

REX1 IN PLURIPOTENT STEM CELLS

THE ROLE AND MOLECULAR MECHANISMS OF REX1 IN PLURIPOTENT STEM
CELLS

By AMANDA HRENCZUK, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Degree
Master of Science

McMaster University © Copyright by Amanda Hrenczuk, August 2017

M.Sc. Thesis – A. Hrenczuk; McMaster University – Biochemistry & Biomedical Sciences

McMaster University MASTER OF SCIENCE (2017) Hamilton, Ontario (Biochemistry
& Biomedical Sciences)

TITLE: The Role and Molecular Mechanisms of Rex1 in Pluripotent Stem Cells

AUTHOR: Amanda Hrenczuk, B.Sc. (Hon., McMaster University)

SUPERVISOR: Jonathan Draper, Ph.D.

NUMBER OF PAGES: xii, 83

ABSTRACT

Pluripotent stem cells (PSCs) are unique in their capability to self-renew and differentiate into cell types of all three embryonic germ layers. Since their discovery, PSCs have become an indispensable tool for modeling development, disease onset/progression, and drug discovery. The pluripotent state is known to be regulated by a core network of transcription factors including Oct4, Sox2, and Nanog. However, the roles of other contributing transcription factors remain understudied. Our research focused on defining the roles and molecular mechanisms of Rex1, a zinc finger transcription factor whose expression is strongly correlated with the pluripotent state. Attempts by our lab to elucidate the role of Rex1 in embryonic stem cells (ESCs) revealed the presence of two smaller protein products that result from the initiation of translation at downstream start codons within the *REX1* open reading frame. We hypothesized that the full-length Rex1 protein and its shorter alternative translation isoforms were acting to regulate the expression of lineage-determining genes in PSCs. To evaluate this hypothesis, we generated mouse embryonic stem cell (mESC) lines expressing FLAG-tagged versions of the full-length Rex1 protein, and its isoforms, from the endogenous locus. Through the use of these lines, we demonstrated the formation of multiple Rex1 isoforms by alternative translation, a novel observation that has yet to be reported. Furthermore, our results indicate that Rex1 is a negative regulator of differentiation-related genes and endogenous retroviral elements, suggesting Rex1 is acting to maintain the tightly regulated transcriptional network of pluripotency, while also maintaining genomic integrity through the repression of repetitive elements.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr. Jonathan Draper. I could not have accomplished what I have if you hadn't given me this opportunity and believed in my potential. Thank you for helping develop my skills and shaping me into the scientist I have become today. I began this journey hoping to learn as much as I possibly could, and can definitely say that this goal was achieved. Thank you for allowing me to contribute to numerous projects varying from breast cancer research to the maintenance of the pluripotent identity. It has been a great learning process.

I would also like to thank my committee members Dr. Bradley Doble, Dr. Kristin Hope, and Dr. Mathieu Lupien. Thank you for taking time out of your busy schedules to provide me with encouragement and experimental insights. Your valuable feedback and guidance over the last two years has contributed greatly to the development of this project.

Thank you to past and present members of the Draper lab, especially Sonam, for your continuous support, contributions, and guidance. I have enjoyed working with you all for the past three years and am grateful for the friendships that I have developed with each of you in this time. I would like to extend this thank you to all members of the Stem Cell and Cancer Research Institute (SCC-RI). Thank you to all whom I have shared numerous meaningful conversations with, laughs, and have developed friendships with.

A special thank you extends to Steven Moreira. Thank you for being my support system over these past two years. I have learned so much from you, and will always be grateful for your support and guidance.

A heartfelt thank you goes to my mother and father for their unconditional love and support. I could not have accomplished all of this without you. Thank you for always believing in me and giving me the opportunity to pursue my dreams. Words cannot express how thankful I am for having you both by my side all these years.

A very special thank you extends to my brother. You have been a huge support system for me over the years and until this day continue to be my inspiration. I wouldn't be the person I am today, if not for you.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS.....	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS.....	x
DECLARATION OF ACADEMIC ACHIEVEMENT	xii
INTRODUCTION	1
1. Pluripotent Stem Cells.....	1
1.1 Mouse Pluripotent Stem Cells.....	1
1.1.1 Derivation.....	1
1.1.2 Functional Assessment of Self-Renewal and Pluripotency.....	3
1.1.3 Heterogeneity in Pluripotent Stem Cell Populations.....	5
2. Transcriptional Networks Underlying the Pluripotent State	6
2.1 Core Transcriptional Network of Pluripotent Stem Cells	6
2.2 Extending the Transcriptional Network in Pluripotent Stem Cells	7
3. Expression and Function of Rex1 in Pluripotent Stem Cells	11
3.1 Rex1 Expression in Pluripotent Stem Cells	11
3.2 Rex1 Function in Pluripotent Stem Cells	12
3.2.1 Reprogramming of X-inactivation	12
3.2.2 Epigenetic Regulation and Genomic Imprinting.....	13
3.2.3 Differentiation of Visceral Endoderm.....	15
3.2.4 Regulation of Endogenous Retroviral Elements	16
4. Project Rationale and Hypothesis.....	19
MATERIALS AND METHODS.....	21
1. Cell culture	21
2. Adaptation of mESCs to LIF2i.....	21
3. Rex1 CRISPR design and construction.....	22
4. Generation of the Rex1 targeting vector	22
5. Generation of endogenously 3XFLAG-epitope tagged Rex1 mESCs	23
6. CRISPR-mediated knockout of Rex1 in mESCs.....	23
7. Chromatin immunoprecipitation sequencing.....	24
8. Antibodies.....	25
9. Cell lysate preparation	26

10. Western blot analyses	27
11. Quantitative RT-PCR	27
12. Statistical analyses	29
13. Generation of plasmids for characterization of Rex1 functional domains and individual isoforms	30
14. Assessment of REX1 isoform expression in HEK293Ts	31
15. Assessment of Rex1 expression in various culture conditions.....	32
16. Immunostaining and high content imaging	32
RESULTS	33
1. Identification of novel Rex1 isoforms in pluripotent stem cells.....	33
2. CRISPR-mediated homologous recombination to epitope tag Rex1 isoforms.....	36
3. Detection of endogenous protein expression of epitope-tagged Rex1 isoforms....	38
4. Mapping the global genomic binding sites of epitope-tagged Rex1 isoforms.....	41
5. ChIP-seq reveals a conserved role of Rex1 in the regulation of endogenous retroviruses in human and mouse embryonic stem cells.....	44
6. Disruption of Rex1 expression via CRISPRs results in the dysregulation of Rex1 target genes	47
7. Characterizing the role of Rex1 isoforms in mESCs	53
DISCUSSION	56
FUTURE DIRECTIONS	62
CONCLUSION.....	66
REFERENCES	68

LIST OF FIGURES

Figure 1: Retrotransposon mediated duplication of Yy1 in placental mammals resulted in the derivation of two Yy1 related proteins, Yy2 and Rex1	10
Figure 2: Evaluating the expression of <i>REXI</i> mutants in hESCs	34
Figure 3: Missense mutations disrupt the downstream translation initiation sites the human <i>REXI</i> open reading frame	35
Figure 4: Sequence alignment of the human and mouse Rex1 protein.....	36
Figure 5: Schematic representation of the targeting strategy used for the generation of C-terminal Rex1 3XFLAG epitope-tagged mESCs.....	38
Figure 6: Detection of endogenously 3XFLAG epitope-tagged Rex1 in isolated clones..	39
Figure 7: Evaluating protein expression levels of endogenously C-terminally 3XFLAG tagged Rex1	40
Figure 8: Immunofluorescence staining of wildtype and Rex1-FLAG mESCs	41
Figure 9: Overview of genomic annotation, Gene Ontology enrichment and motif discovery analyses performed on murine Rex1 ChIP-seq data	43
Figure 10: Overview of Rex1 association with ERVs in mESCs.....	45
Figure 11: Overview of REX1 association with ERVs in hESCs	45
Figure 12: Assessment of Yy1, Yy2 and Rex1 association with ERVs in mESCs	47
Figure 13: Schematic representation of the targeting approach used for the generation of C-terminal Rex1 KO cell lines.....	49
Figure 14: Analysis of Rex1 target gene expression in E14T WT, Rex1 KO C15 and Rex1 KO C25 mESCs	51
Figure 15: Analysis of ERV expression in E14T WT, Rex1 KO C15 and Rex1 KO C25 mESCs.....	51
Figure 16: Schematic representation of targeting approach used for the generation of N-terminal Rex1 KO cell lines.....	52
Figure 17: Assessment of Rex1-FLAG expression in LIF2i versus serum conditions.....	54
Figure 18: Generation of mutant human REX1 expression vectors for the characterization of isoforms	55
Figure 19: Predicted binding of RNA-binding proteins to murine Rex1 mRNA	62

SUPPLEMENTAL FIGURES

Figure S1: Optimization of ChIP-seq analysis parameters for the identification of peaks in murine Rex1 data set.....	67
--	----

LIST OF TABLES

Table 1: Guide RNA oligos used for the construction of Rex1 CRISPR constructs	22
Table 2: In-Fusion® primers for generation of the Rex1 targeting vector	23
Table 3: qRT-PCR primers for Rex1 genic targets.....	28
Table 4: qRT-PCR primers and Universal Probe Library (UPL) probe numbers for Rex1 ERV targets.....	29
Table 5: Primers used for the generation of hsREX1 mutant constructs	31

LIST OF ABBREVIATIONS

Cas9	CRISPR associated protein 9
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CIC	Colony initiating cell
CRISPR	Clustered regularly interspaced short palindromic repeats
DSB	Double stranded break
EB	Embryoid body
EC	Embryonic carcinoma
EpiSC	Epiblast stem cell
ERV	Endogenous retrovirus
ESC	Embryonic stem cells
GREAT	Genomic Regions Enrichment of Annotations Tool
GO	Gene ontology
H3K9me3	Trimethylation of lysine 9 on Histone H3
H3K27me3	Trimethylation of lysine 27 on Histone H3
HDR	Homology directed repair
HERVH	Human endogenous retrovirus subfamily H
hESC	Human embryonic stem cell
hPSC	Human pluripotent stem cell
IAP	Intracisternal A particle
ICM	Inner cell mass
Indel	Insertion/deletion
iPSC	Induced pluripotent stem cell
KO	Knockout
KRAB-ZFP	Krüppel-associated box domain-containing zinc finger protein

LIF	Leukemia inhibitory factor
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
MEF	Mouse embryonic fibroblast
mESC	Mouse embryonic stem cell
mPSC	Mouse pluripotent stem cell
mRNA	Messenger RNA
muERV-L	Murine endogenous retrovirus
musD	Mouse type-D element
NHEJ	Non-homologous end joining
NRT	No reverse transcriptase
PAM	Protospacer adjacent motif
PcG	Polycomb group
PSC	Pluripotent stem cell
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RA	Retinoic acid
RBP	RNA-binding protein
Rex1	Reduced expression 1
RT	Reverse transcriptase
SINE	Short interspersed nuclear element
sgRNA	Single-guide RNA
TBP	TATA binding protein
Tg	Transgene
TSS	Transcriptional start site
Wt	Wildtype
Yy1	Yin Yang 1
Yy2	Yin Yang 2

DECLARATION OF ACADEMIC ACHIEVEMENT

Dr. Jonathan Draper- provided funding support and aided with experimental conception/design.

Sonam Bhatia- generated and performed western for hESC lines expressing various REX1-FLAG constructs, designed/constructed the Rex1-FLAG targeting vector, performed Rex1-FLAG ChIP, ChIPseeker analysis using R, and downstream repetitive element analysis for Rex1, Yy1, and Yy2.

Amanda Hrenczuk- generated the constructs and performed the western for human REX1 alternative translation initiation site assessment in HEK293s, generated the endogenously 3XFLAG-epitope tagged mESC lines, knockout cell lines, rescue plasmids produced by site-directed/truncation mutagenesis, performed the ChIP-seq analysis for the murine Rex1-FLAG samples using the pipeline provided by Dr. Mathieu Lupien, downstream ChIP-seq analysis using GREAT, GO term enrichment analysis, motif enrichment analysis, characterization of epitope-tagged lines by western and immunofluorescence staining, qRT-PCR analyses, and LIF2i versus serum assay. Additionally, a significant amount of work contributing to the development of other projects within our lab including the breast cancer and mitotic bookmarking projects was performed, but will not be discussed in this report.

Dr. Mathieu Lupien- Provided the pipeline for genic ChIP-seq analysis

Dr. Mathew Lorincz-Performed ChIP-seq analysis for repetitive elements

INTRODUCTION

1. Pluripotent Stem Cells

Pluripotent stem cells are defined by their capacity to self-renew and differentiate into specialized cell types of all three primary embryonic germ layers; endoderm, mesoderm, and ectoderm^{1,2}. *In vivo*, pluripotency is a transient state that exists for only a brief period but can be propagated indefinitely *in vitro* through the culture of pluripotent stem cells (PSCs) that additionally maintain their ability to differentiate upon induction by various cytokines^{1,3}. Due to their unique properties, PSCs have become an indispensable research tool and hold tremendous potential for use in therapeutic applications⁴.

1.1 Mouse Pluripotent Stem Cells

1.1.1 Derivation

The first established PSC lines were embryonal carcinoma (EC) cells derived from germ cell tumors⁵. These cells could be continuously cultured in the undifferentiated state and induced to differentiate into tissues of all three embryonic germ layers⁶. Although a useful model system in the laboratory, their cancer-related origin prevents their use in clinical applications¹. Mouse embryonic stem cells (mESCs) were later derived in 1981 from the ICM of the pre-implantation blastocyst^{7,8}. Early protocols of mESC derivation described the use of feeder layers to prevent the spontaneous differentiation of mESCs *in vitro*^{9,10}. However, it was later established that the coculture of mESCs with feeders was not required for the prolonged maintenance of all cell lines when the culture medium was supplemented with leukemia inhibitory factor (LIF)¹¹. Since the discovery of stem cells,

mESCs have remained a convenient and reliable platform to study developmental pathways and the molecular mechanisms fundamental to the acquisition and maintenance of the pluripotent state^{1,12}.

In 2006, the first induced pluripotent stem cells (iPSCs) were derived through the ectopic expression of Oct4, Sox2, c-Myc, and Klf4 in mouse embryonic and adult fibroblasts¹³. This discovery represented a tremendous leap forward towards the use of PSCs, not only in the study of mammalian development and epigenetic reprogramming, but also their use in applications such as patient-specific disease modeling and drug discovery^{14,15}. The generation of iPS cells provided a way for researchers to take patient-derived somatic cells and reprogram them into iPSCs, which could then be differentiated into specialized cell types not otherwise easily attained, and used to recapitulate the disease of interest within a culture dish^{4,14,15}. Such models could then be further used to study the progression of a disease from a very early stage, in addition to the screening and identification of novel drugs for treatment^{4,14}. Since their derivation, substantial development has been made in optimizing the methods by which iPSCs are obtained, such as the use of non-integrative viruses^{16,17}. Studies have additionally uncovered differences between iPS and ES cells in terms of gene expression and epigenetic profiles that arise from epigenetic memory or improper establishment of methylation patterns^{4,18-21}. The identification of such differences that bias the differentiation potential of iPSCs will aid in optimizing both derivation and differentiation related techniques, furthering the use of iPSCs in various applications^{4,18-21}.

1.1.2 *Functional Assessment of Self-Renewal and Pluripotency*

A range of assays currently exist for assessing the self-renewal and developmental potential of PSCs, including colony initiating cell (CIC) assays, *in vitro* differentiation, teratoma formation, and the formation of chimeric mice^{2,22–29}. The ability to quantify the proportion of undifferentiated cells present within PSC populations is instrumental when monitoring the quality of PSC cultures and evaluating changes in self-renewal and differentiation capabilities following manipulation of PSCs by methods such as genome engineering²⁸. CIC assays quantify the proportion of undifferentiated cells within a culture by evaluating the number of alkaline phosphatase-positive cells after having cultured a specified number of PSCs in defined conditions, fixing resultant colonies, and staining for alkaline phosphatase^{27,28}. To date, CIC assays remain a simple and efficient method for quantifying the frequency of self-renewing cells within a culture^{27,28}.

In vitro differentiation of mouse pluripotent stem cells (mPSCs) into derivatives of all three embryonic germ layers represents one of the most basic tests for characterizing pluripotency^{2,22}. *In vitro* differentiation typically consists of replacing culture conditions known to maintain pluripotency with combinations of differentiation-inducing cytokines for specified durations of time and subsequently assessing for markers of specific tissue types². This is often also accomplished through the generation of embryoid bodies (EBs), followed by the assessment of markers of each germ layer by western blot or quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analyses²³.

While *in vitro* assays can be used as an initial assessment of pluripotency, *in vivo* assays, such as the formation of teratomas and chimeras, are more robust measures of pluripotency^{2,22,24,25}. Teratoma assays assess developmental potential at a population-based level, through the formation of a tumour containing cells representative of all three embryonic germ layers, following the injection of cells into immune-compromised mice^{24,25}. However, teratoma assays remain a preliminary assay for testing pluripotency of mPSCs *in vivo*, as histological analysis is subjective and solely qualitative, possibly resulting in misinterpretation². A more stringent assay for testing pluripotency of mPSCs is the generation of chimeric mice through the injection of cells into a host blastocyst^{2,26}. The relative contribution of PSCs to the chimera can be subsequently assessed through histological, proteomic or transcriptional analyses². Additionally, only high-quality PSCs will contribute to the germ line, and the resulting chimeras can be further bred to demonstrate the cells capability of producing functional germ cells^{2,26}. Chimeric mice can additionally be generated through tetraploid complementation, the most stringent assay of pluripotency, in which two blastocysts isolated at the 2-cells stage are fused to generate a 4n host blastocyst^{2,29}. As 4n blastocysts are only capable of contributing to the formation of extra-embryonic tissues, the injection of PSCs into these blastocysts allows for the complementation of their developmental potentials^{2,29}. Thus, while the 4n compartment provides the necessary extra-embryonic tissue to sustain growth and development, the donor PSCs, which are restricted in their potential to forming only lineages of the epiblast, will form the embryo proper^{2,29}.

1.1.3 Heterogeneity in Pluripotent Stem Cell Populations

The balance between stem cell self-renewal and differentiation in mPSC populations results in heterogeneity^{30,31}. Cells within a single population will exhibit functional differences arising from fluctuations in gene expression, resulting in the formation of two distinct sub-populations of cells; naïve and primed pluripotent stem cells^{30,32}. Embryonic stem cells (ESCs) isolated from the inner cell mass (ICM) of the pre-implantation blastocyst are representative of the naïve pluripotent state, in which cells are at the apex of pluripotency and express higher levels of naïve pluripotent markers including Nanog, Klf4 and Stella^{3,30,32,33}. Naïve PSCs have the ability to self-renew indefinitely, differentiate into all three embryonic germ layers, and contribute to chimeras^{3,30,32,33}. Epiblast stem cells (EpiSCs), derived from the epiblast of the post-implantation embryo, are representative of the primed state of pluripotency^{3,30,32,33}. EpiSCs maintain the capability to self-renew and differentiate into all three embryonic germ layers. However, they express higher levels of lineage-related genes, and are unable to contribute to chimeras indicating their more developed nature^{3,30,32,33}.

In serum conditions, the addition of LIF activates Stat3 which inhibits differentiation and promotes self-renewal, while mPSCs stably transition between naïve and lineage-primed pluripotent states^{30,32,34,35}. In contrast, the culture of cells in serum-free conditions, with the addition of LIF and two inhibitors (2i) of MAPK and GSK3 signaling, PD0325901 and CHIR99021, respectively, results in attainment of the naïve state^{34,35}. Within the naïve state, cells express naïve pluripotent markers homogeneously, have decreased methylation, and express lower levels of lineage-related genes, more closely

resembling cells of the ICM³⁵. This eliminates any contaminating primed pluripotent or lineage-committed cell types, resulting in a population of cells with the most unrestricted potential of the pluripotent states^{30,36}. The attainment of the naïve pluripotent state in PSC cultures has provided insight into the underlying mechanisms controlling pluripotency and has furthered the use of PSCs in the study of development and disease³⁶.

2. Transcriptional Networks Underlying the Pluripotent State

2.1 Core Transcriptional Network of Pluripotent Stem Cells

Substantial progress has been made in understanding the molecular mechanisms underlying pluripotency, including that the establishment and maintenance of the pluripotent state is dependent on a network of core transcription factors including Oct4, Sox2, and Nanog^{13,37,38}. Studies have demonstrated extensive co-binding of the core factors to key pluripotency genes, in addition to their ability to positively regulate their own expression³⁹⁻⁴². Furthermore, unbiased proteomic screens conducted in mESCs have identified a highly integrated protein network in which pluripotency related factors including Dax1 and Nac1, a loss of which results in differentiation and reduced growth, interact with core regulators such as Nanog^{38,39,43-45}. Although the transcriptional and protein networks of Oct4, Sox2, and Nanog have been extensively studied, it is evident that a high degree of interdependence exists between factors contributing to the maintenance of the pluripotent identity³⁸⁻⁴¹. Thus, further investigation into the interactomes and roles of other transcription factors within this network is required before a complete understanding of the mechanisms that control the self-renewal and differentiation of PSCs is attained³⁹.

2.2 Extending the Transcriptional Network in Pluripotent Stem Cells

Until recently, most studies have focused on a small subset of core pluripotency factors and their contribution to the regulation of the pluripotent state^{38,39,43}. However, the proper maintenance of pluripotency is dependent on other transcription factors that currently remain understudied, in addition to epigenetic regulators that work in conjunction with the core transcriptional network, to control cell fate^{38,46}. Among the proteins identified within the highly integrated protein network in PSCs was Yin Yang 1 (Yy1), a zinc finger transcription factor regarded as a master regulator of development^{43,47}.

Yy1 is a ubiquitously expressed zinc finger transcription factor that has been demonstrated to play numerous essential roles as a transcriptional regulator involved in cell growth, development, and cellular differentiation⁴⁸⁻⁵¹. Depending on the context in which it binds DNA, Yy1 will either act as an activator or repressor of target genes^{48,50,52}. Specifically in PSCs, Yy1 is a known member of the polycomb group complex (PcG) proteins known to play critical roles in development and to repress differentiation associated genes including Hox, Dlx, and Pou families⁵³⁻⁵⁶. PcG proteins cooperate with transcription factors to maintain pluripotency through chromatin remodeling of differentiation related genes by depositing repressive histone modifications such as trimethylation of lysine 27 on histone H3 (H3K27me3)⁵⁷. A loss of numerous PcG proteins including Rnf2, Ezh2, and Yy1 has been demonstrated to result in lethal phenotypes, all of which cause a failure of blastocyst development to the gastrulation stage, confirming their essential role in development^{51,55,56}. Additional studies have further demonstrated PcG-independent roles of Yy1, in which Yy1 binds to promoter regions of genes that are highly transcriptionally active

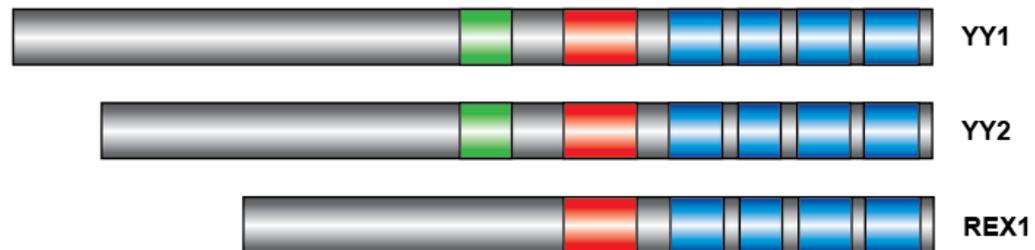
in ES cells⁴⁷. The identified Yy1 binding sites within promoter regions were often co-bound by factors such as Myc and E2F1, which were demonstrated to cooperate with Yy1 and aid in the up-regulation of target genes such as Surf-1 and Cdc6, involved in oxidative phosphorylation and cell cycle progression, respectively^{47,58,59}. Thus, in addition to its role as a repressor of differentiation-related genes, Yy1 is a member of the Myc transcriptional network and aids in the activation of genes required for RNA and protein synthesis, cell cycle progression, and mitochondrial functions required for proper development⁴⁷.

The Yy1 protein consists of two N-terminal domains, Domain I and Domain II, and four C-terminal zinc finger motifs that contribute to its function^{50,60}. Deletion analyses have revealed the transcriptional repressive domain of Yy1 to lie within the C-terminal zinc finger region, whereas two regions of the N-terminus within the first 100 amino acids are required for maximal transcriptional activation⁶¹. Yy1 protein structure is evolutionarily well conserved and more recently, two other Yy1 related genes, Yin Yang 2 (Yy2) and Reduced Expression 1 (Rex1), have been identified in mammals as having emerged through the retroposition-mediated duplication of Yy1⁶⁰. Rex1 and Yy2 both share high levels of sequence homology with Yy1 primarily within the C-terminal zinc finger region responsible for their DNA binding abilities (Figure 1)^{60,62}. As a result, there is a significant overlap in the DNA-binding targets of these proteins⁶⁰.

Similarly to Yy1, Yy2 contains domain I and domain II but is not entirely homologous to Yy1 within its N-terminal region, suggesting that Yy2 displays some differences in function from Yy1⁶⁰. On the contrary, Rex1 only contains Domain II and is much less conserved in its N-terminal sequence in comparison to Yy1 and Yy2, suggesting

that Rex1 has evolved under different functional constraints⁶⁰. A loss of Yy2 results in a depletion of mESC cultures, suggesting a role for Yy2 in the self-renewal of PSCs⁶³. Furthermore, the inability of Yy2-null blastocysts to maintain their inner cell mass suggests that a loss of Yy2 may also result in peri-implantation lethality, demonstrating a redundancy in function between family members⁶³. Interestingly, unlike Yy1 and Yy2 which are expressed in various tissues, Rex1 expression has been demonstrated to be mostly limited to pluripotent cell types suggesting an important role for Rex1 in PSCs^{64,65}. However, unlike its family members, a loss of Rex1 does not result in peri-implantation lethality, perhaps due to compensation in function by Yy1 or Yy2⁶⁶. Furthermore, investigation of Rex1's function in PSCs has revealed numerous roles for this protein associated with pluripotency including the reprogramming of X-inactivation, the establishment and maintenance of epigenetic modifications, the differentiation of visceral endoderm, and the regulation of endogenous retroviral elements⁶⁶⁻⁶⁹. Though, in comparison to its family members, the mechanisms by which Rex1 carries out its roles is understudied.

A)



B)

```

sp|P22227|ZFP42_MOUSE      -----MNEQKMNQMKKTAKTSG
sp|Q00899|YY1_MOUSE       --MASGDTLYIATDGSEMPAIEIVLHEIEV-ETIPVETIETTIVVGEEDDDDEDDGGGG
sp|Q3TTC2|YY2_MOUSE       MASETEKLLCLNTESAEIPADFVELLPDNI GDIEAVSLETS-VGQTIE-----
                               :  :

sp|P22227|ZFP42_MOUSE      QKGP-----G
sp|Q00899|YY1_MOUSE       DHGGGGGGHGHAGHHHHHHHHHPPMIALQPLVTD DPTQVHHHQEVILVQTRE-EVVG
sp|Q3TTC2|YY2_MOUSE       -----VYGDVG--VDWAHGSQYHSPVIALQPLVGSLSR DHDKEMFVVQTRREEVVG
                               *

sp|P22227|ZFP42_MOUSE      GRALDRLTLLKQD---EARPVQ-----NTRVEAPRVTYTIRD-----ESEISPET
sp|Q00899|YY1_MOUSE       GDDSDGLRAEDGFEDQILIPVPAPAGGDDDYIEQTLVTVAAAGKSGGGASSGGGRVKKGG
sp|Q3TTC2|YY2_MOUSE       YQSDNLLFSPEFGSQMVLNVN----EDDYLQPTTASFTGFL----AAENGQGELSPYE
                               * *      **          :  :  :  :
                               :  :  :  :

sp|P22227|ZFP42_MOUSE      EEDGFDPDGYLEC-----IIRGEFSEPILEEDFLFKSFESLEEVEQNLSR
sp|Q00899|YY1_MOUSE       GKKS GKKS YLGGGAGAAGGGGADPGNKKWEQKQVQIKTLEGEFSVTMWSSDEKKDIDHET
sp|Q3TTC2|YY2_MOUSE       GNLCGLTTFIEAG--AEESVNADLGDKQWEQKQ--IDGLDGEFPPTMWDVNEKEDP---
                               :  :  :  :  * : *  . . . . :  :

sp|P22227|ZFP42_MOUSE      QV--LEASSLLESSLEYMTKGTQEKTEVTQETPPLRVGASSLLAGGPAEKPEGGVYCG
sp|Q00899|YY1_MOUSE       VVEEQIIGENSPPDYSEYMTGKKLP-----PGGIPGIDLSDPKQLAEFARMKPRKIKEDD
sp|Q3TTC2|YY2_MOUSE       IA--EEQAGESPPDYSEYMTGKKFP-----PEGIPGIDLSDPKQLAEFTSMRPKKPK-GD
                               . . . . * * * *      * : . . . . * *
                               :  :  :  :

sp|P22227|ZFP42_MOUSE      VLSMLECPQAGCKKKLRGKTALRKHMLVHGPRRHVCAECGKAFTESSKLRHFLVHTGEK
sp|Q00899|YY1_MOUSE       APRTIACPHKGCTKMF RDNSAMRKHHLH THGPRVHVCAECGKAFVSSSKLRHQLVHTGEK
sp|Q3TTC2|YY2_MOUSE       FPRPIACSHRGEKMF RDNSAMRKHHLH IHGPRVHVCAECGKAFVSSSKLRHQLVHTGEK
                               : * : * * * : : : : * * * * : * * * * * * * * * * * * * * * * *

sp|P22227|ZFP42_MOUSE      PYQCTFEGCGKRFSLDNLRTHIRIHTGERRFVCPFDGCEKSF IQSNNQKIHLTHAKAG
sp|Q00899|YY1_MOUSE       PFQCTFEGCGKRFSLDNLRTHVRIHTGDRPYVCPFDGCNKKFAQSTNLKSHILTHAKAK
sp|Q3TTC2|YY2_MOUSE       PYQCTFEGCGRRFSLDNLRTHVRIHTGDKPFVCPFDACNKKFAQSTNLKSHILTHVKNK
* : * * * * * : * * * * * * * * * * * : : * * * * * : * * * * * * *

sp|P22227|ZFP42_MOUSE      KKC
sp|Q00899|YY1_MOUSE       NNQ
sp|Q3TTC2|YY2_MOUSE       NDQ
                               :
    
```

Figure 1: Retrotransposon mediated duplication of Yy1 in placental mammals resulted in the derivation of two Yy1 related proteins, Yy2 and Rex1. A) Schematic representation of Yy1, Yy2 and Rex1 protein structures. Domains are specified by different colours: green for domain I, red for domain II, and blue for the zinc finger domains. Figure adapted from Kim et al (2007). B) Sequence alignment of Yy1, Yy2, and Rex1 performed by using Clustal Omega^{70,71}. Red asterisks denote regions of conservation.

3. Expression and Function of Rex1 in Pluripotent Stem Cells

3.1 *Rex1* Expression in Pluripotent Stem Cells

Rex1 was first discovered as a result of its decreased expression upon inducing the differentiation of F9 teratocarcinoma cells by retinoic acid (RA) treatment⁶⁴. Due to its virtually exclusive expression in pluripotent cell types, Rex1 has become a widely used marker of the pluripotent state^{33,64,65}. As previously discussed, PSCs within a population fluctuate in gene expression levels, transitioning between the naïve and primed states of pluripotency^{30,32}. The culture of PSCs in serum-free media supplemented with two inhibitors (2i) of MAPK signalling and GSK3 in addition to LIF results in the attainment of the ground state, in which pluripotent genes are expressed uniformly displaying transcriptional and epigenetic profiles similar to that of the pre-implantation blastocyst^{30,34,72–74}. It is in this state that Rex1 is homogeneously expressed and is at its highest levels^{30,33,74}. In serum conditions, PSCs no longer exist in a uniform state, but stably transition between the naïve and primed states, resulting in heterogeneity within the cell population^{30,33}. This heterogeneity is also observed in the expression of Rex1 in PSCs cultured in serum-conditions⁷⁵. As cells attain a more EpiSC phenotype, Rex1 expression is lost, and cells transition towards the primed state in which there is an increase in

expression of lineage specific markers^{3,30,33,74}. Thus, unlike *Yy1* which has been demonstrated to be ubiquitously expressed, *Rex1* has evolved to be developmentally regulated⁷⁶.

3.2 Rex1 Function in Pluripotent Stem Cells

3.2.1 Reprogramming of X-inactivation

X-chromosome inactivation, the process by which one of the X-chromosomes is silenced, occurs in early development and is maintained throughout life to ensure proper dosage compensation of X-linked genes among female and male mammals⁷⁷⁻⁷⁹. Reactivation of the X-chromosome in female ES cells after imprinted paternal X-chromosome inactivation within the cleavage stage embryo is a hallmark of the pluripotent naive state and is regulated by numerous factors that also have known roles in the self-renewal and differentiation of PSCs^{67,73,77,79,80}. Thus, disruption of this process leads to a loss of the naive state of pluripotency and more restricted developmental potential of PSCs^{80,81}. Moreover, improper X-inactivation can lead to the expression of X-linked genes at higher levels than normal, which is known to disrupt processes within ES cells required for proper cell function⁸².

The reprogramming of X-chromosome inactivation during the acquisition of pluripotency is dependent on the repression of *Xist*, a noncoding RNA responsible for X-inactivation, and the upregulation of its antisense transcript, *Tsix*, another noncoding RNA responsible for maintaining an active X-chromosome^{67,83}. In female mPSCs, *Oct4*, *Sox2* and *Nanog* inhibit *Xist* expression by directly binding to intron 1, whereas *Rex1*, *c-Myc*,

and Klf4 bind to *DxPas34*, the strong enhancer of *Tsix*, promoting its expression⁶⁷. Studies have established Rex1's binding to the *DxPas34* repeat is a requirement for the efficient elongation of *Tsix* transcription and therefore expression, as maximal *Tsix* levels are required for proper establishment of a second active X-chromosome⁶⁷. Rex1 depletion, by genic knockout or targeted degradation by Rnf12, coincides with a downregulation of *Tsix*, upregulation of *Xist*, and subsequent inactivation of the X-chromosome, thus, linking reactivation of the X-chromosomes to the establishment of pluripotency and implicating a role for Rex1 in the establishment of X-inactivation upon differentiation of PSCs^{67,83}. Furthermore, due to the select expression of Rex1 and *Xist* within placental mammals, studies have suggested that the co-evolution of these two genes has contributed to the evolution of random X-inactivation in placental mammals^{67,83}.

3.2.2 Epigenetic Regulation and Genomic Imprinting

The majority of genes within any given cell type are either biallelically expressed or repressed, as a result of inheriting two sets of chromosomes, one from each parent⁸⁴⁻⁸⁶. A small subset of genes, however, are monoallelically expressed in a parental specific manner and are termed imprinted genes^{84,85}. Since their discovery, imprinted genes have been identified as crucial for proper embryonic development, development of cell lineages, normal brain function, and postnatal energy homeostasis⁸⁴. As a result, the dysregulation of imprinted genes has been implicated in the development of behavioural and neurodevelopmental syndromes, neonatal diabetes, and cell transformation leading to oncogenic phenotypes^{84,85}. The process of genomic imprinting occurs in a number of developmental stages including the acquisition of imprints in gametes, establishment of

imprinted expression in the embryo, the elimination of imprints in primordial germ cells, and the maintenance of imprinted expression in differentiated cells throughout life^{85,86}. As a number of developmental steps of imprinting occur in the embryo, PSCs have provided a suitable model system for studying the regulation and function of genomic imprinting *in vitro*⁸⁶.

Rex1 has been demonstrated to be nonessential for the maintenance of the pluripotent state, however, embryonic loss of Rex1 results in deviation from the expected Mendelian ratio of late stage embryos and neonates which suggested a role for correct Rex1 gene dosage in normal development, potentially via the establishment/maintenance of epigenetic modifications^{66,68}. Chromatin immunoprecipitation (ChIP) identified two imprinted genes, *Nespas* and *Peg3*, which also contain known Yy1 binding sites, as DNA-binding targets of Rex1⁶⁸. Depletion of Rex1 results in DNA hypermethylation of the differentially methylated regions of the two loci, but does not result in any major changes in DNA methylation levels of other imprinted loci⁶⁸. This suggests that Rex1 may bind the unmethylated alleles of *Peg3* and *Nespas* to prevent Yy1 binding and the subsequent recruitment of DNA methyltransferases, acting as a protector of DNA methylation during embryogenesis⁶⁸. This implies that although Rex1 and Yy1 are evolutionarily related proteins and share numerous DNA-binding sites, their roles may not be entirely redundant⁶⁸. Although both Yy1 and Rex1 have been identified as epigenetic regulators in PSCs, such a role has yet to be described for Yy2⁸⁷.

3.2.3 Differentiation of Visceral Endoderm

Previous analysis of Rex1-null teratocarcinoma cells, which are homologous to PSCs in their ability to self-renew and differentiate into all three embryonic germ layers, suggested that Rex1 may play a role in the differentiation of visceral endoderm⁸⁸. In early embryonic development, the pre-implantation blastocyst consists of an ICM from which pluripotent cells are derived, and an outer layer of trophoblast cells, which form the trophectoderm^{88,89}. Shortly after blastocyst formation, the cells of the ICM segregate to form an outer layer of extra-embryonic tissue termed primitive endoderm, while the remaining cells go on to form primitive ectoderm, otherwise known as the epiblast⁸⁹. Upon implantation, the primitive endoderm migrates to cover the inner surface of the trophectoderm and becomes parietal endoderm, while the epiblast segregates to form an outer layer of extra-embryonic tissue known as visceral endoderm⁸⁹. The inner epiblast then forms the embryo proper whereas the extra-embryonic visceral endoderm plays a crucial role in nutrient uptake/transport and embryonic patterning⁸⁸⁻⁹¹. More specifically, visceral endoderm permits the exchange of nutrients and gases between the mother and developing embryo through the synthesis of proteins aiding in nutrient uptake/transport and the differentiation of blood cells and vessel formation in neighbouring mesoderm⁹⁰. Visceral endoderm is additionally a source of signals that are required for correct anterior-posterior and primitive streak patterning^{88,90,91}. Disruption of proper gene expression within the early visceral endoderm results in developmental arrest demonstrating its requirement for proper gastrulation⁹².

To determine whether the differentiation of visceral endoderm was also a function of Rex1 in mESCs, Masui, *et al.* generated EBs from *wildtype (wt)*, *wildtype-Rex1 Transgene (wt-Tg)*, *knockout (KO)*, and *KO-Tg* mESCs⁶⁶. Subsequent evaluation of gene expression by qRT-PCR revealed the reduction of two markers of visceral endoderm, *Transthyretin* and *Indian Hedgehog*, in *KO* derived EBs⁶⁶. As the expression of these markers was restored in *KO-Tg* derived EBs, the observed phenotype was therefore dependent on Rex1 expression⁶⁶. Additionally, no changes were observed in parietal endoderm markers regardless of the *Rex1* genotype, suggesting that Rex1's function in ES cells may involve the differentiation of visceral endoderm⁶⁶. However, as Rex1 expression is limited to pluripotent cells and down-regulated during differentiation, it remains unclear as to how Rex1 modulates the expression of visceral endoderm markers^{64–66,88}.

3.2.4 Regulation of Endogenous Retroviral Elements

In PSCs, regulation of the transcriptional network is dependent on the stringent control of endogenous retroviruses (ERVs)⁹³. ERVs are transposable elements with the ability to move throughout the genome, causing heritable mutations that can potentially alter the integrity and function of a genomic locus^{69,93}. The generalized structure of a replication-competent retrovirus includes four coding domains: Gag, Pro, Pol, and Env, which encode the structural components, viral protease, reverse transcriptase and integrase, and glycoproteins, respectively⁹⁴. Transposable elements can be divided into two main categories: retrotransposons and DNA transposons^{95,96}. Retrotransposons represent the majority of transposable elements present within the genome and require an RNA transcript that is transcribed by a reverse transcriptase (RT) prior to integration into the genome^{95,96}.

Retrotransposons consist of three main families of elements; long terminal repeats (LTRs), and non-LTR retrotransposons such as LINE and SINE elements (long and short interspersed nuclear elements), which differ in their requirements for transposition throughout the genome^{95,96}. The transposition of LTR and LINE elements are dependent on internal RT encoding genes, whereas SINE elements arise from the retroposition of RNA polymerase III transcripts and depend on the RT gene from LINE elements for their propagation^{95,96}. Based on RT gene similarity, ERVs are then further divided into three main classes: Class I (ERVK), Class II (IAP, MusD, ETn), and Class III (ERVL, MaLR)^{69,94,95}. All of which vary in copy number, length, functional components, and activity^{95,96}.

Transcription of ERVs is limited within differentiated tissues due to methylation-dependent silencing, but has been demonstrated to be elevated in the germ line, early embryo, and placenta of various species^{69,97}. At the onset of zygote genome activation in the mouse, murine endogenous retroviral (muERV-L) elements are expressed, but then silenced in the blastocyst stage, whereas intracisternal A particle (IAP) elements and mouse type-D (musD) elements increase in expression until the blastocyst stage, after which they too, are silenced^{69,95,98,99}. Approximately 200 copies of muERV-L and 1000 copies IAP particles exist within the mouse genome, respectively⁹⁵. The tight regulation of IAP and muERVL elements suggests that specific classes of ERVs are responsible for the regulation of stage-specific gene expression throughout development^{69,100}. Thus, despite their pathogenic nature, ERVs play an essential role in embryonic stem cells by acting as regulatory elements, such as promoters and enhancers, bound by pluripotency related

factors such as Oct4 and Nanog to control genes important for maintaining the pluripotent state while also contributing to the stage-specific expression of differentiation related genes^{93,101,102}.

The pluripotent state is closely associated with the increased expression of ERVs as a result of their hypomethylated state¹⁰³. Thus, to fine tune expression and prevent the undesirable propagation of ERVs, maintaining the genomic integrity of the pluripotent state, PSCs have developed various mechanisms for the silencing of these elements including DNA methylation and the recruitment of repressive factors such as Trim28 and SETB1^{104,105}. In PSCs, Trim28 is recruited to repetitive elements by krüppel-associated box domain-containing zinc finger proteins (KRAB-ZFPs) and aids in their silencing through the additional recruitment of the histone methyltransferase SETB1 which deposits the repressive trimethylation of lysine 9 on histone H3 (H3K9me3) mark, thus, protecting the genome integrity of ES cells through preventing spontaneous differentiation, the accumulation of heritable and possibly detrimental mutations, or cell transformation through the activation of oncogenes^{97,105,106}.

Since the expression of Rex1 coincides with transcriptional changes in genes directly regulated by ERVs, this prompted an investigation into the role of Rex1 in the regulation of endogenous retroviral elements in PSCs⁶⁹. Depletion of Rex1 in mPSCs results in an increase in the expression of both muERV-L and musD elements by 2-3 fold⁶⁹. Furthermore, investigation of Rex1-dependent repression of ERVs through histone demethylase recruitment revealed interaction of Rex1 with LSD1, also previously demonstrated to play a role in the silencing of ERVs, and their neighbouring genes, in ES

cells^{69,104}. Thus, Rex1 may also be acting as a protector of the pluripotent state in ES cells through the regulation of ERVs, by recruiting additional chromatin modifying factors necessary to maintain their repression⁶⁹. Furthermore, the observed binding of Yy1 and Yy2, family members of Rex1, to ERVs suggests that Rex1 and its relatives may have evolved as regulators of endogenous retroviral transcription^{69,87,107}.

4. Project Rationale and Hypothesis

Rex1 was first discovered as a result of its decreased expression in teratocarcinoma cells and has since been demonstrated to be strongly correlated with the pluripotent state^{64,74,108}. More specifically, studies have demonstrated Rex1 expression to be closely linked with the “naive state” of pluripotency where cells are at the apex of the pluripotent hierarchy and Rex1 expression is at its highest, suggesting an important role for Rex1 in pluripotent cell types^{30,33,74}. Furthermore, Rex1 has been demonstrated to play a role in the regulation of ERVs, suggesting a role for Rex1 in the maintenance of genomic integrity as the dysregulation of ERVs results in aberrant gene expression and therefore disruption of the stringently regulated transcriptional network in PSCs⁶⁹. Although Rex1 has been extensively used as a marker of pluripotency, the functional elements within Rex1 contributing to its functions in PSCs are not well characterized. In the course of investigating Rex1 in a pluripotent context, we observed multiple previously unidentified isoforms being expressed from the *REX1* locus as a result of alternative translation initiation sites within the *REX1* open reading frame. Upon confirming this phenomenon at the exogenous level in the human context, we sought to investigate the usage of endogenous alternative translation initiation sites in addition to the functional elements of Rex1

contributing its roles in the mouse system as mESCs are a convenient and reliable platform for studying the molecular basis of pluripotency.

Given the previously described roles of Rex1 and the identification of these novel isoforms, we hypothesize that Rex1 and its smaller protein isoforms are functioning to regulate the expression of lineage-determining genes in PSCs. To assess our hypothesis and further characterize the independent roles of the full-length Rex1 protein and its smaller isoforms, the experiments in this thesis describe our effort to map the global genomic binding sites of Rex1 to gain an unbiased understanding of the genes Rex1 is regulating in PSCs, and to investigate the mechanisms by which Rex1 is regulating these genes in PSCs.

MATERIALS AND METHODS

1. Cell Culture

E14Tg2a mESCs were grown in Dulbecco's Modified Eagle Medium (DMEM, Sigma) supplemented with 15% fetal bovine serum (FBS, Sigma), 1× Non-Essential Amino Acids (Gibco), 1× Glutamax (Gibco), 100µM Sodium Pyruvate (Gibco), and 55µM β-mercaptoethanol (Gibco). Media was further supplemented with 1000 U/mL LIF (AMSBIO) after filter-sterilization with a 0.22µm filter (Sigma). Cells were passaged (1:6-1:10) with accutase (Sigma) and grown in a humidified incubator at 37°C and 5% carbon dioxide.

2. 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), 1µM PD0325901 (Tocris), 3µM CHIR99021 (Tocris), 1× Non-Essential Amino Acids (Gibco), 1× Glutamax (Gibco), 100 µM 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.

3. Rex1 CRISPR design and construction

Single guide RNAs (sgRNAs) used to generate epitope-tagged and KO cell lines were designed using the online CRISPR design tool by MIT (<http://crispr.mit.edu/>). All sgRNA oligo inserts were prepared and cloned into the BbsI digested pSpCas9(BB)-2A-Puro (px459) plasmid (Addgene #48139) as previously described¹⁰⁹. MmRex1CprA and mmRexCprB were additionally cloned into the BbsI digested pSpCas9n(BB)-2A-Puro (px462) plasmid (Addgene #48141) as previously described¹⁰⁹. Constructs were verified by Sanger-sequencing (MOBIX; McMaster University). Oligo sequences used for the construction of the sgRNA are as shown in Table 1 below.

Table 1: Guide RNA oligos used for the construction of Rex1 CRISPR constructs

sgRNA	Target	Sequence (5'→3')	
mmRex1CprA	C-terminal Zn finger 1	Forward	CACCGCCTCAGCTTCTTCTTGACC
		Reverse	AAACGGTGCAAGAAGAAGCTGAGGC
mmRex1CprB	C-terminal Zn finger 1	Forward	CACCGACATGCTTGTCACGGGCCC
		Reverse	AAACGGGCCCCGTGGACAAGCATGTC
mmRex1 NtermMIT	N-terminus of Rex1	Forward	CACCGCTCTCCGCCCGGCCCTTTC
		Reverse	AAACGAAAGGGCCGGGCGGAAGAGC

4. Generation of the Rex1 targeting vector

The targeting vector sequence was designed using the Vector NTI® software (Life Technologies) and was ordered from Invitrogen. The GeneArt® product was cloned into NotI digested pBluescript II SK+ using an In-Fusion® HD cloning kit (Clontech). Primers used to amplify the targeting vector sequence for in-fusion cloning are as described in Table 2 below.

Table 2: In-Fusion® primers for the generation of the Rex1 targeting vector

Primer name	Sequence (5'→3')
mmRex1TV_in-fusion_F	AGTTCTAGAGCGGCCGCCAAGTTCCTAGTGAGCCAT
mmRex1TV_in-fusion_R	ACCGCGGTGGCGGCCGCTGGTGTTTGAGGTAGATTTTT

5. Generation of endogenously 3XFLAG epitope-tagged Rex1 mESCs

Transfection of cells was performed using Lipofectamine® LTX with Plus™ Reagent (Invitrogen). For targeting experiments, 2×10^6 mESCs were transfected with 5 µg of px459-mmRex1CrpA or px459-mmRex1CrpB and 5 µg of the targeting vector, or transfected with 5ug of each px462-mmRex1CrpA and px462-mmRex1CrpB plasmid DNA and 5 µg of the targeting vector. The cells were fed 24 hours following transfection with mESC media supplemented with 2µg/mL puromycin. 48 hours post-transfection the cells were plated at clonal density in medium free of puromycin. 24 hours after seeding, puromycin resistant-clones were selected for with media containing 2µg/mL puromycin for 72hours. Colonies were then isolated for western blot analysis to screen for successful homologous recombination events.

6. CRISPR-mediated knockout of Rex1 in mESCs

For the generation of Rex1 knockout cells 2×10^6 mESCs were transfected with 5µg of px459-mmRex1CrpA, px459-mmRex1CrpB, or px459-mmRex1NtermMIT plasmid DNA. Non-homologous end-joining events were enriched for by selecting for puromycin resistant clones with 2µg/mL puromycin for 72hours. Cells were plated at clonal density

and colonies isolated for sequencing based screening to assess for indel formation (MOBIX; McMaster University).

7. Chromatin immunoprecipitation sequencing

Chromatin immunoprecipitation sequencing assays were performed using the SimpleChIP[®] Enzymatic Chromatin IP Kit (#9003, Cell Signaling) and 4.0×10^7 Rex1-FLAG mESCs. Cells were cross-linked with 1% formaldehyde for 10 minutes at room temperature and formaldehyde inactivated by the addition of glycine to a final concentration of 1×. Lysates were treated with micrococcal nuclease to obtain mostly mono-, di-, and tri-nucleosomal bands. Protein-DNA complexes were immunoprecipitated with 10 µg of anti-FLAG antibody (F1804, Sigma) and 60 µL of Protein G Dynabeads (1003D, Thermo Fisher Scientific). The DNA SMART[™] ChIP-Seq kit (#634865, Clontech) was used for the preparation of libraries. Reads were mapped to chromosomal locations in the July 2007 NCBI Build 37 (mm9) mouse genome using bowtie2 and samtools^{110,111}. The July 2007 NCBI Build 37 (mm9) mouse genome was the selected reference genome as a result of available blacklisted genomic regions. Optional de-duplication was not performed. Peaks were called for with MACS2 using a mfold between 3;30 and default bandwidth¹¹².

The analysis pipeline provided by Dr. Mathieu Lupien is as follows:

1. Convert bam files to fastq

```
samtools bam2fq input.bam > output.fastq
```

2. Check fastq file quality

```
fastx_quality_stats -i input.fastq -o output.txt
```

3. Plot the quality score derived from fast_quality_stats .txt file output

```
fastq_quality_boxplot_graph.sh -i input.txt -o output.png -t title shown on graph
```

4. Align to genome using Bowtie2 and samtools, then (optional) remove duplicate reads

```
bowtie2 -x /path/to/genome -U /path/to/fastq |samtools view -bhS - | samtools sort - aligned_file
```

```
samtools rmdup -s /path/to/bam - | samtools view -h - | grep -v chrM | grep -v chrY | grep -v chrUn | grep -v random | samtools view -bS -q 30 - | intersectBed -v -abam stdin -b /path/to/blacklist/bed/file | samtools sort - output_file
```

5. Call peaks using macs2

```
macs2 callpeak -t treatment.bam -c control.bam -n outputname -f BAM -g hs -B - -outdir outputdirectoryname
```

8. Antibodies

The following primary antibodies were used for western blot analysis: mouse anti-FLAG (F1804, Sigma), rabbit anti-Histone H3 (ab1791, Abcam), mouse anti-Oct4 (611203, BD Biosciences), mouse anti-Sox2 (561469, BD Biosciences), rabbit anti-Nanog (A300-397A, Bethyl), and mouse anti- β tubulin-I (T7816, Sigma). The following primary antibodies were used for immunofluorescence analysis: mouse anti-FLAG (F1804, Sigma), mouse anti-Oct4 (sc-9081, Santa Cruz), mouse anti-Sox2 (561469, BD Biosciences), and rabbit anti-Nanog (A300-397A, Bethyl).

Horseshoe peroxidase-conjugated secondary antibodies used for western blot analysis were as follows: goat anti-mouse (1721011, Bio-Rad), goat anti-rabbit (1706515, Bio-Rad). Alexa Fluor® conjugated secondary antibodies used for immunofluorescence staining were as follows: donkey anti-mouse AF546 (A11030; Thermo Scientific), goat anti-rabbit

AF647 (A21244; Thermo Scientific). For immunofluorescence analysis, DNA was stained using Hoechst 33342 (H1399; Life Technologies).

9. Cell lysate preparation

Whole cell lysates were prepared by washing mouse embryonic fibroblast (MEF) depleted mESC cultures three times with 1 × PBS before lysing on ice with 1 × RIPA buffer [50mM Tris-HCl pH 8, 150mM NaCl, 0.1% Sodium Dodecyl Sulfate, 1% NP-40, 0.5% Sodium Deoxycholate and 1× Complete™ Mini Protease Inhibitor Cocktail (Roche)] for 10 minutes. Cells were centrifuged at 16, 800 × 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 × 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× 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× 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 × 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× Complete™ Mini Protease Inhibitor Cocktail (Roche)]. Samples were incubated on ice for 30 minutes and centrifuged at 10,000 × 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× NuPAGE® LDS Sample Buffer (Invitrogen) with 15% TCEP Bond-Breaker solution (Thermo Scientific) and heated at 95°C for 5 minutes prior to electrophoresis.

10. Western blot analyses

Lysates (10-15 µg per well) were separated using 12% Bis-Tris gels and proteins were transferred onto PVDF membranes (Millipore) by wet transfer in 1× Towbins¹¹³. Membranes were blocked in 1× Tris buffered saline containing 5% skim milk for 30 minutes and incubated with primary antibody for 2 hours at room temperature or overnight at 4°C. Primary antibody dilutions are as follows: 3XFLAG (1:5000), Oct4 (1:1000), Sox2 (1:1000), Nanog (1:3000), B-tubulin (1: 400,000), Histone H3 (1: 50,000). Membranes were subsequently incubated with the appropriate secondary antibody (1:20,000) for 1 hour at room temperature. Blots were developed using the Luminata Forte Western HRP substrate (Millipore) and subsequently imaged using the ChemiDoc™ MP Imaging System (Bio-Rad). All antibodies were diluted in 1× Tris buffered saline containing 0.1% Tween 20 and 3% skimmed milk.

11. Quantitative RT-PCR

Total RNA was isolated from MEF depleted mESCs using the RNeasy Mini Kit (Qiagen) and contaminating genomic DNA removed through DNase treatment (Qiagen). cDNA was subsequently generated using the SensiFAST™ cDNA synthesis kit (Bioline). QRT-PCR analysis for Rex1 genic targets was performed on a Bio-Rad CFX96 qPCR instrument using

the SensiFAST™ SYBR® No-ROX Kit (Bioline) and 30ng cDNA per reaction. Reactions preimplantation in triplicate. Gene expression levels normalized to TATA-binding protein (TBP) was determined using the $\Delta\Delta Cq$ method on the CFX™ Manager 2.0 software. Primer sequences were designed using online software provided by Primerbot, obtained from the PrimerBank, or obtained from published studies (Table 3).

Quantitative RT-PCR for Rex1 ERV targets was performed on a Bio-Rad CFX96 qPCR instrument using the PerfeCTa® MultiPlex qPCR SuperMix Low ROX (Quanta Biosciences) and 30ng cDNA per reaction. Reactions were performed in triplicate. Gene expression levels normalized to β -actin and no reverse transcriptase (NRT) controls was determined using the $\Delta\Delta Cq$ method on the CFX™ Manager 2.0 software. Primer sequences were designed using online software provided by ROCHE (lifescience.roche.com).

Table 3: qRT-PCR primers for Rex1 genic targets

Primer	Sequence (5'→3')	Reference
Atf7ip	FWD: GAGTGCCTCAGACAACCACA	N/A
	RVS: CTGTGGGGCTTCTGGTAAGG	
Cited2	FWD: GTTCCGAGCAGAAATCGCAAAG	N/A
	RVS: AAGCGCCCGTGGTTCAT	
Igsf21	FWD: GTATCTACGACCGAGCCACG	N/A
	RVS: GAAGTTCTGGGCCTGGTACC	
Oct4	FWD: AGCTGCTGAAGCAGAAGAGGATCA	[114]
	RVS: TTCATTGTTGTCGGCTTCCTCCA	
Rex1	FWD: GTCCTGCACACAGAAGAAA	[115]
	RVS: GTCTTAGCTGCTTCCTTCTTGA	
TBP	FWD: AAGAGAGCCACGGACAACCTG	N/A
	RVS: AGCCCAACTTCTGCACAACCT	
Trim8	FWD: AGGGACACTCGGTGTGTGA	Primerbank ID: 16716393a1
	RVS: TGTCTGCCGCAAGTCTTCATC	

Unc5a	FWD: AGCAGGTCGAGAAAGTGTTTG	Primerbank ID: 23346571a1
	RVS: GGGCGACAAGGTAGCACAAT	

Table 4: qRT-PCR primers and Universal Probe Library (UPL) probe numbers for Rex1 ERV targets

Primer	Sequence (5'→3')	UPL Probe
Actin (β)	FWD: UPL ACTB specific primer from ROCHE (No sequence available)	B-Actin mouse
	RVS: UPL ACTB specific primer from ROCHE (No sequence available)	
IAPLTR3-int family	FWD: CCCAGTGAGGAGGCTAAAT	16
	RVS: CCATCGGTCAGGGTTATATCTT	
RMER1C	FWD: CCAAATCCATTGATGTCTACTACC	46
	RVS: GGCCAGTCTCATGAATACGAA	
RMER21A Chr6	FWD: TGTACCACAGGAGCTGTCCA	19
	RVS: GGGGCTGCTGAAGTG TAGAG	
RMER21A Chr7	FWD: TCTTGCCATCCTCAGCCTAC	17
	RVS: ACCTTGGACAGTTCCTGTG	
RMER21A Chr9	FWD: GGCACAAAAGTCATTGCATC	40
	RVS: CATAGGCTGGGATGGGAAG	
RMER21A Chr16	FWD: GAGTAGGGCTGAGGCTGCTA	73
	RVS: AGCTCCAGGGGAAGGTAT	

12. Statistical analyses

Error bars show SEM. Statistical analysis (Unpaired t-test) was performed with Prism 7 (Graphpad). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Graphs were generated from independent experiments with three technical replicates per experiment.

13. Generation of plasmids for the characterization of Rex1 functional domains and individual isoforms

The REX1-M123I-FLAG, REX1-M147I-FLAG constructs were generated through inverse PCR amplification of the REX1-FLAG parental construct using site-directed mutagenesis primers described in Table 5 (A&B and C&D respectively) in which the corresponding ATG is converted to ATC. The parental plasmid was digested with DpnI following PCR amplification and linear products self-circularized using a T4 DNA ligase (NEB) prior to transformation into Mach1 competent cells. The REX1-M123I-M147I-FLAG construct was subsequently generated through the PCR amplification of REX1-M147I-FLAG using the M123I primers (A&B; Table 5), as described.

The REX1- Δ M123-FLAG and REX1- Δ M147I-FLAG constructs were generated by the PCR amplification of the REX1-FLAG plasmid using primers containing attB sites (E&G and F&G, respectively, Table 5) to amplify the desired region of the REX1 open reading frame and cloned into a pDONR221 donor vector using a Gateway™ BP Clonase™ II enzyme mix (Invitrogen). The inserts were then subsequently cloned into a pB-TAG destination vector using a Gateway™ LR Clonase™ II enzyme mix (Invitrogen).

The REX1- Δ M123-M147I-FLAG construct was generated through inverse PCR amplification of the REX1- Δ M123-FLAG plasmid using primers C and D (Table 5). The parental plasmid was digested with DpnI following PCR amplification and linear products self-circularized using a 2X Ligation Mix (Takara) prior to transformation into Mach1

competent cells. All constructs were verified by sequencing (MOBIX; McMaster University).

Table 5: Primers used for the generation of hsREX1 mutant constructs

Primer		Sequence (5'→3')
A	REX1-M2I-F	CAAGCTCCCTTGAATGTTCTTTGGAATACATCAAAAAA GGGGTAAAG
B	REX1-M2I-R	GTTCGAGGGAACCTTACAAGAAACCTTATGTAGTTTTTT CCCCATTTC
C	REX1-M3I-F	GAGAGAATTCGCTTGAGTATTCTGAGTACATCACAGGC AAGAAG
D	REX1-M3I-R	CTCTCTTAAGCGAACTCATAAGACTCATGTAGTGTCCG TTCTTC
E	ATTB-REX1- TIS-M2-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAAAAA AGGGGTAAAGAA
F	ATTB-REX1- TIS-M3-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGACAGG CAAGAAGCTTCC
G	ATTB-REX1- FLMUTR	GGGGACCACTTTGTACAAGAAAGCTGGG

14. Assessment of REX1 isoform expression in HEK293Ts

Upon reaching 70-90% confluence, HEK293Ts were transfected using a piggybac based transfection system and Lipofectamine® LTX with Plus™ Reagent (Invitrogen). Cells were transfected with a total of 2µg of DNA including rttA-IPpA (0.63µg), PCYL43 (0.19µg) and a pBTAG expression vector (1.14ug). Cells were subsequently treated with 500ng/mL doxycycline the following day (D1822, Sigma). Protein expression was analyzed by western blot analysis 24 hours after doxycycline treatment.

15. 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 μ M retinoic acid, conditions. Cells were cultured for 2 days and subsequently collected for western blot analyses.

16. Immunostaining and high content imaging

Mouse ESCs were cultured in 96 well plates, fixed with 4% PFA, and permeabilized with 100% ice cold methanol. Cells were stained with primary antibodies for 3XFLAG, Oct4, Sox2, and Nanog in 1% BSA for 2 hours at room temperature or 4°C overnight. Cells were washed with PBS and subsequently stained with Hoechst 33342 (H1399; Invitrogen) and secondary antibodies for 1 hour at room temperature. Plates were imaged using an Operetta High Content Screening System (Perkin Elmer). Primary antibody dilutions are as follows: mouse anti-3XFLAG (1:1000), rabbit anti-Oct4 (1:200), mouse anti-Sox2 (1:400), and rabbit anti-Nanog (1:1000). Secondary antibody dilutions are as follows: donkey anti-mouse AF546 (1:500) and goat anti-rabbit AF647 (1:500).

RESULTS

1. Identification of novel Rex1 isoforms in pluripotent stem cells

Rex1 has been implicated in playing a number of important roles in PSCs, most of which it carries out through the DNA binding abilities of its C-terminal zinc fingers^{60,66–69,108}. However, the functional domains within the N-terminus of the Rex1 protein contributing to its function remain to be fully characterized. To address this issue, our lab had previously generated a series of N-terminal truncation mutants of the human REX1 protein, in addition to a mutant lacking the C-terminal zinc fingers (Figure 2A). Upon introducing these over-expression constructs into hESCs, it was apparent that two unexpected bands were present in all samples in which REX1 was C-terminally 3X-FLAG epitope tagged (Figure 2B). The presence of these additional bands had not been previously reported in any studies, and the absence of these products within the samples containing an N-terminally FLAG tagged REX1 protein (Figure 2B), suggested they may be functional isoforms of the full-length protein.

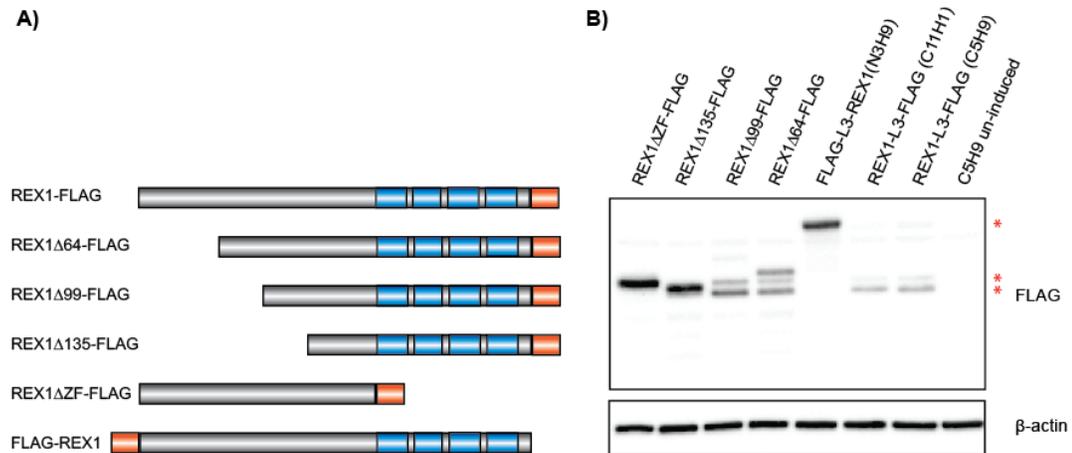


Figure 2: Evaluating the expression of *REX1* mutants in hESCs. A) Human *REX1* truncation mutant constructs generated by inverse PCR B) H9 and H1 hESCs were transfected using a three-plasmid piggybac system allowing for the stable integration of the wildtype and mutant *REX1*-FLAG over-expression constructs. Resultant cell lines were analysed by western blotting. FLAG was detected in whole cell lysates of all samples in which *REX1* expression was induced by doxycycline treatment. No *REX1* expression was observed in the uninduced negative control. Red asterisks denote bands corresponding to *REX1*-FLAG isoforms of interest. β -actin was used as a loading control.

Analysis of the human *REX1* open reading frame revealed the presence of two downstream methionine residues at amino acid positions 123 and 147. To investigate whether the additional start codons were being used as alternative translation initiation sites, we generated a series of mutant constructs by site directed mutagenesis in which the downstream methionine residues were converted to isoleucine (Figure 3A). Western blot analysis of HEK293Ts transiently transfected with these constructs confirmed that upon ablating a specific methionine residue, the corresponding protein product is lost (Figure 3B). Analysis of the murine *Rex1* open reading frame revealed the presence of six

methionine residues, two of which are conserved with the human REX1 protein (Figure 4), suggesting this may be a phenomenon that also occurs in mESCs.

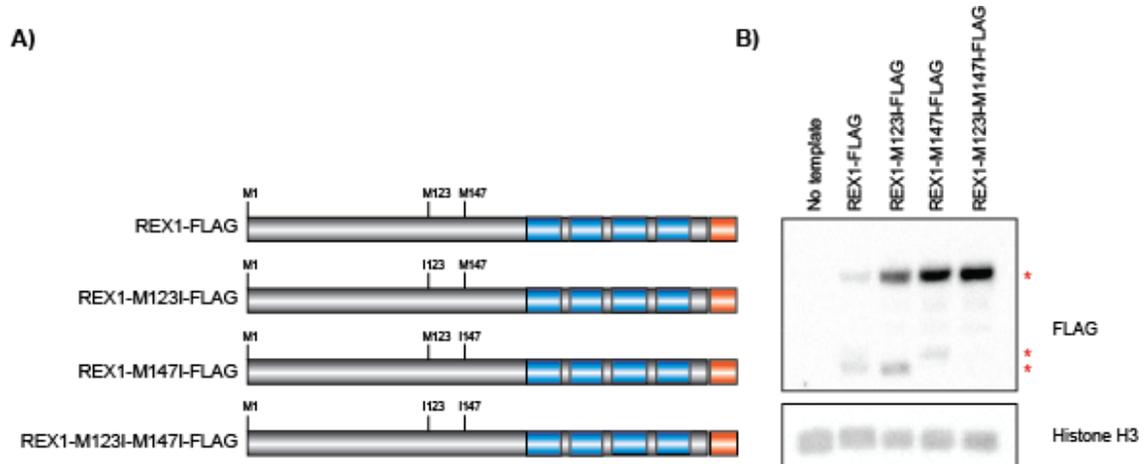


Figure 3: Missense mutations disrupt the downstream translation initiation sites the human *REX1* open reading frame. A) Constructs were generated by whole-plasmid PCR amplification using mutagenic primers to convert downstream methionine residues to isoleucine. B) HEK293Ts were transiently transfected with wildtype or mutant REX1-FLAG expression vectors and treated with 500ng/ml doxycycline the following day. Resultant cell lines were analyzed by western blotting 24 hours after doxycycline treatment. FLAG expression was observed from whole cell lysates of all cell lines with the exception of the no template negative control. Loss of FLAG expression where a methionine residue had been ablated was readily observed. Red asterisks denote bands corresponding to REX1-FLAG. Histone H3 was used as a loading control.

```

sp|P22227|ZFP42_MOUSE      MNEQKMNEQMKKTAKTSGQKGPGRALDRLTLKQDEARPVQNTRV EAPRVTYTIRDES--
sp|Q96MM3|ZFP42_HUMAN      -----MSQQLKKRAKTRHQKGLGGRAPSGAKPRQGKSSQDLQAEIEPVSAVWALCDGYVC
                             *.:*:* * * * * * * * * * . . :*:: :::* . .::: *

sp|P22227|ZFP42_MOUSE      --EISPETEEDGFDPDGYLECIIRGEFSEPILEEDFLFKSFESLEEVE--QNSRQVLEASS
sp|Q96MM3|ZFP42_HUMAN      YEPGPQALGGDDFSDCYIECVIRGEFSQPILEEDSLFESLEYLKKGSEQLSQKVF E-AS
                             * . * * * :*:*:*****:***** **:*: * : : . * :*:*: * : *

sp|P22227|ZFP42_MOUSE      LLESSLEYMTKGTKQEKTEVTQETPPL-----RVGASSLLAGGP AE
sp|Q96MM3|ZFP42_HUMAN      SLECSLEYMKKGVKKELPQKIVGENSLEYSEYMTGKKLPPGGIPGIDLSDPKQLAEFARK
                             ** .*****.*.*:* : * : : . ** :

sp|P22227|ZFP42_MOUSE      KPEGGVYCGVLSMLECPQAGCKKKLRGKTALRKHMLVHGPRRHVCAECGKAFT ESSKLKR
sp|Q96MM3|ZFP42_HUMAN      KPPINKEYDLSAIACPQSGCTRKLRNRAALRKHLLIHGPRDHVCAECGKAFV ESSKLKR
                             ** . . * * : * * : * * : * * : * * : * * : * * : * * : * * : * *

sp|P22227|ZFP42_MOUSE      HFLVHTGEKPYQCTFEGCGKRFSLDFNLRTHIRIHTGERRFVCPFDGCEKSFIQSNNQKI
sp|Q96MM3|ZFP42_HUMAN      HFLVHTGEKPFRC T FEGCGKRFSLDFNLRTHVRIHTGEKRFVCPFQGCNRRFIQSNNLKA
                             *****: : *****:*****:*****:*****:***** *

sp|P22227|ZFP42_MOUSE      HILTHAKAGKKC----
sp|Q96MM3|ZFP42_HUMAN      HILTHANTNKNEQEGK
                             *****: : *

```

Figure 4: Sequence alignment of the human and mouse Rex1 protein. Mouse and human Rex1 protein sequences were aligned using the Clustal Omega multiple sequence alignment tool^{70,71}. Red asterisks denote regions of conservation. Methionine residues within the open reading frame of the human and mouse proteins are highlighted in yellow.

2. CRISPR-mediated homologous recombination to epitope tag Rex1 isoforms

To evaluate the endogenous expression of the Rex1 protein and its isoforms, we generated mESC lines in which the Rex1 protein was C-terminally 3XFLAG-tagged using CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats-CRISPR associated protein 9) technology. The Cas9 nuclease is an RNA-guided genome editing tool that is used to facilitate the generation of a double stranded break (DSB) at target sequences specified by a 20-nucleotide guide RNA^{109,116}. The DSB is then repaired by one of two pathways; the error-prone non-homologous end joining (NHEJ) pathway or precise homology directed repair (HDR) pathway^{109,116}. Alternatively, a nickase mutant of the Cas9 nuclease can be used with paired guide RNA to minimize any off-target cleavage effects¹⁰⁹.

We exploited the HDR pathway to introduce a 3XFLAG epitope-tag into the *Rex1* genomic locus through co-transfection of the Cas9 nuclease or Cas9 nickase pair, 20-nucleotide guide RNA(s), and an exogenous targeting vector repair template with arms of homology flanking the site of the DSB. Guide RNA sequences were designed such that the DSB would occur in the region coding for the four zinc finger motifs responsible for Rex1's DNA binding abilities (Figure 5A). Integration of the strategically designed targeting vector ensured the *Rex1* open reading frame remained intact while allowing for the addition of a 3XFLAG, P2A self-cleaving peptide, and puromycin resistance gene upstream of the stop codon (Figure 5B). To ensure the targeting vector repair template was not cut during the editing process, the protospacer adjacent motif (PAM) sequence, a NGG-nucleotide motif located 3-prime of the 20-base pair target sequence, were silently mutated through a single base pair change guaranteeing no change amino acid sequence and therefore protein function.^{109,116} Addition of puromycin to the culture medium was used to enrich for successful homologous recombination.

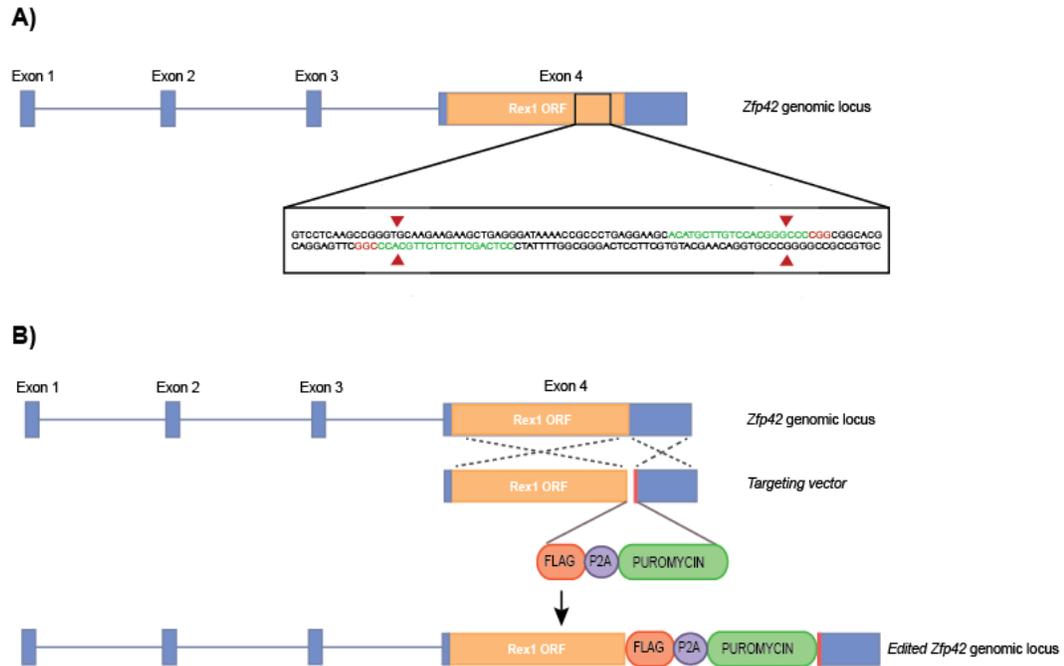


Figure 5: Schematic representation of the targeting strategy used for the generation of C-terminal Rex1 3XFLAG epitope-tagged mESCs. E14Tg2a mESCs were co-transfected with a CRISPR sgRNA(s) and targeting vector repair template to facilitate homology directed repair (dotted lines) and integration of a C-terminal 3XFLAG tag into the *Rex1* locus. A) Wildtype *Rex1* genomic locus annotated with PAM sequences (red) and sgRNA sequences used for the generation of the DSB (green). B) Edited *Rex1* genomic locus containing a C-terminal 3XFLAG epitope-tag, puromycin selectable marker and P2A cleaving peptide, which allows for the simultaneous expression of Rex1-FLAG and puromycin, separately, in successfully targeted mESCs.

3. Detection of endogenous protein expression of epitope-tagged Rex1 isoforms

The presence of the 3XFLAG epitope tag within the targeting vector repair template provided a mechanism for rapid screening of clones in which successful homologous recombination events had occurred. Western blot analysis using an anti-FLAG antibody revealed select isolated clones were successfully epitope tagged, indicated by the presence of a band at approximately 50 kDa corresponding to the 3XFLAG-epitope tagged Rex1

protein in all whole cell lysates with the exception of the E14T wildtype negative control (Figure 6).

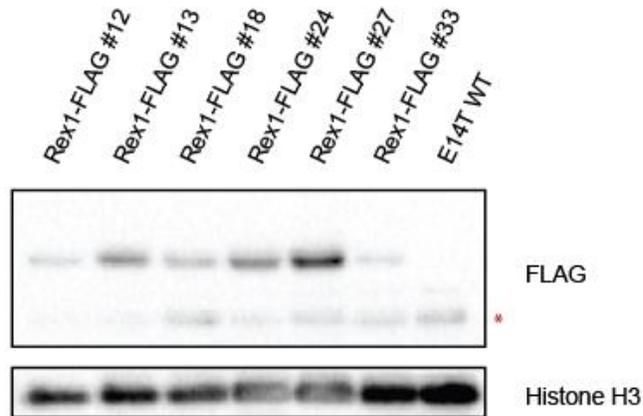


Figure 6: Detection of endogenously 3XFLAG epitope-tagged Rex1 in isolated clones. Rex1-FLAG expression was detected in whole cell lysates of select isolated clones but not in the E14T wildtype negative control. Histone H3 was used as a loading control. Non-specific binding is denoted by the asterisk.

As Rex1 is a nuclear protein and endogenous expression was not readily detectable in whole cell lysates, we attempted to isolate nuclear protein from endogenous 3XFLAG-epitope tagged mESCs through cellular fractionation. Western blot analysis performed on the nuclear fractions of wildtype and Rex1-FLAG mESCs revealed the presence of several Rex1 isoforms (Figure 7).

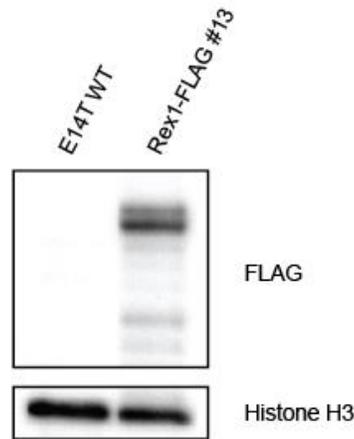


Figure 7: Evaluating protein expression levels of endogenously C-terminally 3XFLAG tagged Rex1. Rex1-FLAG expression was readily detectable in the nuclear lysate of our flag tagged mESC line but not the E14T wildtype negative control. The presence of at least four Rex1 isoforms are observed. Histone H3 was used as a loading control.

In attempt to evaluate the localization of the endogenous Rex1 protein, we performed immunofluorescence staining on wildtype and Rex1-FLAG mESCs. Unfortunately, Rex1-FLAG was not readily detectable by immunofluorescence, likely due to low endogenous protein expression levels. However, the results did demonstrate that the Rex1-FLAG mESCs express Oct4, Sox2, and Nanog, as expected (Figure 8).

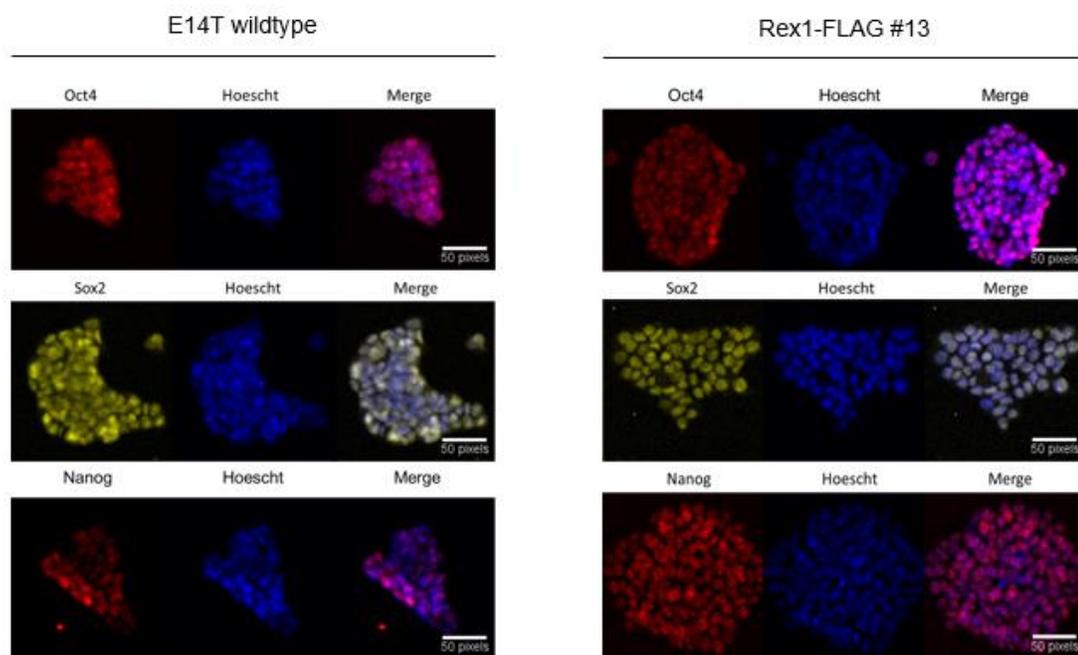


Figure 8: Immunofluorescence staining of wildtype and Rex1-FLAG mESCs. E14T wildtype and Rex1-FLAG mESCs were fixed and stained for FLAG, Oct4, Sox2, and Nanog. Due to low endogenous protein expression levels, Rex1-FLAG expression was not detectable by immunostaining. Rex1-FLAG mESCs express Oct4, Sox2, and Nanog, consistent with wildtype cells. DNA was stained with Hoechst.

4. Mapping the global genomic binding sites of epitope-tagged Rex1 isoforms

Upon successful generation of our endogenously C-terminal 3XFLAG epitope-tagged Rex1 mESC cell lines, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) to investigate the genome-wide binding sites of Rex1 and its isoforms. Immunoprecipitated DNA was prepared from MEF depleted samples using a commercially available anti-FLAG antibody and library samples prepared using the DNA SMART™ ChIP-Seq kit prior to sending samples to Genome Québec for Illumina sequencing. Mapping of sequencing reads followed by peak calling was performed using

the pipeline kindly provided by Dr. Mathieu Lupien, which resulted in the identification of 179 peaks.

To associate our peaks with functionally relevant genomic regions we performed genome annotation using the Bioconductor package ChIPseeker and the Genomic Regions Enrichment of Annotations Tool (GREAT)^{117,118}. Of the peaks identified, 136 correspond to genomic regions associated with more than one gene, 42 with a single gene, and 1 with no genes (Figure 9A), indicating that Rex1 binding was relatively close to gene bodies. Genome annotation revealed that Rex1 binds primarily within distal intergenic and promoter regions, but also displays a significant amount of binding within introns (Figure 9B). Specifically, within genic regions, Rex1 seems to bind primarily between 50-500 kilobases downstream of the transcriptional start site (TSS; Figure 9C). The most significant of which mapped to genes including *Igsf21*, *Unc5a*, and the known Rex1 target gene *Tsix* (Figure 9D). To identify the molecular and biological processes associated with Rex1 in mESCs, we performed Gene Ontology (GO) enrichment analysis. GO term enrichment revealed Rex1 to be associated primarily with molecular functions related to DNA binding and catalytic activity whereas biological processes associated with Rex1 include metabolic and cellular processes^{108,119}. To validate our ChIP, we performed motif discovery analysis. As expected, motif discovery analysis revealed enrichment of motifs bound by Rex1, Yy1, and Yy2 since their C-terminal zinc fingers remain highly conserved and display similarities within the core of their binding motifs (Figure 9E)¹²⁰. Thus, providing confidence in the success of our experiment.

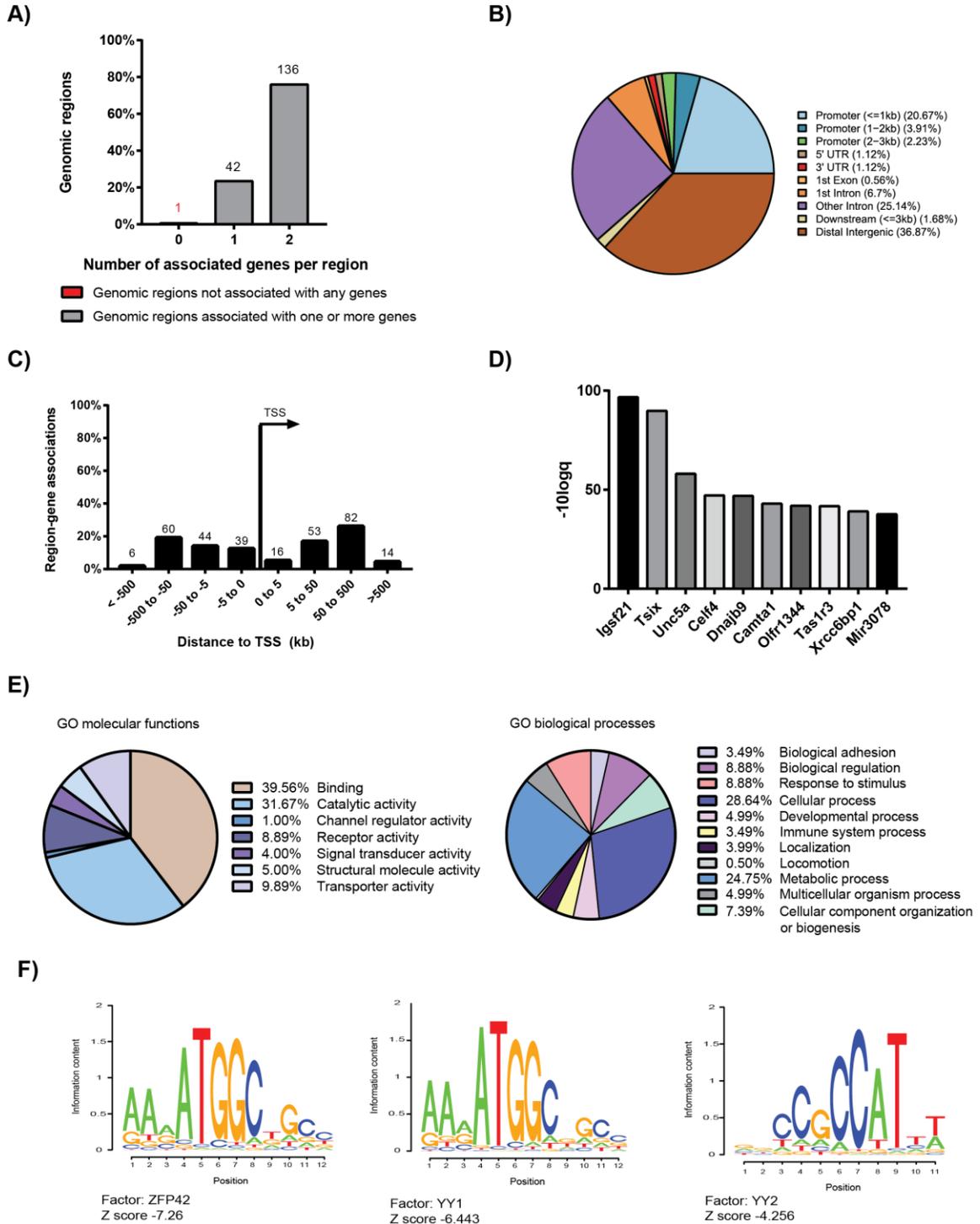


Figure 9: Overview of genomic annotation, Gene Ontology enrichment and motif discovery analyses performed on murine Rex1 ChIP-seq data. A) GREAT analysis revealed of the 179 peaks identified 136 mapped to genomic regions associated with 2 genes, 42 with a single gene, and 1 with no genes. B) Genomic annotation analysis performed using ChIPseeker by Bioconductor displayed high levels of enrichment of Rex1 binding within promoter and distal intergenic regions. C) GREAT analysis demonstrated Rex1 binds primarily between 50-500 kilobases downstream of the TSS. D) Top 10 most significant Rex1 genic targets identified by ChIP-seq analysis include differentiation related genes including *Igsf21* and *Unc5a* in addition to the known Rex1 target gene *Tsix*. Target genes are ordered according to significance as determined by $-10\log q$ values. E) GO term enrichment analysis reveals molecular functions and biological processes associated with Rex1 based on target genes. F) Motif discovery analysis revealed enrichment of motifs recognized by beta-beta-alpha zinc finger family of transcription factors within our dataset including Rex1, Yy1, and Yy2.

5. ChIP-seq reveals a conserved role of Rex1 in the regulation of endogenous retroviruses in human and mouse embryonic stem cells

We sought to determine whether the involvement of Rex1 in the regulation of endogenous retroviral elements was a conserved role amongst the human and mouse Rex1 proteins. Through the assistance of Dr. Matthew Lorincz and his lab members, we were able to identify the families and classes of ERVs which Rex1 associates with in both human and mouse PSCs. Our data suggests that within both human and mouse ESCs, Rex1 primarily binds LTR elements, but will also bind to LINE and SINE elements to a lesser extent (Figure 10A, Figure 11A). This notion was further supported by our mESC data which demonstrated that Rex1 binds preferentially to RMER21A and IAPEY classes, both of which belong to the LTR family (Figure 10B). However, this was not entirely consistent with our human dataset in which we observed preferential binding to HSATII satellite repeats amongst other satellite repeat elements (Figure 11B). Nevertheless, despite this

difference in preferential binding, these data suggest a conserved role for Rex1 in the regulation of ERVs and other repetitive elements in PSCs.

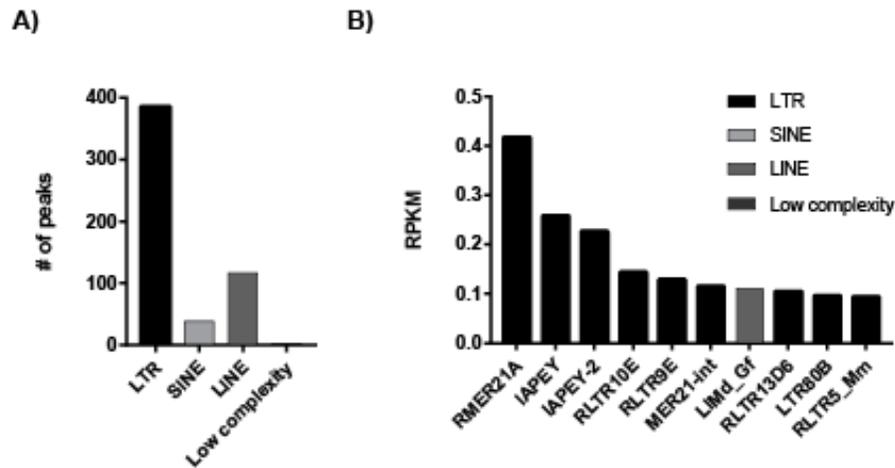


Figure 10: Overview of Rex1 association with ERVs in mESCs. A) Binding preference of Rex1 amongst ERV families B) Rex1 displays highest enrichment and preferential binding to RMER21A and IAPEY elements in mESCs. Enrichment was determined using reads per kilobase of transcript per million mapped reads (RPKM).

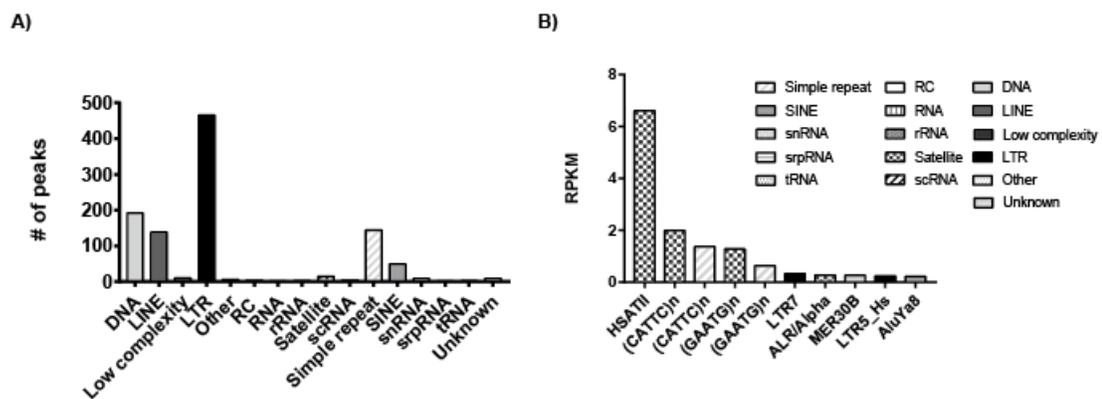


Figure 11: Overview of REX1 association with ERVs in hESCs. A) Binding preference of REX1 amongst ERV families B) Rex1 displays highest enrichment and preferential binding to HSATII and other satellite/simple repeat elements in hESCs. Enrichment was determined using reads per kilobase of transcript per million mapped reads (RPKM).

As Yy1 and Yy2 also play a role in the regulation of ERVs in mESCs, a number of which are shared targets amongst this protein family, our goal was to identify how Rex1 is regulating these elements independently of Yy1 and Yy2. To evaluate which of our ERV targets are specific to Rex1 we took our data and compared it to existing ChIP-seq data sets for Yy1 and Yy2. Comparison of the peaks identified within previously published Yy1 and Yy2 data sets with our Rex1 dataset, and overlaying this data with repeat masker tracks demonstrated in comparison to its family members, Rex1 is found to bind RMER21A and IAPLTR3-int elements at a much higher frequency than Yy1 and Yy2 (Figure 12). Of the peaks identified, 30% and 26% of peaks corresponded to RMER21A and IAPLTR3-int elements within our Rex1 data set, respectively. In contrast, no peaks within the Yy1 dataset corresponded to either of these elements, and only 2% of peaks within the Yy2 dataset corresponded to RMER21A elements. The higher frequency of binding by Yy1 and Yy2 to LIMd_T, ID_B1, and (other) elements, suggests that although Rex1 and its family members display similar binding to some elements, their targets do not overlap entirely. Due to Rex1's preferential binding to RMER21A and IAPLTR-3 elements in comparison to its family members we chose to focus primarily on these classes of ERVs in downstream analyses.

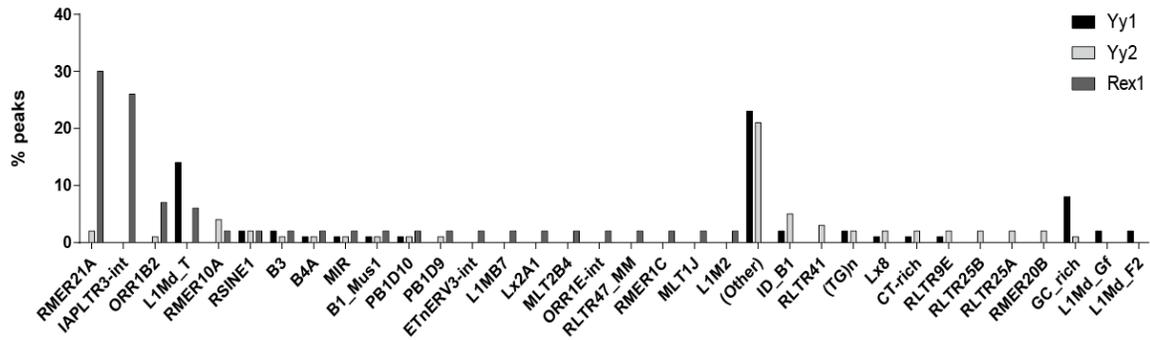


Figure 12: Assessment of Yy1, Yy2 and Rex1 association with ERVs in mESCs. Comparison of peaks identified in previously published Yy1 and Yy2 data sets with our Rex1 dataset demonstrates a higher frequency of Rex1 binding to RMER21A and IAPLTR3 elements whereas Yy1 and Yy2 demonstrate higher frequencies 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.

6. Disruption of Rex1 expression via CRISPRs results in the dysregulation of Rex1 target genes and endogenous retroviral elements

To assess the mechanisms by which Rex1 is regulating its targets, we generated C-terminal Rex1 KO mESC lines using the CRISPR-Cas9 system previously described. Without the presence of a repair template, the DSB generated by the Cas9 nuclease is repaired by NHEJ, which often results in small insertions/deletions (indels) at the site of the DSB (Figure 13A,B).^{109,116} In both C-terminal Rex1 KO lines that we have generated, a premature stop codon was introduced within the first zinc finger and within the second zinc finger, respectively (Figure 13C). It has been previously demonstrated that the second and third zinc finger of Yy1 are required for efficient nuclear targeting and therefore protein function¹²¹. As the zinc fingers of Rex1 remained highly conserved after its generation through retroposition of Yy1, the loss of the second and third zinc fingers provides high

confidence that it should result in the loss of Rex1's DNA binding capacity in addition to its ability to interact with known protein partners. As expected, western blot analysis revealed our C-terminal KO cell lines express Oct4, Sox2 and Nanog (Figure 13D), suggesting pluripotency is not compromised in knockout lines, a result consistent with previous reports^{66,108}.

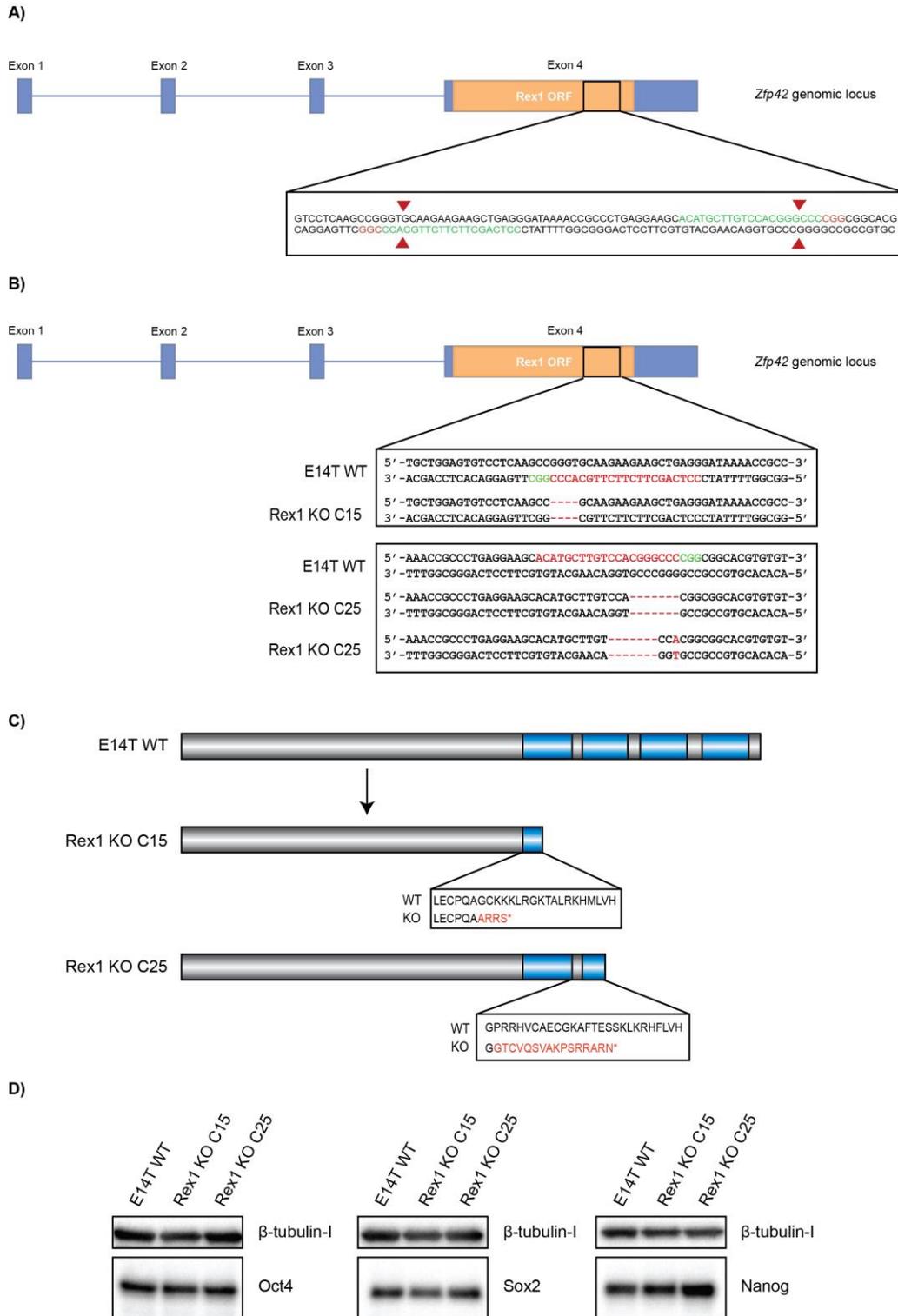


Figure 13: Schematic representation of the targeting approach used for the generation of C-terminal Rex1 KO cell lines. E14Tg2a mESCs were transfected with a CRISPR sgRNA to generate a double stranded break within region of the *Rex1* open reading frame encoding for the zinc finger motifs. A) Wildtype *Rex1* genomic locus annotated with PAM sequences (red) and sgRNA sequences used for the generation of the DSB (green). B) Edited *Rex1* genomic locus in C-terminal KO clones 15 and 25. Sequencing of the loci within these two clones revealed the presence of deletions and an insertion (shown in red) within the *Rex1* open reading frame encoding for the zinc finger motifs. C) Mutations within C-terminal KO clones 15 and 25 resulted in the formation of premature stop codons within the zinc fingers of Rex1. The premature stop codons occurred within the first zinc finger and the second zinc finger, respectively. D) E14T wildtype, Rex1 KO C15 and Rex1 KO C25 cells were cultured in mESC media with feeders. Cells were collected after MEF depletion and subsequently analyzed by western blot. Oct4, Sox2 and Nanog are observed across all samples, consistent with the literature. B-tubulin I was used as a loading control.

Evaluation of target gene expression upon Rex1 depletion by qRT-PCR revealed that in comparison to wildtype E14Ts, our C-terminal Rex1 KO lines display a significant increase in the expression of differentiation-related genes including *Unc5a* and *Igsf21* (Figure 14), known to play roles in axon guidance and immune response, respectively. This suggests a disruption of Rex1 function in our KO cell lines and supports our hypothesis that Rex1 and its isoforms are acting to regulate lineage-related genes in PSCs. Further analyses performed to evaluate the regulation of ERVs revealed that in comparison to wildtype there is a dysregulation of ERV expression in our KO cell lines (Figure 15). Quantitative RT-PCR analyses demonstrated a general trend of ERV repression by Rex1 PSCs, specifically of RMER1C, RMER21A, and IAPLTR3 classes. However, inconsistencies were observed between our KO cell lines at two RMER21A loci, specifically the elements located on chromosome 6 and chromosome 16. Nonetheless, overall these data suggest Rex1 is acting to repress differentiation related genes and

repetitive elements overall contributing to the maintenance of the pluripotent transcriptional network and genomic stability of PSCs.

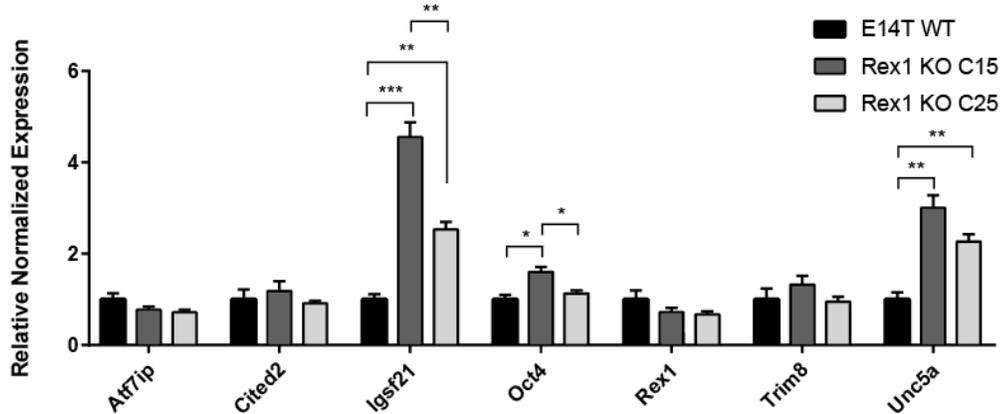


Figure 14: Analysis of Rex1 target gene expression in E14T WT, Rex1 KO C15 and Rex1 KO C25 mESCs. Cells were cultured in the presence of MEFs in mESC media for 2 days. RNA was then isolated from MEF depleted samples and contaminating DNA removed through DNase I treatment. The housekeeping gene TBP was used as a baseline control. Three independent experiments (n=3) with three technical replicates per independent experiment were plotted with standard error mean. Expression was normalized to E14T WT. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (Unpaired t-test).

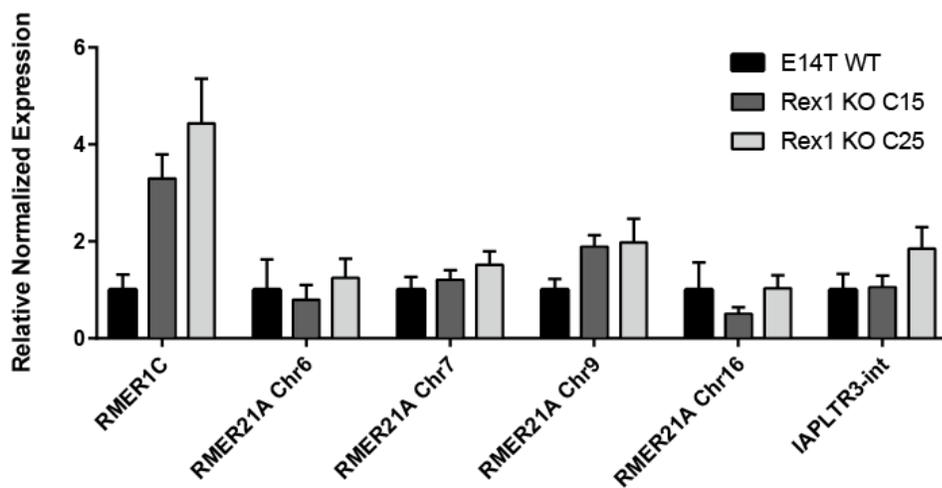


Figure 15: Analysis of ERV expression in E14T WT, Rex1 KO C15 and Rex1 KO C25 mESCs. Cells were cultured in the presence of MEFs in mESC media for 2 days. RNA was then isolated from MEF depleted samples and contaminating DNA removed through DNase I treatment. The housekeeping gene β -actin and NRT controls were used as a baseline. Two independent experiments (n=2) with three technical replicates per independent experiment were plotted with standard error mean. Expression was normalized to E14T WT.

As the functional domains of the Rex1 protein upstream of its C-terminal zinc finger motifs have not been characterized there is no guarantee that any truncated protein product formed within these C-terminal KO cells are completely non-functional. Therefore, to ensure that the changes we observed are a result of a complete loss of Rex1 function we have transfected wildtype and C-terminal knockout cells with a CRISPR targeting Rex1 within the first 100 base pairs of the open reading frame (Figure 16). We are currently screening isolated cell lines for the presence of indels within the selected targeted region.

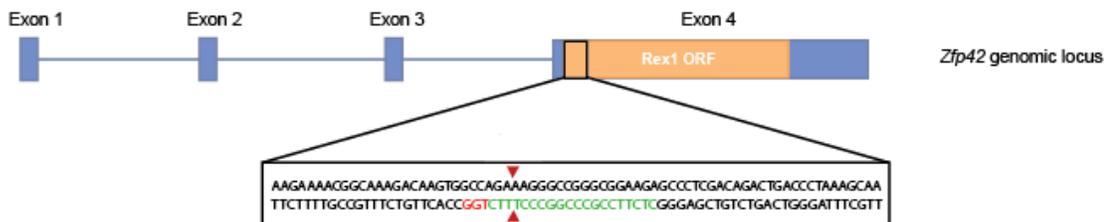


Figure 16: Schematic representation of targeting approach used for the generation of N-terminal Rex1 KO cell lines. E14Tg2a wildtype and Rex1 C-terminal KO mESCs were transfected with a CRISPR sgRNA to generate a double stranded break within the first 100 base pairs of the *Rex1* open reading frame. Schematic displays wildtype and C-terminal KO N-terminal *Rex1* genomic locus annotated with PAM sequences (red) and sgRNA sequences used for the generation of the DSB (green).

7. Characterizing the roles of Rex1 protein isoforms in PSCs

We have previously confirmed the presence of Rex1 isoforms by initiation of translation at alternative start codons within the Rex1 open reading frame at the exogenous level in the human context (Figure 3B). Furthermore, western blot analysis of our Rex1-FLAG lines suggests four or more isoforms of the Rex1 protein may be present within mESCs (Figure 7). To assess whether Rex1 and its isoforms display context-dependent expression in naïve pluripotent, mixed pluripotent, and differentiated states, Rex1-FLAG mESCs were grown in various culture conditions and protein expression assessed by western blot. Analysis of protein expression revealed that Rex1 isoforms display differential expression in different pluripotent states (Figure 17). Of the isoforms identified, each isoform demonstrated variable levels amongst each of the tested conditions, except for the EB conditions in which the expression of all isoforms was down-regulated. No Rex1-FLAG expression was observed in the E14T WT control, as expected. This suggests that the small and long forms of Rex1 act independently of one another in PSCs. To further investigate the independent roles of these isoforms, we have generated a series of constructs that will be transfected into our Rex1 KO cell lines (Figure 18). Through site-directed and truncation mutagenesis techniques, the constructs were generated such that the REX1 open reading frame contains only the desired translation initiation sites to ensure the expression of one or more isoforms.

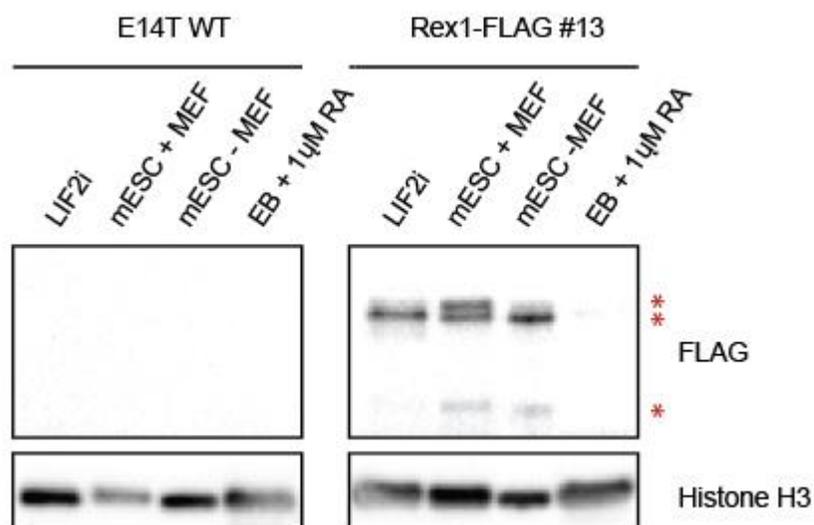


Figure 17: Assessment of Rex1-FLAG expression in LIF2i versus serum conditions. Cells were cultured in LIF2i, mESC with MEF, mESC without MEF, and EB (5% FBS) + 1μM RA medium conditions and subsequently collected for western analyses. Rex1-FLAG expression was readily detectable in the nuclear lysates of Rex1-FLAG samples but not the wildtype negative control. Bands corresponding to Rex1 isoforms are denoted by asterisks. Histone H3 was used as a loading control. n=2.

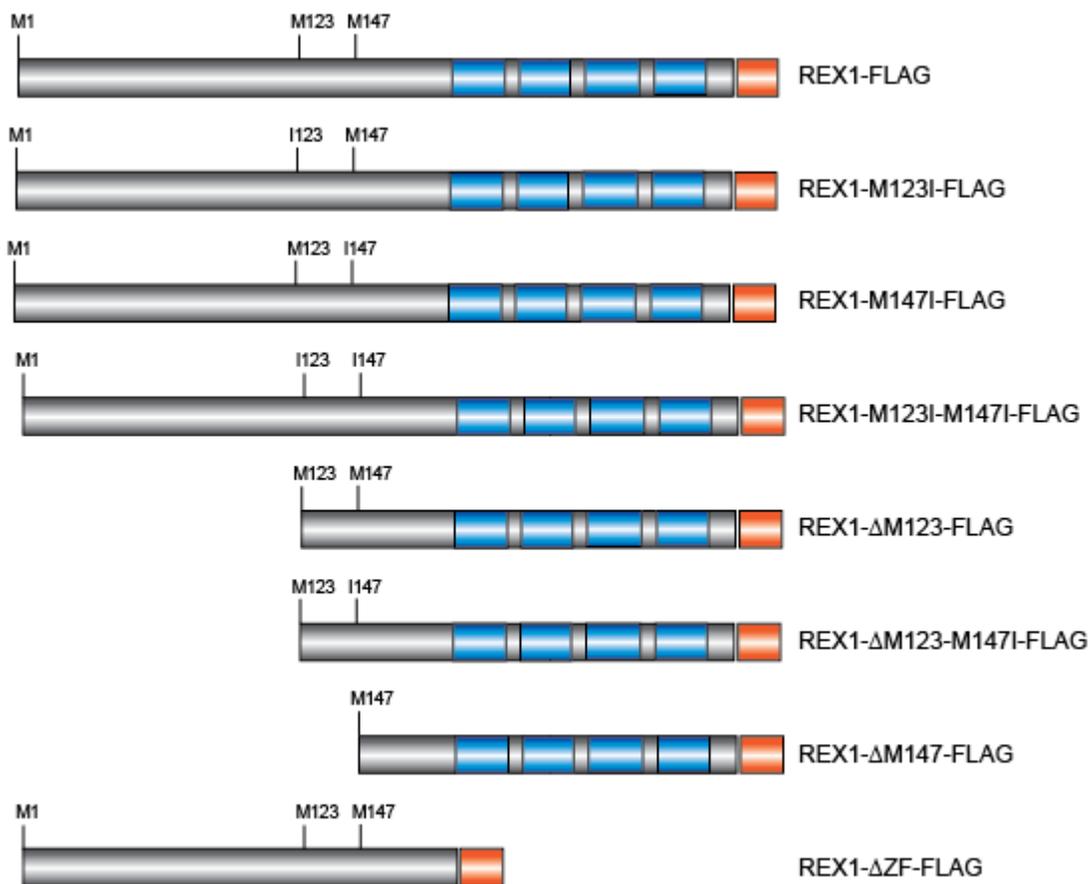


Figure 18: Generation of mutant human REX1 expression vectors for the characterization of isoforms. Truncation mutant constructs were generated through the PCR amplification of the pBTAG-REX1-L3-FLAG expression vector using attb primer sets designed such that the inserts for the pDONR221 vector begin at the second and third translation initiation sites in the *REX1* open reading frame, respectively. The inserts were subsequently cloned from pDONR221 into the pBTAG destination vector through a LR gateway cloning reaction. Undesired methionine residues were converted to isoleucine residues by site-directed mutagenesis.

DISCUSSION

Previous work in our lab established the existence of novel human REX1 protein isoforms that may act independently of the full-length protein. Through site-directed mutagenesis-mediated conversion of methionine residues to isoleucine within the open reading frame, we were able to demonstrate that the ablation of a methionine residue resulted in the loss of its corresponding protein product (Figure 3B), confirming the usage of alternative translation initiation sites in the exogenous context. Since these isoforms had not been described in any previous studies, and this phenomenon had not been established at the endogenous level, we translated this project into mESCs, as they are a convenient and reliable platform for studying the molecular basis of pluripotency. Thus, this thesis aimed to delineate our undertaking in characterizing the roles of Rex1's isoforms and the mechanisms by which Rex1 carries out its functions in PSCs.

As there is currently no high-quality commercially available antibody for the detection of endogenous Rex1 protein expression, we utilized CRISPR-Cas9 technology to successfully generate mESC lines in which Rex1 is C-terminally FLAG-epitope tagged. Through western blot analysis of our Rex1-FLAG lines, we established the presence of several Rex1 isoforms in mESCs, confirming the usage of alternative initiation start sites at the endogenous level (Figure 7). In attempt to evaluate the localization of the Rex1 protein and its isoforms, we performed immunofluorescence staining of wildtype and Rex1-FLAG mESCs for FLAG, Oct4, Sox2, and Nanog (Figure 8). Rex1-FLAG mESCs stained positive for Oct4, Sox2 and Nanog, similar to wildtype, as expected, since the introduction of the targeting vector should not have effected the pluripotent nature of the

cells. However, no positive staining was observed for FLAG as a result of low endogenous Rex1 levels, which were likely further reduced due to the absence of feeders within the culture prior to analysis. Despite the absence of feeders, low endogenous expression was also a limitation when evaluating Rex1 expression in whole cell lysates, which was resolved through the use of cell fractionation techniques. Culturing of cells in LIF2i and performing subsequent immunofluorescence analysis may aid in overcoming limitations due to low and heterogeneous Rex1 expression in mixed pluripotent cultures.

To gain an unbiased understanding of the genes regulated by Rex1 in mESCs, we performed CHIP-seq analysis, which revealed binding of Rex1 at 179 genic regions (Figure S1), which is considerably low given Rex1 is a DNA-binding transcription factor.^{122,123} As our mESC lines are endogenously tagged and Rex1 expression levels are low this likely accounts for the small number of peaks identified within our data set in comparison to other transcription factor based CHIP experiments in which proteins are endogenously more abundant or are over-expressed above normal physiological levels. Furthermore, validation of our dataset by motif enrichment established the presence of motifs corresponding to factors including Yy1, Yy2 and Rex1 (Figure 9F), confirming the success of our experiment, as these proteins are known to share a core motif binding sequence and therefore, gene targets within PSCs⁶⁰. GO term enrichment analysis revealed metabolic processes, developmental processes, cellular processes and locomotion as biological processes associated with Rex1 through our list of target genes (Figure 9E), consistent with the literature, additionally supporting the validity of our dataset¹⁰⁸. It was initially unclear as to why genomic annotation of our identified peaks revealed a significant amount of

binding to distal intergenic regions. However, further analysis our dataset revealed a significant binding of Rex1 to repetitive elements. As these elements are present throughout the genome, this clarified why Rex1 seemed to binding within distal intergenic regions more than promoter regions.

Rex1 and its family members have been previously described as negative regulators of ERVs in mESCs^{69,87}, we therefore sought to determine whether this was a conserved role amongst the human and mouse Rex1 proteins. Our data suggest that, within mESCs, Rex1 preferentially binds LTR elements, specifically RMER21A and IAPEY elements, whereas in hESCs Rex1 is found to bind predominantly within LTR elements but preferentially to HSATII satellite repeats (Figure 10, 11). Within mouse and human PSCs, the most active ERVs are the IAP and human endogenous retrovirus subfamily H (HERVH) elements, respectively, the expression of which have been demonstrated to track closely and aid in the maintenance of the pluripotent identity^{99,124–127}. Although we observed preferential binding to IAPs within our murine Rex1 data, the same does not hold true for HERVH elements within our human REX1 dataset. Alignment of the human and mouse Rex1 protein sequences indicates considerable homology amongst the two proteins, suggesting a conserved role of protein function amongst species (Figure 4). Thus, this difference in enrichment at highly active elements may be due to the fact that REX1 is expressed at higher than physiological levels upon induction by doxycycline in our REX1 hESCs over-expression lines, in comparison to our mESCs in which Rex1 is endogenously FLAG-tagged. As a result, the REX1-FLAG protein may be saturating sites not normally bound by endogenous REX1, enriching for REX1 binding at what are normally weakly

bound sites. Alternatively, the difference in preferential binding may be a result of the inherent differences between human and mouse PSCs, since unlike mPSCs, which associate more closely with the naïve state, hPSCs more closely resemble EpiSCs, differing from mESC not only in global molecular signatures, signalling pathways, colony shape, growth rate, and surface markers, but also in their developmental potential^{33,128}. Nonetheless, the large number of binding sites identified within these elements in both datasets suggests a conserved role of Rex1 in the regulation of ERVs and maintenance of genomic integrity in mouse and human PSCs.

The generation of C-terminal knockout lines disrupting the DNA-binding capabilities of Rex1 allowed for the evaluation of the mechanisms by which Rex1 and its smaller isoforms are regulating genic and ERV targets in mESCs. The disruption of Rex1 function in these cell lines revealed a significant up-regulation of differentiation-related genes *Unc5a* and *Igsf21* in KO cell lines in comparison to wildtype (Figure 14), suggesting that Rex1 is acting as a negative regulator of differentiation-related genes. This supports our hypothesis that Rex1 is acting to regulate lineage-determining genes in PSCs. Although we additionally identified the previously known target gene *Tsix* as one of our 179 genic targets⁵⁷, we did not further pursue this gene in qRT-PCR analyses as our mESCs are male lines. Consistent with the literature, a loss of Rex1 in our KO cell lines had no effect on the expression of other pluripotent markers such as Oct4, Sox2 and Nanog, since Rex1 expression is not required for the maintenance of the pluripotent state (Figure 13D)^{66,108}. However, the increase in differentiation-related genes does suggest that our KO cell lines are representative of a lineage-primed pluripotent state, also as expected³⁶.

We additionally evaluated changes in ERV expression within our KO lines, which indicated that Rex1 may be acting as a repressor of ERVs in PSCs. Since the high frequency of binding to RMERA and IAPEY elements by Rex1 was not observed amongst its family members, it is reasonable to consider that Rex1 has evolved to independently regulate these particular classes of ERVs in PSCs (Figure 12). Thus, we aimed to focus on these particular families in downstream analyses to better investigate the regulation of ERVs by Rex1 independent of its family members. We additionally included two sites, RMER21A Chr6 and RMER21A Chr9, which overlap with loci identified in the Yy1 data set. However, some of the loci we assessed produced inconsistent results between KO cell lines (Figure 15). As ERVs are highly repetitive elements we attempted to ensure the removal of any contaminating genomic DNA within our RNA samples through DNase I treatment. In conjunction with highly specific probe-based qRT-PCRs, we were confident that any changes observed were, in fact, due to changes in expression of the selected ERV targets. We additionally normalized our sample values to both NRT controls and the housekeeping gene β -actin. However, the addition of more experimental replicates is required before these results can be fully interpreted. To evaluate the regulation of ERVs by Rex1, we took a target-based approach and selected loci were chosen based on significance and whether sites were solely occupied by Rex1 or co-bound by Yy1. In attempt to identify loci demonstrating more substantial changes in expression, it may be beneficial to pursue analyses such as RNA-seq which will provide a more global overview of the changes in gene expression within our KO cell lines. However, the lack of fluctuation in expression at

selected loci, may also be due to other ERV regulatory factors binding in the absence of Rex1.

Initial evaluation of endogenous Rex1-FLAG expression demonstrated variability in the observed levels of long and short isoforms of Rex1, suggesting these isoforms may play roles independent of the full-length protein and display context-dependent expression. Western blot analysis of cells cultured in LIF2i versus various serum conditions revealed that Rex1 and its isoforms vary in expression in each condition tested (Figure 17). Thus, Rex1 and its isoforms display differential expression, likely contributing to isoform-specific roles associated with the various sub-populations of pluripotency. In an attempt to further characterize the mechanisms contributing to alternative translation initiation site usage of Rex1 in PSCs, we observed the predicted binding of various RNA-binding proteins (RBPs) and associated factors to mouse Rex1 messenger RNA (mRNA) transcript (Figure 19). RBPs are master regulators of mRNA processing and also contribute to mRNA stability and the initiation of translation^{129,130}. The observed predicted binding of RBPs, including PABPC1, KHDRBS1, FXR1, and IGF2BP2, within the regions coding for the various initiation start sites, particularly those found within the 5-prime region, suggests that RBPs could be one potential mechanism governing the regulation of transcriptional start site usage, and that differential binding of these RBPs in various pluripotent states may contribute to the context-dependent usage of translational start sites of Rex1.

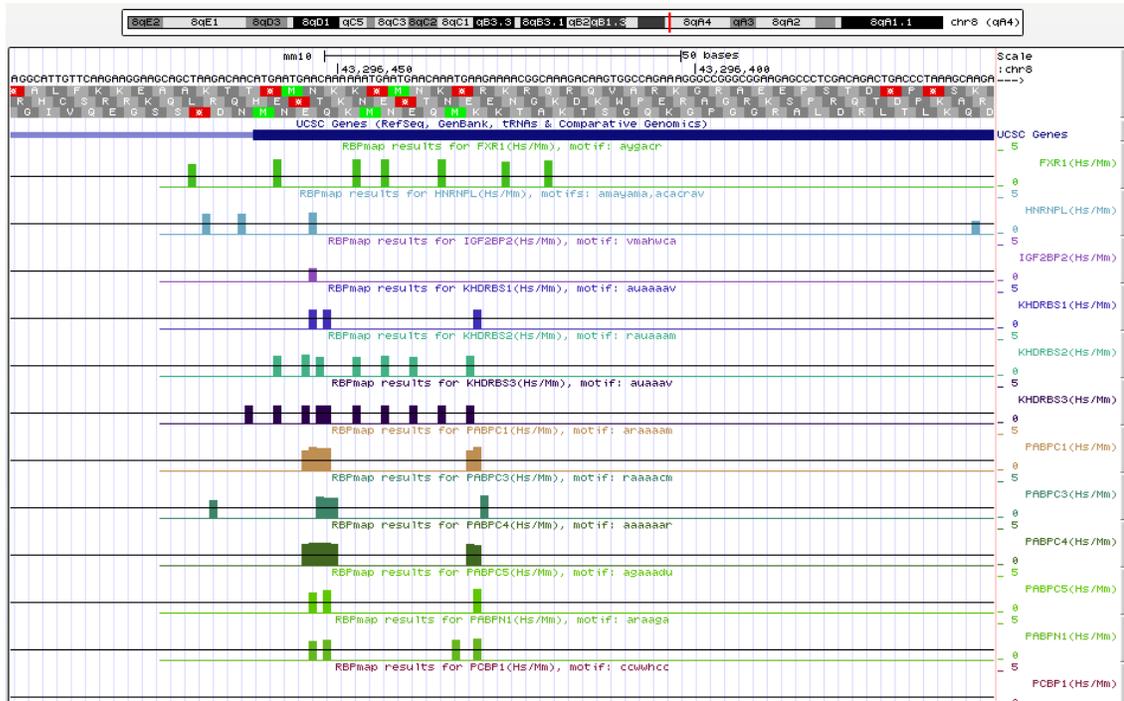


Figure 19: Predicted binding of RNA-binding proteins to murine Rex1 mRNA. Predictive binding of RBPs to the Rex1 mRNA transcript was performed using RBPmap¹³¹. Motif analysis revealed predictive binding sites for numerous RBPs within the most N-terminal region of Rex1 containing methionine residues M1, M6, and M10 including FXR1, KDHRBS1, and PABPC proteins.

FUTURE DIRECTIONS

While this thesis describes an in-depth characterization of the roles and molecular mechanisms of Rex1 in pluripotent stem cells, a few matters remain unresolved. Although we were able to successfully generate Rex1 C-terminal KO lines, there may be residual Rex1 function as a result of uncharacterized functional domains upstream of the zinc fingers. As a result, we are working towards screening C-terminal Rex1 KO mESCs that were subsequently transfected with the N-terminal CRISPR by DNA-based sequencing to assess for the generation of an additional indel that would prevent the generation of protein

products with residual function. The generation of such a line will ensure the loss of all Rex1 function due to the targeted disruption of the alternative translation initiation sites. Characterization of the N and C-terminal KO by qRT-PCR analyses to evaluate genic and ERV target regulation should then clarify whether any uncharacterized domains lie within the N-terminal region of the Rex1 protein, or whether the zinc finger domains are responsible for the majority of Rex1's function in PSCs. BLAST sequence analysis revealed that the N-terminal region of Rex1 does not display sequence similarity to any other known proteins in the genome, thus any functional domain within this region would be specific to Rex1's function in PSCs¹³².

A previous study demonstrated that in addition to the naïve and primed pluripotent states, PSCs exist in a variety of sub-populations within these two states, suggesting heterogeneity may be a fundamental feature of pluripotent stem cell populations⁹⁹. Since ERVs provide a platform for influencing gene expression by acting as promoters, enhancers, or promoting open chromatin states, the stage-specific propagation of ERVs may be an important source of regulatory elements for the generation and regulation of alternate pluripotent states⁹⁹. As Rex1 is a known regulator of ERVs and its expression is closely associated with pluripotency, it would be insightful to perform qRT-PCR analyses on cells cultured in LIF2i versus serum conditions to assess the effects of Rex1 loss on the expression of ERVs in naïve versus primed pluripotent states. This will, hopefully, provide insight into the stage-specific regulation of particular ERV classes by Rex1 and aid in a further understanding of the underlying of the mechanisms contributing to the maintenance of pluripotent states.

Through the generation of our C-terminal KO lines, we were able to elucidate how Rex1 is regulating its targets within PSCs. To further assess the independent functions of the smaller REX1 protein isoforms and to identify any variances in function as a result of structural differences of their N-termini, we have generated a series of human REX1 mutant constructs that we will over-express in double targeted KO lines. We will then subsequently assess the changes in Rex1 target gene expression through qRT-PCR analyses. Although we would ideally rescue our cell lines with murine Rex1 constructs, our human REX1 constructs were readily available. As previously mentioned, alignment of the human and mouse Rex1 protein sequences demonstrates considerable conservation in the amino acid sequence amongst the two proteins, particularly within the zinc finger region responsible for DNA-binding. Furthermore, two of the three human methionine residues, M1 and M123 are conserved with M6 and M124 within the murine Rex1 open reading frame, respectively (Figure 4). Overall this suggests conservation in structure and likely protein function amongst the human and mouse Rex1 proteins. Although they may have context-dependent roles due to the inherent differences between human and mouse ESCs, the human REX1 constructs should be able to rescue the observed phenotypes when introduced into Rex1 KO mESCs.

Finally, preliminary analyses regarding the underlying mechanisms contributing to the usage of alternative translation initiation sites within the Rex1 open reading frame suggests that RBPs may be contributing to the context-dependent regulation of Rex1 protein expression in PSCs. To investigate which RBPs are regulating the translation of Rex1 protein isoforms in an unbiased manner, we would perform a biotinylated RNA-

pulldown followed by mass spectrometry. Biotinylated RNA-pulldowns require the *in vitro* tagging of the mRNA with biotin through the use of biotinylated CTP¹³³. The biotin-tagged mRNA is then incubated with nuclear extracts allowing RBPs to bind regions of affinity within the mRNA^{133,134}. RNA-protein complexes are then purified through the use of streptavidin coated beads and RBPs identified through mass spectrometry^{133,134}. This technique can additionally be used to identify specific regions of the mRNA transcript bound by the RBPs through the use of smaller mRNA fragments¹³³. However, this approach is not without limitations. The observed binding of RBPs to the mRNA of interest may not be representative as *in vitro* methods require the formation of complexes between cell lysates and synthetic target RNAs and they do not allow for the identification of interactions that are formed in response stimuli within their environment¹³⁵.

Alternatively, a highly-specific and *in vivo* approach free of any cross-linking, RNA modification, and extensive purification strategies could be employed, but would require the modification of the existing CasID method¹³⁶. Previous studies identified the CRISPR-Cas effector C2c2 and demonstrated its ability to cleave target RNA sequences through the use of a sgRNA¹³⁷. Thus, an enzymatically dead C2c2 could be used to direct the mutant biotin ligase, BirA*, to the RNA of interest to biotinylate associated RBPs for identification by mass spectrometry as previously demonstrated with enzymatically dead Cas9 for the identification of chromatin associated proteins in CasID¹³⁶. However, although probable, such a mechanism of translational control has yet to be described¹³⁸. Thus, our study would be the first to report not only the identification of novel Rex1

isoforms, but would also be the first to describe a novel mechanism of translational control by RBPs in the pluripotent context.

CONCLUSION

Since their discovery, PSCs have become an indispensable tool for disease modeling and drug discovery. Although substantial progress has been made in elucidating the molecular mechanisms underlying the pluripotent state, a complete understanding of the regulation of self-renewal and differentiation of PSCs has yet to be attained. The experiments and results delineated in this thesis provide insight to the roles and mechanisms of Rex1, a factor closely associated with the pluripotent state. Through this work, we have confirmed the presence of numerous isoforms at the endogenous level in mESCs through the generation of endogenously FLAG tagged cell lines. We have also demonstrated that the long and short forms of Rex1 act as repressors of lineage-related genes and endogenous retroviral elements, likely aiding in maintenance of the pluripotent state and genomic integrity. Furthermore, we have generated tools that will allow for the evaluation of the independent roles of the long and short forms of the Rex1 protein, in addition to identifying a possible novel mechanism contributing to the context-dependent expression of the Rex1 isoforms in varying pluripotent states. Thus, the completion of experiments outlined previously, will aid in a further understanding of the mechanisms governing the regulation of the pluripotent state. Further characterization of such mechanisms will hopefully aid in improved techniques for the generation of iPSCs and further their use in the modeling of disorders and development of novel therapies in addition to furthering the use of PSCs, including both iPSCs and ESCs, in unforeseen forthcoming applications.

SUPPLEMENTAL MATERIAL

A)

	De-duplication		No de-duplication	
	mmCl13Flaginputn2	mmCl13Flagn2	mmCl13Flaginputn2	mmCl13Flagn2
QC-passed reads +QC-failed reads	3652334 + 0	3926417 + 0	29518386 + 0	33718108 + 0
Duplicates	0 + 0	0 + 0	0 + 0	0 + 0
Mapped	3652334 + 0 mapped (100.00%;-nan%)	3926417 + 0 mapped (100.00%;-nan%)	26591480 + 0 mapped (90.70%;-nan%)	30760456 + 0 mapped (91.23%;-nan%)
macs2 mfold and bandwidth	5,50; 300		5,50;300 3,30; 300	
#of peaks from macs2	14		163 179	

B)

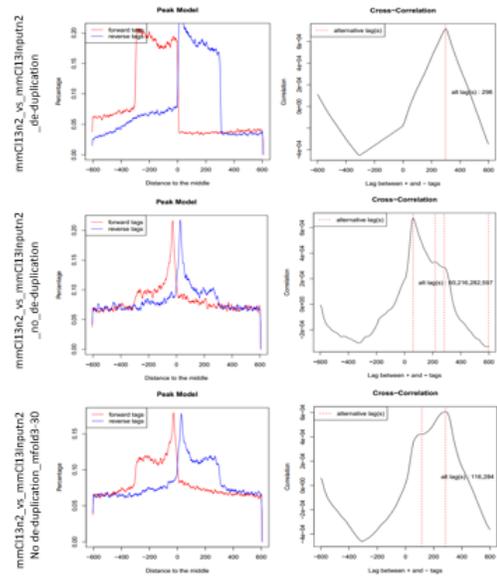


Figure S1: Optimization of ChIP-seq analysis parameters for the identification of peaks in murine Rex1 data set. Identification of peaks representative of Rex1 binding required optimization of de-duplication and mfold parameters. A) ChIP-seq parameters used in analysis B) Peak models and cross-correlation analysis for each analysis condition used. Maximal number of peaks with an acceptable peak model was obtained upon not performing optional de-duplication step and changing mfold from default of 5,50 to 3,30. Default bandwidth of 300 was used as this was the fragment size of our library preps used for sequencing.

REFERENCES

1. Rippon, H. J. & Bishop, A. E. Embryonic stem cells. *Cell Prolif.* **37**, 23–34 (2004).
2. De Los Angeles, A. *et al.* Hallmarks of pluripotency. *Nature* **525**, 469–478 (2015).
3. Li, M. & Belmonte, J. C. I. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* **18**, 180–191 (2017).
4. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **24**, 2239–2263 (2010).
5. Finch, B. W. & Ephrussi, B. Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture in vitro and their extinction upon hybridization with cells of permanent lines*. *Proc. Natl. Acad. Sci. U. S. A.* **57**, 615–621 (1967).
6. Kleinsmith, L. J. & Pierce, G. B. Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res.* **24**, 1544–1551 (1964).
7. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci.* **78**, 7634–7638 (1981).
8. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
9. Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687 (1988).
10. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).

11. Smith, A. G. Culture and differentiation of embryonic stem cells. *J. Tissue Cult. Methods* **13**, 89–94 (1991).
12. Czechanski, A. *et al.* Derivation and characterization of mouse embryonic stem cells (mESCs) from permissive and non-permissive strains. *Nat. Protoc.* **9**, 559–574 (2014).
13. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676
14. Wilson, K. D. & Wu, J. C. Induced Pluripotent Stem Cells. *JAMA* **313**, 1613–1614 (2015).
15. Singh, V. K., Kalsan, M., Kumar, N., Saini, A. & Chandra, R. Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. *Front. Cell Dev. Biol.* **3**, (2015).
16. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**, 949–953 (2008).
17. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322**, 945–949 (2008).
18. Okita, K. iPS cells for transplantation. *Curr. Opin. Organ Transplant.* **16**, 96–100 (2011).
19. Liang, G. & Zhang, Y. Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. *Cell Res.* **23**, 49–69 (2013).
20. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* **28**, 848–855 (2010).

21. Robinton, D. A. & Daley, G. Q. The promise of induced pluripotent stem cells in research and therapy. *Nature* **481**, 295–305 (2012).
22. Martí, M. *et al.* Characterization of pluripotent stem cells. *Nat. Protoc.* **8**, 223–253 (2013).
23. Itskovitz-Eldor, J. *et al.* Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Mol. Med.* **6**, 88–95 (2000).
24. Nelakanti, R. V., Kooreman, N. G. & Wu, J. C. Teratoma Formation: A Tool for Monitoring Pluripotency in Stem Cell Research. *Curr. Protoc. Stem Cell Biol.* **32**, 4A.8.1-4A.8.17 (2015).
25. Wesselschmidt, R. L. The teratoma assay: an in vivo assessment of pluripotency. *Methods Mol. Biol. Clifton NJ* **767**, 231–241 (2011).
26. Daley, G. Q. *et al.* Broader Implications of Defining Standards for the Pluripotency of iPSCs. *Cell Stem Cell* **4**, 200–201 (2009).
27. 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).
28. O'Connor, M. D., Kardel, M. D. & Eaves, C. J. Functional Assays for Human Embryonic Stem Cell Pluripotency. *SpringerLink* 67–80 (2011). doi:10.1007/978-1-60761-962-8_4
29. Wen, D., Saiz, N., Rosenwaks, Z., Hadjantonakis, A.-K. & Rafii, S. Completely ES Cell-Derived Mice Produced by Tetraploid Complementation Using Inner Cell Mass (ICM) Deficient Blastocysts. *PLoS ONE* **9**, (2014).

30. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* **38**, 1027–1032 (2010).
31. Dynamic stem cell states: naive to primed pluripotency in rodents and humans : Nature Reviews Molecular Cell Biology: Nature Research. Available at: <http://www.nature.com/nrm/journal/v17/n3/full/nrm.2015.28.html>. (Accessed: 5th August 2017)
32. Kumari, D. States of Pluripotency: Naïve and Primed Pluripotent Stem Cells. (2016). doi:10.5772/63202
33. Loh, K. M., Lim, B. & Ang, L. T. Ex uno plures: molecular designs for embryonic pluripotency. *Physiol. Rev.* **95**, 245–295 (2015).
34. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
35. 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).
36. Hackett, J. A. & Surani, M. A. Regulatory Principles of Pluripotency: From the Ground State Up. *Cell Stem Cell* **15**, 416–430 (2014).
37. Boyer, L. A., Mathur, D. & Jaenisch, R. Molecular control of pluripotency. *Differ. Gene Regul.* **16**, 455–462 (2006).
38. 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

39. Orkin, S. H. *et al.* The Transcriptional Network Controlling Pluripotency in ES Cells. *Cold Spring Harb. Symp. Quant. Biol.* sqb.2008.72.001 (2008). doi:10.1101/sqb.2008.72.001
40. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).
41. 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).
42. Ng, H.-H. & Surani, M. A. The transcriptional and signalling networks of pluripotency. *Nat. Cell Biol.* **13**, 490–496 (2011).
43. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, (2006).
44. 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 Dayt. Ohio* **27**, 1529–1537 (2009).
45. Ruan, Y. *et al.* Nac1 promotes self-renewal of embryonic stem cells through direct transcriptional regulation of c-Myc. *Oncotarget* **8**, 47607–47618 (2017).
46. 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).
47. 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).

48. 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).
49. 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).
50. Gordon, S., Akopyan, G., Garban, H. & Bonavida, B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **25**, 1125–1142 (2006).
51. 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).
52. Park, K. & Atchison, M. L. Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta), that binds to the immunoglobulin kappa 3' enhancer and the immunoglobulin heavy-chain mu E1 site. *Proc. Natl. Acad. Sci.* **88**, 9804–9808 (1991).
53. Morey, L. & Helin, K. Polycomb group protein-mediated repression of transcription. *Trends Biochem. Sci.* **35**, 323–332 (2010).
54. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
55. O'Carroll, D. *et al.* The Polycomb-Group Gene *Ezh2* Is Required for Early Mouse Development. *Mol. Cell. Biol.* **21**, 4330–4336 (2001).
56. 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).

57. 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).
58. Ballabeni, A. *et al.* Cell cycle adaptations of embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19252–19257 (2011).
59. 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).
60. 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).
61. Bushmeyer, S., Park, K. & Atchison, M. L. Characterization of Functional Domains within the Multifunctional Transcription Factor, YY1. *J. Biol. Chem.* **270**, 30213–30220 (1995).
62. 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).
63. 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).
64. Hosler, B. A., LaRosa, G. J., Grippo, J. F. & Gudas, L. J. Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. *Mol. Cell. Biol.* **9**, 5623–5629 (1989).

65. Rogers, M. B., Hosler, B. A. & Gudas, L. J. Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. *Development* **113**, 815–824 (1991).
66. Masui, S. *et al.* Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev. Biol.* **8**, 45 (2008).
67. Navarro, P. *et al.* Molecular coupling of Tsix regulation and pluripotency. *Nature* **468**, 457–460 (2010).
68. Kim, J. D. *et al.* Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. *Hum. Mol. Genet.* **20**, 1353–1362 (2011).
69. Guallar, D. *et al.* Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. *Nucleic Acids Res.* **40**, 8993–9007 (2012).
70. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
71. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695-699 (2010).
72. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130540 (2014).
73. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* **4**, 487–492
74. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909 (2008).

75. Tanaka, T. S. Transcriptional heterogeneity in mouse embryonic stem cells. *Reprod. Fertil. Dev.* **21**, 67–75 (2009).
76. Shi, W. *et al.* Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. *J Biol Chem* **281**, (2006).
77. Minkovsky, A., Patel, S. & Plath, K. Pluripotency and the transcriptional inactivation of the female mammalian X chromosome. *Stem Cells Dayt. Ohio* **30**, 48–54 (2012).
78. Payer, B., Lee, J. T. & Namekawa, S. H. X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells. *Hum. Genet.* **130**, 265–280 (2011).
79. Ohhata, T. & Wutz, A. Reactivation of the inactive X chromosome in development and reprogramming. *Cell. Mol. Life Sci.* **70**, 2443–2461 (2013).
80. Morey, R. & Laurent, L. C. Getting Off the Ground State: X Chromosome Inactivation Knocks Down Barriers to Differentiation. *Cell Stem Cell* **14**, 131–132 (2014).
81. Schulz, E. G. *et al.* The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14**, 203–216 (2014).
82. Barakat, T. S. & Gribnau, J. X chromosome inactivation and embryonic stem cells. *Adv. Exp. Med. Biol.* **695**, 132–154 (2010).
83. Gontan, C. *et al.* RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**, 386–390 (2012).
84. Bartolomei, M. S. & Ferguson-Smith, A. C. Mammalian Genomic Imprinting. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

85. Plasschaert, R. N. & Bartolomei, M. S. Genomic imprinting in development, growth, behavior and stem cells. *Development* **141**, 1805–1813 (2014).
86. Latos, P. A. *et al.* An in vitro ES cell imprinting model shows that imprinted expression of the *Igf2r* gene arises from an allele-specific expression bias. *Dev. Camb. Engl.* **136**, 437–448 (2009).
87. 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).
88. Thompson, J. R. & Gudas, L. J. Retinoic acid induces parietal endoderm but not primitive endoderm and visceral endoderm differentiation in F9 teratocarcinoma stem cells with a targeted deletion of the *Rex-1* (*Zfp-42*) gene. *Mol Cell Endocrinol* **195**, (2002).
89. Familiarì, M. Characteristics of the endoderm: embryonic and extraembryonic in mouse. *ScientificWorldJournal* **6**, 1815–1827 (2006).
90. Bielinska, M., Narita, N. & Wilson, D. B. Distinct roles for visceral endoderm during embryonic mouse development. *Int. J. Dev. Biol.* **43**, 183–205 (2002).
91. Rossant, J. Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. *Semin. Cell Dev. Biol.* **15**, 573–581 (2004).
92. Duncan, S. A., Nagy, A. & Chan, W. Murine gastrulation requires HNF-4 regulated gene expression in the visceral endoderm: tetraploid rescue of *Hnf-4*(*-/-*) embryos. *Development* **124**, 279–287 (1997).
93. Schlesinger, S. & Goff, S. P. Retroviral Transcriptional Regulation and Embryonic Stem Cells: War and Peace. *Mol. Cell. Biol.* **35**, 770–777 (2015).

94. Gifford, R. & Tristem, M. The Evolution, Distribution and Diversity of Endogenous Retroviruses. *Virus Genes* **26**, 291–315 (2003).
95. 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).
96. Stocking, C. & Kozak, C. A. Murine endogenous retroviruses. *Cell. Mol. Life Sci. CMLS* **65**, 3383–3398 (2008).
97. Robbez-Masson, L. & Rowe, H. M. Retrotransposons shape species-specific embryonic stem cell gene expression. *Retrovirology* **12**, 45 (2015).
98. Gifford, W. D., Pfaff, S. L. & Macfarlan, T. S. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* **23**, 218–226 (2013).
99. Hackett, J. A., Kobayashi, T., Dietmann, S. & Surani, M. A. Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. *Stem Cell Rep.* **8**, 1645–1658 (2017).
100. Peaston, A. E. *et al.* Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Dev. Cell* **7**, 597–606 (2004).
101. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634 (2010).
102. Fort, A. *et al.* Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* **46**, 558–566 (2014).

103. Reik, W., Dean, W. & Walter, J. Epigenetic Reprogramming in Mammalian Development. *Science* **293**, 1089–1093 (2001).
104. Macfarlan, T. S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
105. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
106. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931 (2010).
107. Yang, B. X. *et al.* Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* **163**, 230–245 (2015).
108. Scotland, K. B., Chen, S., Sylvester, R. & Gudas, L. J. Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. *Dev. Dyn.* **238**, 1863–1877 (2009).
109. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* **8**, 2281–2308 (2013).
110. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
111. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
112. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
113. Kurien, B. T. & Scofield, R. H. Introduction to Protein Blotting. in *Protein Blotting and Detection* 9–22 (Humana Press, Totowa, NJ, 2009).

114. Doble, B. W., Patel, S., Wood, G. A., Kockeritz, L. K. & Woodgett, J. R. Functional redundancy of GSK-3alpha and GSK-3beta in Wnt/beta-catenin signaling shown by using an allelic series of embryonic stem cell lines. *Dev. Cell* **12**, 957–971 (2007).
115. ten Berge, D. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.* **13**, 1070–1075 (2011).
116. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotech* **32**, 347–355 (2014).
117. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinforma. Oxf. Engl.* **31**, 2382–2383 (2015).
118. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
119. Mi, H. *et al.* PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
120. Liu, T. *et al.* Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83 (2011).
121. Austen, M., Lüscher, B. & Lüscher-Firzlaff, J. M. Characterization of the Transcriptional Regulator YY1 THE BIPARTITE TRANSACTIVATION DOMAIN IS INDEPENDENT OF INTERACTION WITH THE TATA BOX-BINDING PROTEIN, TRANSCRIPTION FACTOR IIB, TAFII55, OR cAMP-RESPONSIVE

- ELEMENT-BINDING PROTEIN (CBP)-BINDING PROTEIN. *J. Biol. Chem.* **272**, 1709–1717 (1997).
122. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
123. Bailey, T. *et al.* Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data. *PLoS Comput. Biol.* **9**, (2013).
124. Santoni, F. A., Guerra, J. & Luban, J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9**, 111 (2012).
125. Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21**, 423–425 (2014).
126. Ramírez, M. A. *et al.* Transcriptional and post-transcriptional regulation of retrotransposons IAP and MuERV-L affect pluripotency of mice ES cells. *Reprod. Biol. Endocrinol.* **4**, 55 (2006).
127. Göke, J. *et al.* Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. *Cell Stem Cell* **16**, 135–141 (2015).
128. Schnerch, A., Cerdan, C. & Bhatia, M. Distinguishing Between Mouse and Human Pluripotent Stem Cell Regulation: The Best Laid Plans of Mice and Men. *STEM CELLS* **28**, 419–430 (2010).
129. Abaza, I. & Gebauer, F. Trading translation with RNA-binding proteins. *RNA* **14**, 404–409 (2008).

130. Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* **582**, 1977–1986 (2008).
131. Paz, I., Kosti, I., Ares, M., Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **42**, W361–W367 (2014).
132. Mongan, N. P., Martin, K. M. & Gudas, L. J. The putative human stem cell marker, Rex-1 (Zfp42): Structural classification and expression in normal human epithelial and carcinoma cell cultures. *Mol. Carcinog.* **45**, 887–900 (2006).
133. Panda, A. C., Martindale, J. L. & Gorospe, M. Affinity Pulldown of Biotinylated RNA for Detection of Protein-RNA Complexes. *Bio-Protoc.* **6**, (2016).
134. Marín-Béjar, O. & Huarte, M. RNA pulldown protocol for in vitro detection and identification of RNA-associated proteins. *Methods Mol. Biol. Clifton NJ* **1206**, 87–95 (2015).
135. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. & Krzyzosiak, W. J. Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. *Nucleic Acids Res.* **44**, 9050–9070 (2016).
136. Schmidtman, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. Determination of local chromatin composition by CasID. *Nucl. Austin Tex* **7**, 476–484 (2016).
137. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).

138. Re, A., Waldron, L. & Quattrone, A. Control of Gene Expression by RNA Binding Protein Action on Alternative Translation Initiation Sites. *PLoS Comput. Biol.* **12**, (2016).