140122	2 %	2.5 ml
BME	Sigma / M6250-100ML	10^{-4} M	0.9 ul
NEAA	Thermo Fisher / 11140035	1 %	1.25 ml
Supplement for 50 ml			
LIF	Millipore / ESG1107	10^3 units/ml	5 µl

Table A.2. *The recipe of nucleoside mix.*

Reagent	*Brand/ Cat. No.*	*Amount (mg)*
Adenosine	Sigma / A4036-5G	85
Guanosine	Sigma / G6264-1G	85
Uridine	Sigma / U3003-5G	73
Cytidine	Sigma / C4654-1G	73
Thymidine	Sigma / T1895-1G	24

Table A.3. *The recipe of 2i4 medium and its supplement.*

Reagent	***Brand/ Cat. No***	***Final concentration/ percentage***	***Volume for needed for 100 ml medium***
Neurobasal	Thermo Fisher / 21103049	50%	50 ml
DMEM/F12	Thermo Fisher / 11320074	50%	50 ml
N2	Thermo Fisher / 17502048	0.5%	500 µl
B27 with Vit A	Thermo Fisher / 17504044	1%	1 ml
10 %BSA (in PBS)	Sigma / A3311-50G	0.5%	500 µl
GLUTAMAX	Thermo Fisher / 35050038	1%	1 ml
PEN/STREP	Thermo Fisher / 15140122	1%	1 ml
MTG	Sigma / M6145-25ML	1.5 x 10-4 M	1.3 µl
FBS	Gibco/10500064	4%	4 ml
Supplement for 50 ml			
LIF	Millipore / ESG1107	10^5 units/ml	5 µl
CHIR-99021	S1263-5MG	3 µM	15 µl
PD0325901	Selleckchem / S1036-5MG	1 µM	5 µl

B. The Information of Primers That Is Used In qRT-PCR

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

Primer Name	***Sequence (5'-> 3')***	***PMID***
B-actin-F-qPCR	ATGAAGATCCTGACCGAGCG	14998924
B-actin-R-qPCR	TACTTGCGCTCAGGAGGAGC	14998924
Pdx1-F-qPCR	CAACATCACTGCCAGCTCCACC	27715254
Pdx1-R-qPCR	TCACCTCCACCACCACCTTCCA	27715254
Prrx2-F-qPCR	AGTGAGGCACGTGTCCAAGTC	24770895
Prrx2-R-qPCR	GTAGCCAGCATGGCACGTT	24770895
Bach1-F-qPCR	TGAGTGAGAGTGCGGTATTTGC	26084661
Bach1-R-qPCR	GTCAGTCTGGCCTACGATTCT	26084661
MafK-F-qPCR	GACAGGGCCCGGGTTATG	10409670
MafK-R-qPCR	AGCTCATCATCGCTAAGAACAGG	10409670
Znf354c-F-qPCR	CCGGCGTCCGCATATTT	27723809
Znf354c-R-qPCR	CCCTTCTTAGTTTTTCTGCCAAAG	27723809
Mzf1-F-qPCR	GAGGCTGCTGCCCTAGTAGA	22246292
Mzf1-R-qPCR	GAGGGCTCCATCTTCTCTGA	22246292
Ddit3-F-qPCR	GTCAGTTATCTTGAGCCTAACACG	26634309
Ddit3-R-qPCR	TGTGGTGGTGTATGAAGATGC	26634309
Tfap2c-F-qCPR	GGGCTTTTCTCTCTTGGCTGGT	23913270
Tfap2c-R-qCPR	TCCACACGTCACCCACACAA	23913270
Atoh1-F-qPCR	ATGCACGGGCTGAACCA	26786414
Atoh1-R-qPCR	TCGTTGTTGAAGGACGGGATA	26786414
Stat6-F-qPCR	GCATCTATCAGAGGGACCCC	26899911
Stat6-R-qPCR	ACTTGTCCAGTCTTAGGCCC	26899911

C. The Information of Antibodies Used in Western Blot and co-IP

Table C.1. *The information of antibody used in western blot in Figure 3.4.*

Protein Name	***Brand/ Cat. No***	***Host***	***Reactivity***	***Dilution***
Bach1	Abcam/ ab124919	rabbit	mouse, human	1:1000
Znf354c	LifeSpan/LS-C145172-100	rabbit	mouse	1:500
Tfap2c	Sigma/sab2102408	rabbit mouse	horse, human, mouse, sheep, rabbit, dog, bovine, guinea pig, rat	1:1000
Ddit3	Santa Cruz/sc-7351		mouse, rat, human	1:1000
Prrx2	Abcam/ab156096	rabbit	mouse	1:100
Actin	Millipore/mab1501	mouse	all	1:3000
Anti-rabbit IgG H&L (HRP)	Abcam/ab97051	goat	rabbit	1:5000
Anti-mouse IgG H&L (HRP	Abcam/ab97023	goat	mouse	1:5000

Table C.2. *The information of antibody used in western blot of direct (Arid4b-IP) and reciprocal co-IP (Tfap2c-IP). ** Figure 3.5 and 3.6*

Protein Name	***Host***	***Primary ab dilution / Cat. No.***	***Secondary ab dilution / Cat. No.***	***Tertiary ab dilution / Cat. No.***
Arid4b	rabbit	1:1000 / A302-233A	1:5000/ ab97051	
Sin3a	rabbit	1:1000 / 39866	1:5000/ ab97051	
Hdac1	rabbit	1:1000 / 06-720	1:5000/ ab97051	
Tfap2c	rabbit	1:1000 / ab2102408	1:2000 / 3678S	1:4000 / R05071-500
Bach1	rabbit	1:1000 / ab124919	1:5000 / ab97051	
Prrx2	rabbit	1:100 / ab156096	1:2000 / 3678S	1:4000 / R05071-500
Ddit3	mouse	1:1000 / sc-7351	1:5000 / ab97023	

Table C.3. *The information of antibody used in direct (Arid4b-IP) and reciprocal co-IP (Tfap2c-IP)*

Protein Name	***Host***	***Cat. No.***	***Stock concentration/ Cat. No.***	***Amount of Antibody used in IP***
Arid4b	rabbit	A302-233A	1 mg/ml	5 µg / 400 µg
Normal IgG	rabbit	12-370	1 mg/ml	5 µg / 400 µg
Normal IgG	mouse	sc-2025	200 µg /0.5 ml	2 µg / 350 µg
Tfap2c	mouse	sc-12762	200 µg / ml	2 µg / 350 µg