

THE MECHANISMS GOVERNING SELF-RENEWAL AND
DIFFERENTIATION IN PLURIPOTENCY

By Mohammad (Ishmam) Alam, B. Sc. Hons

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment
of the Requirements for the Degree Master of Science

McMaster University

© Copyright by Mohammad (Ishmam) Alam, June, 2019

Descriptive Note

McMaster University Master of Science (2019), Hamilton, Ontario

TITLE: THE ROLE AND MOLECULAR MECHANISMS OF PLURIPOTENT STEM CELLS

AUTHOR: Mohammad (Ishmam) Alam, B. Sc. Hons (McMaster University)

SUPERVISOR: Dr. Jonathan S. Draper

NUMBER OF PAGES: ix, 72

Acknowledgement

The last two years have been a rollercoaster of emotions and struggle, more than what I thought to experience; the time has been a fantastic learning opportunity both from a scientific perspective and a self-directed one. Certainly, the experience will have a long-term impact on the way I think and the work that I accomplish in my entrepreneurial journey.

I would like to express my sincere gratitude to my Master's supervisor Dr. Jonathan Draper for the endless support and even though I have at many times been difficult to work with because of my solitude and sometimes a disappointment, I would like to say that I sincerely appreciate the commitment he has to his students, even through his own life's turmoil, and that commitment will never be forgotten. The opportunity to join his lab at a time of distress for me was a miracle and because of him I was provided some great friendships, intensive learning opportunities, and, most importantly, personal growth. I have developed extensively in my own independent critical thinking, learned how to learn, especially within science, and triumphed in building self-discipline. The experience will be put to excellent use and has already drastically improved my level of thinking and progress in independent entrepreneurial projects.

I would also like to thank my committee members, Dr. Bradley Doble and Dr. Kristin Hope, who have been very supportive and encouraging throughout my studies. Dr. Kristin Hope has further supported me in Jon's absence and strived to make my unique experience in the lab more comforting. I would like to thank Kristin for her sincere efforts and initiative to provide assistance, without which I would have been lost at several points in time.

I truly appreciated all the help and supports from my lab colleagues, especially Ava Keyvani Chahi, who has helped me extensively, not only in the technical aspects but also in providing

great friendship in a time where my lab was gone. In addition, I would like to thank Daisy Deng for all her support and friendship through my Master's, without whom I would be lost in my projects. Furthermore, I thank my lab members, Rohan Nadkarni, who has been a great friend and colleague throughout my years in SCC-RI, and Amos Lim, who I have struggled side by side with through our Master's journey and have shared great times and adventures; I will always think of Amos when rock climbing.

Finally, I would like to thank my family, especially my mother, without whom my journey would be grim; thank you for all the emotional support, encouragement, and for being understanding of my ambitions, you truly are an amazing human being.

Thank you also to all of SCC-RI and McMaster Biochemistry, who have provided professional support and great conversations throughout the years, all of which has made my time more enjoyable at McMaster University.

Table of Contents

Descriptive Note	i
Acknowledgement	ii
Table of Contents	iv
List of Figures and Tables	vi
Part 1	vi
Part 2	vi
List of Abbreviations and Symbols	viii
CHAPTER 1: Investigating the Molecular Mechanisms of REX1 in Pluripotency Maintenance	1
1.1 Abstract	1
1.2 Preface	2
1.3 Introduction	3
1.3.1 Introduction to pluripotency	3
1.3.2 Heterogeneity in Pluripotent Stem Cell Populations	3
1.3.3 Core Transcriptional Network of Pluripotent Stem Cells	4
1.3.4 Extending the Transcriptional Network in Pluripotent Stem Cells	5
1.3.5 REX1 Expression in Pluripotent Stem Cells	6
1.3.6 Regulation of Endogenous Retroviral Elements	6
CHAPTER 1.4: Characterizing REX1 Isoforms and YY1/YY2 Independent Role of REX1	9
1.4.1 Summary of Intent	9
1.4.1.1 Identification of Novel REX1 Isoforms in PSCs	10
1.4.2 Material & Methods	11
1.4.3 Results & Discussion	14
1.4.3.1 Generation of REX1 Knockout Lines does not affect pluripotency related protein expression levels	14
1.4.3.2 REX1 Knockout possibly effects expression of Endogenous Retroviral Elements	15
1.4.3.3 REX1 isoforms display differential expression under different pluripotent sub-states	15
1.4.3.4 Transfection with full-length REX1 does not affect REX1 target expression in REX1-KO cells	16
1.4.4 Conclusion	17
1.5 Figures	19
1.6 Tables	29
CHAPTER 2: Characterizing Putative Mitotic Bookmarking Factors in Pluripotency Maintenance	30

2.1 Abstract.....	30
2.2 Preface	31
2.3 Introduction	32
2.3.1 Cell Cycle	32
2.3.2 Pluripotent Cell Cycle.....	32
2.3.3 The mitotic chromatin.....	33
2.3.4 Mitotic bookmarking.....	34
2.3.5 ChIP-MS reveals putative pluripotency associated mitotic bookmarks	36
2.3.6 ATAC-seq reveals bookmarked sites during mitosis	37
CHAPTER 2.4: Mitotic Chromatin Interaction Dynamics of PARP1.....	38
2.4.1 Summary of Intent	38
2.4.2 Materials & Methods	38
2.4.3 Results & Discussion	39
2.4.3.1 PARP1 displays distinct mobility on mitotic chromosome	39
CHAPTER 2.5: Characterization of NFY α Interaction with Putative MBFs	41
2.5.1 Summary of Intent	41
2.5.2 Materials & Methods	42
2.5.3 Results & Discussion	45
2.5.3.1 ChIP-WB suggests the interaction of NFYA with PARP1 and, possibly, DNMT1	45
CHAPTER 2.6	46
2.6.1 Summary of Intent	46
2.6.2 Materials & Methods	47
2.6.3 Results & Discussion	49
2.6.3.1 HDGF and PSIP1 KO in mouse PSCs results in reduced self-renewal capacity	49
2.6.3.2 Mitotic specific degradation of HDGF does not affect self-renewal in mouse PSCs	50
2.6.3.3 Mitotic specific degradation of PSIP1 in mouse PSCs	51
2.6.4 Conclusion.....	52
2.7 Figures.....	54
2.8 Tables	62
References	63

List of Figures and Tables

Part 1

Figure 1: Overview of REX1 association with ERVs in mESCs

Figure 2. Assessment of Yy1, Yy2 and REX1 association with ERVs in mESCs

Figure 3. REX1 protein expression levels in hESCs

Figure 4. Point mutations disrupting downstream initiation sites.

Figure 5. Schematic representation of REX1 KO clones

Figure 6. Pluripotency related protein expression levels in REX1 KO clones

Figure 7. ERV expression levels in E14T WT and REX1 KO 15 cells

Figure 8. Targeting strategy for generating 3XFLAG epitope C-terminally tagged REX1 mESCs

Figure 9. REX1 expression in LiF2i vs Serum conditions

Figure 10. ERV expression of mESCs in LiF2i vs Serum conditions

Figure 11. Generation of truncated mutant REX1 expression vectors for the characterization of isoforms

Figure 12. ERV expression in E14T WT, KO, and KO transfected human REX1 variants

Table 1: qRT-PCR primers and Universal Probe Library (UPL) probe numbers for REX1 ERV targets

Table 2: Primers used for generating human REX1 truncation mutants

Part 2

Figure 1. Identification of putative MBFs from ChIP-MS data

Figure 2. ATAC-seq reveals putatively bookmarked gene loci

Figure 3. Chromosomal association of PARP1 in interphase vs mitosis

Figure 4. Association of NFYA with Mitotic Chromatin and Bookmarking Factors

Figure 5. Self-renewal difference upon loss of HDGF and PSIP1

Figure 6. Pluripotency maintenance effects of mitotic specific degradation of HDGF

Figure 7. Construction of PSIP1-mko2-MD* plasmid

Figure 8. E14T WT and PSIP1-KO cells lines transfected with PSIP1 and MD/MD* constructs

Table 1: Primers used for generating MD and MD* constructs

Table 2: List of antibodies used for western blot and immunoprecipitation

List of Abbreviations and Symbols

PSCs	Pluripotent Stem Cells
ESCs	Embryonic Stem Cells
TFs	Transcription Factors
Yy1	Yin Yang 1
Yy2	Yin Yang 2
ERVs	Endogenous Retroviral
muERV-L	Murine endogenous retrovirus
musD	Mouse type-D
IAP	Intracisternal A-particle
mESCs	Mouse Embryonic Stem Cells
KO	Knock-Out
DSB	Double Stranded Break
Indels	Insertions/Deletions
RA	Retinoic Acid
LIF	Leukemia Inhibitory Factor
MEFs	Mouse Embryonic Fibroblasts
PB	Piggybac
MBFs	Mitotic Bookmarking Factors
ChIP-Seq	Chromatin Immunoprecipitation followed by DNA Sequencing
RT-qPCR	Reverse Transcriptase Quantitative Real-Time PCR
MD	Mitotic Degradation
ChIP-MS	Chromatin Immunoprecipitation followed by Mass Spectrometry
RNA-Seq	RNA Sequencing
NLS	Nuclear Localization Signal
KD	Knock-Down
CIC	Colony-Initiating Cell
FLIP	Fluorescence Loss in Photobleaching
GFP	Green Fluorescent Protein
ATAC-Seq	Assay for Transposase-Accessible Chromatin using Sequencing
FLIP	Fluorescence loss in photobleaching
G1	Gap 1
G2	Gap 2
G1t20	populations collected after 20 minutes of release into G1 from mitosis
G1t35	G1 populations collected after 35 minutes of release into G1 from mitosis
G2M	G2 and Mitotic mixed populations
GO	Gene Ontology
H2B	Histone2B
H3	Histone 3
H3K27Ac	Acetylation of histone H3 at lysine 27
H3K36me3	Trimethylation of histone H3 at lysine 36

H3K79me3	Trimethylation of histone H3 at lysine 79
H3S10P	Phosphorylated serine 10 on histone 3
H4K20me3	Trimethylation of histone H4 at lysine 20
hESCs	Human embryonic stem cells
HDGF	Hepatoma-derived growth factor
IgG	Immunoglobulin
IP	Immunoprecipitation
KO	Knockout
LIF	Leukemia inhibitor factor
M	Mitosis
MD	Mitotic degradation domain or mitotic degron
MD*	Inactive MD, Arginine at amino acid 42 is substituted with Alanine
mESCs	Mouse embryonic stem cells
mKO2	monomeric Kusabira orange 2
MLL	Mixed Lineage Leukemia
NLS	Nuclear localization signal
RA	Retinoic acid
pADPr	Poly(ADP-ribose)
PARP1	Poly(ADP ribose) polymerase I
PFA	Paraformaldehyde
pRb-E2F	Retinoblastoma protein and E2F transcription factor
PSCs	Pluripotent stem cells
PSIP1	PC4 and SFRS1 interacting protein
PWWP	Proline-Tryptophan-Tryptophan-Proline
RIPA	Radioimmunoprecipitation assay buffer
RNA-seq	RNA-sequencing
RT	Room temperature
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
S	Synthesis
SD	Standard deviation
TADs	Topologically associated domains
WT	Wild-type

CHAPTER 1: Investigating the Molecular Mechanisms of REX1 in Pluripotency Maintenance

1.1 Abstract

The pluripotent state is maintained by a network of “core” transcription factors (TF). REX1 (Reduced Expression-1) is a pluripotency related TF derived from retrotransposon-mediated duplication of the zinc finger TF Yin Yang 1 (YY1). Furthermore, expression of REX1 and YY1 induces changes in genes regulated by endogenous retroviral elements (ERV), suggesting an evolutionary origin of REX1 for ERV regulation. Studies suggest that murine REX1 may act in epigenetic regulation of gene expression and ERVs, but the precise mechanism remains unelucidated, so we generated FLAG-tagged REX1 pluripotent stem cell (PSC) lines, as well as a series of truncation mutants to explore the REX1 function. Our studies indicate the presence of previously undescribed isoforms of the full-length REX1 protein, suggesting that regulation by REX1 may be more complex than initially appreciated. We hypothesize that REX1 regulates the expression of a sub-set of ERVs and REX1 isoforms regulate REX1 target genes in pluripotent stem cells. Previously, we performed REX1 ChIP-seq and found enrichment for REX1 binding at specific ERVs. Here, we show that differential expression of REX1 isoforms do not change the expression of ERVs. Furthermore, our REX1 KO lines show changes in expression of ERV family members and together with the ChIP data, suggest that REX1 may act as a negative regulator of some retroviral elements. However, further experiments reveal a potential compensation of REX1 KO, possibly by the homologous factors YY1 and YY2. Due to the limited nature and time constrain of our study, we did not find conclusive evidence to further elucidate the potential compensation mechanism and the characteristics of the REX1 isoforms.

Our work provided a new avenue for exploring the functional importance of REX1 isoforms and the potential, YY1 and YY2 independent, regulatory role REX1.

1.2 Preface

The first part of my thesis focuses on the role of REX1, independent of YY1 and YY2, in the context of pluripotency regulation in pluripotent stem cells. The project started off characterizing the conserved role of REX1 in human and mouse in the context of pluripotency, later transitioning into characterizing the binding regions of REX1 and the conserved role of ERV regulation in both human and mouse; in particular, a subset of ERVs were specifically bound by REX1, independent of YY1/YY2. In the process of characterizing REX1 binding, a novel phenomenon was observed where REX1 was observed to express multiple isoforms in both human and mouse. Based off the analysis of the REX1 ChIP-seq data, I first focus on the ERV binding targets reasoned to be regulated specifically by REX1 and analyzing the differences in target expression in REX1 KO cells. Concurrently, I looked to quantify ERV expression level changes in naïve vs primed mouse PSCs and REX1 KO cells transfected with human REX1 truncation mutants, based on the different isoform expression patterns in naïve vs primed states. Amanda Hrenczuk, from the Draper Lab, had planned the experiments, generated the REX1 KO cell lines and constructed the human REX1 truncation mutants. I have carried on the work in conducting the qRT-PCR assays investigating differential expression in the generated lines and the impact of REX1 isoform expression.

1.3 Introduction

1.3.1 Introduction to pluripotency

Pluripotent Stem Cells (PSCs) are derived from the inner cell mass of a blastocyst, characterized by their ability to self-renew indefinitely and can differentiate into all three germ layers: endoderm, ectoderm, and mesoderm. (Evans & Kaufman, 1981) The ability to self-renew and differentiate make PSCs excellent for applications such as drug discovery, cell therapy, and disease modelling. The discovery of PSCs has led to many scientists uncovering the molecular basis of pluripotency and embryonic development in both mouse (Evans & Kaufman, 1981) and human (Thompson et al, 1998). Mechanistic studies have revealed that pluripotency is regulated via multiple signal transduction pathways governed by a network of core pluripotency transcription factors, including OCT4 (Nichols et al, 1998), NANOG (Mitsui et al, 2003), SOX2 (Masui et al, 2007). Furthermore, reintroducing the pluripotency transcription factors reprograms somatic cells into an induced pluripotent state (Takahashi & Yamanaka, 2006).

1.3.2 Heterogeneity in Pluripotent Stem Cell Populations

Heterogeneity within cellular populations refers to distinct groups or sub-populations with varying transcription levels of self-renewal and differentiation genes, which causes spontaneous cell fate decisions to occur (Bhatia et al, 2013; Kalkan & Smith, 2014; Nichols & Smith, 2009; Toyooka et al., 2008). PSCs tend to fluctuate between a naïve state, expressing more pluripotency related genes, and a primed state, expressing more differentiation related genes (Bhatia et al, 2013; Kalkan & Smith, 2014; Nichols & Smith, 2009; Toyooka et al., 2008). The naïve PSC state, at the apex of pluripotency, is analogous to embryonic stem cells (ESCs) isolated from the inner cell mass (ICM) of pre-implantation embryos and can self-renew

indefinitely, differentiate into all three embryonic germ layers, and contribute to chimeras (Li & Belmonte, 2017; Wray et al., 2010; Kumari, 2016; Loh et al., 2015). The primed state is analogous to epiblast stem cells (EpiSCs) derived from the epiblast of post-implantation embryos, maintain the ability to self-renew, differentiate into all three germ layers, but express higher levels of lineage-specific genes and are unable to contribute to chimeras (Li & Belmonte, 2017; Wray et al., 2010; Kumari, 2016; Loh et al., 2015). In vitro culturing conditions of mouse PSCs are divided into serum, with the addition of LIF to activate the self-renewal promoting factor STAT3 and fluctuate between both naïve and primed states, and non-serum, with the addition of LIF alongside two inhibitors (2i) of MAPK and GSK3 signalling to attain a relatively homogenous naïve state (Wray et al., 2010; Kumari, 2016; Ying et al., 2008; Tosolini & Jouneau, 2016; Hackett & Surani, 2014).

1.3.3 Core Transcriptional Network of Pluripotent Stem Cells

Studying the pluripotent state has revealed the existence of a core network of pluripotency factors, including OCT4, SOX2, and NANOG, that mediate the establishment and maintenance of the pluripotent state by co-binding to key pluripotency genes (Takahashi & Yamanaka, 2006; Boyer et al., 2006; Kim et al., 2008; Orkin et al., 2008; Loh et al., 2006; Chew et al., 2005; Ng & Surani, 2011). Although essential to pluripotency, the core network is not the whole picture and further studies have suggested an interdependence between factors contributing to the pluripotent identity. For instance, the loss of pluripotency related factors, DAX1 and NAC1 result in differentiation and reduced growth, and is suggested to interact with core regulators like NANOG (Kim et al., 2008; Orkin et al., 2008; Wang et al., 2006; Khalfallah et al., 2009; Ruan et al., 2017).

1.3.4 Extending the Transcriptional Network in Pluripotent Stem Cells

The pluripotent state is regulated not only the core network of transcription but also by interactions with other transcription factors and epigenetic regulators (Kim et al., 2008; Morey et al., 2015). Within the extended pluripotency transcription factor network is a zinc finger protein YY1 (Wang et al. 2006; Vella et al., 2012), suggested to transcriptionally regulate cell growth, development, and differentiation (Shi et al., 1991; Seto et al., 1991; Gordon et al., 2006; Donohoe et al., 1999). Furthermore, YY1 is part of the polycomb group complex (PcG) proteins, known for their critical roles in development through repression of differentiation associated genes, such as HOX, DLX, and POU family, via chromatin remodeling by repressive histone modifications (Morey & Helin, 2010; Boyer et al., 2006; O'Carroll et al., 2001; Voncken et al., 2003; Rajasekhar & Begemann, 2007). Additionally, YY1 has been shown to co-bind, often with MYC and E2F1, promoter regions of highly expressed genes in PSCs, such SURF-1 and CDC6, involved in oxidative phosphorylation and cell cycle progression (Vella et al., 2012; Ballabeni et al., 2011; Vernon & Gaston, 2000). Recently, YY1 has been shown to directly interact with OCT4 in ESCs and is suggested to promote ESC proliferation via interaction with the BAF complex (Wang et al., 2018). In terms of structure, YY1 has two N-terminal domains, where the first 100 amino acids are required for activation, and four C-terminal zinc finger motifs, which are involved in repression (Gordon et al., 2006; Kim et al., 2007; Bushmeyer et al., 1995). Interestingly, YY1 has two other homologues, YY2 and Reduced Expression 1 (REX1), suggested to have emerged via retrotransposition-mediated duplication of YY1, and having high sequence homology in the DNA-binding and repressive C-terminal (Kim et al., 2007; Nguyen et al., 2004).

Structurally, YY2 is similar to YY1 as it contains N-terminal Domains I and II, whereas REX1 only contains Domain II, suggesting different evolutionary constraints for the function of each protein (Kim et al., 2007). Knockout (KO) of YY2 results in deletion of mESC cultures, suggesting a role in self-renewal of PSCs, but KO of REX1 does not result in the same phenomenon, possibly due to the compensation by YY1 or YY2 (Tahmasebi et al., 2016; Hosler et al., 1989; Rogers et al., 1991; Masui et al., 2008). Furthermore, in comparison to its family members, REX1 is understudied and would benefit from functional studies independent of YY1 or YY2, especially due to studies suggesting a pluripotency specific role for REX1 in contrast to its family members (Hosler et al., 1989; Rogers et al., 1991).

1.3.5 REX1 Expression in Pluripotent Stem Cells

First discovered due to its decreased expression in F9 teratocarcinoma under retinoic acid conditions, REX1 has been widely used as a marker for the pluripotent state, due to its pluripotent cell type exclusive expression (Loh et al., 2015; Hosler et al., 1989; Rogers et al., 1991). REX1 expression is both high and uniform within the transcriptionally and epigenetically homogenous naïve ESCs, when contrasted to serum-conditions, in which cells fluctuate between naïve and primed states and express REX1 in a mosaic manner (Ray et al., 2010; Loh et al., 2015; Toyooka et al., 2008; Tanaka, 2009). Cumulatively, studies suggest that REX1 has evolved to be developmentally regulated, as expression is lost with as cells transition into EpiSCs, whereas YY1 is expressed ubiquitously through these stages (Ray et al., 2010; Loh et al., 2015; Toyooka et al., 2008; Shi et al., 2006).

1.3.6 Regulation of Endogenous Retroviral Elements

Endogenous Retroviral Elements (ERVs) are viral genes with the ability to transpose through the genome potentially causing a loss of genomic locus integrity and function (Guallar et al., 2012; Yang et al., 2015; Rowe et al., 2013; Schoorlemmer et al., 2014; Schlesinger & Goff, 2015). The generalized structure of a replication competent retrovirus includes four coding domains, which are Gag, Pro, Pol, and Env, encoding the structural components, viral protease, reverse transcriptase and integrase, and glycoproteins, respectively (Gifford & Tristem, 2003). Transposable elements (TE) can be divided into DNA transposon and retrotransposons, of which retrotransposons are the majority, consisting of three main families: long terminal repeats (LTRs), and, non-LTR retrotransposons, LINE and SINE (long and short interspersed nuclear elements), which differ in retrotransposition requirements. Furthermore, LTR and LINE elements depend on internal reverse transcriptase (RT), whereas SINE elements arise via retrotransposition of RNA polymerase III and the RT genes from LINE elements (Schoorlemmer et al., 2014; Stocking & Kozak, 2008). Additionally, ERVs can be subdivided, by RT gene similarity, into Class I (ERVK), Class II (IAP, MusD, ETn), Class III (ERVL, MaLR) (Gifford & Tristem, 2003; Schoorlemmer et al., 2014).

In humans, Class I ERVs have been found to be highly conserved, which had likely integrated into Old World primates some 35 million years ago. Following the divergence of humans some 6 million years ago, the insertional rate of new integrations has remained constant at about 70 human-specific viral insertions, in which 15-20 are highly preserved full-length proviruses (Ono et al., 1987; Bannert & Kurth, 2006; Macfarlane & Simmonds, 2004; Belshaw et al., 2005; Belshaw et al., 2004). The high degree of conservation found in some classes of human ERVs, suggests a beneficial role, namely as genomic architectural elements, reverse transcriptases, and promoter functional elements (Buzdin et al., 2006; Eickbush, 1997; Brandt et

al., 2005; Xing et al., 2006). Furthermore, some ERV glycoproteins have been suggested to provide protection against superinfection by exogenous retroviruses *in vitro* (Ponferrada et al., 2003). However, ERVs have not primarily evolved for the benefits of the host and possess detrimental effects, mainly being implicated as co-factors in multi-step development of tumors and autoimmunity (Rasheed, 1995; Contreras-Galindo et al., 2008; Ruprecht et al., 2008).

In PSCs, ERVs are under stringent regulation to maintain the core transcriptional network and, over time, have assumed roles in pluripotency cell fate decision making, often bound by core transcription factors to maintain control of the pluripotent state (Guallar et al., 2012; Yang et al., 2015; Rowe et al., 2013; Schoorlemmer et al., 2014; Schlesinger & Goff, 2015). Furthermore, RNA-seq data of PSCs in different pluripotent sub-states, showed significant differences in expression of transposable elements, including LINEs and LTRs, with some negatively correlating and some positively correlating with developmental progression (Hackett et al., 2017). Analysis of splice junctions in the RNA-seq data showed chimeric transcripts composed of TEs upstream of pluripotency related transcripts, which, together with the observation of highly correlative expression of said TEs and pluripotency related genes, suggests a mechanistic role of TEs in controlling pluripotency related gene expression (Hackett et al., 2017).

Expression of REX1 induces changes in genes directly regulated by ERVs and taken together with the binding of its relatives, YY1 and YY2, to ERVs suggests, for REX1, an evolutionary origin of ERV regulation (Guallar et al., 2012; Rowe et al., 2013; Schoorlemmer et al., 2014; Schlesinger & Goff, 2015; Pérez-Palacios et al., 2016). In support, depletion of REX1 in mouse PSCs leads to a 2-3 fold increase in murine endogenous retrovirus (muERV-L) and

mouse type-D elements (musD), while intracisternal A-particle (IAP) elements remain unaffected (Guallar et al., 2012).

Conflicting reports suggest a possible dispensable role of REX1 in pluripotency, but several findings are in support of the contrary. Indeed, REX1 expression and knockdown in mouse and human PSCs is linked to increased pluripotency and loss of self-renewal, respectively (Takahashi & Yamanaka, 2006; Toyooka et al., 2008; Climent et al., 2013; Zhang et al., 2006). Furthermore, REX1 knockout (KO) in mouse ESCs (mESCs), although viable, leads to altered lineage related marker expression, perturbed visceral endoderm differentiation, and deviation in Mendelian ratios during heterozygous crosses, suggesting REX1 dosage to be critical in development (Kim et al., 2011; Masui et al., 2007; Scotland et al., 2009). Although, murine REX1 may act in epigenetic regulation of gene expression and ERVs, precise mechanisms remain unelucidated.

CHAPTER 1.4: Characterizing REX1 Isoforms and YY1/YY2 Independent Role of REX1

1.4.1 Summary of Intent

Our lab sought to determine the involvement of REX1 in the regulation of endogenous retroviral elements in mouse PSCs. Through CHIP-Seq, the data suggested that REX1 primarily binds not only LTR elements but can also bind LINE and SINE elements to a lesser extent, further supported by data showing preferential binding of REX1 to RMER21 and IAPEY classes of ERVs, both part of the LTR family (Figure 1A, 1B). Furthermore, overlapping previously published CHIP-seq data sets for YY1 and YY2, revealed REX1 specific binding sites to RMER21A (30% of peaks) and IAPLTR3-int (26% of peaks) elements at high frequencies than YY1 and YY2, which show <2% of peaks corresponding to RMER21A (Figure 2). The higher

frequencies of REX1 specific binding to RMER21 and IAPLTR3-int elements suggests preferential regulation by REX1 independent of YY1 and YY2.

1.4.1.1 Identification of Novel REX1 Isoforms in PSCs

The function of REX1 has been viewed from the perspective of the DNA binding C-terminus as carrying out the important roles in PSCs. However, our lab has generated a series of N-terminal truncation mutants of human REX1 protein and a C-terminal zinc finger lacking mutant. Interestingly, overexpression of these human REX1 protein truncation mutants and subsequent western blot analysis, leads to the presence of two additional bands in samples with C-terminally 3X-FLAG tagged REX1 protein but absence in samples with N-terminally FLAG tagged REX1 protein, suggesting potential isoforms of the full-length protein (Figure 3).

Further analysis of the human REX1 open reading frame revealed methionine residues at amino acid positions 127 and 147. Interestingly, ablation of these methionine residues, via site directed mutagenesis converting methionine into isoleucine, generated by our lab, resulted in loss of the corresponding protein in western blot analysis (Figure 4A, Figure 4B). Furthermore, two of six methionine residues in murine REX1 are conserved in human REX1 protein, suggesting the presence of the phenomenon in mESCs.

We hypothesize that REX1 regulates the expression of a sub-set of ERVs and REX1 isoforms differentially regulate REX1 target genes in pluripotent stem cells

The long-term goal of this project is to elucidate the function and mechanisms of REX1 in pluripotent stem cells. To assess whether REX1 is functionally regulating the expression of lineage-determining genes, the first aim of the project is to evaluate the regulation of REX1

target genes via REX1 KO cells. In relation, the second aim is to understand the function of REX1 protein isoforms in different pluripotent sub-states.

1.4.2 Material & Methods

Cell culture: E14TG2A mouse embryonic stem cells were cultured on 0.1% gelatin coated culture dishes in mESC media: DMEM (Sigma Aldrich, D5796), 15% FBS, 1X non-essential amino acids (Life technologies: 11140-050), 1X glutamax (Life technologies: 35050-061), 1X sodium pyruvate (Life technologies: 11360-070), and 1X betamercaptoethanol (Gibco: 21985-023). Media was further supplemented with 1000 U/mL LIF (Amsbio, AMS-263-100) after filter-sterilization with a 0.22 μ m filter (Sigma) Cells were maintained at 37°C, 5% CO₂ and passage every three days using accutase® (Sigma Aldrich: A6964). PARP1 knockout cells were routinely maintained on a layer of x-ray irradiated mouse embryonic fibroblasts (xMEFs) seeded at a density of 1x10⁶ cells/60 cm². mESCs were pre-plated on gelatin coated dishes for 30 minutes to deplete xMEFs.

Adaptation of mESCs to LIF2i: Mouse ESCs were grown on feeders in serum conditions for a single passage, feeder depleted, and grown in feeder-free serum conditions for two subsequent passages. Cells were dissociated with accutase (Sigma), split at a 1:3 ratio, and washed once in 1 × Phosphate buffered saline (PBS) to remove serum. Cells were plated in LIF2i media [Neurobasal medium (Gibco), DMEM/F12 (Gibco), 0.5 × N2-supplement (Gibco), 1 × B27-supplement without Vitamin A (Gibco), 0.05% BSA (Sigma), 1 μ M PD0325901 (Tocris), 3 μ M CHIR99021 (Tocris), 1× Non-Essential Amino Acids (Gibco), 1× Glutamax (Gibco), 100 μ M Sodium Pyruvate (Gibco), 100 mM β -mercaptoethanol (Gibco), 1000 U/mL LIF (AMSBIO)] on 0.1% gelatin coated plates. Cells were passaged at a 1:3 ratio and grown in a humidified incubator at 37°C and 5% carbon dioxide.

Cell lysate preparation: Whole cell lysates were prepared by washing mouse embryonic fibroblast (MEF) depleted mESC cultures three times with $1 \times$ PBS before lysing on ice with $1 \times$ RIPA buffer [50mM Tris-HCl pH 8, 150mM NaCl, 0.1% Sodium Dodecyl Sulfate, 1% NP-40, 0.5% Sodium Deoxycholate and $1 \times$ Complete™ Mini Protease Inhibitor Cocktail (Roche)] for 10 minutes. Cells were centrifuged at $16,800 \times g$ at 4°C for 10 minutes. The supernatant was collected and quantified using the DC Protein Assay II kit (BioRad). Samples were normalized in $1 \times$ NuPAGE® LDS Sample Buffer (Invitrogen) with 15% TCEP Bond-Breaker Solution (Thermo Scientific) and heated at 95°C for 5 minutes prior to electrophoresis. Fractionated cell lysates were prepared by washing MEF depleted mESC cultures three times with $1 \times$ PBS before lysing on ice with cytoplasmic extraction buffer [10mM HEPES pH 7.9, 10mM KCl, 0.1mM EDTA pH 8.0, 0.1mM EGTA, 1mM DTT, and $1 \times$ Complete™ Mini Protease Inhibitor Cocktail (Roche)] for 15 minutes. NP-40 was added to a final concentration of 0.3% and samples centrifuged at $10,000 \times g$ at 4°C for 1 minute. The cytoplasmic supernatant was collected and the nuclear pellet was washed twice with cytoplasmic extraction buffer prior to being resuspended in nuclear extraction buffer [20mM Tris pH 7.9, 400mM NaCl, 0.2mM EDTA pH 8.0 and $1 \times$ Complete™ Mini Protease Inhibitor Cocktail (Roche)]. Samples were incubated on ice for 30 minutes and centrifuged at $10,000 \times g$ at 4°C for 5 minutes. The nuclear supernatant was collected, and the pellet discarded. Protein concentration was quantified using the DC Protein Assay II kit (BioRad). Samples were normalized in $1 \times$ NuPAGE® LDS Sample Buffer (Invitrogen) with 15% TCEP Bond-Breaker solution (Thermo Scientific) and heated at 95°C for 5 minutes prior to electrophoresis.

Western blot: Single-cell suspensions harvested during passage of mESC cultures were lysed in $1 \times$ radioimmunoprecipitation assay buffer (RIPA, Sigma) with $1 \times$ Protease Inhibitor Cocktail

(Roche). The protein concentrations were quantified using the DC Protein Assay kit (Biorad) with the bovine serum albumin standard curve ranging from 0 to 2 mg/ml concentrations. The quantified protein extracts were transferred into 1X NuPAGE LDS Sample Buffer (Thermo Fisher) with 15% Bond-Breaker™ TCEP Solution (Thermo Fisher). The prepared extracts were heated at 95°C for 5 minutes and separated on a 12% polyacrylamide gel at 180V for 45 minutes. The separated protein was transferred onto a PVDF membrane using constant-current electrophoresis at 200 mA for 2 hours, followed by blocking in 5% milk in 1X TBS. The proteins were then blotted with monoclonal antibodies (Table 3) in 3% milk in 1X Tris-buffered saline (TBS) containing 0.1 % Tween20 (1X TBST; Bio Shop) at 4°C overnight. After washing, the blot was incubated with the secondary anti-rabbit HRP-conjugated antibody (1:20,000; Biorad) in 3% milk in 1X TBST for 1 hour at RT, and was developed for 10 minutes using an HRP substrate (1:5 diluted in ultrapure water; Lumina). The blot was visualized using the ChemiDoc™ MP Imaging System (BioRad) with its associated ImageLab analysis software (Biorad).

qRT-PCR: RNA was isolated using Trizol™ LS (Thermo Fisher 10296028) according to manufacture's protocols. For nascent RNA q-RT-PCR, RNA was DNaseI treated in solution and purified using Qiagen RNeasy Micro Kit (Qiagen Cat # 74004). cDNA was prepared for 1µg of RNA using SensiFAST cDNA Synthesis Kit (Froggabio BIO-65054), and q-PCR was performed using SensiFAST SYBR No-ROX Kit (Froggabio CSA-01194).

Statistical analysis: All statistical analyses were performed using Prism 7 (GraphPad) and Microsoft Excel software. Error bars reflect the standard error of mean, unless otherwise stated. The paired two-tailed Student's t-test was used to compute all p-values in cell tracking experiments for comparison of the intensities under distinct channels for individually tracked

cells; all other p-values were computed with the unpaired two-tailed student's t-test. Unless otherwise stated error bars represent standard error of the mean and alpha of 0.05 was used as a cut-off for statistical significance

Assessment of REX1 expression in various culture conditions: 4.0×10^5 - 8.0×10^5 mESCs were seeded on 0.1% gelatin coated wells of a 6 well plate in LIF2i medium, mESC medium with MEFs, mESC medium without MEFs, or EB (5% FBS) medium supplemented with $1\mu\text{M}$ retinoic acid, conditions. Cells were cultured for 2 days and subsequently collected for western blot analyses.

1.4.3 Results & Discussion

1.4.3.1 Generation of REX1 Knockout Lines does not affect pluripotency related protein expression levels

Our lab has generated REX1 KO lines via transfection of CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-CRISPR associated protein 9) nuclease to generate Double stranded breaks (DSB) and guide RNA to specify the genomic location of the DSB, which induces the non-homologous end-joining pathway and results in small insertion/deletions (indels) at the site of the DSB (Ran et al., 2013; Sander et al., 2014). The indels introduced a premature stop codon in two KO lines either within the first zinc finger or just after the first zinc finger (Figure 5). Subsequently, the loss of the second and third zinc finger, due to the premature stop codon, should result in the loss of REX1 DNA binding capacity as well as its ability to interact with known protein partners, as shown previously (Bushmeyer et al., 1995). Consistent with literature, western blot analysis revealed the KO cell lines express OCT4, SOX2, and NANOG comparable to wildtype cells suggesting minimal loss of pluripotency (Figure 6) (Hackett et al., 2014). However, since the deletion is introduced

downstream of the start site, the REX1 truncated gene may be expressed and retain partial activity. Indeed, the preservation of OCT4, SOX2, and NANOG expression in the, at least, partial loss of REX1 activity (Figure 6) agrees with previous REX1 KO studies, suggesting REX1 is dispensable for pluripotency, with Masui *et al* showing minimal dysfunction in fully homozygous REX1 KO embryos (Masui et al., 2008).

1.4.3.2 REX1 Knockout possibly effects expression of Endogenous Retroviral Elements

Evaluation of target gene expression in RMER21A and IAP elements in the REX1 KO samples, via qRT-PCR, revealed statistically significant changes in RMER21A Ch6 and, possibly, in RMER21A Ch7 (Figure 7). The increase in the expression of these ERVs is consistent with the hypothesis that REX1 acts to negatively regulate ERVs in pluripotent stem cells. However, further replicates of the qRT-PCR experiment showed inconsistencies in the trends observed (data not shown). Interestingly, the earlier passage numbers of REX1 KO samples correspond to higher expressions of RMER21A Ch6 and Ch7 (Figure 7) but with the higher passage numbers used in later experiments, the trend dissipates (Figure 12), suggesting that with increased culture time, hence increased selection over time, the function of REX1 may be compensated by other factors, likely YY1 and YY2. Hypothetically, if REX1 activity is reduced suddenly, the regulated sites would change in expression and cells with compensatory activity from YY1 and YY2 could be selected for and outcompete cells with little compensatory activity, over time, which could explain the inconsistencies in the qRT-PCR data.

1.4.3.3 REX1 isoforms display differential expression under different pluripotent sub-states

Our lab had generated REX1 tagged with a 3X FLAG epitope, using CRIPSR-Cas9 and guide RNA to generate a DSB, as described in Aim 1, and simultaneously transfected a repair

template with the 3X FLAG epitope, inducing the homology directed repair pathway and insertion of the 3X FLAG epitope downstream of REX1 (Figure 8). Protein expression using Western blot analysis of REX1-FLAG in conditions with LIF2i, gelatin with mouse embryonic fibroblasts (MEFs), gelatin only, and RA, consistently revealed two additional bands in the gelatin conditions as opposed to the LIF2i conditions (Figure 9). qPCR analysis using samples cultured in the same four conditions, showed a consistent decrease in RMERA and IAP transcript levels under non-RA conditions (Figure 10). Furthermore, gelatin conditions show differing RMERA and IAP transcript levels in comparison to LIF2i conditions (Figure 10). Inferring from both western blot and qPCR data, suggests that the presence of REX1 in non-RA conditions is enough to induce changes in the RMERA and IAP transcript levels, and, interestingly, suggests there may be a link between the additional presence of one or more of the isoforms and the changes in RMERA and IAP transcript levels (Figure 10).

1.4.3.4 *Transfection with full-length REX1 does not affect REX1 target expression in REX1-KO cells*

Our lab has generated a library of human REX1 mutants with the N-terminal domain only, C-terminal domain only, and truncation mutants at different methionine sites within the REX1 open reading frame (Figure 11). These truncation mutants were cloned into a pCAG backbone vector and transfected into the REX1-KO mouse PSC line, essentially constitutively expressing the REX1 human full-length and mutant proteins (Figure 11). REX1 target expression levels, measured by qRT-PCR, show no statistically significant variation in expression between the KO and mouse PSCs transfected with human REX1 proteins (Figure 12). Although there is no measurable variation in REX1 target gene expression, it is possible that overtime, YY1 or YY2 compensated for REX1-KO, especially with both REX1 and YY1 coding for the same

C2H2 zinc finger (Kim et al., 2007), and since the human REX1 full-length and mutants were transfected in the REX1-KO lines, there would be little to no measurable difference to begin with. Alternatively, since the REX1 KO line is generated via introducing a premature stop codon in the first DNA binding domain (Figure 5), the translated REX1 truncated protein may retain partial DNA binding activity leading to masking of a true KO phenotype. The study by Guallar et al., used two shRNA sequences to target REX1 mRNA near the 5' end (shRNA1) and 3' end (shRNA2) and measured Class II and III ERV expression. Notably, Class II ERV expression significantly increased only during REX1 mRNA cleavage by shRNA1, suggesting that 3' end cleavage by shRNA2, which would preserve the majority of the 5' end, preserved partial REX1 function (Guallar et al., 2012). Indeed, the concept of partially retained protein activity could explain the inconsistency in the ERV expression seen in the REX1 KO line and might benefit from introducing a stop codon further upstream of the REX1 protein to ensure more complete ablation.

1.4.4 Conclusion

Here we have given some insight into the epigenetic regulation mechanisms of REX1 in PSCs. Using CRISPR-Cas9 mediated REX1 KO, we have shown the possible negative regulation of subsets of ERVs, in addition to the possibility of differentially expressed functional REX1 isoforms. However, further experiments are needed to confirm the REX1 isoform phenomenon and whether REX1 is truly regulating the ERVs. One phenomenon that requires further investigation is the possibility of the isoforms being post-translational modifications of the REX1 protein, which would lead to slightly different protein sizes. Furthermore, CHIP-MS could be employed to purify REX1 proteins from PSCs in different pluripotent sub-states, to examine differences in isoform expression.

However, the data does pose some questions, mainly around the redundancy of the YY1 family of proteins, since REX1 has considerable homology with YY1 and YY2, particularly in the DNA binding domain C-terminus, in both murine and human cells. Furthermore, KO of REX1 has been suggested previously to be dispensable for pluripotency, at least *in vitro*, and this seems to be supported by our data with ERVs, where the apparent increase in ERV expression in REX1 KO cells is rescued over time, possibly by YY1 or YY2. With the possibility of compensation by YY1 and YY2, it is difficult to truly establish the epigenetic functions of REX1 in PSCs and the solution could be to maintain an early passage in the REX1 KO cells, to reduce the probability of compensation on the population level. As such, the mechanism by which REX1 KO may be rescued over time is unclear, especially due to different N-terminal functional domains between REX1 and YY1/YY2.

1.5 Figures

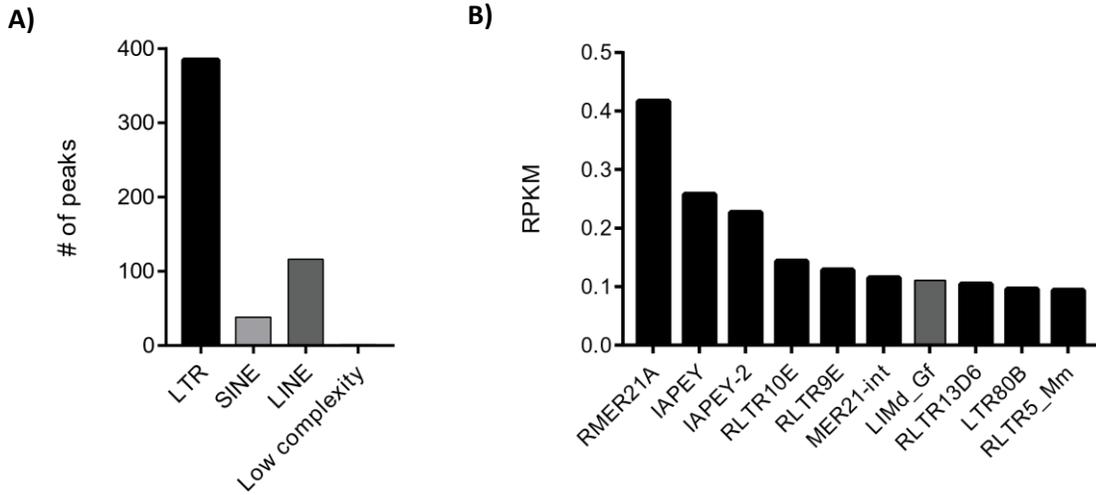


Figure 1. Overview of REX1 association with ERVs in mESCs. A) Binding preference of REX1 amongst ERV families B) REX1 displays highest enrichment and preferential binding to RMER21A and IAPEY elements in mESCs. Enrichment was determined using reads per kilobase of transcript per million mapped reads (RPKM). Figure generated by Amanda Hrenczuk.

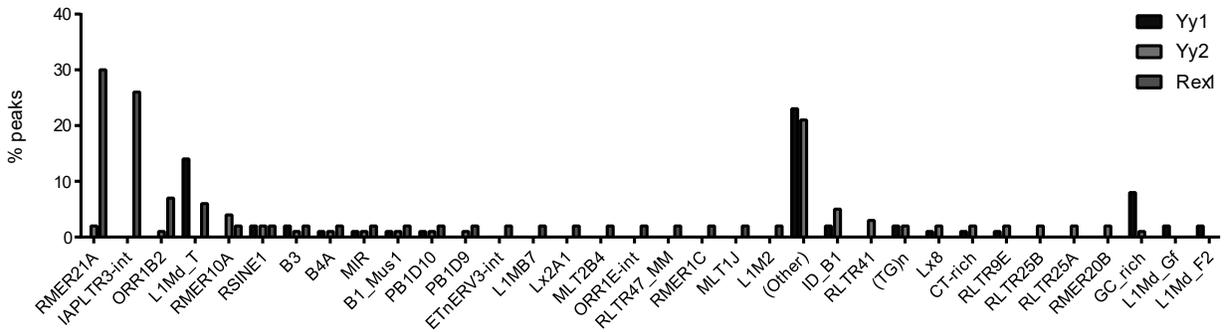


Figure 2. Assessment of Yy1, Yy2 and REX1 association with ERVs in mESCs. Comparison of peaks identified in previously published Yy1 and Yy2 data sets with our REX1 dataset demonstrates a higher frequency of REX1 binding to RMER21A and IAPLTR3 elements whereas Yy1 and Yy2 demonstrate higher frequencies of binding to (other) in addition to L1Md_T and ID_B1 elements, respectively. Frequency of binding is represented by the percentage of peaks per class in each data set. Figure generated by Amanda Hrenczuk.

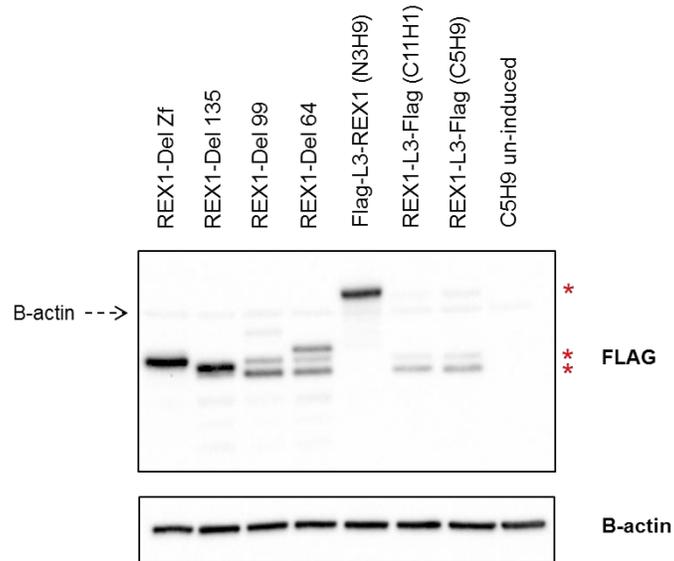


Figure 3. REX1 protein expression levels in hESCs. REX1 overexpression constructs were transfected into H9 and H1 cell lines using a three-plasmid piggyback system, allowing for stable integration. Plasmids were induced using doxycycline treatment for 24 hours. Figure shows western blot analysis of cell lysates probed for FLAG and B-actin was used as the loading control. Red asterisks denote REX1 protein isoforms. Figure generated by Sonam Bhatia.

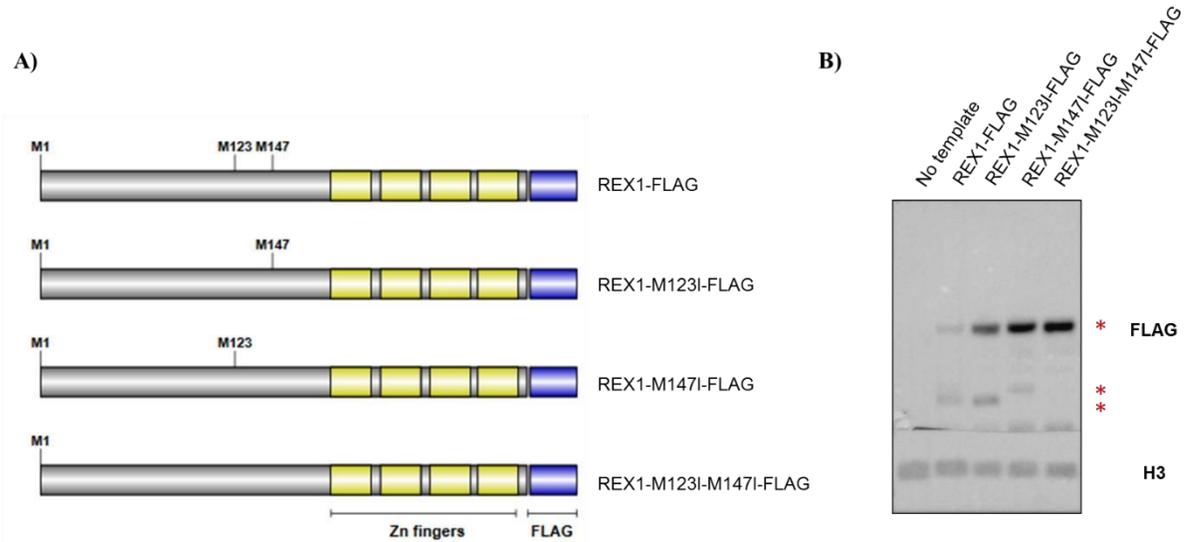


Figure 4. Point mutations disrupting downstream initiation sites. A) Schematic of mutated REX1 proteins, generated by PCR amplification using mutagenic primers converting methionine residues into isoleucine. B) HEK293 cells were transfected with REX1-FLAG or mutated REX1-FLAG vectors and treated with doxycycline for 24 hours. Figure shows western blot of whole cell lysates probed for FLAG and H3 as a loading control. Red asterisks denote flag expression and loss of REX1 isoforms can be seen to correlate with cells transfected with mutated REX1-FLAG vectors. Figure generated by Amanda Hrenczuk.

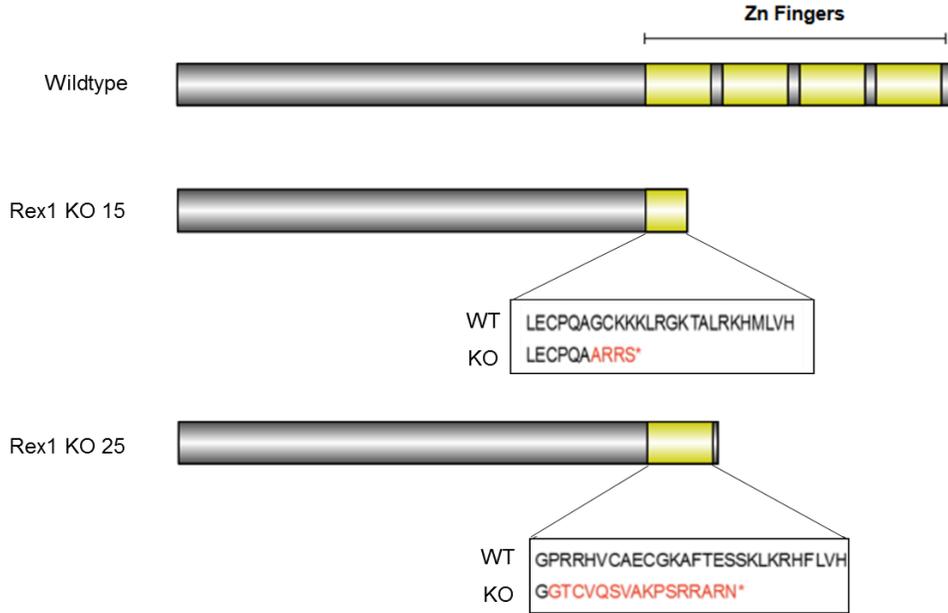


Figure 5. Schematic representation of REX1 KO clones. Sequencing shows premature stop codons in the zinc-finger domains responsible for DNA-binding. Mutations occurred in the first zinc-finger and just after the first zinc-finger in REX1 KO 15 and KO 25, respectively. Figure generated by Amanda Hrenczuk.

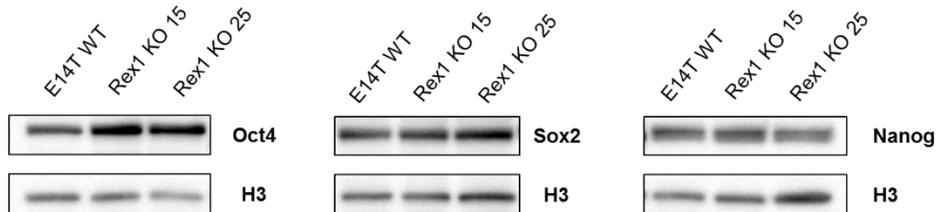


Figure 6. Pluripotency related protein expression levels in REX1 KO clones. Cells were cultured in mESC media with LIF on MEFS. Figure shows western blot of MEF depleted lysates probed for OCT4, SOX2, NANOG, and H3 as a loading control. Similar levels of protein are seen across all samples, suggesting no significant affects on pluripotency in KO lines, consistent with literature. Figure generated by Amanda Hrenczuk.

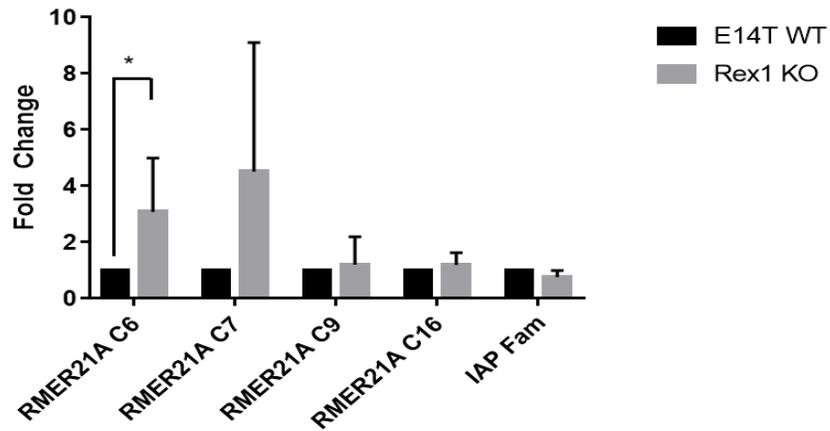


Figure 7. ERV expression levels in E14T WT and REX1 KO 15 cells. Cells were cultured in mESC media with LIF for 3 days. RNA was extracted and Dnase treated in extraction column. To completely remove genomic contamination, the eluate was Dnase treated again, and purified to remove the Dnase, with another final column Dnase treatment. qPCR analysis was performed with three technical replicates using the cDNA. The expression levels of REX1 KO cells were normalized to WT, with RPL13 as a reference gene. (n=5)

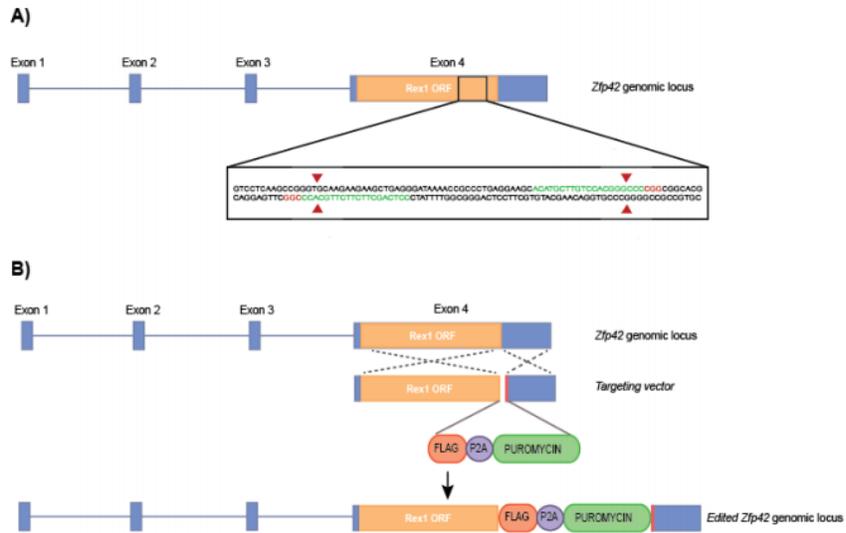


Figure 8. Targeting strategy for generating 3XFLAG epitope C-terminally tagged REX1 mESCs. E14T mESCs were cultured under serum conditions and transfected with CRISPR sgRNA(s) and targeting vector to insert 3XFLAG-2A-Puro into the REX1 locus, via induce homology directed repair. A) Schematic shows the sgRNA with a PAM sequence (red) and DSB targeting sequence (green). B) Cleavage and insertion of 3XFLAG-2A-Puro, allowing for simultaneous expression of 3XFLAG and puromycin selectable marker via the cleavable p2A peptide. Figure generated by Amanda Hrenczuk.

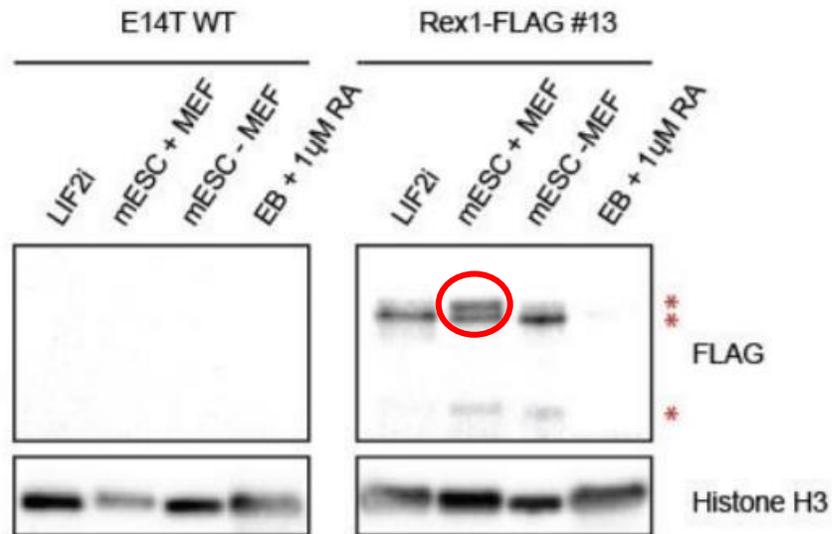


Figure 9. REX1 expression in LiF2i vs Serum conditions. REX1-FLAG mESCs were cultured in LiF2i, mESC media with MEFs, mESC media without MEFs, and EB (5% FBS) + 1µM RA for 2 days. Western blot was performed on the nuclear lysates, probed for FLAG and H3 as a loading control. FLAG expression was detected in all conditions except with Retinoic Acid, with varying isoform expression. Asterisks denote REX1 isoforms.

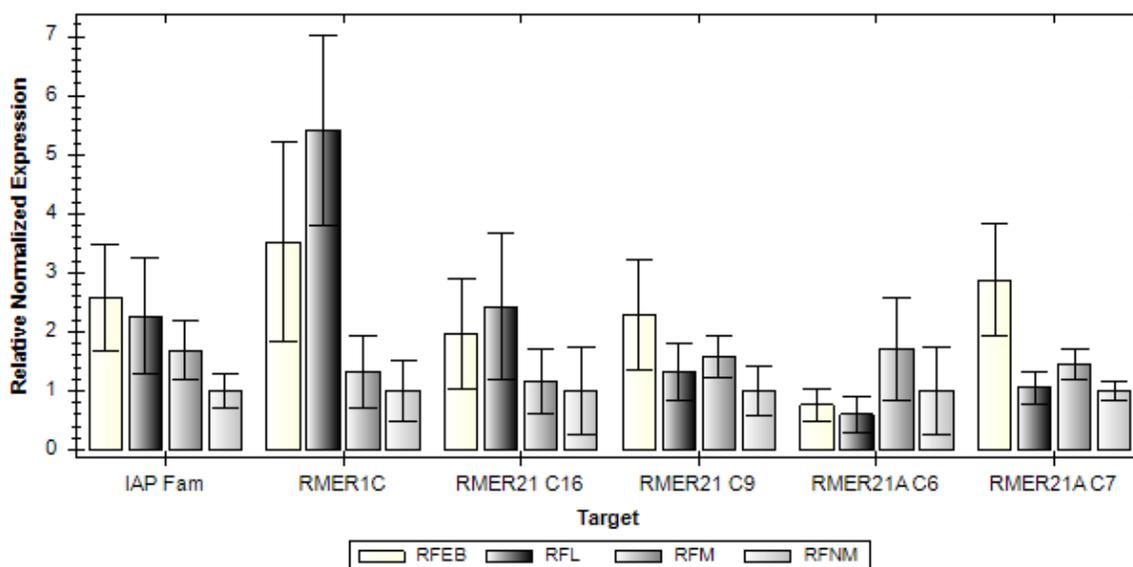


Figure 10. ERV expression of mESCs in LiF2i vs Serum conditions. REX1-FLAG mESCs were cultured in LiF2i (RFL), mESC media with MEFs (RFM), mESC media without MEFs (RFNM), and EB (5% FBS) + 1uM RA (RFEB) for 3 days. qPCR was performed with the extracted RNA converted to cDNA, in three technical replicates. The expression levels were normalized to mESCs cultured in mESC media without MEFs, with GAPDH as a reference gene. Figure shows the variable expression of ERV targets in mESCs culture under different pluripotent conditions (n=3).

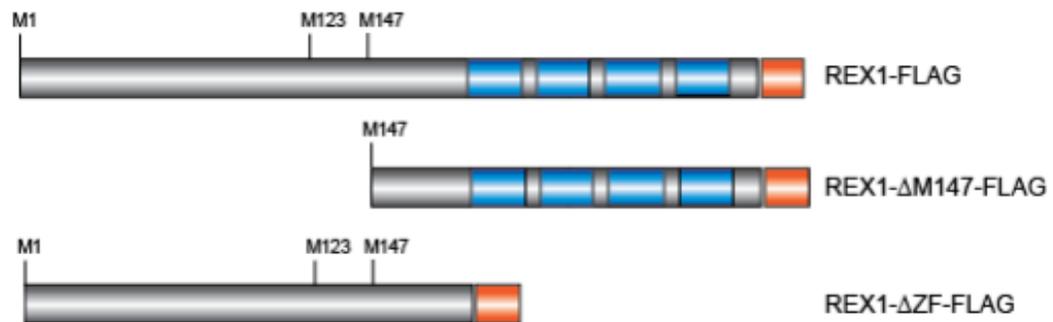


Figure 11. Generation of truncated mutant REX1 expression vectors for the characterization of isoforms. Inserts were generated via PCR amplification of pBTAG-REX1-L3-FLAG expression vector using attb primers designed to amplify from the third initiation site (REX1-M147-FLAG) and end prior to the first zinc finger (REX1-ZF-FLAG). The inserts were cloned into a pCAG backbone with an mKO2 fluorescent protein tag and puromycin selectable marker, using T4 ligase. Figures were generated by Amanda Hrenczuk.

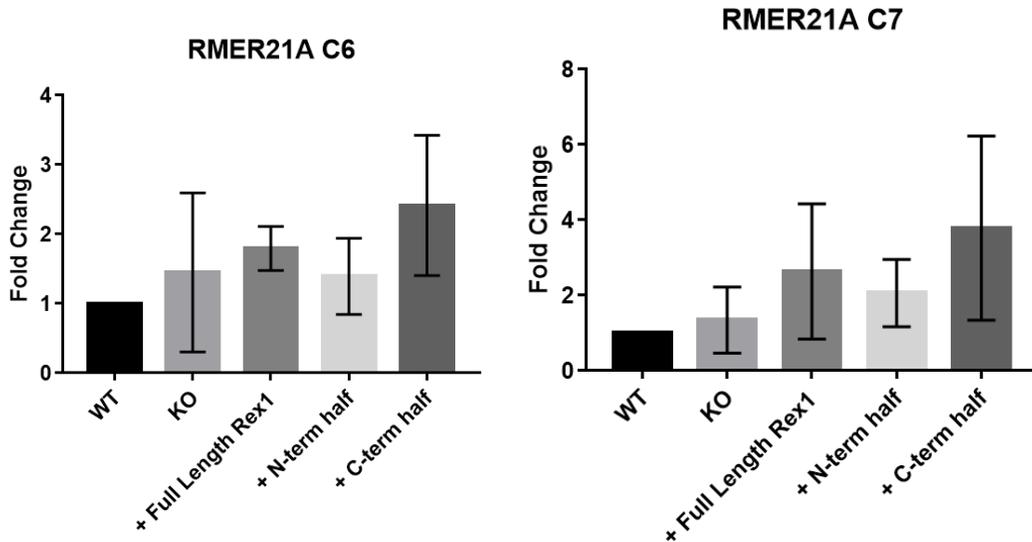


Figure 12. ERV expression in E14T WT, KO, and KO transfected human REX1 variants. Cells were cultured in mESC media with LIF for 3 days. RNA was extracted and Dnase treated in extraction column. To completely remove genomic contamination, the eluate was Dnase treated again, and purified to remove the Dnase, with another final column Dnase treatment. qPCR analysis was performed with three technical replicates and the expression levels of REX1 KO and transfected REX1 variants were normalized to WT, with RPL13 as a reference gene. (n=3)

1.6 Tables**Table 1: qRT-PCR primers and Universal Probe Library (UPL) probe numbers for REX1 ERV targets**

Primer	Sequence (5'→3')	UPL Probe
RPL13	FWD: GCCGGACTCCCTACAAGC	16
	RVS: GCTTCAGTATCATGCCATTCC	
IAPLTR3-int family	FWD: CCCAGTGAGGAGGCTAAAT	16
	RVS: CCATCGGTCAGGGTTATATCTT	
RMER1C	FWD: CCAAATCCATTTGATGTCTACTACC	46
	RVS: GGCCAGTCTCATGAATACGAA	
RMER21A Chr6	FWD: TGTACCACAGGAGCTGTCCA	19
	RVS: GGGGCTGCTGAAGTGTAGAG	
RMER21A Chr7	FWD: TCTTGCCATCCTCAGCCTAC	17
	RVS: ACCTTGGACAGTTCCTGTG	
RMER21A Chr9	FWD: GGCACAAAAGTCATTGCATC	40
	RVS: CATAGGCTGGGATGGGAAG	
RMER21A Chr16	FWD: GAGTAGGGCTGAGGCTGCTA	73
	RVS: AGCTCCCAGGGGAAGGTAT	

Table 2: Primers used for generating human REX1 truncation mutants

Name	Sequence (5'→3')	Note
Inf_attB_F	CAA AGA ATT CCT CGA GGG GAC AAG TTT GTA CAA AAA AGC	Human REX1 truncation mutants
Inf_Flag_R2	CCA GCT TCA TCT CGA GCT TGT CAT CGT CAT CCT TGT AGT CG	Human REX1 truncation mutants

CHAPTER 2: Characterizing Putative Mitotic Bookmarking Factors in Pluripotency Maintenance

2.1 Abstract

Mitotic bookmarking describes a potential mechanism involved in the stable propagation of cellular identity through the cell cycles. Candidate based studies have identified mitotic bookmarking factors (MBFs) that are retained on the mitotic chromatin and preserve the transcriptional memory of the cell. Nevertheless, there is a poor understanding of which proteins can serve as MBFs, as well as the chromatin dynamics of bookmarked sites during mitosis and the start of G1 phase. Previously, we designed a chromatin immunoprecipitation followed by mass spectrometry (ChIP-MS) assay to develop a global unbiased approach for identifying and characterizing novel MBFs. Using ChIP-MS, we identified putative MBFs associated with the mitotic chromatin in pluripotent stem cells (PSCs) and used ATAC-seq to identify subsets of pluripotency-associated accessible gene regions that appear to be bookmarked by a variety of transcription factors, including PARP1, PSIP1, and HDGF. Here, we characterize the interaction of a putative MBF, not found in our ChIP-MS screen, NFYA, with PARP1 and, inconclusively, another putative MBF, DNMT1. Furthermore, we found that PWWP containing putative MBF, HDGF, has a potential role in pluripotency maintenance but it is not mitosis-specific. Due to the limited nature and time constrain of our study, we did not find conclusive evidence to establish the role of PSIP1 in PSC mitotic bookmarking. Our work provided a new avenue for exploring the functional importance of mitotic bookmarks in pluripotent maintenance.

2.2 Preface

The second part of the thesis covers the characterization of novel mitotic bookmarking factors obtained from our ChIP-MS screen. The project started with the ChIP-MS screen designed and conducted by Sonam Bhatia, which revealed a list of putative bookmarking factors related to pluripotency. Out of the list, PARP1 was selected based on previous data suggesting bookmarking activity in HEK293 cell line, and HDGF and PSIP1 were selected due to the histone binding activity of the PWWP domain of both proteins. Furthermore, NFYa was reasoned to be a candidate, due to the overlap of DNA binding motifs of NFYa with the bookmarked regions eluded to from the ATAC-seq data. Importantly, PARP1 was characterized extensively by Sonam Bhatia and Daisy Deng, and the objective was to close the chapter on PARP1 with final supportive data. The design of the experiments was done by Sonam Bhatia, from the Draper Lab, including the experiments for NFYa. Furthermore, the experiments related to HDGF and PSIP1 as well as the plasmids used in the PARP1, HDGF, and PSIP1 experiments were mostly done and constructed by Daisy Deng, with input from Dr. Jonathan Draper. I executed on the fluorescence bleaching experiments on PARP1, with assistance from Daisy Deng, conducted all CIC related experiments with HDGF, constructed plasmids used for PSIP1, and performed all ChIP-WB experiments for NFYa, with the exception of the first ChIP-WB, which was performed by Ava Keyvani Chahi.

2.3 Introduction

2.3.1 Cell Cycle

Studies have suggested that cell division (Hanna et al, 2009) and a high proliferation rate (Ruiz et al, 2011) tends increase reprogramming efficiency of somatic cells into an induced pluripotent state, hinting towards the involvement of cell cycle regulation in the maintenance of pluripotency. The eukaryotic cell cycle is composed of four tightly regulated phases, being G1, S, G2 , and M phase. The interphase consists of G1, which is the first gap phase where a decision is made and prepared to undergo DNA replication, S phase is where DNA replication occurs, and G2 phase is where DNA repair takes place, and mitosis consists of M phase where cell division occurs. Cell cycle transitions are governed by fluctuations in expression and phosphorylation of different molecules called Cdks and cyclins (Amon et al., 1993; Nigg, 2001). Fluctuating levels of Cdk activity (Breedon et al, 2003), which requires assembly with target cyclins (Resnitzky et al, 1994; Jeffrey et al, 1995), activate or inactivate proteins to promote cell cycle progression. Cyclins are unstable proteins that create waves of Cdk activity through precise temporal synthesis and degradation.

2.3.2 Pluripotent Cell Cycle

PSCs exhibit unusual cell cycle features in comparison to somatic cells. It has been shown that mESCs, a type of PSC, have a truncated G1 phase and spend most of their time in S phase (Ballabeni et al, 2011), which leads to a shorter cell cycle period (Dalton et al, 2015) compared to most somatic cells. PSCs lack extensive G1 phases and possess longer S and M phases leading to its characteristic rapid self-renewal during the cell cycle, resulting in two daughter cells with conserved pluripotent states.

The difference between PSC and somatic cell cycles can be partially explained by differences in cyclin and Cdk expression and activity. PSCs have been shown to have atypically high Cdk activity throughout the cell cycle (Stead et al., 2002) and lack fluctuations during cell cycle progression in Cdk-cyclin complexes, except for Cdk1-cyclin B (Stead et al., 2002; White and Dalton, 2005). In contrast, in somatic cells, Cdk activity is reduced during G1 phase to allow for the establishment of pre-replicative complexes, which is a requirement for progression into S-phase (Maiorano et al., 2000; Prasanth et al., 2004). Studies knocking down Cdk activity slows the pluripotent cell cycle but does not affect the cell cycle structure (Stead et al., 2002), suggesting the structure is governed by the lack of Cdk periodicity. For instance, the high Cdk activity during G1 leads to hyperphosphorylation of pRb (Retinoblastoma protein), which inhibits interaction with E2F (E2F transcription factor) leading to repression of E2F gene expression, crucial to establishing full length G1 (Dyson, 1998), and shortening of the G1 phase (Stead et al., 2002). Furthermore, as PSCs undergo differentiation, constitutive Cdk activity collapses and becomes cell cycle dependent, suggesting that the pluripotent state and identity is linked to the cell cycle machinery (Faast et al, 2004; Li et al, 2012; Neganova & Lako, 2008; Coronado et al, 2013).

2.3.3 The mitotic chromatin

During mitosis, several changes occur for all cell types including universal chromatin condensation, dissociation of transcriptional machinery from interphase binding sites, and deposition of mitotic specific histone modification (Naumova et al, 2013; Bernardi, 2015; Oomen & Dekker, 2017). Despite this upheaval, the chromatin returns to its uncondensed state and interphase-specific histone marks are re-established, transmitting the parental cell-type specific gene expression and epigenetic state to each daughter cell. The mechanisms that

govern the preservation of cell identity throughout cell division is not fully understood.

The mitotic chromatin folds into a three-dimensional organization distinct from the interphase chromatin (Naumova et al., 2013). In interphase, the chromosomes spatially fold by long-range chromatin interactions into topologically associated domains (TADs) (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Markaki et al., 2010).. TADs are cell-type specific, folding patterns are dependent on the locus, which is correlated with gene expression, and they are lost during mitosis (Nora et al., 2012). In contrast, mitotic chromatin is not cell-type specific and has a generic folding pattern independent of gene loci (Naumova et al., 2013). Interestingly, the mitotic chromatin remains relatively accessible at broad regions of loci, particularly near promoters, with some accessibility reduction at narrow DNase sensitive distal regulatory regions, as determined by DNase I sensitivity assays coupled with sequencing (DNase-seq). Studies have suggested the potential role of transcription factor binding that could generate the narrow DNase sensitive regions by spatially manipulating nucleosomes and generate broad regions by recruiting other factors and remodeling the chromatin. The maintenance of the accessible regions present in interphase and mitosis, specially at the promoter regions, suggests the presence of chromatin features that keep the regions open throughout mitosis and available for binding upon mitotic exit (Liu et al., 2017a; Xu et al., 2017). Furthermore, open chromatin regions during mitosis can be different between cell-types which may be important to transmitting cell-type specific transcriptional information, but the exact mechanisms here must be resolved.

2.3.4 Mitotic bookmarking

Cells undergoing mitosis cease transcription and a subset of DNA associated proteins dissociate from the chromosome, but transcription resumes upon mitotic exit, suggesting the

presence of chromatin associated factors involved in preserving interphase transcriptional information as the cells exit mitosis. Two observations from the 1990s, the presence of a greater proportion of single stranded DNA in mitosis (Juan et al., 1996) and these single stranded regions were correlative with active gene expression (Michelotti et al., 1997), sprung to life the concept of mitotic bookmarking as a mechanism of transmitting epigenetic memory from parent to daughter cell. Currently, the concept is proposed as ‘memory signatures’ preserved throughout mitosis at key fate maintaining genomic sites and passed on as a blueprint for daughter cells to re-initiate parental transcription state (Hsiung and Blobel, 2016; Hsiung et al., 2015; Sarge and Park-Sarge, 2009). The ‘memory signatures’ include DNA methylation profiles, mitotically retained transcription factors, mitotically stable histone variants, and architectural components of the chromatin (Hsiung et al., 2015). Studies over the past decade have revealed retention of some transcription factors and chromatin regulators on the mitotic chromatin, MLL (Mixed Lineage Leukemia), being one of the first shown to be associated with the mitotic chromatin and induce rapid reactivation of bookmarked genes in HeLa cells (Blobel et al., 2009). Furthermore, the behavior of mitotic bookmarking has been shown in transcription factors, such as GATA1 (Kadauke et al., 2012), FOXA1 (Caravaca et al., 2013), PARP1 (Lodhi et al., 2016), ESSRB (Festuccia et al., 2016), SOX2 (Deluz et al., 2016), and KLF4 (Liu et al., 2017b), all of which facilitate rapid transcriptional reactivation of bookmarked genes. Interestingly, epigenetic marks, such as H4K5Ac and H3K27Ac, have been suggested to work in conjunction with transcription factors, such as BRD4 (Zhao et al, 2011), OCT4, SOX2, and KLF4 (Liu et al. 2017b). More recent studies continue to reveal the role of mitotic bookmarking factors like SOX2, shown to be required specifically during the M to G1 phase transition to drive neuroectodermal differentiation (Deluz et al., 2016), and RUNX1,

suggested to control mammary cell proliferation by binding to genes controlling growth and proliferation during mitosis (Rose et al., 2019).

2.3.5 ChIP-MS reveals putative pluripotency associated mitotic bookmarks

Our Lab developed an unbiased technique, using ChIP followed by MS to identify proteins associated with mitotic chromosomes in PSCs at a global proteome level, as opposed to single target-based approaches (Figure 1A). Essentially, the PSCs were enriched for mitotic populations using double thymidine block (S-phase arrest) and nocodazole (G2/M phase arrest) treatment followed by mitotic shake off (Teves et al., 2016). Then ChIP-MS was performed utilizing antibodies against histone 3 (H3), used to precipitate non-mitotic-specific proteins, to precipitate interphase proteins, and against histone 3 phosphorylated at serine 10 (H3S10), a mitotic specific histone modification, to precipitate mitotic proteins (Hendzel et al., 1997). The peptide sequences obtained from ChIP-MS were filtered based on threshold values and replicated commonality between H3, interphase proteins, and H3S10, mitotic-specific proteins, to generate the putative MBF list (Figure 1A, Figure 1B). Dividing the putative MBFs into two groups, high expression in PSCs with LIF vs RA, using published RNA-sequencing (RNA-seq) datasets, revealed 1 RA-specific MBF, 124 generic MBFs, and 10 LIF/pluripotency specific MBFs (Figure 1B) (Terranova et al., 2015). Further filtration based on GO analysis and KO lethality revealed a final 31 putative MBF list, which were validated by fusing mKO2, a fluorescent protein, to the MBFs and using immunofluorescence to confirm mitotic chromosome association.

Several proteins, including UTF1, DNMT3a, UHRF1, and PARP1, have been previously shown to be associated with the mitotic chromatin with PARP1 recently suggested to associate with rapid reactivation of genes upon mitotic exit (Lodhi et al. 2014). However,

further studies are required to reveal the roles of MBFs in chromatin remodeling, epigenetic inheritance, and gene expression during mitosis in PSCs.

2.3.6 ATAC-seq reveals bookmarked sites during mitosis

Our lab has also established chromatin accessibility profiles for PSCs in interphase (asynchronous), during mitosis (G2M), and during mitotic exit (G1t20 and G1t35), using transposase-accessibility chromatin followed by sequencing (ATAC-seq), which involves treatment with hyperactive Tn5 transposase to cleave DNA and insertion of DNA adapters for downstream sequencing (Bhatia, 2017). With mitotically bookmarked sites defined as maintenance of chromatin accessibility throughout mitosis into G1, we found that most sites were overlapped between interphase and mitotic populations and located within gene promoters and at distal intergenic regions, along with a large percentage of interphase proteins that lose accessibility during mitosis and early G1 (Figure 2A). Furthermore, we examined the overlapped the ATAC-seq data for the putative mitotic bookmarking factors with mitotic specific sites of H3K27Ac, previously suggested to be retained on mitotic chromatin and correlate to mitotic binding sites of OCT4, SOX2, and KLF4 (Figure 2B). We found strong co-localization between H3K27Ac marks with our putative mitotic bookmarked sites, from the ATAC-seq data, and no co-localization with non-mitotically bookmarked sites. Amongst the co-localization data, KLF4 and PARP1 showed the highest co-localization with H3K27Ac, and a subset of their bookmarked genes showed higher expression in pluripotency, suggesting a role in pluripotency maintenance through mitotic bookmarking.

CHAPTER 2.4: Mitotic Chromatin Interaction Dynamics of PARP1

2.4.1 Summary of Intent

PARP1, one of the putative MBFs, belongs in the poly(ADP-ribose) polymerase family of proteins mainly involved in DNA repair pathways by transferring ADP-ribose to various substrates (Amé et al., 2004). PARP1 contains a DNA-binding domain composed of two zinc-fingers, a nuclear localization signal (NLS) domain, a protein interaction domain, and a catalytic domain (Roper et al., 2014; Ji & Tulin, 2010). Interestingly, PARP1 and its family are highly expressed in PSCs and regulate “core” transcription factors, including NANOG and OCT4, and have even been demonstrated to remain bound to mitotic chromatin in human embryonic kidney cells, suggesting its involvement in mitotic bookmarking (Roper et al., 2014; Ji & Tulin, 2010). Furthermore, PARP1 knockdown (KD) studies demonstrated PARP1 necessity for rapid transcriptional reactivation of mitotic target genes in early G1 phase, validating its role as a MBF in human embryonic kidney cells (Lodhi et al., 2014). Since MBFs remain bound to a subset of target genes during mitosis, *we hypothesize that PARP1 displays distinct mobility during mitosis in comparison to interphase.*

The aim of this part is to provide support to our lab’s in progress MBF screening project, by measuring the dynamics of PARP1, a putative MBF from the screen, mobility during mitosis compared to interphase, which potentially will support PARP1 as a MBF retained on the mitotic chromatin.

2.4.2 Materials & Methods

Cell culture: E14TG2A mouse embryonic stem cells were cultured on 0.1% gelatin coated culture dishes in mESC media: DMEM (Sigma Aldrich, D5796), 15% FBS, 1X non-essential

amino acids (Life technologies: 11140-050), 1X glutamax (Life technologies: 35050-061), 1X sodium pyruvate (Life technologies: 11360-070), and 1X betamercaptoethanol (Gibco: 21985-023). Media was further supplemented with 1000 U/mL LIF (Amsbio, AMS-263-100) after filter-sterilization with a 0.22 μ m filter (Sigma) Cells were maintained at 37°C, 5% CO₂ and passage every three days using accutase® (Sigma Aldrich: A6964). PARP1 knockout cells were routinely maintained on a layer of x-ray irradiated mouse embryonic fibroblasts (xMEFs) seeded at a density of 1x10⁶ cells/60 cm². mESCs were pre-plated on gelatin coated dishes for 30 minutes to deplete xMEFs.

Fluorescence loss in photobleaching (FLIP): FLIP experiments were performed on Leica SP5 confocal microscopy at McMaster Biophotonics Facility. mESCs with stable PARP1-mKO2 expression was imaged and bleached using 561-nm laser. To reduce background fluorescence, the cell culture medium was exchanged before imaging for mESC media made with DMEM lacking phenol red (Gibco™). In mitotic cells, we measured time traces of average fluorescence from a square region in the area containing condensed chromosomes while continuously bleaching another region containing condensed chromosomes within the same cell. In interphase cells, both bleaching and measurement regions were inside the nucleus.

2.4.3 Results & Discussion

2.4.3.1 PARP1 displays distinct mobility on mitotic chromosome

Our lab has generated PARP1 KO mouse PSCs by disrupting exon 2 within the DNA-binding domain of PARP1, using CRISPR/Cas9, and further clonal selection revealed complete disruption of open reading frame and loss of protein product. Furthermore, we evaluated the mitotic specific perturbation of PARP1 chromosomal association on pluripotency maintenance

by transfecting mitotically degraded PARP1, as described previously with SOX2, into established PARP1 KO mESC lines. For validation purposes, we have used Fluorescence Loss in Photobleaching (FLIP) to validate the association of PARP1, fused with a fluorescent tag mKO2 protein, with the mitotic chromosome. As a PARP1-mKO2 fused cell is photobleached, using a laser, in a small cytoplasmic area, the rate of fluorescence recovery of the bleached area is negatively correlated with the chromosomal association of said fluorescently tagged protein (Figure 3). Thus, the rate of fluorescence recovery can be measured via the half-life of fluorescence loss in a neighbouring part of the cell, which indicates diffusion of fluorescence across the cell to compensate for photobleaching, to quantify the chromosomal association of the fluorescently tagged protein (Figure 3). Essentially, the greater the fluorescence half-life in the measured area, the greater the chromosomal association of the fused protein. Indeed, the half-life of PARP1-mKO2 is statistically measured to be higher during mitosis vs interphase in both wildtype and KO lines, validating that PARP1 is associated with the mitotic chromosome (Figure 3).

Furthermore, we are currently in the process of validating the mitotic specific degradation of PARP1-mKO2 fused with Cdk1-Cyclin B using time-lapse live cell imaging. Using Histone H2B fused with green fluorescent protein (GFP), we can track both PARP1-mKO2 and Histone H2B through mitosis across several mESCs and validate the functionality of the degradation domain via the loss of mKO2 fluorescence during the mitosis to G1 phase.

CHAPTER 2.5: Characterization of NFY α Interaction with Putative MBFs

2.5.1 Summary of Intent

NFY, Nuclear Factor Y, is a trimeric protein with three subunits, composed of NFYA, NFYB, and NFYC (Nardini et al., 2013). Functionally, NFY binds to promoters, enhancers, and LTRs of human ERVs and, moreover, can access polycomb repressed chromatin domains and co-associate with growth controlling factors (Fleming et al., 2013). NFYA binds to the CCAAT motif, located in promoter of a variety of genes, including genes involved in cell cycle progression, whereas, NFYB and NFYC form a dimer, with histone folding domains similar to Histone 2A and 2B, to bind the DNA backbone (Nardini et al., 2013; Fleming et al., 2013). Together, NFY contains both specific binding, via NFYA binding to the CCAAT motif, and non-specific binding, with the NFYB/NFYC dimer (Nardini et al., 2013). Interestingly, NFY has been shown, through KO studies, to have both activation marks, such as ubiquitination of Lys138 by NFYB, and repressive activity, by interacting with and enhancing zinc and homeobox protein 1, and even dnmt3b activity (Chen et al., 2013). Importantly, in antigen presenting cells, NFYA has been shown to participate in “bookmarking” activity, for example, via the interaction with PP2A to occupy the DRA gene promoter (Arampatzi et al., 2013). Furthermore, in a human colon cancer cell line, shRNA K/D of NFYA led to a delay of S-phase progression, whereas, NFYB depletion led to a stall in G2/M phase but not S-phase (Benatti et al., 2011).

Interestingly, ATAC-Seq data from our lab had revealed the enrichment of CCAAT motif, bound by NFYA, in multiple mitotically bookmarked regions, suggesting that NFY associates with the mitotic chromatin via NFYA binding the CCAAT motif. Hence, *we hypothesize that NFY*

interacts with the mitotic bookmarking factors that we have identified based on the mass spec data.

The aim is to investigate protein-protein interactions between NFYA and putative MBF candidates generated from our ChIP-MS MBF screen, in the context of mitosis. Interactions with putative MBFs would support the case for NFY being a novel MBF, even though it was not a candidate from our screen, which can be explained by low protein abundance.

2.5.2 Materials & Methods

Cell culture: E14TG2A mouse embryonic stem cells were cultured on 0.1% gelatin coated culture dishes in mESC media: DMEM (Sigma Aldrich, D5796), 15% FBS, 1X non-essential amino acids (Life technologies: 11140-050), 1X glutamax (Life technologies: 35050-061), 1X sodium pyruvate (Life technologies: 11360-070), and 1X betamercaptoethanol (Gibco: 21985-023). Media was further supplemented with 1000 U/mL LIF (Amsbio, AMS-263-100) after filter-sterilization with a 0.22 μ m filter (Sigma) Cells were maintained at 37°C, 5% CO₂ and passage every three days using accutase® (Sigma Aldrich: A6964). PARP1 knockout cells were routinely maintained on a layer of x-ray irradiated mouse embryonic fibroblasts (xMEFs) seeded at a density of 1x10⁶ cells/60 cm². mESCs were pre-plated on gelatin coated dishes for 30 minutes to deplete xMEFs.

Cross-linked chromatin immunoprecipitation: 1x10⁷ mESCs were used per IP with 1 μ g/1x10⁷ cells of antibody. Cells were cross-linked in 1% paraformaldehyde (PFA) for 10 minutes at room temperature with shaking and then washed 2X in large volumes of PBS. Cross-linked cells were lysed in RIPA buffer (50mM Tris-Cl pH7.45, 50mM NaCl, 0.1%SDS, 2%NP-40, 1% Sodium deoxycholate) supplemented with protease inhibitors (11836153001

cComplete™ mini-tablets Roche, Sigma) for 30 at 4°C. Nuclear and chromatin pellet was collected by gentle centrifugation at 2500 g for 5min at 4°C. The pellet was resuspended in RIPA dilution buffer (RDB, 50mM Tris- Cl pH7.45, 150mM NaCl) and supplemented with protease inhibitors, and homogenized with 18G needle. The lysate was gently sonicated to release shearing chromatin (6 pulses each with 20s ON, 30s OFF at 30% amplitude). Sheared chromatin supernatant was collected by spinning at 10,000 g for 10 min at 4°C. Chromatin was diluted in RDB to get a final SDS concentration of 0.025% to assist in IP. The samples were reverse cross-linked at 95°C for 10 min and run on a precast 10% bis-tris gel (Cat# NP0322, ThermoFisher) following the western blot protocol.

Western blot: Single-cell suspensions harvested during passage of mESC cultures were lysed in 1X radioimmunoprecipitation assay buffer (RIPA, Sigma) with 1X Protease Inhibitor Cocktail (Roche). The protein concentrations were quantified using the DC Protein Assay kit (Biorad) with the bovine serum albumin standard curve ranging from 0 to 2 mg/ml concentrations. The quantified protein extracts were transferred into 1X NuPAGE LDS Sample Buffer (Thermo Fisher) with 15% Bond-Breaker™ TCEP Solution (Thermo Fisher). The prepared extracts were heated at 95°C for 5 minutes and separated on a 12% polyacrylamide gel at 180V for 45 minutes. The separated protein was transferred onto a PVDF membrane using constant-current electrophoresis at 200 mA for 2 hours, followed by blocking in 5% milk in 1X TBS. The proteins were then blotted with monoclonal antibodies (Table 3) in 3% milk in 1X Tris-buffered saline (TBS) containing 0.1 % Tween20 (1X TBST; Bio Shop) at 4°C overnight. After washing, the blot was incubated with the secondary anti-rabbit HRP-conjugated antibody (1:20,000; Biorad) in 3% milk in 1X TBST for 1 hour at RT, and was developed for 10 minutes using an HRP substrate (1:5 diluted in ultrapure water; Lumina). The blot was visualized using the

ChemiDoc™ MP Imaging System (BioRad) with its associated ImageLab analysis software (Biorad).

Alternatively, the Licor Odyssey (B446) imaging system was used for western blots prepared according to the IRDye 680 Quick Western Kit.

Quantification of Western Blot: Western Blot images were saved as TIF file images. The TIF images were converted to a grayscale image JPEG image using GIMP 2.10.10 image editing software. The JPEG images were opened on ImageJ with measurements set to “mean gray value” only. A square range of interest (ROI) was drawn to fully encompass the largest protein band in the JPEG western blot image. The same ROI was used to take measurements across the row for all lanes and the measurements were transferred to excel. The IgG measurement was subtracted from the NFY α immunoprecipitated measurement and either normalized, as a fraction by the input measurement, or non-normalized. The calculations were transferred to Prism 7.00 and graphs were generated with Standard Error of the Mean.

Statistical analysis: All statistical analyses were performed using Prism 7 (GraphPad) and Microsoft Excel software. Error bars reflect the standard error of mean, unless otherwise stated. The paired two-tailed Student’s t-test was used to compute all p-values in cell tracking experiments for comparison of the intensities under distinct channels for individually tracked cells; all other p-values were computed with the unpaired two-tailed student’s t-test. Unless otherwise stated error bars represent standard error of the mean and alpha of 0.05 was used as a cut-off for statistical significance.

2.5.3 Results & Discussion

2.5.3.1 ChIP-WB suggests the interaction of NFYA with PARP1 and, possibly, DNMT1

To validate the interaction between NFYA and putative MBFs from the ChIP-MS screen, ChIP-WB is conducted with mitotic cells enriched in WT mouse PSC cultures and para-formaldehyde crosslinked protein-protein interactions. Essentially, antibody-bound NFYA is purified using magnetic protein G dynabeads and the proteins of interest are probed in the purified NFYA protein solution. Concurrently, an IgG control is used to control for the over-saturation of residual IgG from the ChIP. Indeed, from the multiple biological replicates, NFYA seems to be interacting not only with PARP1, but also with H3S10ph, a mitotic chromatin marker, and, possibly, DNMT1, another mitotic bookmarking factor from the ChIP-MS screen (Figure 4A). Furthermore, image analysis quantifying the band intensities shows higher enrichment of the NFYA IP sample, in comparison to IgG control, for PARP1 and possibly for DNMT1, however, the interaction with DNMT1 is inconsistent and requires the use of a fresh antibody for ChIP-WB (Figure 4B). Together, the ATAC-seq and ChIP-WB data suggests that NFYA may function as a MBF, possibly through interaction with PARP1 and other MBFs (Figure 2, Figure 4). From previous studies, NFYA is suggested to bind the CCAAT motif in the promoter regions of genes involved in cell-cycle regulation, which could be the case for the NFYA and PARP1 interaction. However, further experiments need to be conducted where IP samples are near 100% enrichment for mitotic cells to confirm the mitotic-specific interaction between NFYA and PARP1, which would establish a stronger case for NFYA as a MBF.

CHAPTER 2.6

2.6.1 Summary of Intent

Among the list of putative MBFs were a subfamily of proteins, HDGF (hepatoma-derived growth factor) and PSIP1 (PC4 and SFRS1 interacting protein), containing the PWWP domain, a highly conserved N-terminal DNA binding domain, but differing in the gene-specific C-termini (Qin & Min., 2014; Lukasik et al., 2006; Thakar et al., 2012; Nameki et al., 2005). HDGF is a mitotic phosphoprotein that acts as a mitogen in different cell types and is associated with organ development and tissue differentiation in the intestine, kidney, liver, and cardiovascular system, but is dispensable in mouse development (Everett et al., 2011; Gallitzendoerfer et al., 2008; Oliver & Al-Awqati, 1998; Enomoto et al., 2002; Cilley et al., 2000; Everett, 2001). Interestingly, HDGF may also be a regulator of ERVs based on the HDGF PWWP domain being a component of an Alu element repeat (Yang & Everett, 2007).

PSIP1 encodes for two tissue dependants differentially expressed protein isoforms, p52 and p75, and is associated with transcriptional regulation of stress genes, alternative splicing, DNA repair, and HIV integration in host genome (Ganapathy et al., 2003; Pradeepa et al., 2012; Daugaard et al., 2012; Ciuffi et al., 2005; Nishizawa et al., 2001). Interestingly, PSIP1 KO lines display homeotic skeletal transformations akin to Hox deficient mice and further *in vitro* transcriptional profiling revealed that Hox gene expression remains unaffected in mouse embryonic fibroblasts but changes in human embryonic kidney cells, suggesting lineage specific Hox gene regulation by PSIP1 (Sutherland et al., 2006).

Most importantly, HDGF and PSIP1 have been demonstrated to associate with both interphase and mitotic chromatin and PSIP1 has also been shown to preferentially bind, using the

PWWP domain, H3K36me3, a transcriptionally active gene marker, marked genome regions (Thakar et al., 2012; Pradeepa et al., 2012; Nishizawa et al., 2001; Van Nuland et al., 2013).

Importantly, both PSIP1 and HDGF were candidates from our ChIP-MS MBF screen and since PWWP domain containing proteins have not been explored in the context of mitotic bookmarking, we aimed to investigate the potential role of PWWP containing proteins, PSIP1 and HDGF, as MBFs. *We hypothesize that HDGF and PSIP1 are retained at a subset of their binding sites throughout mitosis, to ensure the proper regulation of lineage-specific genes in early G1 phase, that contribute to the maintenance of PSC identity.*

The aim is to characterize the self-renewal of HDGF and PSIP1 KO cells and evaluate the effect of mitosis-specific loss of HDGF and PSIP1 on the self-renewal of pluripotent stem cells, using the same experimental design as the prior SOX2 paper. The data generated will be a testament to using the ChIP-MS screen as a confirmed methodology to study other MBFs in PSCs in the context of pluripotency regulation.

2.6.2 Materials & Methods

Cell culture: E14TG2A mouse embryonic stem cells were cultured on 0.1% gelatin coated culture dishes in mESC media: DMEM (Sigma Aldrich, D5796), 15% FBS, 1X non-essential amino acids (Life technologies: 11140-050), 1X glutamax (Life technologies: 35050-061), 1X sodium pyruvate (Life technologies: 11360-070), and 1X betamercaptoethanol (Gibco: 21985-023). Media was further supplemented with 1000 U/mL LIF (Amsbio, AMS-263-100) after filter-sterilization with a 0.22 μ m filter (Sigma) Cells were maintained at 37°C, 5% CO₂ and passage every three days using accutase® (Sigma Aldrich: A6964). PARP1 knockout cells were routinely maintained on a layer of x-ray irradiated mouse embryonic fibroblasts (xMEFs) seeded

at a density of 1×10^6 cells/60 cm². mESCs were pre-plated on gelatin coated dishes for 30 minutes to deplete xMEFs.

Colony initiation cell assay: mESCs were seeded at a density of 250 cells/well onto a well of a 12-well plate with x-MEFs at a density of 1×10^6 cells/plate. For each experiment, cells were seeded in a technical triplicate and cultured for 5 days. At day 5, colonies were fixed in the dish with 250 ul of 4% PFA (Electron microscopy sciences, Cat # 15710) for 1-2 minutes and washed with water. Alkaline phosphatase (AP) staining was performed as described in Sigma AP staining kit (86R-1KT, Sigma). Plates were scanned on EPSON Scanner with 3200 dpi and 24-bit colour and analyzed on ImageJ (Schneider et al., 2012). Dense colonies with intense AP staining were characterized as AP positive (AP+) while the less dense ones with dispersed pink staining around the edges were characterized as mixed colonies.

Mitotic degron constructs: The DNA sequence encoding the peptide corresponding to residues 13-91 of murine cyclin B1 (Kadauke et al.) was subcloned into pCAG-mKO2 using MD primers (Table 1) to generate the mitotic degron construct (MD). The (MD) and inactive mutant MD* (Kadauke et al.) were fused to the C-terminus of pCAG-HDGF-mKO2 and pCAG-PSIP1-mKO2 to create MD and MD* construct.

Statistical analysis: All statistical analyses were performed using Prism 7 (GraphPad) and Microsoft Excel software. Error bars reflect the standard error of mean, unless otherwise stated. The paired two-tailed Student's t-test was used to compute all p-values in cell tracking experiments for comparison of the intensities under distinct channels for individually tracked cells; all other p-values were computed with the unpaired two-tailed student's t-test. Unless otherwise stated error bars represent standard error of the mean and alpha of 0.05 was used as a cut-off for statistical significance.

Western blot: Single-cell suspensions harvested during passage of mESC cultures were lysed in 1X radioimmunoprecipitation assay buffer (RIPA, Sigma) with 1X Protease Inhibitor Cocktail (Roche). The protein concentrations were quantified using the DC Protein Assay kit (Biorad) with the bovine serum albumin standard curve ranging from 0 to 2 mg/ml concentrations. The quantified protein extracts were transferred into 1X NuPAGE LDS Sample Buffer (Thermo Fisher) with 15% Bond-Breaker™ TCEP Solution (Thermo Fisher). The prepared extracts were heated at 95°C for 5 minutes and separated on a 12% polyacrylamide gel at 180V for 45 minutes. The separated protein was transferred onto a PVDF membrane using constant-current electrophoresis at 200 mA for 2 hours, followed by blocking in 5% milk in 1X TBS. The proteins were then blotted with monoclonal antibodies (Table 3) in 3% milk in 1X Tris-buffered saline (TBS) containing 0.1 % Tween20 (1X TBST; Bio Shop) at 4°C overnight. After washing, the blot was incubated with the secondary anti-rabbit HRP-conjugated antibody (1:20,000; Biorad) in 3% milk in 1X TBST for 1 hour at RT, and was developed for 10 minutes using an HRP substrate (1:5 diluted in ultrapure water; Lumina). The blot was visualized using the ChemiDoc™ MP Imaging System (BioRad) with its associated ImageLab analysis software (Biorad).

2.6.3 Results & Discussion

2.6.3.1 HDGF and PSIP1 KO in mouse PSCs results in reduced self-renewal capacity

Our lab had generated HDGF and PSIP1 KO cells via CRISPR-Cas9, targeting regions of the proteins within or downstream of the PWWP domain. Consistent with literature, the KO lines were morphologically identical to wildtype cells, and to understand the effects on pluripotency, our lab uses the alkaline phosphatase, shown to recognize undifferentiated cells, colony-initiating

cell (CIC) assay to evaluate self renewal of these KO lines (Gallitzendoerfer et al., 2008; Sutherland et al., 2006; O'Connor et al., 2008).

Essentially, the CIC assay allows for quantification of self-renewal based on the number of alkaline phosphatase (AP) positive colonies that are generated from initial seeding. Indeed, both HDGF and PSIP1 KO lines can be observed to have decreased alkaline phosphatase colonies in comparison to wildtype cells, suggesting both are involved in the maintenance of pluripotency in mESCs (Figure 5).

2.6.3.2 Mitotic specific degradation of HDGF does not affect self-renewal in mouse PSCs

The CIC assay using HDGF KO cells suggests that HDGF is involved in maintaining self-renewal but whether the mechanism is mitosis specific, requires further experimentation. To test the mitosis specific effect of HDGF, the CIC assay was designed following a previous paper using PARP1 and the mitotic degron (MD) domain of Cdk1-Cyclin B. Constructs with HDGF fused an MD domain and a mutated non-functional MD domain (MD*) were transfected into mouse PSCs and the CIC assay was conducted using the WT, KO, MD, and MD* versions of HDGF, alongside western blots confirming corresponding expected expression patterns. The rescue experiments using HDGF overexpression in the corresponding KO shows a recovery in the number of generated colonies, validating their importance in pluripotency maintenance (Figure 6). However, if HDGF is degrading specifically during mitosis into early G1 phase, by fusing an MD domain (HDGF-MD) to HDGF, the KO line phenotype can be recapitulated but cannot be rescued, by fusing a non-functional mitotic degradation domain of cdk1-cyclin B (HDGF-MD*), suggesting HDGF does not have a mitosis specific pluripotency maintenance function in mESCs (Figure 6). Although there was no significant change in self-renewal when HDGF is degraded in mitosis, it could affect the differentiate potential of the PSCs, which is not

captured by the CIC assay. Furthermore, there is a difference in the number of AP positive colonies during the initial CIC experiments compared to the later experiments, suggesting an epigenetic compensatory mechanism, which could have biased the CIC assay data. A previous study had shown that HDGF knockdown but not KO lead to reduced proliferation, which could be a result of selection overtime, where the KO colonies with higher self-renewal capacity produce larger colonies and are overrepresented during rounds of passaging. Future experiments repeating the CIC assay using HDGF KO lines adapted to naïve pluripotent cell culture conditions, which increases the transcriptional homogeneity of the cell population that could potentially decrease the rate of selection and circumvent the selection of highly proliferative cells within the KO population. Additionally, HDGF expression levels during mitosis is not measured, which is an important control since the MD and MD* constructs may not be functional, which would result in a full rescue.

2.6.3.3 Mitotic specific degradation of PSIP1 in mouse PSCs

The CIC assay for PSIP1 KO cells had suggested that it is involved in self-renewal but, similar to HDGF, if it is mitosis specific is not known. The plan was to conduct the same experiment designed previously with PARP1, from a previous paper, and with HDGF, but although the constructs for WT, MD, and MD* PSIP1 are constructed (Figure 7), the transfection shows a few problematic technical issues. Interestingly, in contrast to HDGF, the PSIP1 MD and MD* transfected mouse PSCs do not show stable expression of the construct over time, even though the pCAG promoter is known to be a strong constitutive promoter, and only maintains expression under strong puromycin selection. Furthermore, even with strong selection, the cultures are extremely unstable and, more importantly, can result in potential silencing of MD and MD* constructs resulting in false positive colonies that are puromycin resistant (Figure 8).

2.6.4 Conclusion

Here, we have eluded to a few points, including further cementing the use of our ChIP-MS screen to reveal putative MBFs. Providing supportive evidence, we have shown PARP1 to be quantifiable dynamically associated with mitotic chromatin, using established ChIP-WB methods we have eluded to a novel MBF, NFYa, and characterized its interaction with PARP1, and using the CIC assay we have showed the absence of quantifiable effect of the PWWP containing putative MBF, HDGF, on *in vitro* self-renewal capacity, a fundamental component of pluripotency. However, further experiments are required to follow up on the NFYa as a MBF and whether the interactions characterized in this study are specific to mitosis. Additionally, regarding the PWWP containing domains, HDGF may contribute to pluripotency states *in vivo* and potentially impact the differentiation capacity of PSCs through its mitosis specific absence, however, PSIP1 may need an experimental re-design due to the issues experienced in the experimental design.

Nonetheless, our results open interesting questions, especially regarding the novel potential MBF, NFYa. Due to the constraints of the experimental design, whether NFYa interactions with other MBFs are mitosis specific is up for debate until the experiment is conducted with a fully mitotically enriched sample. Furthermore, the function of NFYa may be to act as a recruiting factor for other MBFs as suggested with PP2A in a previous study, which would be interesting to test through KO followed by ChIP-seq for NFYa interacting factors, such as PARP1. Moreover, the level of impact mitotic bookmarking actually plays in cellular identity preservation is still in question, even with studies of MBFs, it is difficult to conclude that bookmarking is the sole mechanism by which generational identity information is transmitted.

To establish mitotic bookmarking as a fundamental mechanism of cellular identity transmission, there needs to be a direct connection to nascent transcripts in the M to G1 phase and cellular fate.

2.7 Figures

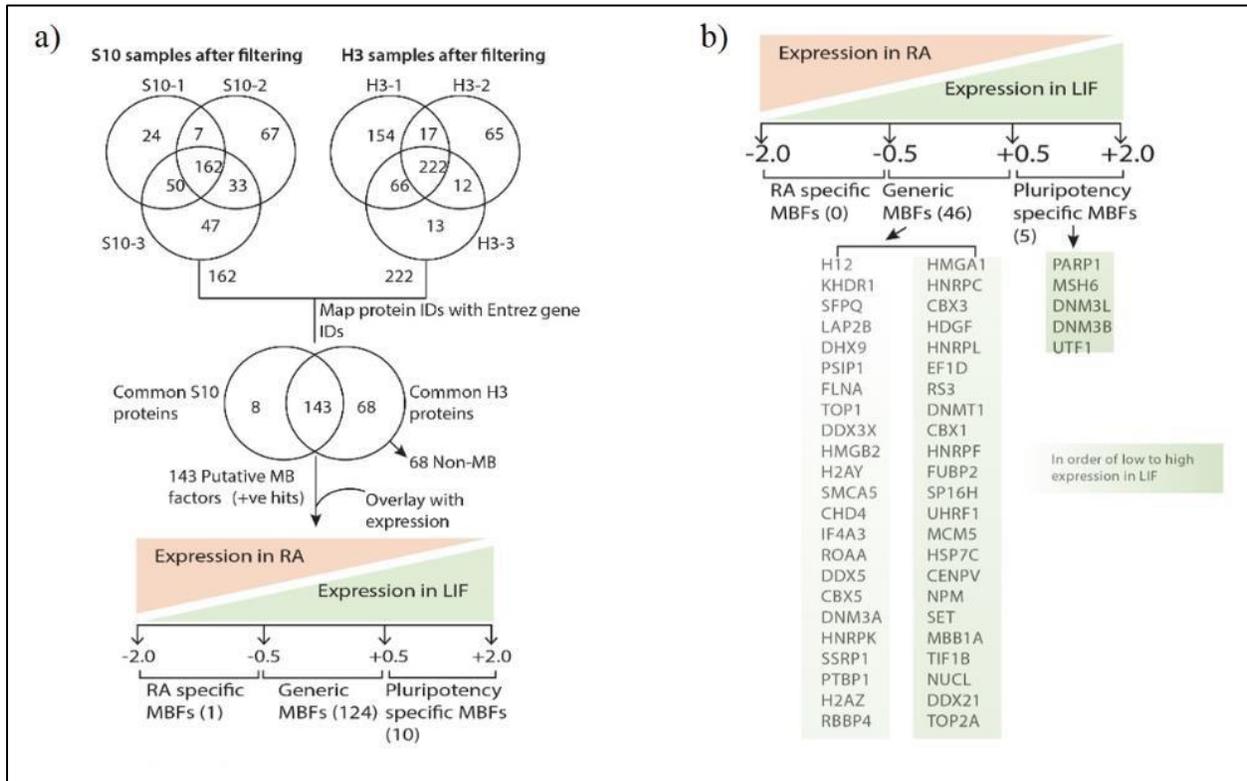


Figure 1. Identification of putative MBFs from ChIP-MS data. A) Figure shows the strategy used to identify and classify putative MBFs based on ChIP-MS data and expression levels in pluripotent conditions. B) List of putative MBFs after filtration. Figure generated by Sonam Bhatia

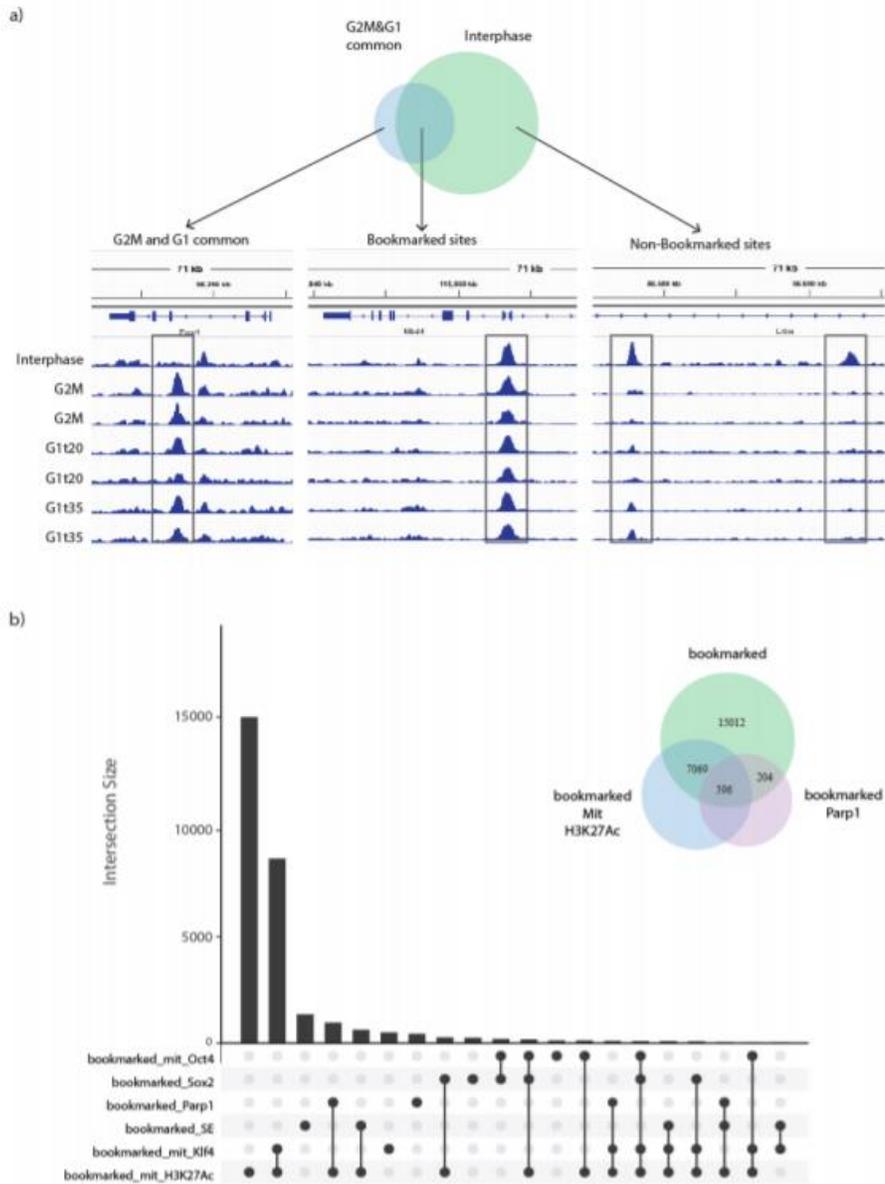


Figure 2. ATAC-seq reveals putatively bookmarked gene loci. a) Venn diagram (top) and representative signal track (bottom) showing the sites common between G2M, G1t20 and G1t25 (G2M&G1 common) and interphase, identifying the bookmarked and non-bookmarked sites. B) UpSet plot showing the relationship between occupancy of bookmarked sites with mitotic specific binding of key pluripotency related factors, epigenetic modification associated bookmarked sites, and a hit identified in ChIP-MS screen. Filled circles represent overlap between the different datasets. Set size indicate the total site of the data set. The inset Venn diagram shows the extensive overlap between mitotic PARP1 binding, bookmarked site, and mitotic H3K27Ac occupancy. (Figure adapted from Sonam Bhatia)

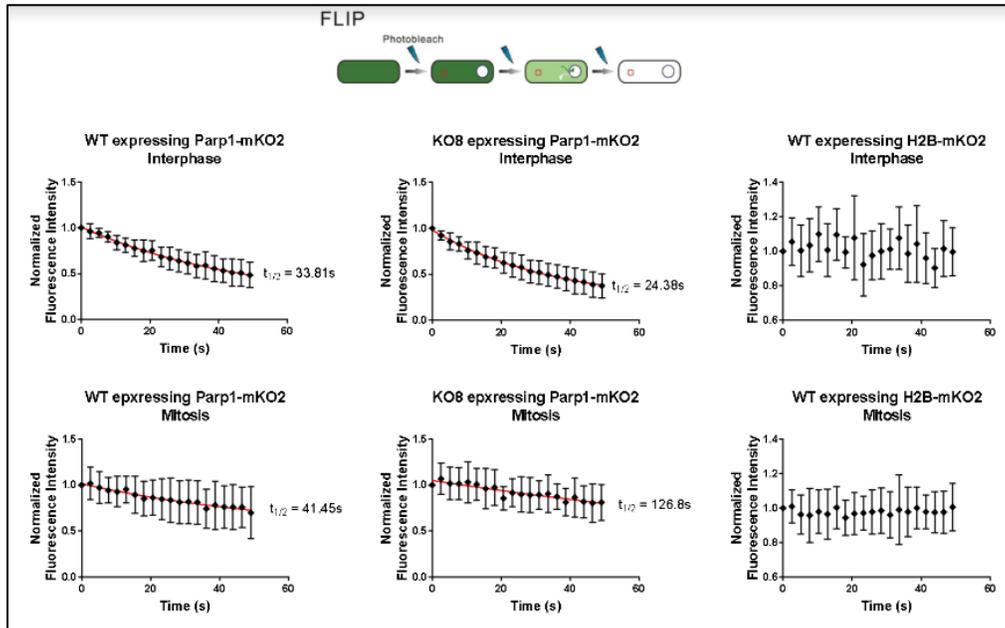
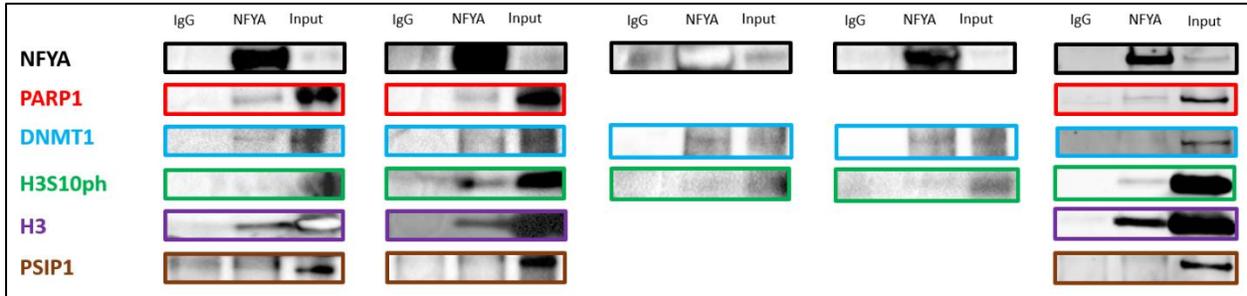


Figure 3. Chromosomal association of PARP1 in interphase vs mitosis. PARP1-mKO2 WT and KO mESCs were cultured in mESC media with LIF and used for FLIP. Figure shows the half-life of PARP1-mKO2, after photobleaching using an Argon laser, in interphase and mitotic cells, using H2B-mKO2 as a control. Figure generated by me and Daisy Deng.

A)



B)

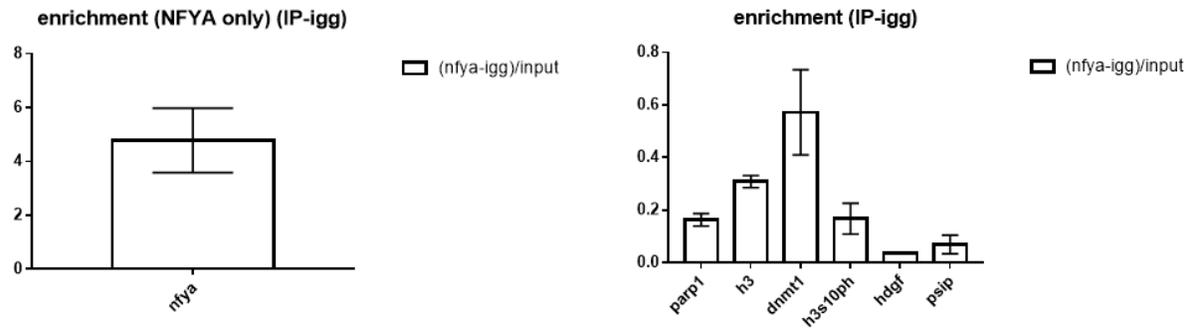


Figure 4. Association of NFYA with Mitotic Chromatin and Bookmarking Factors. E14T WT mESCs were cultured in mESC media with LIF, mitotically enriched for 7 hours using nocodazole, fixed using 1% PFA, and harvested for ChIP. Cross-linked ChIP is performed using the NFYA antibody, IgG, which is used as a control, and protein G Dynabeads. Figure shows the western blot, in which, target antibodies are used to probe for interacting proteins in the NFYA ChIP samples (a). Images were quantified by measuring band intensity, using ImageJ software, and normalizing to both input and IgG controls (b). Graphs were generated using Prism.

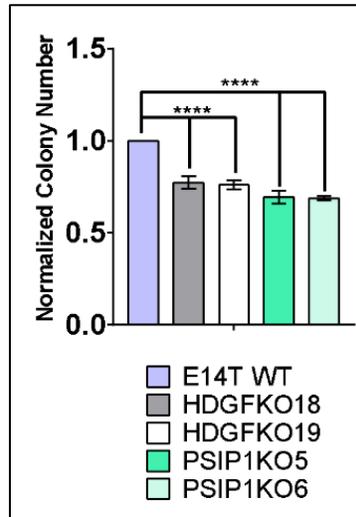


Figure 5. Self-renewal difference upon loss of HDGF and PSIP1. HDGF and PSIP1 KO mESCs were cultured in mESC media with LIF and CIC assay was performed. The colony numbers were normalized to E14T WT and shows a reduction in colony number with the loss of HDGF and PSIP1. Figure generated by Daisy Deng.

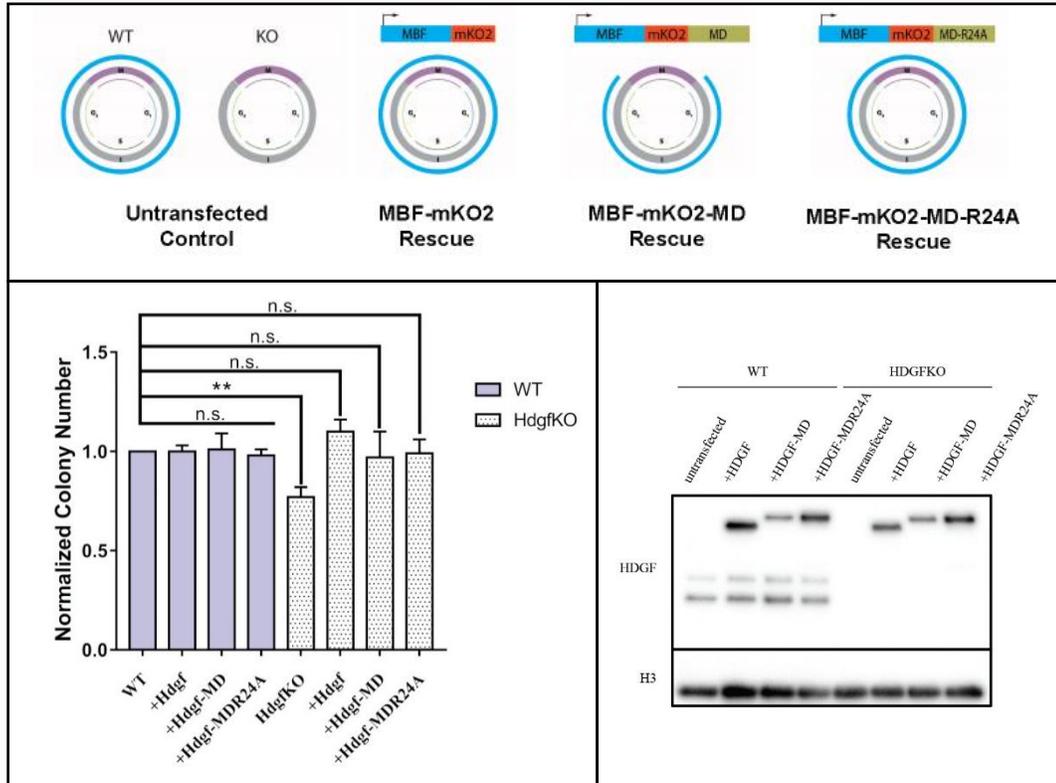


Figure 6. Pluripotency maintenance effects of mitotic specific degradation of HDGF. HDGF KO mESCs were cultured in mESC media with LIF and subsequently transfected with HDGF-mKO2 without an MD domain, with an MD domain (MD), and with a mutated MD domain (MD-R24A). Figure shows the various constructs and controls along with the presence/absence of said vector during the cell cycle (a), the decrease in colony number with mitotic specific degradation of HDGF as well as its rescue with the mutated MD domain (b), and confirmation of the presence of each protein corresponding to the transfected lines (c). (n=3)

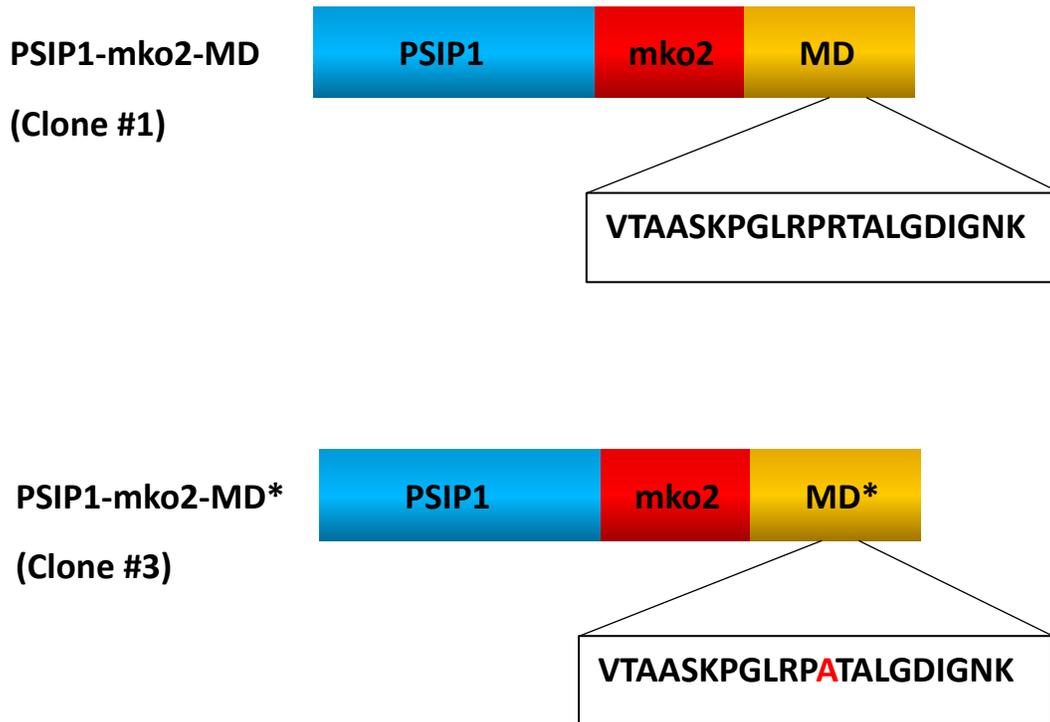


Figure 7. Construction of PSIP1-mko2-MD* plasmid. PSIP1-mko2 and HDGF-mko2-MD* plasmids were digested using AgeI and SacI. Fragments containing PSIP1 within the backbone vector and mko2-MD* were purified and ligated to generate the PSIP1-mko2-MD* plasmid. Image shows amino acid sequence of PSIP1-mko2-MD and PSIP1-mko2-MD* clones, purified using Mini-prep kit.

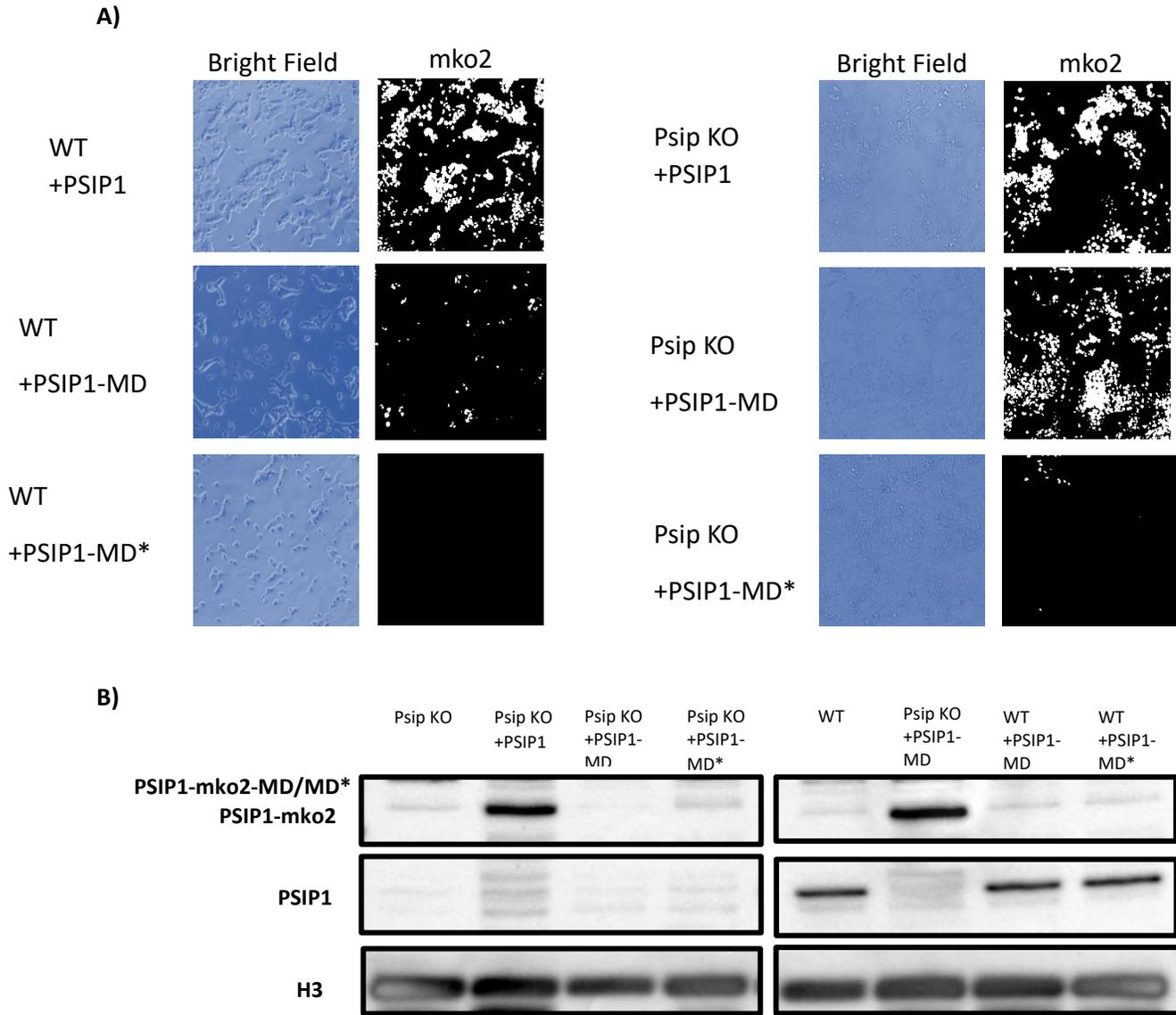


Figure 8. E14T WT and PSIP1-KO cells lines transfected with PSIP1 and MD/MD* constructs. HDGF KO mESCs were cultured in mESC media with LIF and subsequently transfected with PSIP1-mKO2 or PSIP1-mKO2 without an MD domain, with an MD domain (MD), and with a mutated MD domain (MD-R24A). Figure shows bright field (Left) and mko2 fluorescence (Right) following standard passaging (a). Western blot diagram shows PSIP1 expression following standard RIPA buffer based protein extraction. Protein concentration was quantified using Bradford Assay and 15ug of protein was loaded into each corresponding lane. Western blot was divided into two and probed for PSIP1 or H3, used as a loading control (b).

2.8 Tables

Table 1: Primers used for generating MD and MD* constructs

Name	Sequence (5'→3')	Note
CrSprPARP1_gRNA-F	CACCGGGACTTTCCCATCGAACAT	sgRNA for CRISPR/Cas9 mediate KO
CrSprPARP1_gRNA-R	aaacATGTTCGATGGGAAAGTCCC	sgRNA for CRISPR/Cas9 mediate KO
Age1-mKO2_F	GTACCGGTCATGGTGAGTGTGATT	MD fusion
Not1-MD_R	ATATAGCGGCCGCTTAGAATTGTG GTTCGCACACAGG	MD fusion

Table 2: List of antibodies used for western blot and immunoprecipitation

Target	Source	Identifier
PARP1	Abcam	ab6079
NFYa	Santa Cruz	SC-17753
HDGF	Abcam	ab128921
PSIP1	Abcam	A300-847A
H3	Millipore	06-570

References

1. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156 (1981).
2. Thompson, J., Itskovitz, J., Shapiro, S.S., et al. Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145-7 (1998).
3. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379–391 (1998).
4. Mitsui, K. et al. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* 113, 631–642 (2003).
5. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* 9, 625–635 (2007).
6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676 (2006).
7. Bhatia, S., Pilquill, C., Roth-Albin, I. & Draper, J. S. Demarcation of Stable Subpopulations within the Pluripotent hESC Compartment. *PLoS ONE* 8, e57276 (2013).
8. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. B Biol. Sci.* 369, (2014).
9. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* 4, 487–492 (2009).
10. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* 135, 909–918 (2008).
11. Li, M. & Belmonte, J. C. I. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* 18, 180–191 (2017).
12. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* 38, 1027–1032 (2010).
13. Kumari, D. States of Pluripotency: Naïve and Primed Pluripotent Stem Cells. (2016).
14. Loh, K. M., Lim, B. & Ang, L. T. Ex uno plures: molecular designs for embryonic pluripotency. *Physiol. Rev.* 95, 245–295 (2015).
15. Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523 (2008).
16. Tosolini, M. & Jouneau, A. Acquiring Ground State Pluripotency: Switching Mouse Embryonic Stem Cells from Serum/LIF Medium to 2i/LIF Medium. *Methods Mol. Biol.* Clifton NJ 1341, 41–48 (2016).

17. Hackett, J. A. & Surani, M. A. Regulatory Principles of Pluripotency: From the Ground State Up. *Cell Stem Cell* 15, 416–430 (2014).
18. Boyer, L. A., Mathur, D. & Jaenisch, R. Molecular control of pluripotency. *Differ. Gene Regul.* 16, 455–462 (2006).
19. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* 132, 1049–1061 (2008).
20. Orkin, S. H. et al. The Transcriptional Network Controlling Pluripotency in ES Cells. *Cold Spring Harb. Symp. Quant. Biol.* 73, 195–202 (2008).
21. Loh, Y.-H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* 38, 431–440 (2006).
22. Chew, J.-L. et al. Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Mol. Cell. Biol.* 25, 6031–6046 (2005).
23. Ng, H.-H. & Surani, M. A. The transcriptional and signalling networks of pluripotency. *Nat. Cell Biol.* 13, 490–496 (2011).
24. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444, (2006).
25. Khalfallah, O., Rouleau, M., Barbry, P., Bardoni, B. & Lalli, E. Dax-1 knockdown in mouse embryonic stem cells induces loss of pluripotency and multilineage differentiation. *Stem Cells* 27, 1529–1537 (2009).
26. Ruan, Y. et al. Nac1 promotes self-renewal of embryonic stem cells through direct transcriptional regulation of c-Myc. *Oncotarget* 8, 47607–47618 (2017).
27. Morey, L., Santanach, A. & Croce, L. D. Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. *Mol. Cell. Biol.* 35, 2716–2728 (2015).
28. Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T. & Pasini, D. Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res.* 40, 3403–3418 (2012).
29. Shi, Y., Seto, E., Chang, L.-S. & Shenk, T. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* 67, 377–388 (1991).
30. Seto, E., Shi, Y. & Shenk, T. YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* 354, 241–245 (1991).
31. Gordon, S., Akopyan, G., Garban, H. & Bonavida, B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* 25, 1125–1142 (2006).
32. Donohoe, M. E. et al. Targeted Disruption of Mouse Yin Yang 1 Transcription Factor Results in Peri-Implantation Lethality. *Mol. Cell. Biol.* 19, 7237–7244 (1999).

33. Morey, L. & Helin, K. Polycomb group protein-mediated repression of transcription. *Trends Biochem. Sci.* 35, 323–332 (2010).
34. Boyer, L. A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353 (2006).
35. O'Carroll, D. et al. The Polycomb-Group Gene *Ezh2* Is Required for Early Mouse Development. *Mol. Cell. Biol.* 21, 4330–4336 (2001).
36. Voncken, J. W. et al. *Rnf2* (*Ring1b*) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci. U. S. A.* 100, 2468–2473 (2003).
37. Rajasekhar, V. K. & Begemann, M. Concise Review: Roles of Polycomb Group Proteins in Development and Disease: A Stem Cell Perspective. *STEM CELLS* 25, 2498–2510 (2007).
38. Ballabeni, A. et al. Cell cycle adaptations of embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19252–19257 (2011).
39. Vernon, E. G. & Gaston, K. *Myc* and *YY1* mediate activation of the *Surf-1* promoter in response to serum growth factors. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* 1492, 172–179 (2000).
40. Kim, J. D., Faulk, C. & Kim, J. Retroposition and evolution of the DNA-binding motifs of *YY1*, *YY2* and *REX1*. *Nucleic Acids Res.* 35, 3442–3452 (2007).
41. Bushmeyer, S., Park, K. & Atchison, M. L. Characterization of Functional Domains within the Multifunctional Transcription Factor, *YY1*. *J. Biol. Chem.* 270, 30213–30220 (1995).
42. Nguyen, N., Zhang, X., Olashaw, N. & Seto, E. Molecular Cloning and Functional Characterization of the Transcription Factor *YY2*. *J. Biol. Chem.* 279, 25927–25934 (2004).
43. Tahmasebi, S. et al. Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of *YY2*. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12360–12367 (2016).
44. Hosler, B. A., LaRosa, G. J., Grippo, J. F. & Gudas, L. J. Expression of *REX1*, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. *Mol. Cell. Biol.* 9, 5623–5629 (1989).
45. Rogers, M. B., Hosler, B. A. & Gudas, L. J. Specific expression of a retinoic acid-regulated, zinc-finger gene, *REX1*, in preimplantation embryos, trophoblast and spermatocytes. *Development* 113, 815–824 (1991).
46. Masui, S. et al. *REX1/Zfp42* is dispensable for pluripotency in mouse ES cells. *BMC Dev. Biol.* 8, 45 (2008).
47. Tanaka, T. S. Transcriptional heterogeneity in mouse embryonic stem cells. *Reprod. Fertil. Dev.* 21, 67–75 (2009).
48. Shi, W. et al. Regulation of the pluripotency marker *REX1* by *Nanog* and *Sox2*. *J Biol Chem* 281, (2006).

49. Guallar, D. *et al.* Expression of endogenous retroviruses is negatively regulated by the pluripotency marker REX1/Zfp42. *Nucleic Acids Res.* 40, 8993–9007 (2012).
50. Yang, B. X. *et al.* Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* 163, 230–245
51. Rowe, H. M. *et al.* TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* 23, 452–461 (2013).
52. Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. *Front. Oncol.* 4, 14 (2014).
53. Schlesinger, S. & Goff, S. Retroviral Transcriptional Regulation and Embryonic Stem Cells: War and Peace. *Molecular and Cellular Biology* 35, 770–777 (2015).
54. Gifford, R. & Tristem, M. The Evolution, Distribution and Diversity of Endogenous Retroviruses. *Virus Genes* 26, 291–315 (2003).
55. Stocking, C. & Kozak, C. A. Murine endogenous retroviruses. *Cell. Mol. Life Sci. CMLS* 65, 3383–3398 (2008).
56. Ono M, Kawakami M, Takezawa T. A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* 15, 8725–37 (1987).
57. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet* 7, 149–73 (2006).
58. Macfarlane C, Simmonds P. Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. *J Mol Evol* 59: 642–56 (2004).
59. Belshaw R, Katzourakis A, Paces J, Burt A, Tristem M. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol Biol Evol* 22, 814–7 (2005).
60. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci USA* 101, 4894–9 (2004).
61. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E. At least 50% of human-specific HERV-K(HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. *J Virol* 80, 10752–62 (2006).
62. Eickbush TH. Telomerase and retrotransposons: which came first? *Science* 277: 911–2 (1997).
63. Brandt J, Schrauth S, Veith AM, Froschauer A, Haneke T, Schultheis C, Gessler M, Leimeister C, Volff JN. Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. *Gene* 345: 101–11 (2005).

64. Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci USA* 103, 17608–13 (2006).
65. Ponferrada VG, Mauck BS, Wooley DP. The envelope glycoprotein of human endogenous retrovirus HERV-W induces cellular resistance to spleen necrosis virus. *Arch Virol* 148, 659–75 (2003).
66. Rasheed, S. Retroviruses and Oncogenes. *The Retroviridae*, 293–408 (1995).
67. Contreras-Galindo R, Kaplan MH, Leissner P, Verjat T, Ferlenghi I, Bagnoli F, Giusti F, Dosik MH, Hayes DF, Gitlin SD, Markovitz DM. Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *J Virol* 82, 9329–36 (2008).
68. Ruprecht K, Mayer J, Sauter M, Roemer K, Mueller-Lantzsch N. Endogenous retroviruses and cancer. *Cell Mol Life Sci* 65, 3366–82 (2008).
69. Pérez-Palacios, R. *et al.* In Vivo Chromatin Targets of the Transcription Factor Yin Yang 2 in Trophoblast Stem Cells. *PLoS ONE* 11, e0154268 (2016).
70. Climent, M. *et al.* Functional Analysis of REX1 During Preimplantation Development. *Stem Cells Dev.* 22, 459–472 (2013).
71. Zhang, J.-Z. *et al.* Screening for Genes Essential for Mouse Embryonic Stem Cell Self-Renewal Using a Subtractive RNA Interference Library. *STEM CELLS* 24, 2661–2668 (2006).
72. Kim, J. D. *et al.* REX1/Zfp42 as an epigenetic regulator for genomic imprinting. *Hum. Mol. Genet.* 20, 1353–1362 (2011).
73. Scotland, K. B., Chen, S., Sylvester, R. & Gudas, L. J. ANALYSIS OF REX1 (ZFP42) FUNCTION IN EMBRYONIC STEM CELL DIFFERENTIATION. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 238, 1863–1877 (2009).
74. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* 8, 2281–2308 (2013)
75. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotech* 32, 347–355 (2014).
76. Bushmeyer, S., Park, K. & Atchison, M. L. Characterization of Functional Domains within the Multifunctional Transcription Factor, YY1. *J. Biol. Chem.* 270, 30213–30220 (1995).
77. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 462, 595–601 (2009).
78. Ruiz, S. *et al.* A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Curr Biol* 21, 45–52 (2011).

79. Amon, A., Tyers, M., Futcher, B., and Nasmyth, K. Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell* 74, 993-1007 (1993).
80. Nigg, E.A. Mitotic kinases as regulators of cell division and its checkpoints. *Nature reviews Molecular cell biology* 2, 21-32 (2001).
81. Breeden, L. L. Periodic transcription: a cycle within a cycle. *Curr. Biol.* 13, R31-38 (2003).
82. Resnitzky, D., Gossen, M., Bujard, H. & Reed, S. I. Acceleration of the G1/S phase transition by expression of cyclins D1 and E with an inducible system. *Mol. Cell. Biol.* 14, 1669–1679 (1994).
83. Jeffrey, P. D. et al. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 376, 313–320 (1995).
84. Ballabeni, A. et al. Cell cycle adaptations of embryonic stem cells. *PNAS* 108, 19252–19257 (2011).
85. Dalton, S. Linking the Cell Cycle to Cell Fate Decisions. *Trends in Cell Biology* 25, 592–600 (2015).
86. Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D., and Dalton, S. Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* 21, 8320-8333 (2002).
87. White, J., and Dalton, S. Cell cycle control of embryonic stem cells. *Stem cell reviews* 1, 131-138 (2005).
88. Maiorano, D., Moreau, J., and Mechali, M. XCDT1 is required for the assembly of pre-replicative complexes in *Xenopus laevis*. *Nature* 404, 622-625 (2000).
89. Prasanth, K.V., Sacco-Bubulya, P.A., Prasanth, S.G., and Spector, D.L. Sequential entry of components of the gene expression machinery into daughter nuclei. *Molecular biology of the cell* 14, 1043-1057 (2003).
90. Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* 12, 2245–2262 (1998).
91. Faast, R. et al. Cdk6-cyclin D3 activity in murine ES cells is resistant to inhibition by p16(INK4a). *Oncogene* 23, 491–502 (2004).
92. Li, V. C., Ballabeni, A. & Kirschner, M. W. Gap 1 phase length and mouse embryonic stem cell self-renewal. *PNAS* 109, 12550–12555 (2012).
93. Neganova, I. & Lako, M. G1 to S phase cell cycle transition in somatic and embryonic stem cells. *Journal of Anatomy* 213, 30–44 (2008).
94. Coronado, D. et al. A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Research* 10, 118–131 (2013).

95. Naumova, N. et al. Organization of the Mitotic Chromosome. *Science* 342, 948–953 (2013).
96. Bernardi, G. Genome Organization and Chromosome Architecture. *Cold Spring Harb Symp Quant Biol* 80, 83–91 (2015).
97. Oomen, M. E. & Dekker, J. Epigenetic characteristics of the mitotic chromosome in 1D and 3D. *Critical Reviews in Biochemistry and Molecular Biology* 52, 185–204 (2017).
98. Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. Organization of the mitotic chromosome. *Science (New York, NY)* 342, 948-953 (2013).
99. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380 (2012).
100. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293 (2009).
101. Markaki, Y., Gunkel, M., Schermelleh, L., Beichmanis, S., Neumann, J., Heidemann, M., Leonhardt, H., Eick, D., Cremer, C., and Cremer, T. Functional nuclear organization of transcription and DNA replication: a topographical marriage between chromatin domains and the interchromatin compartment. *Cold Spring Harbor symposia on quantitative biology* 75, 475-492 (2010).
102. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381-385 (2012).
103. Liu, Y. et al. Transcriptional landscape of the human cell cycle. *PNAS* 114, 3473–3478 (2017).
104. Xu, J. et al. Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nat Genet* 49, 377–386 (2017).
105. Juan, G., Pan, W., and Darzynkiewicz, Z. DNA segments sensitive to singlestrand-specific nucleases are present in chromatin of mitotic cells. *Experimental cell research* 227, 197-202 (1996).
106. Michelotti, E.F., Sanford, S., and Levens, D. Marking of active genes on mitotic chromosomes. *Nature* 388, 895-899 (1997).
107. Hsiung, C.C., Bartman, C.R., Huang, P., Ginart, P., Stonestrom, A.J., Keller, C.A., Face, C., Jahn, K.S., Evans, P., Sankaranarayanan, L., et al. A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. *Genes & Development* 30, 1423-1439 (2016).
108. Hsiung, C.C., Morrissey, C.S., Udugama, M., Frank, C.L., Keller, C.A., Baek, S., Giardine, B., Crawford, G.E., Sung, M.H., Hardison, R.C., et al. Genome accessibility is

- widely preserved and locally modulated during mitosis. *Genome research* 25, 213- 225 (2015).
109. Sarge, K.D., and Park-Sarge, O.K. Mitotic bookmarking of formerly active genes: keeping epigenetic memories from fading. *Cell cycle* 8, 818-823 (2009).
 110. Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Molecular cell* 36, 970-983 (2009).
 111. Kadauke, S. et al. Tissue-Specific Mitotic Bookmarking by Hematopoietic Transcription Factor GATA1. *Cell* 150, 725–737 (2012).
 112. Caravaca, J. M. et al. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev.* 27, 251–260 (2013).
 113. Lodhi, N., Kossenkov, A. V. & Tulin, A. V. Bookmarking promoters in mitotic chromatin: poly(ADP-ribose)polymerase-1 as an epigenetic mark. *Nucl Acids Res* 42, 7028–7038 (2014).
 114. Festuccia, N. et al. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat Cell Biol* 18, 1139–1148 (2016).
 115. Deluz, C. et al. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev.* 30, 2538–2550 (2016).
 116. Liu, Y. et al. Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Reports* 19, 1283–1293 (2017).
 117. Zhao, R., Nakamura, T., Fu, Y., Lazar, Z., and Spector, D.L. Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. *Nature cell biology* 13, 1295-1304 (2011).
 118. Rose, J. T., Boyd, J. R., Gordon, J. A., Kang, M., Moskovitz, E., Bouffard, N. A., et al. Mitotic gene bookmarking by RUNX1 contributes to stabilization of the normal mammary epithelial phenotype. *bioRxiv*, (2019).
 119. Teves, S. S. et al. A dynamic mode of mitotic bookmarking by transcription factors. *eLife* 5, e22280 (2016).
 120. Hendzel, M. J. et al. Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* 106, 348–360 (1997).
 121. Terranova, C. et al. Global Developmental Gene Programming Involves a Nuclear Form of Fibroblast Growth Factor Receptor-1 (FGFR1). *PLOS ONE* 10, e0123380 (2015).
 122. Bhatia, S. Transcription Regulation of Pluripotency and Cell Fate. PhD thesis, (2017).
 123. Amé, J., Spenlehauer, C. & de Murcia, G. The PARP superfamily. *BioEssays* 26, 882–893 (2004).

124. Roper, S. J. et al. ADP-ribosyltransferases PARP1 and Parp7 safeguard pluripotency of ES cells. *Nucleic Acids Res.* 42, 8914–8927 (2014).
125. Ji, Y. & Tulin, A. V. The roles of PARP1 in gene control and cell differentiation. *Curr. Opin. Genet. Dev.* 20, 512–518 (2010).
126. Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., et al. Sequence-Specific Transcription Factor NF-Y Displays Histone-like DNA Binding and H2B-like Ubiquitination. *Cell* 152, 132-143 (2013).
127. Fleming, J. D., Pavesi, G., Benatti, P., Imbriano, C., Mantovani, R., & Struhl, K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome Research* 23, 1195-1209 (2013).
128. Chen, S., Yu, X., Lei, Q., Ma, L., & Guo, D. The sumoylation of zinc-fingers and homeoboxes 1 (ZHX1) by ubc9 regulates its stability and transcriptional repression activity. *Journal of Cellular Biochemistry* 114, 2323-2333 (2013).
129. Arampatzi, P., Gialitakis, M., Makatounakis, T., & Papamatheakis, J. Gene-specific factors determine mitotic expression and bookmarking via alternate regulatory elements. *Nucleic Acids Research*, 41, 2202-2215 (2013).
130. Benatti, P., Dolfini, D., Viganò, A., Ravo, M., Weisz, A., & Imbriano, C. Specific inhibition of NF-Y subunits triggers different cell proliferation defects. *Nucleic Acids Research* 39, 5356-5368 (2011).
131. Qin, S. & Min, J. Structure and function of the nucleosome-binding PWWP domain. *Trends Biochem. Sci.* 39, 536–547 (2014).
132. Lukasik, S. M. et al. High-resolution structure of the HDGF PWWP domain: A potential DNA binding domain. *Protein Sci. Publ. Protein Soc.* 15, 314–323 (2006).
133. Thakar, K. et al. Interaction of HRP-2 isoforms with HDGF: chromatin binding of a specific heteromer. *FEBS J.* 279, 737–751 (2012).
134. Nameki, N. et al. Solution structure of the PWWP domain of the hepatoma-derived growth factor family. *Protein Sci. Publ. Protein Soc.* 14, 756–764 (2005).
135. Everett, A. D., Yang, J., Rahman, M., Dulloor, P. & Brautigan, D. L. Mitotic phosphorylation activates hepatoma-derived growth factor as a mitogen. *BMC Cell Biol.* 12, 15–15 (2011).
136. Gallitzendoerfer, R. et al. Hepatoma-derived growth factor (HDGF) is dispensable for normal mouse development. *Dev. Dyn.* 237, 1875–1885 (2008).
137. Oliver, J. A. & Al-Awqati, Q. An endothelial growth factor involved in rat renal development. *J. Clin. Invest.* 102, 1208–1219 (1998).
138. Enomoto, H., Yoshida, K., Kishima, Y., Okuda, Y. & Nakamura, H. Participation of hepatoma-derived growth factor in the regulation of fetal hepatocyte proliferation. *J. Gastroenterol.* 37 Suppl 14, 158–161 (2002).

139. Cilley, R. E., Zgleszewski, S. E. & Chinoy, M. R. Fetal lung development: airway pressure enhances the expression of developmental genes. *J. Pediatr. Surg.* 35, 113–118; discussion 119 (2000).
140. Everett, A. D. Identification, cloning, and developmental expression of hepatoma-derived growth factor in the developing rat heart. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 222, 450–458 (2001).
141. Yang, J. & Everett, A. D. Hepatoma derived growth factor binds DNA through the N-terminal PWWP domain. *BMC Mol. Biol.* 8, 101–101 (2007).
142. Ganapathy, V., Daniels, T. & Casiano, C. A. LEDGF/p75: a novel nuclear autoantigen at the crossroads of cell survival and apoptosis. *Autoimmun. Rev.* 2, 290–297 (2003).
143. Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R. & Bickmore, W. A. PSIP1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet.* 8, e1002717 (2012).
144. Daugaard, M. et al. LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat. Struct. Mol. Biol.* 19, 803–810 (2012).
145. Ciuffi, A. et al. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11, 1287–1289 (2005).
146. Nishizawa, Y., Usukura, J., Singh, D. P., Chylack, L. T. J. & Shinohara, T. Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens epithelium-derived growth factor (ledgf/p75) and p52, in the nucleus. *Cell Tissue Res.* 305, 107–114 (2001).
147. Sutherland, H. G. et al. Disruption of Ledgf/PSIP1 results in perinatal mortality and homeotic skeletal transformations. *Mol. Cell. Biol.* 26, 7201–7210 (2006).
148. Van Nuland, R. et al. Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin* 6, 12 (2013).
149. O'Connor, M. D. et al. Alkaline Phosphatase-Positive Colony Formation Is a Sensitive, Specific, and Quantitative Indicator of Undifferentiated Human Embryonic Stem Cells. *STEM CELLS* 26, 1109–1116 (2008).
150. Hackett, J. A., Kobayashi, T., Dietmann, S., & Surani, M. A. Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. *Stem Cell Reports* 8, 1645–1658 (2017).
151. Wang, J. et al. YY1 Positively Regulates Transcription by Targeting Promoters and Super-Enhancers through the BAF Complex in Embryonic Stem Cells. *Stem Cell Reports*, 10, 1324–1339 (2018).