TRANSCRIPTIONAL REGULATION OF PLURIPOTENCY AND CELL FATE

TRANSCRIPTIONAL REGULATION OF PLURIPOTENCY AND CELL FATE

BY
SONAM BHATIA, B.Sc. (Honours)

A Thesis Submitted to the School of Graduate Studies In Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

McMaster University

Descriptive Note

McMaster University Doctor of Philosophy (2017) Hamilton, Ontario

TITLE: Transcriptional regulation of pluripotency and cell fate

AUTHOR: Sonam Bhatia, B.Sc. (Honours) (University of Toronto)

SUPERVISOR: Dr. Jonathan S. Draper

NUMBER OF PAGES: xvi, 218

ABSTRACT

Embryonic stem cells are pluripotent in nature, in that they can self-renew indefinitely, while maintaining the capability to give rise to all adult cell types. This characteristic makes them an attractive avenue for various therapeutic purposes. Therefore, many studies have been devoted to understanding the fundamental nature of these cells and the processes that govern and maintain their pluripotent cell fate. We hypothesized that *cell fate is intrinsically regulated by the underlying chromatin and transcriptional machinery of the cells*. To test this, we first studied the heterogeneity that exists within pluripotent cell cultures. We showed that pluripotent sub-populations demarcated by expression of REX1, a pluripotency transcription factor, have distinct differentiation propensities. Additionally, we found that chromatin modification via DNA methylation was the underlying cause of this heterogeneity, providing evidence for the major roles that chromatin and transcription play in regulating cell fate.

We next studied a fate maintaining mechanism, mitotic bookmarking, as a method of cell fate preservation. During cell division, chromatin structure undergoes significant remodeling, various proteins uncouple from the DNA, and there is a temporary hiatus in transcription. Despite this restructuring the transcriptional memory of the parent cell is faithfully transmitted to daughter cells. We hypothesized that, in ES cells, several chromatin bound factors would be retained on the mitotic chromatin and would act as bookmarks preserving the underlying pluripotent chromatin structure. Using a global unbiased approach, we found that a large number of chromatin regulators are indeed bound to the mitotic chromatin. Additionally, a chromatin accessibility assay revealed that a large number of accessible promoter sites are preserved during mitosis and into G1

of daughter cells. The mitotic chromatin bound factors likely play a role in the maintenance of these protected DNA sites. Our data suggest that, preservation of these sites by various chromatin regulators during mitosis underlies the faithful transmission of cell identity from parent to daughter cells, ergo maintaining the cell's fate.

Acknowledgements

Who you are today is a sum of everything that has happened until that moment, and everyone you've met has had a role to play in it. The past few years of graduate school have been momentous in a variety of ways and couldn't have been the same without all the amazing people I had the chance to work with.

First, and foremost I would like to thank my supervisor Dr. Jon Draper for giving me the opportunity to pursue my graduate studies in his lab. It's been a journey! I started in your lab with the aim to enhance my knowledge and understanding and I can honestly say I have achieved that. The amount that I have learnt in the past few years was beyond what I had in mind when I started. Of course, there are things that I wish had gone differently and that I wish I'd done better (like my publication record) but overall it's been a productive journey! You gave me the freedom to conduct research with independent thought, which despite being a bit of a struggle at times, made for a great graduate education. So, thank you for that! As part of the team I also had the opportunity to mentor graduate and undergraduate students, and a special thanks to them, especially Amanda and Daisy, for helping out with the project.

I'd also like to thank my committee members Drs. Brad Doble, and Andre Bedard who have been very supportive and encouraging all these years; and have been there to discuss the project whenever need be.

A special 'Arigatou' to my Japanese lab team Dr. Aki Minoda, Jen, Ye, Maddie, Chung, and Jay. Aki, you have been a great mentor, and I'm very glad we had lunch in Whistler otherwise I would have missed out on an amazing opportunity to visit Japan, and get to know and learn from you all. Chung, thanks for being so patient with my early days of bioinformatics struggle! Jay, a special thanks to your intro to R classes that have changed the way I look at data.

I'd also like to acknowledge the many people (past and present) here at SCC-RI that I have had the chance to work with (there are too many of you to name ;)), I'm glad our paths crossed. A special thanks to my crew of McMaster friends, Akshita, Jen, Ryan, Aji, who I met during the first few weeks of grad school, and who have been there since, for dinners, board-games, writing sessions and the general wows and woes of grad school!

Lastly, I'd like to thank my amazing family who have put up with me and my absence all these years. Mum and Dad, you are the most amazing parents one could ask for. Your support, encouragement and patience has meant the world to me. Every time I had doubts if I could do this, you'd unknowingly say something and I knew I could. So, thank you for always being in our corner! A special thanks to my awesome siblings Akash, Payal and Ranjit (yes! You are one of them ;)), you guys are the best! I know I have been a part-time sibling lately, but thanks for being so understanding, supportive and for always being there. I promise I'd make up for lost times, I know how much you miss me bothering you ;)!

Table of Contents

List of figures and tables

Section 1

Chapter 1.1

Chapter 1.2

Supplementary Material

List of Abbreviations

3D	3-Dimensional
7AAD	7 aminoactinomycin D
ATAC-seq	Assay for Transposase-Accessible Chromatin by sequencing
bFGF	Basic Fibroblast Growth Factor
BME	Betamercaptoethanol
BMPs	Bone Morphogenic Proteins
Br-UTP	Bromouridine 5'-triphosphate
BSA	Bovine Serum Albumin
Cas9	CRISPR associated protein 9
Cdk	Cyclin Dependent Kinase
ChIP	Chromatin Immunoprecipitation
ChIP-MS	Chromatin Immunoprecipitation followed by Mass Spectrometry
Chr	Chromosome
CIC	Colony Initiating Cell
CpG	5'—C—phosphate—G—3'
Cpm	Counts per million
CRISPR	Clustered regularly interspaced short palindromic repeats
EDTA	Ethylenediaminetetraacetic acid
EpiSCs	Epiblast Stem Cells
ES	Embryonic Stem
ESC	Embryonic Stem Cell
F2A	FMDV 2A peptide
FACS	Fluorescent activated cell sorting
FC	fold change
FDR	False Discovery Rate
G1	Gap-1
G1t20	G1 populations collected after 20 minutes of release into G1 from mitosis
G1t35	G1 populations collected after 35 minutes of release into G1 from mitosis
G2	Gap-2
G2M	G2 and Mitotic mixed populations
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GO	Gene Ontology
GSK3	Glycogen synthase kinase-3
H2B	Histone2B
H3	Histone 3
H3K27Ac	Acetylation of histone H3 at lysine 27
H3K27me3	Trimethylation of histone H3 at lysine 27
H3K4me3	Trimethylation of histone H3 at lysine 4
H3K5Ac	Acetylation of histone H3 at lysine 27
H3K9me3	Trimethylation of histone H3 at lysine 9
H3S10P	Phosphorylation of serine 10 of histone 3
H3T3P	Phosphorylation of threonine 3 of histone 3
hESCs	human Embryonic Stem Cells

HR	Homologous recombination
ICM	Inner Cell Mass
IF	Immunofluorescence
IgG	Immunoglobulin
IP	Immunoprecipitation
IPpA	IRES-Puromycin polyA
IPS	Induced Pluripotent Stem
LADs	Lamin associated domains
LIF	Leukemia Inhibitory Factor
M	Mitosis
mAb	monoclonal antibody
Macs2	Model based analysis of ChIP-seq
MBF	Mitotic Bookmarking Factor
MD	Mitotic Degron
MEFs	Mouse embryonic fibroblasts
MEK	MAPK/ERK Kinase
mESCs	mouse Embryonic Stem Cells
mKO2	monomeric Kusabira orange 2
Mm9/Mm10	Mouse genome assemblies 9/10
MNase	Micrococcal Nuclease
MS	Mass Spectrometry
Mso	Mitotic shake off
NAD	Nicotinamide adenine dinucleotide
NORs	Nucleolus organizing regions
ORF	Open reading frame
pAb	polyclonal antibody
Parp1	poly(ADP-ribose)polymerase 1
Parp1KO	Parp1 knock out
PBS	Phosphate buffer saline
PcG	Polycomb Group
PCR	Polymerase chain reaction
PECAM1	platelet endothelial cell adhesion molecule 1
PePr	Peak-calling and Prioritization pipeline
PFA	Paraformaldehyde
Postmso	Post mitotic shake off
Pre-RC	Pre-Replicative Complexes
PSCs	Pluripotent Stem Cells
qRT-PCR	Quantitative Reverse transcription polymerase chain reaction
RA	Retinoic Acid
REX1	Reduced Expression 1
RNA pol	RNA polymerase
rRNA/DNA	Ribosomal RNA/DNA
S	Synthesis
S10	Phosphorylation of serine 10 of histone 3
SSEA1	stage-specific embryonic antigen-1
TADs	Topologically associated domains

TCAG	The Centre for Applied Genomics
TF	Transcription factor
TRA	TRA-1-60
TSS	Transcription Start Site
TV	Targeting vector
VEN	Venus
VF2Pu	Venus-F2A-Puromycin-polyA
xMEFs	x-ray irradiated Mouse embryonic fibroblasts
Zfp42	Zinc finger protein 42

Preface

My dissertation is focused on stem cell fate regulation and maintenance in pluripotent cells. The thesis is divided into two related sections focusing on cell fate regulation and maintenance.

Section I is a study of the heterogeneity that exists within pluripotent cell cultures, resulting from cell fate decisions. Pluripotent cells can make a variety of decisions based on culture conditions, environmental factors, intra-cellular communication, and even spatial localization of the cells. As a result, a pluripotent culture is rarely homogeneous, and is a composite of cells with varying molecular phenotypes. The study focuses on how the expression of transcription factor REX1 affects cell fate in human pluripotent cells, and assays the phenotypic outcomes associated with it.

Section II is focused on the mechanisms of fate maintenance in pluripotent cells, specifically investigating and establishing the role of mitotic bookmarking in pluripotency. The basis of this study was an interesting observation made with a fluorescent REX1 fusion protein, which seemed to associate with mitotic chromatin. Mitotic chromatin association of transcription regulators in pluripotency was largely unexplored when we commenced the project. This piqued our interest, and provided us with an opportunity to fill some gaps that existed in the field of study.

SECTION 1: Stable subpopulations constitute the pluripotent compartment in human embryonic stem cell cultures

SECTION 1: Stable subpopulations constitute the pluripotent compartment in human embryonic stem cell cultures

Preface

This section focuses on studying heterogeneity within human pluripotent cultures, based on differences in transcription factor expression. This project was already underway when I joined Dr. Jon Draper's lab, and I had the chance to contribute significantly towards it. The section starts off (Chapter 1.1) with an introduction to pluripotency, the known states of pluripotency, and the current knowledge of heterogeneity within pluripotent cultures. Chapter 1.2 is a published research study that was a result of this project. Chapter 1.3 is a brief conclusion and provides an update on the current research status of the project. This chapter also briefly outlines the findings that led to the design of research study that is the focus of Section 2.

Chapter 1.1: Introduction

1.1.1 Introduction to pluripotency: Embryonic stem cells (ESCs) are characterized by their ability to self-renew indefinitely and are pluripotent in nature, in that, they can differentiate into cells of all adult lineages: ectoderm, mesoderm and endoderm. ESCs, are tissue culture artifacts that are considered to be *in vitro* counterparts of the inner cell mass (ICM) of a blastocyst stage embryo, from which they are derived[1]. The self-renewal and differentiation capabilities of ESCs make them a powerful tool for various clinical applications, including cell transplants, *in vitro* organ/tissue generation, and tissue specific drug discovery. They also present a unique opportunity to study mechanisms governing early development. Due to their wide array of potential applications, an in-depth understanding of mechanisms that maintain ESC pluripotency is crucial. Therefore, ESCs have been a hot spot for study since their initial derivation both in mouse[2] and humans[1]. Pluripotent stem cells (PSCs), are characterized by the expression of key transcription factors that include OCT4, NANOG, KLF4, SOX2 and REX1 and cell surface antigens such as TRA-1-60 and SSEA3 (in humans) or SSEA1 (in mouse) (reviewed in Hoffman and Carpenter, 2005[3]). Due to various ethical concerns surrounding the use of embryonic stem cells, Yamanaka and colleagues showed that adult somatic cells could be re-programmed to embryonic-like stem cells by re-expressing key pluripotency transcription factors, Oct4, Sox2, Klf4 and c-Myc to produce induced pluripotent stem (IPS) cells[4,5]. These observations highlight the importance of core pluripotency-associated transcription factors in maintaining a stem like state. OCT4, SOX2 and NANOG form key components of the pluripotency transcription factor

network in PSCs[6] and further interact with transcription factors such as, Klf4, c-myc and Rex1[7,8] that could, therefore, play important roles in maintaining pluripotency.

Recently, another type of murine pluripotent stem cell was derived; Epiblast Stem cells (EpiSCs)[9]. As opposed to pre-implantation blastocyst derived ESCs, EpiSCs are derived from a post-implantation stage embryo. Although still pluripotent in nature, EpiSCs differ from ICM-derived mESCs in their transcriptional profile and phenotypic characteristics. EpiSCs, like ESCs, express Oct4 and Nanog, but lack the expression of pluripotency markers such as Rex1, Klf4 and Fgf4, that are associated with the earlier state of pluripotency observed in the preimplantation ICM[9]. Like ESCs, EpiSCs can differentiate into all three lineages and can contribute to chimeras, but at a much lower frequency[10]. EpiSCs also express epiblast differentiation markers, such as Fgf5, and are therefore considered to represent a more advanced state of development than ESCs[9]. Importantly, in order to proliferate *in vitro* these two mouse pluripotent cell types require distinct, almost antagonistic, growth conditions: mESCs maintain pluripotency in the presence of LIF (Leukemia inhibitory factor) and BMPs (Bone Morphogenic Proteins); EpiSCs maintain pluripotency in bFGF- (basic fibroblast growth factor) and Activin- supplemented media[9]. Similar to mESCs,

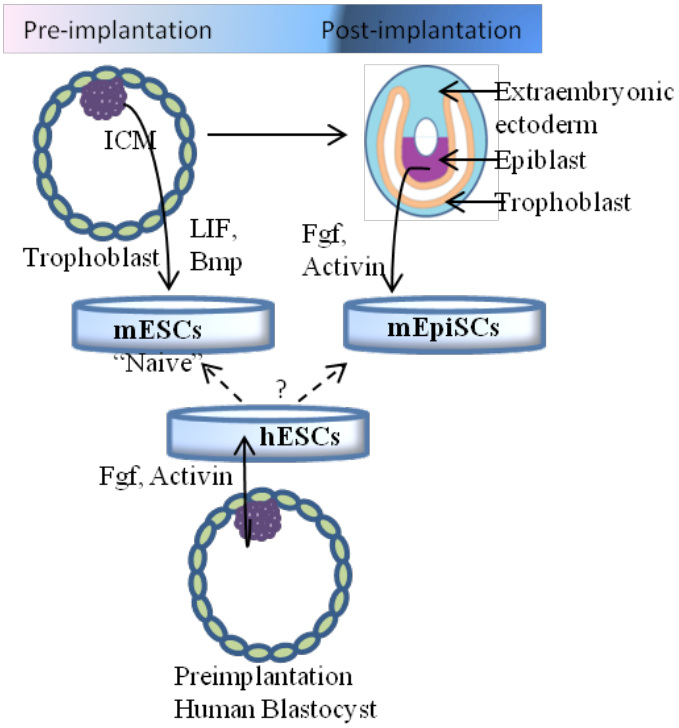


Figure 1: An overview of derivation and culturing of mESCs, EpiSCs and hESCs

human ESCs (hESCs) are derived from the pre-implantation blastocyst[1], and they express mESC-like pluripotency genes such as REX1, KLF4 and FGF4. In contrast to mESCs, hESCs differentiate in the presence of LIF and BMP4, and require EpiSC-like growth conditions (bFGF and Activin) to maintain pluripotency *in vitro*[1]. Several attempts have been made to characterize hESCs on the pluripotency axis as defined by the mESCs and EpiSCs (Summarised in Figure 1)

1.1.2 Heterogeneity in murine ESCs: Pluripotency-associated makers such as SSEA1, Nanog, Rex1 and Stella are known to divide the mouse pluripotent compartment into at least two sub-types of pluripotent stem cells[12-14]. Analyses of these populations have shown that mESCs are a heterogeneous mix of naïve and primed pluripotent cells, which exist in a dynamic state of equilibrium with each other. It is now understood that pluripotency is not a single state, but represents a hierarchy, with cells at the apex representing the ground or "naïve" state of pluripotency. The phenotypic heterogeneity in stem cell cultures appears to arise, at least in part, from the heterogeneous expression of pivotal transcription factors and is often associated with varying cell fates. Phenotypic heterogeneity has real-world consequences. In ESCs, it is a limitation for efficiently differentiating stem cells into cell populations of desired lineages for clinical applications[11].

Studying transcription factor heterogeneity has provided much needed insights into the expression kinetics of these factors in maintaining pluripotency *in vitro,* and has shed light into their *in vivo* functions. Furusawa et al in 2004, first described heterogeneity in mESCs by isolating different mESC populations based on the expression of the stem cell specific surface markers PECAM1 (platelet endothelial cell adhesion

molecule 1) and SSEA1 (stage-specific embryonic antigen-1)[15]. PECAM1+ cells marked the pluripotent cells, while SSEA1 sub-divided this pluripotent compartment into PECAM1+SSEA1- and PECAM1+SSEA1+. These subtypes were still pluripotent, as marked by their ability to contribute to chimeras, but PECAM+SSEA1- cells when transplanted in mouse embryos disappeared in post-implantation stages suggesting that these cells have lost the competence for later organ development, and therefore represent an altered cell fate compared to pluripotent SSEA+ cells[15,16].

Dissecting heterogeneity in ESCs, is limited by the handful of stem cell specific surface antigens yet discovered. Additionally, immunostaining offers a static image of populations and hence limits studying the dynamic interactions between these sub-populations. To overcome these limitations, several groups have undertaken gene knock-in transcriptional reporter based approaches. Nanog is considered a core element important for pluripotency *in vivo* and *in vitro,* as its deletion causes embryonic lethality while its over-expression can maintain pluripotency in a cytokine independent manner *in vitro*[12]. Despite its important role in pluripotency, Nanog expression in mESCs was found to be highly unequal within pluripotent cells when compared to Oct4 expression. Two independent studies in 2007, using Nanog-GFP-[17]and Nanog-β-geo-[18] based transcriptional reporter systems discovered a mosaic pattern of Nanog expression in mESC cultures. Upon in-depth analysis, it was shown that Nanog low cells were enriched for early endoderm markers, priming them for a primitive endoderm fate, while Nanog high cells showed a greater expression of pluripotency genes[18]. Similar to Nanog expression, the expression of StellaGFP reporter also showed an uneven distribution in mESCs. The Stella gene is a germ cell marker, but its expression is also observed in pre-

implantation blastocysts and, therefore, is considered an embryonic stem cell marker[19]. StellaGFP+ cells showed greater expression of pluripotency genes such as Pecam1 and Rex1, while StellaGFP- cells were enriched for differentiation gene transcripts such as Fgf5 and Gbx2[13].

Additionally, Rex1 (Reduced Expression 1), an early ICM marker, is also expressed heterogeneously in mESCs. Using a Rex1GFP reporter, along with an Oct4CFP reporter, it was shown that Oct4+ cells can be segregated into functionally distinct Rex1-/Oct4+ and Rex1+/Oct4+ sub-populations[14]. In addition to similar gene expression differences observed with NanogGFP and StellaGFP positive/negative cells, Rex1-/Oct4+ cells were considerably less efficient in generating mouse chimeras as opposed to Rex1+/Oct4+ cells. The Rex1-/Oct4+ population also differentiated efficiently into early ectoderm more easily compared to Rex+Oct4+ cells and was, therefore, characterized with a primitive ectoderm cell fate as opposed to the more ICM-like Rex1+/Oct4+ population[14].

These studies highlight that at least two types of pluripotency states exist within the murine embryonic stem cell compartment. What has not yet been identified is the overlapping relationship between these sub-categories marked by the loss of expression of individual markers. The loss of expression of Nanog, Stella and Rex1 in pluripotent cells (as marked by Oct4 expression) correlates with cells primed for differentiation, and they represent a pre-committed state slightly more developmentally advanced[12] compared the ground state of ESCs (Figure 2).

Naïve and Primed Pluripotent States: Heterogeneity in ESCs has led to the concept of naïve (or ground) and primed states of pluripotency. “Stemness” is thought to

be a default state of ESCs, in which they exist and proliferate without any external stimuli for differentiation[20,21]. Indeed, when grown in growth conditions supplemented by small molecule inhibitors of MEK (MAPK/ERK Kinase) and GSK3 (glycogen synthase kinase-3), the two pathways that promote differentiation in mESCs, it was shown that mESCs enter a stable ground state, which is characterized by a relatively homogeneous gene expression[21]. In the ground state, heterogeneity in Nanog and Rex1 expression was substantially decreased[21], suggesting that heterogeneity is a function of the culture conditions ESCs are grown in. Similarly, when Nanog is over-expressed in ESCs, the resulting culture is more homogenous in nature, with a decrease in expression of early differentiation markers, maintaining cells in the more stem-like state[18] .The naïve state, therefore, is a stable state represented by homogenous cultures, and maintained by the core transcriptional circuitry, which is very sensitive to the exogenous factors it is exposed to[21]. Mouse ESCs in the ground state are at the apex of pluripotency, and homogeneously express Rex1 and Nanog. Loss of expression of these markers in mESCs coincides with descent though this hierarchy, and priming for differentiation.

1.1.4 Heterogeneity in EpiSCs and hESCs: Human embryonic stem cells (hESCs) and EpiSCs share similar culture conditions (Figure 1), but also share other phenotypic traits, including a flat pancake like morphology and lower colony forming capabilities when seeded as single cells, when contrasted to mESCs[1,2,9]. The concept of heterogeneity is understudied in hESCs and EpiSCs, however, recently, using an Oct4-GFP reporter, Schöler and colleagues showed that EpiSCs are a heterogeneous mix of Oct4+ and Oct4- cells, with the former occupying a very small percentage of the total cell population[10]. Using gene expression analysis with two different counterpart *in vivo*

epiblast stages (E5.5 and E6.5) it was shown that Oct4+ and Oct4- cells were developmentally distinct with the former being similar to the early (E5.5) epiblast. Oct4+ cells, as opposed to Oct4- cells and similar to mESCs, were capable of contributing to chimeras[10]. Oct4+ EpiSCs were therefore higher up in the pluripotency hierarchy.

1.1.5 Epigenetic fluidity and stability of stem cell subpopulations: A common denominator in the pluripotent subpopulations of all murine models is their metastablity, which defines the existence of these populations in a dynamic equilibrium. While Nanog+, Rex1+ and Stella+ populations spontaneously converted to Nanog-, Rex1- and Stella- populations, respectively, in culture, the reverse was also true[13,14,18]. Even though the loss of these transcription factors is speculated to be marked by less stringent epigenetic modifications and is, therefore, reversible, the process has not been studied in detail. However, it was shown that the loss of expression of Stella in mESCs was accompanied by histone modifications[13]. Since the underlying culture environment is a common factor, it has been speculated that the mESC culture environment, LIF and BMPs, promotes epigenetic stability and resists major changes[22].

Supporting this hypothesis, Bao and colleagues showed that EpiSCs could spontaneously revert to an ESC like state when moved from their standard bFgf+Activin containing media into LIF+BMP containing media[22], despite the EpiSC displaying divergent DNA methylation mechanisms compared to mESCs. Together, these observations imply that if culture conditions were the only underlying factor governing stability and metastability of the pluripotent subpopulations, then Oct4+/- EpiSCs would also represent stable populations. However, it was shown that the Oct4+/- EpiSCs were

metastable in nature[10], perhaps suggesting that interchangeable dynamics between closely related pluripotent sub-populations is a common factor in early developmental cells.

1.1.6 Research Hypothesis

Heterogeneity in transcription factor expression in hESCs, although widely speculated and expected, is barely studied. To address the phenomena of heterogeneity and metastability in the context of human pluripotency, we assayed the expression of the REX1 gene locus in hESCs using a REX1Venus reporter system in conjunction with staining for the pluripotency marker TRA-1-60. We *hypothesized that like mESCs, the pluripotent compartment of hESCs is divided by the expression of REX1, and that REX1+ and REX1- pluripotent sub-populations are functionally different, with the latter being more primed for differentiation*. We tested the hypothesis by characterizing REX1Venus expression in hESCs, and used fluorescence activated cell sorting (FACS) to isolate the different populations and to assay their phenotypic properties, including self-renewal and differentiation capabilities.

Chapter 1.2: Demarcation of stable subpopulations within the pluripotent hESC compartment.

Preface

This chapter is an original published article from the journal PLoS ONE, and is presented in its published form.

Bhatia, S., Pilquil, C., Roth-Albin, I., Draper, J.S. - Demarcation of stable subpopulations within the pluripotent hESC compartment. - *PLoS One.2013;8(2):e57276.doi: 10.1371/journal.pone.0057276*

This study was designed by Dr. Jonathan Draper and I helped in the execution of it. Ivana Roth-Albin, and Dr. Carlos Pilquil generated the REX1Venus reporter lines. I performed self-renewal, and differentiation assays, and designed and executed the DNA methylation assay. I contributed towards the following figures: Fig. 2, 5, 6, S5 and S6 and the writing of the manuscript.

TITLE: Demarcation of stable subpopulations within the pluripotent hESC compartment.

Authors
Sonam Bhatia[1,3,], Carlos Pilquil[1], Ivana Roth-Albin[1] & Jonathan S. Draper[1,2, 3*]

[1] McMaster Stem Cell and Cancer Research Institute, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON L8N 3Z5, Canada.
[2] Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON L8N 3Z5, Canada.
[3]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada.

* To whom correspondence should be addressed (draperj@mcmaster.ca)

ABSTRACT

Heterogeneity is a feature of stem cell populations, resulting from innate cellular hierarchies that govern differentiation capability. How heterogeneity impacts human pluripotent stem cell populations is directly relevant to their efficacious use in regenerative medicine applications. The control of pluripotency is asserted by a core transcription factor network, of which Oct4 is a necessary member. In mouse embryonic stem cells (ESCs), the zinc finger transcription factor Rex1 (Zfp42) closely tracks the undifferentiated state and is capable of segregating Oct4 positive mESCs into metastable populations expressing or lacking Rex1 that are inter-convertible. However, little is currently understood about the extent or function of heterogeneous populations in the human pluripotent compartment.

Human ESCs express *REX1* transcripts but the distribution and properties of *REX1* expressing cells have yet to be described. To address these questions, we used gene targeting in human ESCs to insert the fluorescent protein Venus and an antibiotic selection marker under the control of the endogenous *REX1* transcription regulatory elements, generating a sensitive, selectable reporter of pluripotency. *REX1* is co-expressed in OCT4 and TRA-1-60 positive hESCs and rapidly lost upon differentiation. Importantly, *REX1* expression reveals significant heterogeneity within seemingly homogenous populations of OCT4 and TRA-1-60 hESCs. *REX1* expression is extinguished before OCT4 during differentiation, but, in contrast to the mouse, loss of *REX1* expression demarcates a stable, OCT4 positive lineage-primed state in pluripotent hESCs that does not revert back to *REX1* positivity under normal conditions. We show that loss of *REX1* expression correlates with altered patterns of DNA methylation at the

REX1 locus, implying that epigenetic mechanisms may interfere with the metastable phenotype commonly found in murine pluripotency.

INTRODUCTION

Heterogeneity describes mixtures of distinct sub-populations of cells with functional differences that arise due to a balance of stem cell self-renewal and differentiation. In pluripotent stem cells, the cells at the apex of potency make discreet fate decisions, committing to one of numerous, but finite lineage choices, and descend through stages of cellular potential towards differentiated somatic phenotypes. Heterogeneity is a feature of stem cell systems throughout development, including intestinal, neural and hematopoietic stem cells (Graf and Stadtfeld, 2008), and the fluctuations in gene expression that comprise the heterogeneity in stem cell populations may be a necessary feature, presenting "windows of opportunity", during which cellular fate choices can be made (Chang et al., 2008; Graf and Stadtfeld, 2008; Huang et al., 2007). The identification and characterization of the cellular hierarchies that distinguish the differentiation capability of cells during development enables control over these processes, permitting the efficient differentiation of cells into tissues suitable for regenerative medicine applications.

In the early mouse embryo, a network of genes, including Oct4, Sox2 and Nanog, establish and maintain the pluripotent state (Chambers et al., 2003; Chambers et al., 2007; Masui et al., 2007; Mitsui et al., 2003; Niwa et al., 2000). Pluripotent cells can differentiate into all tissues of the adult organism and represent the highest level of potency from which permanent *in vitro* cell lines, embryonic stem cells (ESCs), have been established. Mouse ESCs closely resemble the "naïve" inner cell mass (ICM) of the blastocyst both in gene expression and differentiation capability (Beddington and Robertson, 1989; Guo et al., 2010) but display measurable differences from later mouse

epiblast stem cells (EpiSC) (Brons et al., 2007; Najm et al., 2011; Tesar et al., 2007), which are still considered pluripotent and capable of generating tissues comprising all three germ layers. These observations suggested the existence of a hierarchy within the pluripotent compartment that has recently been explored by several elegant genetic experiments. Mouse ESCs carrying fluorescent reporter proteins under the control of pluripotency-associated transcription factors such as Rex1 (Toyooka et al., 2008), Nanog (Chambers et al., 2007) and Stella (Hayashi et al., 2008) have described an unappreciated level of heterogeneity present in pluripotent Oct4 expressing ESC cultures. These reports have described the phenomena of metastability within the pluripotent compartment, in which ESCs fluctuate the expression of pluripotent markers as they transit between a naïve and lineage primed state. In particular, expression of the zinc finger transcription factor Rex1 (Zfp42) is exquisitely controlled during early embryogenesis and is sufficient to distinguish cells with an earlier ICM phenotype, capable of re-entering development and contribution in chimeric assays, from cells with later epiblast-like characteristics, that show poor chimeric contribution but good in vitro differentiation (Toyooka et al., 2008).

To date, the expression and necessity of genes such as *OCT4*, *SOX2* or *NANOG* have been investigated in undifferentiated hESCs(Adachi et al., 2010; Babaie et al., 2007; Hyslop et al., 2005) but attempts to explore the presence of a hierarchy within the pluripotent compartment have been limited to extant antibodies to cell surface markers(Enver et al., 2005; Stewart et al., 2006).

We previously identified the human *REX1* gene and showed that *REX1* transcripts are expressed in human ESCs and are associated with an undifferentiated phenotype (Henderson et al., 2002). To gain insight into *REX1* transcript expression, distribution and

the nature of pluripotency in hESCs, we used homologous recombination to target the human *REX1* locus with the Venus fluorescent reporter gene(Nagai et al., 2002). The REX1$^{Ven/w}$ hESC reporter cell lines not only allow a functional enrichment for undifferentiated cells but also describe a subpopulation of *REX1* expressing cells within heterogeneous populations of pluripotent OCT4 or TRA-1-60 expressing hESCs. Fractionation of hESC based on REX1Venus expression reveals a previously hidden hierarchy within the pluripotent compartment, comprising undifferentiated and differentiation primed cells, which lacks the metastability observed in murine ESCs

MATERIALS AND METHODS

Human ESC culture and differentiation.

Human ESC line H1(Thomson et al., 1998) (WiCell) was grown on mitotically-inactivated MEFs in hESC media (Knockout DMEM supplemented with 15% Knockout SR, 1x Non Essential Amino Acids, 1x Glutamax, 1x 2ME (all Invitrogen) and 16ng/ml bFGF (Peprotech) and passaged with Collagenase type IV (Invitrogen). For antibiotic selection experiments, cells were cultured in hESC media with or without the addition of 1.5ug/ml puromycin. For monolayer differentiation, hESCs were grown in hESC media on MEF coated 48 well plates and differentiation initiated by substituting the hESC media for DMEM supplemented with 10% FBS, 1x Non Essential Amino Acids and 1x Glutamax. 10uM retinoic acid (RA) was added to the differentiation media in some experiments. To evaluate hematopoietic differentiation in REX1$^{Ven/w}$ hESCs, EBs were generated by suspension culture methods as previously described(Cerdan et al., 2007). Briefly, undifferentiated REX1$^{Ven/w}$ hESCs were grown on Matrigel to confluence and then treated with Collagenase IV and mechanically scraped off into clumps and incubated overnight in 6-well ultralow attachment plates to allow EB formation (Cornings). For endoderm differentiation of cells isolated using fluorescence activated cell sorting (FACS), cells were grown in hESC media supplemented with Y27632 (Tocris Bioscience) for 24hrs and then placed in DMEM/F12 media with 1% FBS + 100ng/ml Activin A (Peprotech)+ 100ng/ml BMP4 for three days. For hematopoietic differentiation, EBs were cultured in StemPro34 serum-free medium (Invitrogen) supplemented with cytokines as follows: 300 ng/ml stem cell factor (SCF; Amgen), 50 ng/ml granulocyte colony stimulating factor (G-CSF; Amgen), 25 ng/ml bone

morphogenic protein-4 (BMP-4; R&D systems), 10 ng/ml interleukin-3 (IL-3; R&D systems), 10 ng/ml interleukin-6 (IL-6; R&D systems), and 300 ng/ml Flt-3 ligand (Flt-3 L: R&D systems). The EBs were cultured for 15 days with medium changes every 3 days. For mesoderm differentiation, FACS isolated populations were cultured in hESC media with Y27632 for 24hrs, followed by a 48hr treatment with 10ng/ml BMP4 (Peprotech) and 20ng/ml bFGF (Peprotech) in DMEM/F12 with 1% NEAA, 2% B27(Invitrogen), 1% ITS(Invitrogen) and 90uM 2-ME(Bernardo et al., 2011).

Vector construction and homologous recombination (HR).

The REX1-VF2Pu targeting vector was generated by recombineering. Briefly, a SalI/EcoRI cut Venus-F2A-Puro-pA cassette was cloned into SalI/EcoRI cut pL451 (NCI-Frederick) to create pL451+VF2Pu. 50bp REX1 locus specific homology arms were added to the region spanning Venus to the 3' Flp site of pL451 by PCR amplification (PrimeStar, Takara) with the REXVenus-F and REXVenus-pL451-R primers (Table S1; primers ordered from Sigma Genosys), producing the REX1VEN PCR product. Bacteria carrying the Human BAC RP11-713C19 and the pSC101-BAD-gba plasmid(Wang et al., 2006) (containing the Red/ET recombineering genes) were then electroporated with the REXVEN PCR product and correct replacement of the ATG of the *REX1* open reading frame (ORF) within exon 4 by the Venus-F2A-Puromycin cassette was confirmed by PCR and sequencing. REX-Gap-Rep-R and REX-Gap-Rep-F primers were used to add REX 5' and 3' specific 50bp homology arms onto EcoRI/NotI linearised pBS2SK (Stratgene), producing the REXGAP PCR product. Gap repair was performed on the REXVEN BAC with the REXGAP PCR product to generate the

pREX1-VF2Pu-TV targeting vector with 2.5kb 5' and 4.5kb 3' *REX1* specific homology arms, and confirmed by sequencing across HR junctions. The pREX1-VF2Pu-TV plasmid was transferred to the EL250 recombineering strain bacteria (NCI-Frederick) containing an inducible Flp and the FRT flanked PGK-Neo-pA excised.

HESC cell line H1 was pre-treated with 10uM Y27632 in hESC media for 1 hour and electroporated with 30ug of AloI linearised pREX1-VF2Pu-TV as previously described(Costa et al., 2007). After electroporation, cells were replated on 4DR(Tucker et al., 1997) MEFs in hESC media containing Y27632. 72 hours after electroporation, homologous recombination events were selected for by the addition of 1ug/ml puromycin for 10 days. Colonies were picked by hand under a dissecting microscope and transferred to MEF coated 4 well plates prior to expansion. Southern blot was performed on 10ug of PvuII digested hESC genomic DNA with a 470bp 5' probe generated by PCR with REXprb-F and REXprb-R, producing an 8.9kb band from the wild type allele and a 6.8kb band from the targeted allele.

Immunofluorescence, high content imaging and analysis.

HESCs were cultured in 48 well plates, washed with PBS, fixed with 4% PFA, washed with PBS, permeablised with 100% ice cold methanol and washed with PBS. Cells were stained with Hoechst 33342 and primary antibodies for OCT4 (mouse monoclonal 1:200, BD #611203), NANOG (rabbit monoclonal 1:400, Cell Signaling #4903), GATA4 (rabbit polyclonal 1:300, Sana Cruz #sc-9053) and p21 (rabbit monoclonal 1:400, Cell Signaling #2947) in 1% BSA in PBS for 2 hours at room temperature or 4^{o}C overnight, washed with PBS and stained with secondary antibodies (Goat anti Mouse AF546 1:500,

Invitrogen # A-11030; Donkey anti Rabbit AF647 1:500, Invitrogen #A-31573). Analysis was performed as previously described(Calder et al., 2012), briefly: plates were imaged on a Cellomics ArrayScan HCS reader (Thermo Scientific) or an Operatta High Content Screening System (Perkin Elmer) and images uploaded to a Columbus database (Perkin Elmer) and image analysis of immunofluorescence and reporter fluorescence was performed using Acapella high content and analysis software (Perkin Elmer). Cell nuclei were identified by Hoechst 33342 staining and the fluorescence intensity of the same nuclei in the VENUS, Cy3 and Cy5 channels measured. Custom MatLab (Mathworks) scripts were then used to quantify the fluorescent intensity of each nuclei in all channels and output statistics.

Flow Cytometry, FACS and CIC assay.

HESCs were dissociated to single cells, counted and $2x10^5$ cells co-stained with antibodies (or their corresponding isotype controls) diluted in staining buffer (1% BSA in PBS with 2mM EDTA) for 30 minutes on ice and then washed 2x with staining buffer. Cells were stained with the viability dye 7 aminoactinomycin D (7-AAD) (Immunotech) to exclude dead cells and analysed on a FACSCalibur or LSRII (BD Biosciences). For FACS, populations were fractionated using an Aria II (BD Biosciences) or a MoFlo (BD). Antibody dilutions were as follows: TRA1-1-60 @ 1:2000 (AF647 conjugated, BD Biosciences #560122), E-Cadherin @ 1:100 (PE conjugated ,Santa Cruz #sc-21791-PE), A2B5 @ 1:100 (APC conjugated, Miltenyi Biotec #130-093-58), CXCR4 @ 1:100 (APC conjugated, R&D# FAB170A) , CD31-APC @ 1:100 (BD Pharmingen), CD34-APC @ 1:100 (Miltenyi Biotech), and CD45-APC @ 1:100 (Miltenyi Biotech). For the colony

initiating cell (CIC) assay, cells were deposited at 25k and 50k cells by the ARIA II directly into wells of a 6 well plate containing mitotically inactivated hDFs(Bendall et al., 2007) (hESC derived fibroblasts, 200k per well) and hESC media and then re-cultured for 12 days. Plates were then fixed with 100% methanol, washed in PBS and stained with OCT4 (mouse monoclonal 1:200, BD #611203). Plates were imaged using a flatbed scanner (Canon) and colonies enumerated using a custom macro written for ImageJ (rsbweb.nih.gov/ij/).

mRNA extraction and PCR

mRNA and genomic DNA was isolated from hESCs with a RNA/DNA/Protein Purification kit (Norgenbiotek). mRNA was reverse transcribed into cDNA with an iScript kit (BIORAD), and subject to SYBR Green chemistry based QRT-PCR (Quantitative Real-Time PCR) (GoTaq master mix, Promega). Target genes were quantified relative to the house keeping genes TBP and/or CYCG. The presence of the REX1 targeting vector in genomic DNA was ascertained by gDNA PCR using a common forward primer (REXgDNA-F) in the 5' UTR of exon 4 combined with either a reverse primer (REXgDNA-R) in the endogenous REX1 ORF (recognizing endogenous REX1; 643bp band) or a reverse primer (Venus-R) in the Venus ORF (recognizing the REX1-VF2Pu targeting vector; 545bp band). QRT-PCR primers used in this study are listed in Table S2.

Bisulphite DNA methylation assay

Genomic DNA was isolated using All-in-one purification kit (Norgen Biotek Corp., Cat #: 24200), and was subjected to bisulfite conversion and treatment as per manufacturer's instructions (EZ DNA Methylation-Gold™ kit, Zymoresearch). Bisulfite converted DNA was PCR amplified using IMMOLASE™ DNA Polymerase (Bioline) cycling at: 95°C for 1min, [95°C for 30s, 58°C for 30s and 72°C for 1min]x40. Primers used in this study were generated elsewhere (Deb-Rinker et al., 2005; Takahashi et al., 2007) and are listed in Table S3. PCR products were cloned into pGEM®-T Easy Vector System I (Promega), purified and sequenced using T7 primer. The sequences were analyzed using QUMA analysis tool (http://quma.cdb.riken.jp/) (Kumaki et al., 2008).

Statistical analysis.

Error bars show SEM. Statistical analysis (t-test) was performed with Prism 5 (Graphpad). *= $p<0.05$, Graphs generated from the automated image analysis are derived from an n of between 3 and 6. Each n involved the analysis of >10,000 cells.

RESULTS

1. Generation of REX1$^{Ven/w}$ human embryonic stem cells.

REX1 is highly expressed in undifferentiated cultures of hPSC(Henderson et al., 2002), so we used homologous recombination to replace the start codon of an endogenous REX1 allele with a Venus-F2A-puromycin cassette (Fig. 1A) and enriched for HR events by puromycin selection. Two clones with a correctly targeted *REX1* allele (REX1$^{Ven/w}$) were confirmed by southern blotting (Fig. 1B) and displayed bright REX1Venus expression restricted to colonies with an undifferentiated hPSC morphology (Fig. 1C). Karyotyping

revealed that the REX1$^{Ven/w}$ clones retained a normal 46 XY chromosomal count (Fig. S1).

2. REX1 expression delineates a subpopulation of pluripotent hESCs.

REX1Venus expression was confined to a subpopulation of cells co-stained with the human pluripotency-associated cell surface marker TRA-1-60(Draper et al., 2002) (Fig. 1D), confirming the association of our *REX1* reporter expression with pluripotency in hPSCs but also establishing that *REX1* displays heterogeneous expression in hPSCs. All REX1Venus-positive (herein referred to as VEN+) cells co-stained with the epithelial marker E-CADHERIN but showed virtually no reactivity with differentiation markers such as A2B5 and CXCR4 (Fig. 1D & Fig. S2A), expressed on ectoderm and endoderm cells, respectively. Functional enrichment for VEN+ cells by the addition of puromycin to REX1$^{Ven/w}$ cultures depleted virtually all spontaneous differentiation (Fig. 1D & Fig. S2B), as measured by loss of A2B5 and CXCR4 expressing cells and TRA-1-60 negative cells, being composed almost uniformly of TRA-1-60 and VEN double positive cells (Fig. 1D). VEN+ cells isolated by FACS display a 13-fold enrichment for *REX1* transcript, when compared to VEN- cells, but less than 3-fold increase in other pluripotency markers such as *OCT4*, *NANOG* and *SOX2* (Fig. 1E), validating the fidelity of the REX1Venus reporter to enrich REX1-expressing cells. Markers of early differentiation, including *N-CAD*, *EOMES, FOXA2* and *CDX2* were all enriched in the VEN- cells (Fig. 1F).

3. REX1 expression marks a high level domain in the pluripotent hierarchy.

We next tested the hypothesis that *REX1* expressing cells occupy a position towards the top of the pluripotency hierarchy. FACS-fractionated (to 99.9% purity) VEN+ or VEN- populations were separately re-cultured in undifferentiated hPSC conditions and the profile of TRA-1-60 and REX1Venus expression was evaluated over time. Ten days after sorting, the VEN+ fraction contain not only TRA-1-60 expressing VEN+ (TRA+VEN+) cells but had also re-constituted the TRA+VEN- population (Fig. 2A). Cultures derived from the VEN- fraction contained a large proportion of TRA-1-60 positive cells but, in contrast to the behaviour of the VEN+ hPSCs and the murine Rex1-GFP reporter(Toyooka et al., 2008), did not re-establish a VEN+ population, even after 2 months of continuous culture (data not shown), despite the demonstrable presence of the REX1-VF2Pu targeting vector in the genomic DNA (Fig. S3). A second serial round of FACS purification performed on the VEN+ cultures derived from the first sort displayed the same pattern of REX1Venus distribution (Fig. 2A), confirming that REX1-expressing VEN+ cells can produce VEN- cells but not the converse, implying that VEN+ cells occupy a higher level in the pluripotent hierarchy than VEN- hPSCs.

The lack of reversion from VEN- to VEN+ cells prompted an investigation of the epigenetic mechanisms that might be regulating the expression of *REX1*. We performed bisulfite genome sequencing on the *REX1* and *OCT4* gene loci on FACS isolated populations demarcated by the expression of both TRA-1-60 and REX1Venus. Assay of the three main populations demonstrated that only the TRA+VEN+ populations displayed hypo-methylation at both the REX1 and OCT4 promoters (Fig. 2B). Therefore, the lack of metastable reversion from VEN- to VEN+ in hESCs could be due to epigenetic

changes at the REX1 locus. This suggests that epigenetic modification to the DNA may be responsible for the stability of the VEN- population.

4. REX1 expression is rapidly lost upon differentiation.

Antibody co-staining of REX1$^{Ven/w}$ hPSCs revealed that both OCT4 and NANOG marked virtually all cells within colonies that were morphology identifiable as undifferentiated, but VEN+ cells were often distributed in a mosaic pattern (Fig. 3A). By contrast, p21, a cell cycle inhibitor associated with the differentiation of hPSCs(Egozi et al., 2007), surrounded the colonies and did not overlap with REX1Venus and OCT4. In addition, differentiation of REX1$^{Ven/w}$ hPSCs in a hematopoietic differentiation assay showed that REX1Venus intensity is not detected in populations of cells that express CD31, CD34 or CD45, all markers of hematopoietic cell lineages(Cerdan et al., 2007) (Fig. S4). Like murine Rex1(Toyooka et al., 2008), our data shows that expression of human *REX1* is associated with the pluripotent state and is lost upon differentiation.

We used automated high-content imaging combined with cell annotation software analysis to quantify the overlap of REX1Venus, OCT4 and NANOG expression in undifferentiated and differentiating REX1$^{Ven/w}$ hPSCs. In undifferentiated cultures, just under half of the cells were VEN+ and NANOG+ or VEN+ and OCT4+, with a fifth expressing only NANOG or OCT4 (Fig. 3B). In comparison, over half of undifferentiated hPSCs were NANOG and OCT4 double positive, and fewer were solely OCT4 or NANOG positive. After two days of culturing in conditions that antagonize pluripotency (DMEM + 10% FBS & retinoic acid)(Draper et al., 2002), REX1Venus and NANOG

were virtually absent from REX1$^{Ven/w}$ hPSCs cultures, despite a fifth of the population continuing to express OCT4 (Fig. 3B). Nearly all VEN+ cells were NANOG+ or OCT4+, compared with two thirds of NANOG+ or OCT4+ cells that were VEN+ (Fig. 4A), showing that OCT4 is the most widespread pluripotency marker in hPSC cultures. Upon induction of differentiation, REX1Venus was quickly lost from NANOG+ and OCT4+ cells, and nearly all remaining NANOG-positive cells continued to express OCT4 after 2 days of differentiation (Fig. 4A). Together, these data imply that both REX1 and NANOG mark a subset of cells within the more abundant OCT4-positive population, and that the expression of REX1 and NANOG are extinguished before OCT4 during differentiation. In contrast, the differentiation marker, p21, displayed no appreciable overlap with either OCT4+ or VEN+ cells (Fig. 4B).

5. Colony forming capacity is not confined to *REX1* expressing hESCs.

A colony initiating cell (CIC) assay, a measure of self-renewal capacity(Wray et al., 2010), was used to evaluate whether a functional advantage was associated with *REX1* expression (Fig 5A). The REX1Venus were co-stained with TRA-1-60 and fractionated by FACS into TRA+VEN+, TRA+VEN- and TRA-VEN- populations, seeded back into undifferentiated hESC growth conditions(Draper et al., 2002) at two defined dilutions and emerging pluripotent colonies quantified by OCT4 expression. The CIC activity of both TRA+VEN+ and TRA+VEN- fractions were comparable, at ~0.1%, for all dilutions tested, in accordance with the anticipated CIC efficacy for karyotypically normal hPSCs(Enver et al., 2005; Watanabe et al., 2007), and the loss of both REX1Venus and TRA-1-60 marks a differentiated state containing negligible CIC activity. We then

analyzed the REX1Venus expression in colonies that emerged from FACS isolated VEN+ or VEN- fractions. The VEN+ derived colonies expressed both REX1Venus and OCT4 protein but the VEN- derived colonies expressed OCT4 and remained uniformly negative for REX1Venus expression (Fig. 5B), a result consistent with our previous findings that VEN- cells are unable to re-establish VEN+ cells (Fig. 2A). These data suggest that the TRA+VEN+ and TRA+VEN- fractions are essentially equivalent in their ability to regenerate OCT4 expressing hESC colonies, and that the higher levels of *REX1* expression associated with the TRA+VEN+ populations are not a requisite for hESC colony formation.

6. REX1-negative hPSCs are lineage primed.

To understand if there was a functional outcome associated with loss of REX1 expression in hPSCs, we then used FACS to isolate the TRA+VEN+ and TRA+VEN- populations and asked whether they had distinct phenotypes. QRT-PCR analysis demonstrated that pluripotency genes such as *OCT4*, *NANOG* and *SOX2* were expressed at equivalent levels in both the TRA+VEN+ and TRA+VEN- populations despite the enrichment for *REX1* transcripts in the TRA+VEN+ population (Fig. 6A). In contrast, the TRA+VEN- population displayed a marked increase in the transcripts of early definitive endoderm specification such as *EOMES* and *SOX17* when compared to the TRA+VEN+ cells (Fig 6A), as well as the pan or extraembryonic endoderm markers *FOXA2*, *AFP*, *GATA6* and *HNF1B* (Fig 6A and Fig. S5A). We next tested whether the differential in expression of early lineage marker transcripts between the TRA+VEN+ and TRA+VEN- cells were maintained after guided differentiation in endoderm or mesoderm inducing conditions for

3 or 2 days respectively (Fig. 6B). The TRA+VEN- population displayed a two-fold increase over the TRA+VEN+ population in the expression of mesoderm genes, *BRACHYURY* and *MIXL1*, after 2 days in mesoderm inducing conditions (Fig. 6C). Similarly, the TRA+VEN- cells showed higher transcript expression for the endoderm markers, *EOMES*, *SOX17* and FOXA2 in TRA+VEN- cells after 3 days of differentiation towards the endoderm lineage, when compared to TRA+VEN+ cells (Fig. 6D). The TRA+VEN- cells also generated a higher number of cells expressing GATA4 protein compared to the TRA+VEN+ cells after 3 days of endoderm differentiation (Fig. 6E & Fig. S5B). Similar trends were also evident in an endoderm time course differentiation (Fig. S6) performed on VEN+ cells that were purified by puromycin drug selection or VEN- cells that had been isolated by FACS (Fig. 2A) and then subsequently cultured for several months, implying that the VEN- cells represent a stable lineage primed state.

DISCUSSION

We have targeted an allele of the transcription factor, *REX1*, in human embryonic stem cells with a fluorescent protein and used this reporter to investigate the pluripotent compartment present in cultures of hESCs. Our *REX1* reporter hESCs have, for the first time, enabled the tracking of *REX1* expression during the culture and differentiation of hESCs. Although it has been understood for some time that undifferentiated hESC cultures often contain cells that have arisen by spontaneous differentiation (Draper et al., 2002), markers like TRA-1-60 or OCT4 are frequently accepted as faithful markers of seemingly equivalent populations of pluripotent hESCs. Our analysis of REX1$^{Ven/w}$ hESCs revealed that TRA-1-60 and OCT4 expressing cells contain a subpopulation that

is demarcated by *REX1* expression, and provides evidence that a similar heterogeneity exists within hESCs that has previously been observed within the undifferentiated murine ESC compartment using a mouse Rex1 ESC reporter line(Toyooka et al., 2008). We have demonstrated the connection between *REX1* and pluripotency by using the puromycin antibiotic selection cassette in our REX1$^{Ven/w}$ hESCs as a mechanism for enriching pluripotent hESCs at the expense of differentiated cells. In addition, we have used prospective isolation of VEN+ cells to annotate the molecular phenotype associated with *REX1* expression. VEN+ cells displayed a comparative enrichment for *REX1* transcripts and lower levels of differentiation markers. However, two distinct subpopulations are evident within TRA-1-60 expressing *REX1* reporter hESCs: TRA+VEN+ and TRA+VEN- cells, with the latter showing similar levels of *OCT4*, *SOX2* and *NANOG* gene transcripts to the TRA+VEN+ but higher levels of early lineage associated markers. In addition, the TRA+VEN- cells showed a greater expression of lineage markers when challenged for differentiation, providing compelling evidence that TRA+VEN- cells are primed for differentiation. Importantly, our finding that distinct subpopulations of hESCs display different aptitudes at differentiation is not without precedent, with others showing that clonal tracking can unmask significant contribution variability in seemingly identical hESCs(Stewart et al., 2010). Notwithstanding, there are some important caveats to the data we present here. We undertook gene targeting of *REX1* in hESCs principally due to the paucity of commercially available REX1 antibodies that accurately detect this protein. The lack of connection between the *REX1* transcript data that our reporter provides and endogenous REX1 protein levels negates the drawing of conclusions concerning REX1 function in the TRA/VEN populations that we have isolated. In addition, the act of

targeting one of the *REX1* alleles could impact the levels of REX1 protein expressed within the hESCs, potentially disturbing the ability of this transcription factor to properly function. The impact of heterozygosity at the Rex1 locus in murine pluripotency is poorly defined, but heterozygous Rex1-GFP reporter murine ESCs continue to participate in the formation of chimeric animals (Toyooka et al., 2008), suggesting that any impact on pluripotency appears to be minimal. Interestingly, although Rex1 heterozygous adult mice are viable, the expected Mendelian ratio of their litters is disturbed, a phenotype which has been speculated to occur due to the role of Rex1 in later gametogenesis(Kim et al., 2011). With these concerns in mind, knock-in reporter lines continue to provide important insights into biological processes when other conventional reagents are lacking.

Our study features the first in depth investigation of heterogeneity and the stability of sub-populations within the pluripotent compartment of hESC cultures. Using our reporter, we have isolated discrete pluripotent fractions and then mapped the inter conversion between phenotypes. VEN+ cells occupy the top domain of hPSC pluripotency, giving rise to the TRA+VEN- and differentiated TRA-VEN- cells when re-cultured. Unexpectedly, the VEN- cells could only re-generate the TRA+VEN- and TRA-VEN- populations but not the TRA+VEN+, despite the demonstrable presence of the REX1 targeting vector in the gDNA of these cells. To rule out the presence of contaminating wild type hESCs in the TRA+VEN- population we performed serial FACS fractionation and found the same pattern of provenience for the TRA+VEN- cells from the TRA+VEN+ population but not vice versa. The finding that the loss of *REX1*

expression, but retention of TRA-1-60, signals commitment to a stable intermediate lineage primed state is in stark contrast to the mouse(Toyooka et al., 2008). Notably, experiments with Nanog, Stella and Rex1 (Chambers et al., 2007; Hayashi et al., 2008; Toyooka et al., 2008) all suggest that murine pluripotent heterogeneity is comprised of dynamic, metastable states(Cherry and Daley, 2010); GFP+ cells from each of these reporters can re-establish a GFP- population when isolated and re-cultured and, importantly, vice versa. Why there exists a discrepancy between the human and mouse REX1 reporters remains uncertain, but might describe either the manifestation of distinct growth conditions or species related differences in how early fate allocation is managed. The DNA methylation data presented here suggests a stable epigenetic regulation may govern the irreversible loss of *REX1* expression in hESCs. The metastable nature of other murine pluripotency associated factors, like Stella, is associated with a more plastic regulatory mechanism, which appears to involve histone modifications (Hayashi et al., 2008). Our findings provide some clarity to the ongoing debate concerning the similarity of pluripotent stem cells between mouse and human. Human ESCs share several key features with mESCs that distinguish them from mouse epiblast stem cells (EpiSC), including the expression of two ICM markers, REX1(Henderson et al., 2002) and KLF4(Chan et al., 2009), in combination with a lack of the FGF5 expression, a well-characterised epiblast marker(Pelton et al., 2002). Whilst hESCs reflect many mESC properties, they do display a growth factor dependency that is more akin to EpiSC than mESC(Brons et al., 2007; Rossant, 2008; Tesar et al., 2007). Human ESCs do not self-renew in response to LIF, as observed in mESCs, possibly due to the lack of diapause in humans (Humphrey et al., 2004; Nichols et al., 2001; Renfree and Shaw, 2000). FGF2

and Activin A can maintain undifferentiated cultures of both hESC and mouse EpiSCs(Vallier et al., 2009a), strengthening the speculation that hPSCs may represent a later embryonic stage than LIF-dependent mouse ESCs. However, it has been shown that mouse EpiSCs can spontaneously revert to an ES cell-like state when cultured in media containing LIF and serum (Bao et al., 2009), which is accompanied by a reset of DNA methylation at Rex1 and Stella promoters. It has been demonstrated that FGF-based signaling blocks reversion of mouse EpiSC to an ESC-like state (Greber et al., 2010), a finding that may help to explain our data showing that the TRA-1-60 and OCT4 positive VEN- hESCs do not convert back to a VEN+ state in normal hESC culture conditions. FGF2 is a common denominator in virtually all hESC media compositions(Greber et al., 2010; Lanner and Rossant, 2010), and appears necessary to maintain the long-term self-renewal of hESCs(Amit et al., 2000; Vallier et al., 2005; Vallier et al., 2009b). Thus, we speculate that FGF signaling may play a role in protecting and/or causing DNA methylation of the *REX1* locus in VEN- hESCs and Rex1- EpiSC, implying that the culture conditions that support undifferentiated hESC propagation may limit the metastable gene expression observed in murine ESCs. Recently "naïve" LIF-dependent hESC-like lines have been derived, albeit in an unstable state, that mimic more closely some of the properties of mESCs (Hanna et al., 2010). Evaluating subpopulations identified by our human *REX1* reporter in the context of "naïve" LIF-dependent hESC-like growth conditions may provide significant insight into how signaling pathways mediate metastability in gene expression.

Why *REX1* transcripts are asymmetrically expressed within the pluripotent hESC compartment is likely linked to the function of *REX1*, of which there is currently a limited understanding. The culture of mESC in signaling conditions that promote a pluripotent ground state yields uniformity of Rex1 expression(Wray et al., 2010), suggesting that Rex1 is closely associated with the naïve pluripotent state. Rex1 deletion in the mouse embryos and ESCs perturbs both gene expression and differentiation(Masui et al., 2008; Rezende et al., 2011; Scotland et al., 2009). This phenotype exerted in Rex1-null embryos and ESCs may at least, in part, be due to the role of Rex1 as an epigenetic regulator. Rex1-null blastocysts display hypermethylation of imprinted genes, such as Peg3 and Nespas, and chromatin immunoprecipitation demonstrated that Rex1 binds only to the unmethylated allele of these genes(Kim et al., 2011). Rex1 appears to share a common evolutionary ancestor with Ying Yang 1 (Yy1)(Kim et al., 2007), a ubiquitously expressed zinc finger transcription factor that has a proven role as a mediator of epigenetic regulation(Atchison et al., 2003), including interactions with Polycomb Group (PcG) proteins(Garcia et al., 1999; Kim et al., 2007). PcG proteins are known to repress gene expression by interacting with and changing chromatin structure(Schuettengruber et al., 2007), and are believed to aid in the modulation of PSC fate by inhibition of lineage specific markers(Boyer et al., 2006), with a recent study indicating that Rex1 may also interact with PcGs(Garcia-Tunon et al., 2011). More recently, it has been demonstrated that REX1 is an integral part of the mechanism that prevents lyonization in female mouse embryonic stem cells by directly interacting with both the Xist and Tsix loci (Deuve and Avner, 2010; Gontan et al., 2012; Navarro et al., 2010). A conserved function for *REX1* in human X-inactivation remains to be discovered, although it has been questioned

whether X-inactivation via the governance of *XIST* expression by *TSIX* even occurs in human cells(Migeon, 2003, 2011). Notwithstanding, the hESCs targeted in this study are male, and only recently have advances permitting the derivation of female hESCs with two active X chromosomes(Hanna et al., 2010; Lengner et al., 2010), affording an opportunity to explore REX1 function in this area. Significantly, LIF-based signaling appears to play a strong role in the generation of human induced pluripotent stem cells (iPSC) that contain two active X chromosomes (Tomoda et al., 2012), providing an intriguing model for exploring REX1 function in human lyonization.

It is clear that identifying and characterising the subpopulations that occur within hESC culture heterogeneity can yield significant increases in our understanding of these important cells and have direct impact on their future utility in drug discovery and therapeutic applications. The *REX1* reporter lines described here represent a powerful new tool for understanding human pluripotency.

ACKNOWLEDGMENTS: We thank the technicians (Dr Marilyne Levadoux-Martin, Jamie McNicol and Jennifer Russell) at the SCCRI, Ashley Calder and Seok-Ho Hong for assistance in the acquisition and analysis of data. We are indebted to Tony Collins for custom image analysis scripts and help with data analysis, and to Dr Sheila Singh and Nicole McFarlane for FACS assistance. We are extremely grateful to Drs Janet Rossant and Brad Doble for critical reading of the manuscript. We thank Dr Mickie Bhatia for thoughtful discussions and generous infrastructure and technical support.

REFERENCES:

1. Graf T, Stadtfeld M (2008) Heterogeneity of embryonic and adult stem cells. Cell Stem Cell 3: 480-483.

2. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453: 544-547.

3. Huang S, Guo YP, May G, Enver T (2007) Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental biology 305: 695-713.

4. Chambers I, Colby D, Robertson M, Nichols J, Lee S, et al. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113: 643-655.

5. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, et al. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113: 631-642.

6. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics 24: 372-376.

7. Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, et al. (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature cell biology 9: 625-635.

8. Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, et al. (2007) Nanog safeguards pluripotency and mediates germline development. Nature 450: 1230-1234.

9. Beddington RS, Robertson EJ (1989) An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development 105: 733-737.

10. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, et al. (2010) Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental cell 18: 675-685.

11. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, et al. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448: 196-199.

12. Najm FJ, Chenoweth JG, Anderson PD, Nadeau JH, Redline RW, et al. (2011) Isolation of epiblast stem cells from preimplantation mouse embryos. Cell Stem Cell 8: 318-325.

13. Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, et al. (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448: 191-195.

14. Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H (2008) Identification and characterization of subpopulations in undifferentiated ES cell culture. Development 135: 909-918.

15. Hayashi K, Lopes SM, Tang F, Surani MA (2008) Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell 3: 391-401.

16. Hyslop L, Stojkovic M, Armstrong L, Walter T, Stojkovic P, et al. (2005) Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. Stem Cells 23: 1035-1043.

17. Adachi K, Suemori H, Yasuda SY, Nakatsuji N, Kawase E (2010) Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes to cells : devoted to molecular & cellular mechanisms 15: 455-470.

18. Babaie Y, Herwig R, Greber B, Brink TC, Wruck W, et al. (2007) Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells 25: 500-510.

19. Stewart MH, Bosse M, Chadwick K, Menendez P, Bendall SC, et al. (2006) Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. Nature methods 3: 807-815.

20. Enver T, Soneji S, Joshi C, Brown J, Iborra F, et al. (2005) Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. Human molecular genetics 14: 3129-3140.

21. Henderson JK, Draper JS, Baillie HS, Fishel S, Thomson JA, et al. (2002) Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens. Stem Cells 20: 329-337.

22. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, et al. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nature biotechnology 20: 87-90.

23. Thomson J, Itskovitz-Eldor J, Shapiro S, Waknitz M, Swiergiel J, et al. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282: 1145.

24. Cerdan C, Hong SH, Bhatia M (2007) Formation and hematopoietic differentiation of human embryoid bodies by suspension and hanging drop cultures. Current protocols in stem cell biology Chapter 1: Unit 1D 2.

25. Bernardo AS, Faial T, Gardner L, Niakan KK, Ortmann D, et al. (2011) BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. Cell Stem Cell 9: 144-155.

26. Wang J, Sarov M, Rientjes J, Fu J, Hollak H, et al. (2006) An improved recombineering approach by adding RecA to lambda Red recombination. Molecular biotechnology 32: 43-53.

27. Costa M, Dottori M, Sourris K, Jamshidi P, Hatzistavrou T, et al. (2007) A method for genetic modification of human embryonic stem cells using electroporation. Nature protocols 2: 792-796.

28. Tucker KL, Wang Y, Dausman J, Jaenisch R (1997) A transgenic mouse strain expressing four drug-selectable marker genes. Nucleic acids research 25: 3745-3746.

29. Calder A, Roth-Albin I, Bhatia S, Pilquil C, Lee JH, et al. (2012) Lengthened G1 Phase Indicates Differentiation Status in Human Embryonic Stem Cells. Stem cells and development.

30. Bendall SC, Stewart MH, Menendez P, George D, Vijayaragavan K, et al. (2007) IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. Nature 448: 1015-1021.

31. Deb-Rinker P, Ly D, Jezierski A, Sikorska M, Walker PR (2005) Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. The Journal of biological chemistry 280: 6257-6260.

32. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131: 861-872.

33. Kumaki Y, Oda M, Okano M (2008) QUMA: quantification tool for methylation analysis. Nucleic acids research 36: W170-175.

34. Draper JS, Pigott C, Thomson JA, Andrews PW (2002) Surface antigens of human embryonic stem cells: changes upon differentiation in culture. J Anat 200: 249-258.

35. Egozi D, Shapira M, Paor G, Ben-Izhak O, Skorecki K, et al. (2007) Regulation of the cell cycle inhibitor p27 and its ubiquitin ligase Skp2 in differentiation of human embryonic stem cells. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology 21: 2807-2817.

36. Wray J, Kalkan T, Smith AG (2010) The ground state of pluripotency. Biochemical Society transactions 38: 1027-1032.

37. Watanabe K, Ueno M, Kamiya D, Nishiyama A, Matsumura M, et al. (2007) A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nature biotechnology 25: 681-686.

38. Stewart MH, Bendall SC, Levadoux-Martin M, Bhatia M (2010) Clonal tracking of hESCs reveals differential contribution to functional assays. Nature methods 7: 917-922.

39. Kim JD, Kim H, Ekram MB, Yu S, Faulk C, et al. (2011) Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. Human molecular genetics 20: 1353-1362.

40. Cherry A, Daley GQ (2010) Another horse in the meta-stable state of pluripotency. Cell Stem Cell 7: 641-642.

41. Chan KK, Zhang J, Chia NY, Chan YS, Sim HS, et al. (2009) KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells. Stem Cells 27: 2114-2125.

42. Pelton TA, Sharma S, Schulz TC, Rathjen J, Rathjen PD (2002) Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. Journal of cell science 115: 329-339.

43. Rossant J (2008) Stem cells and early lineage development. Cell 132: 527-531.

44. Humphrey RK, Beattie GM, Lopez AD, Bucay N, King CC, et al. (2004) Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells 22: 522-530.

45. Nichols J, Chambers I, Taga T, Smith A (2001) Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development 128: 2333-2339.

46. Renfree MB, Shaw G (2000) Diapause. Annual review of physiology 62: 353-375.

47. Vallier L, Mendjan S, Brown S, Chng Z, Teo A, et al. (2009) Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development 136: 1339-1349.

48. Bao S, Tang F, Li X, Hayashi K, Gillich A, et al. (2009) Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature 461: 1292-1295.

49. Greber B, Wu G, Bernemann C, Joo JY, Han DW, et al. (2010) Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. Cell Stem Cell 6: 215-226.

50. Lanner F, Rossant J (2010) The role of FGF/Erk signaling in pluripotent cells. Development 137: 3351-3360.

51. Amit M, Carpenter MK, Inokuma MS, Chiu CP, Harris CP, et al. (2000) Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. Developmental biology 227: 271-278.

52. Vallier L, Alexander M, Pedersen RA (2005) Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci 118: 4495-4509.

53. Vallier L, Touboul T, Brown S, Cho C, Bilican B, et al. (2009) Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. Stem Cells 27: 2655-2666.

54. Hanna J, Cheng AW, Saha K, Kim J, Lengner CJ, et al. (2010) Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proceedings of the National Academy of Sciences of the United States of America 107: 9222-9227.

55. Rezende NC, Lee MY, Monette S, Mark W, Lu A, et al. (2011) Rex1 (Zfp42) null mice show impaired testicular function, abnormal testis morphology, and aberrant gene expression. Developmental biology 356: 370-382.

56. Masui S, Ohtsuka S, Yagi R, Takahashi K, Ko MS, et al. (2008) Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. BMC developmental biology 8: 45.

57. Scotland KB, Chen S, Sylvester R, Gudas LJ (2009) Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. Developmental dynamics : an official publication of the American Association of Anatomists 238: 1863-1877.

58. Kim JD, Faulk C, Kim J (2007) Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. Nucleic acids research 35: 3442-3452.

59. Atchison L, Ghias A, Wilkinson F, Bonini N, Atchison ML (2003) Transcription factor YY1 functions as a PcG protein in vivo. The EMBO journal 22: 1347-1358.

60. Garcia E, Marcos-Gutierrez C, del Mar Lorente M, Moreno JC, Vidal M (1999) RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. The EMBO journal 18: 3404-3418.

61. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G (2007) Genome regulation by polycomb and trithorax proteins. Cell 128: 735-745.

62. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349-353.

63. Garcia-Tunon I, Guallar D, Alonso-Martin S, Benito AA, Benitez-Lazaro A, et al. (2011) Association of Rex-1 to target genes supports its interaction with Polycomb function. Stem cell research 7: 1-16.

64. Deuve JL, Avner P (2010) The Coupling of X-Chromosome Inactivation to Pluripotency. Annual review of cell and developmental biology.

65. Gontan C, Achame EM, Demmers J, Barakat TS, Rentmeester E, et al. (2012) RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature 485: 386-390.

66. Navarro P, Oldfield A, Legoupi J, Festuccia N, Dubois A, et al. (2010) Molecular coupling of Tsix regulation and pluripotency. Nature 468: 457-460.

67. Migeon BR (2003) Is Tsix repression of Xist specific to mouse? Nature genetics 33: 337; author reply 337-338.

68. Migeon BR (2011) The single active X in human cells: evolutionary tinkering personified. Human genetics 130: 281-293.

69. Lengner CJ, Gimelbrant AA, Erwin JA, Cheng AW, Guenther MG, et al. (2010) Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell 141: 872-883.

70. Tomoda K, Takahashi K, Leung K, Okada A, Narita M, et al. (2012) Derivation conditions impact x-inactivation status in female human induced pluripotent stem cells. Cell Stem Cell 11: 91-99.

Figure Legends

Figure 1. Generation of REX1$^{Ven/w}$ human embryonic stem cells. A) Schematic of the wild type human REX1 allele, targeting vector (REX1-VF2Pu TV) and targeted allele. B) Southern blot confirmation of targeting. C) Phase and fluorescence images of a REX1Ven/w hESCs. Scale bar = 100 microns. D) Flow cytometry on REX1$^{Ven/w}$ cells grown for 7 days in undifferentiated hESC conditions with or without the addition of puromycin co-stained with TRA -1-60 or A2B5. Control inset. E & F) QRT-PCR analysis of pluripotency (E) and differentiation (F) gene transcript expression VEN+ populations isolated by FACS from undifferentiated REX1$^{Ven/w}$ hESCs. All values are normalised relative to the VEN- population =1.

Figure 2. Serial fractionation of REX1$^{Ven/w}$ cultures based upon REX1Venus expression. A) FACS fractionation, re-culture and TRA-1-60 flow analysis of REX1$^{Ven/w}$ hPSCs. B) Bisulphite DNA sequencing on TRA-1-60/REX1Venus hESC populations isolated by FACS for the OCT4 and REX1 promoters. Empty circles designate unmethylated CpG residues and filled circles denote methylated residues. CpG position is provided with reference to transcription start site.

Figure 3. Distribution of pluripotent markers in undifferentiated REX1$^{Ven/w}$ hPSCs. A) Immunocytochemistry for OCT4 (red), NANOG (blue, bottom row) or p21 (blue, top row) in REX1$^{Ven/w}$ cells. Scale bar = 120 microns. B) Quantification of REX1Venus, OCT4 and NANOG expression by high content imaging and automated cell level analysis in undifferentiated cultures (Day 0) and during a time course of retinoic acid induced differentiation (n=4). V= REX1Venus, O = OCT4, N=NANOG, + = positive, - = negative.

Figure 4. Co-incidence of pluripotency markers in undifferentiated REX1$^{Ven/w}$ hPSC cultures. A) Output of imaging analysis measuring the co-expression of REX1Venus (VEN), OCT4 or NANOG (NAN) pluripotency markers in undifferentiated (Undiff) hESCs and cells treated with retinoic acid (RA) for 2 days, n=4. B) Output of cell level analysis of p21 co-expression with REX1Venus (VEN) or OCT4 positive cells, n=4.

Figure 5. Phenotype of REX1Venus positive and negative populations within REX1$^{Ven/w}$ hESC cultures. A) CIC activity of FACS purified TRA+VEN+ (T+V+), TRA+VEN- (T+V-) and TRA-VEN- (T-V-) populations isolated from undifferentiated cultures of H1 REX1$^{Ven/w}$ cells. B) OCT4 immunocytochemistry and REX1Venus expression in FACS isolated REX1Venus positive (VEN+) or negative (VEN-) populations after 12 days culture. Scale = 120 microns.

Figure 6. Loss of REX1 within the pluripotent population primes cells for differentiation. A) QRT-PCR analysis of gene transcript expression in FACS separated TRA+VEN+ and TRA+VEN- populations. The TRA+VEN- fraction is normalised relative to the TRA+VEN+ population =1. B) Schematic showing the differentiation treatment of hESCs C & D) QRT-PCR data showing the expression of mesoderm (C) and endoderm (D) lineage associated markers after the TRA+VEN+ and TRA+VEN-

fractions were subject to 2 or 3 days of differentiation in mesoderm or endoderm conditions, respectively. C) *BRACHYURY* and *MIXL1* were used as mesoderm associated markers. Undifferentiated TRA+VEN+ population was used as a control. n=2 D) *EOMES*, *SOX17* and *FOXA2* were used as endoderm specific markers, and gene expression was normalized to TRA+VEN+ day 3 differentiated cells. n=3 E) Fold enrichment of the percentage of GATA4 positive endoderm cells generated from TRA+VEN- cells relative to those from TRA+VEN+ population after 3 days of treatment with Activin A and BMP4 in low serum media, as observed by immunocytochemistry.

Supplementary Material

Figure S1. Normal 46XY karyotype, assayed by WiCell Institute, of two H1 subclones expressing REX1-VF2Pu targeting vector.

Figure S2. A) Flow cytometric analysis of REX1$^{Ven/w}$ cells grown for 7days in undifferentiated hESC conditions with or without puromycin co-stained with E-CADHERIN (E-CAD) or CXCR4. B) Histograms of REX1Venus expression with (green line) or without (blue line) 7 day puromycin treatment. Control H1 hESCs (red line).

Figure S3. PCR on genomic DNA for the presence of the REX1-VF2Pu targeting vector (REX1 TV) versus control endogenous REX1 locus (REX1 END). Samples assayed: Wild type H1 hESC (H1 wt), TRA-1-60/REX1Venus fractions (TRA VEN) and VEN- cultures after 7 passages (RXVen-).

Figure S4. Hematopoietic differentiation of REX1$^{Ven/w}$ cells. REX1 reporter cells were differentiated in embryoid bodies in conditions that induce blood formation and assayed at day 4, 10 and 15 for A) REX1Venus expression and B) markers of hematopoietic specification CD31, CD34 and CD45.

Figure S5. A) QRT-PCR of undifferentiated FACS isolated TRA+VEN+ and TRA+VEN- cells for extraembryonic endoderm markers. Gene expression is normalized to the housekeeping gene *TBP*, and is relative to TRA+VEN+ fraction (n=2) B) Cells were isolated by FACS, re-seeded and the next day treated with endoderm-inducing conditions for 3 days before fixation and staining with GATA4 and OCT4. Scale = 120 microns.

Figure S6. QRT-PCR of several endoderm markers, *SOX17, EOMES, FOXA2, Goosecoid (GSC), Cerberus-like (CER)* and *GATA4*, over a three day (d0-d3) time-course analysis of puromycin selected VEN+ (dashed-line) cells, and VEN- (solid-line) cells (n=1). Single cells were seeded in Y27632 for 24hrs (d=0) before treating for endoderm differentiation for three days. Gene expression is normalized to housekeeping gene *TBP*, and is relative to d0 VEN- control; n=1.

Table S1. Recombineering primers used to generate the pREX1-VF2Pu-TV targeting vector.

Table S2. QRT-PCR primers used in the study

Table S3. *REX1* and *OCT4* primers for amplifying bisulfite converted gDNA for DNA methylation analysis.

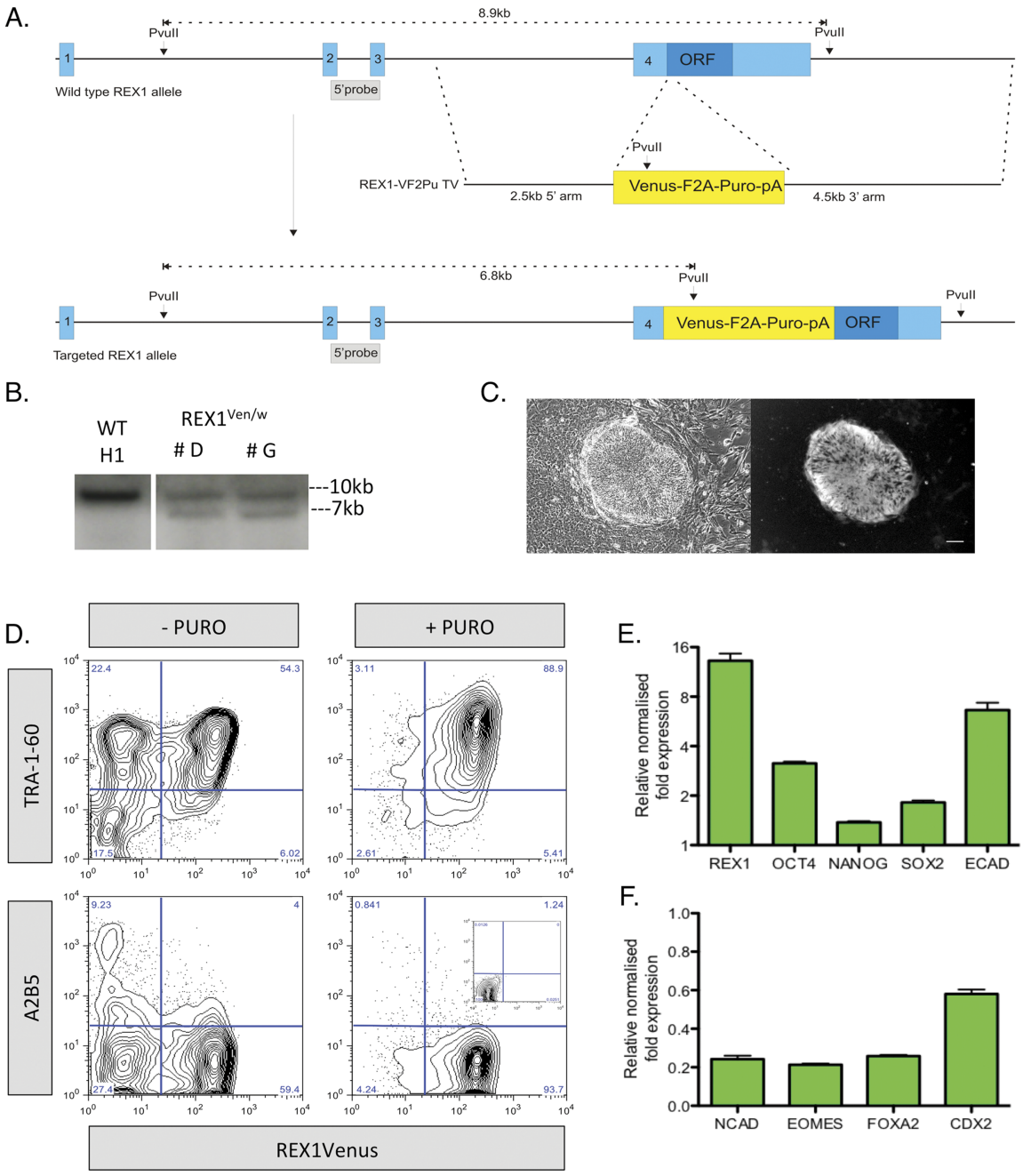

A.
8.9kb
PvuII
PvuII
1
2
3
4
ORF
Wild type REX1 allele
5'probe
PvuII
REX1-VF2Pu TV
2.5kb 5' arm
Venus-F2A-Puro-pA
4.5kb 3' arm
6.8kb
PvuII
PvuII
PvuII
1
2
3
4
Venus-F2A-Puro-pA
ORF
Targeted REX1 allele
5'probe
B.
WT
H1
REX1Ven/w
D
G
---10kb
---7kb
C.
D.
- PURO
+ PURO
TRA-1-60
A2B5
22.4
54.3
17.5
6.02
3.11
88.9
2.61
5.41
9.23
4
27.4
59.4
0.841
1.24
4.24
93.7
REX1Venus
E.
Relative normalised fold expression
16
8
4
2
1
REX1
OCT4
NANOG
SOX2
ECAD
F.
Relative normalised fold expression
1.0
0.8
0.6
0.4
0.2
0.0
NCAD
EOMES
FOXA2
CDX2

Figure 1

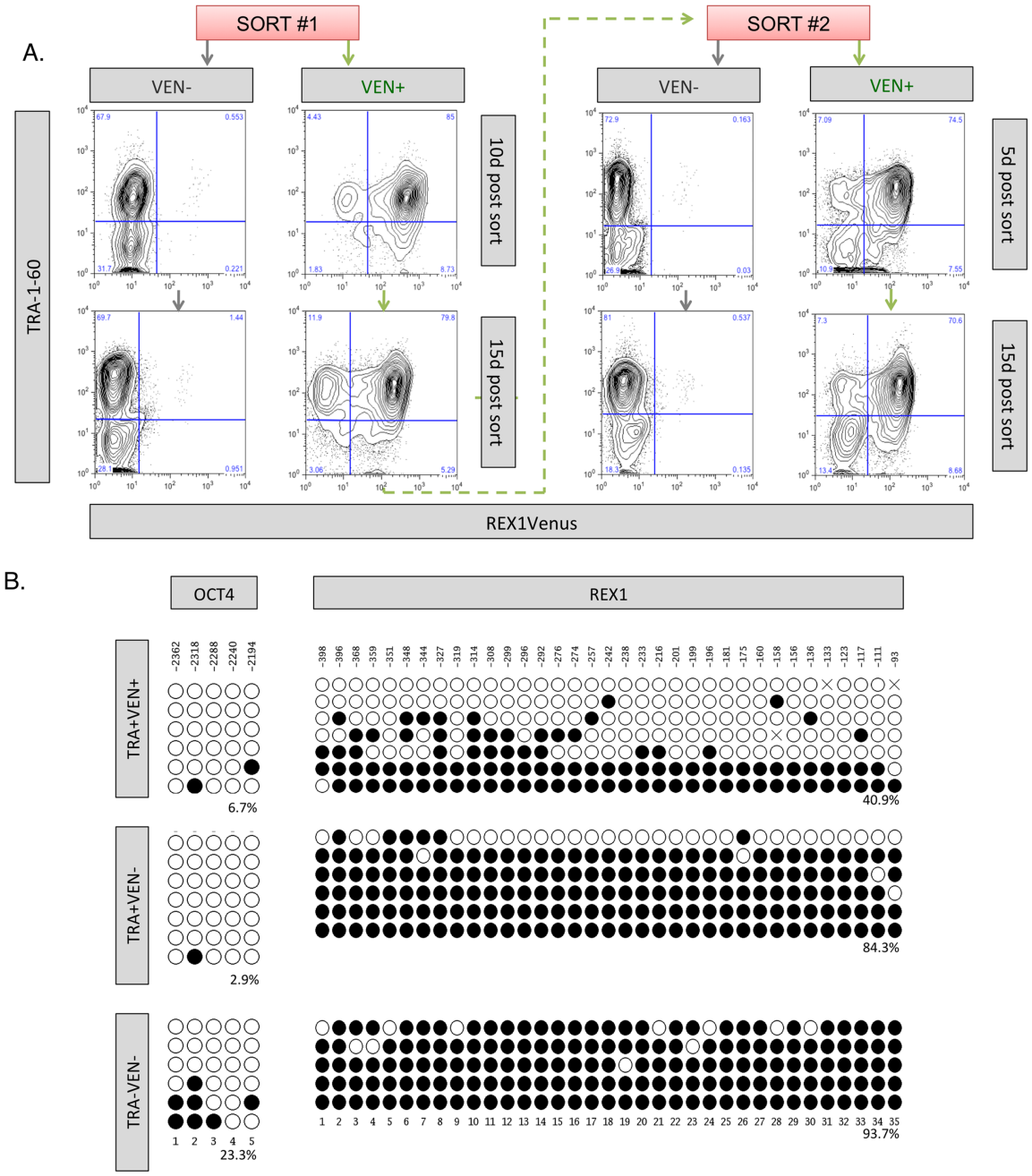

A.
SORT #1
SORT #2
VEN-
VEN+
VEN-
VEN+
TRA-1-60
10d post sort
15d post sort
5d post sort
15d post sort
REX1Venus
B.
OCT4
REX1
TRA+VEN+
TRA+VEN-
TRA-VEN-
6.7%
2.9%
23.3%
40.9%
84.3%
93.7%

Figure 2

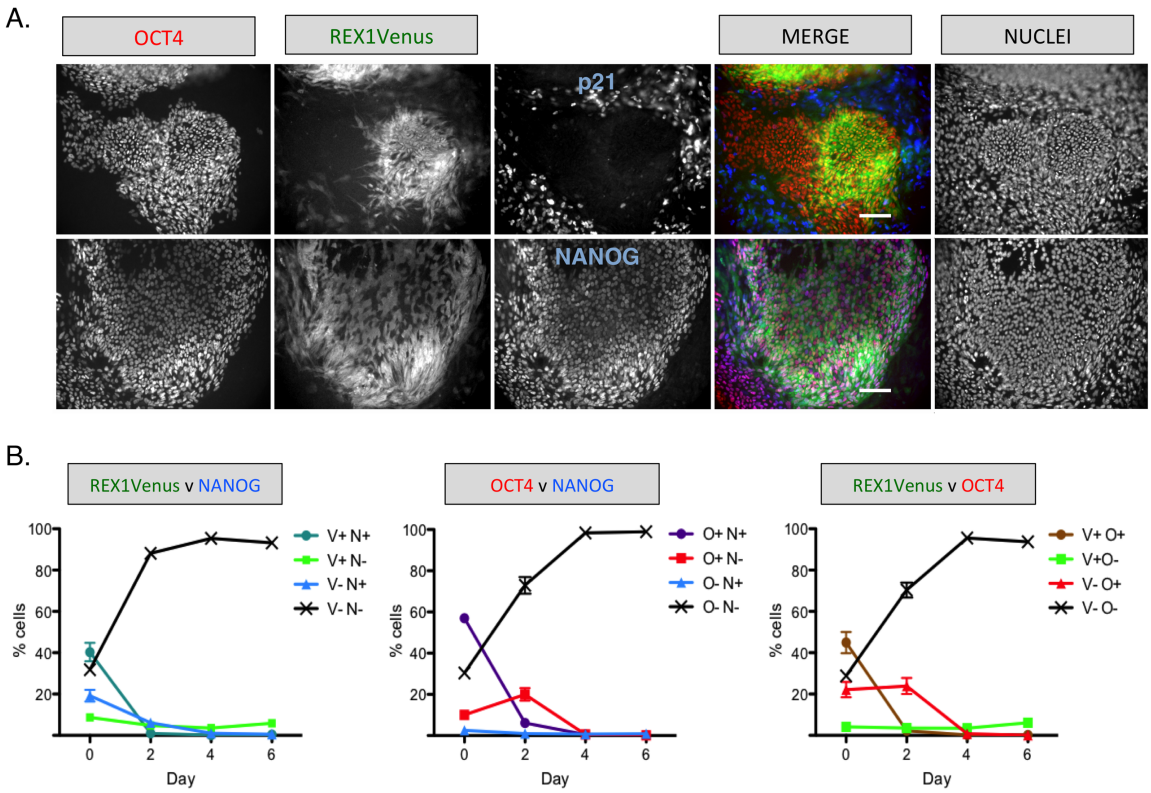

A.
OCT4
REX1Venus
MERGE
NUCLEI
p21
NANOG
B.
REX1Venus v NANOG
OCT4 v NANOG
REX1Venus v OCT4
% cells
Day
V+ N+
V+ N-
V- N+
V- N-
O+ N+
O+ N-
O- N+
O- N-
V+ O+
V+O-
V- O+
V- O-

Figure 3

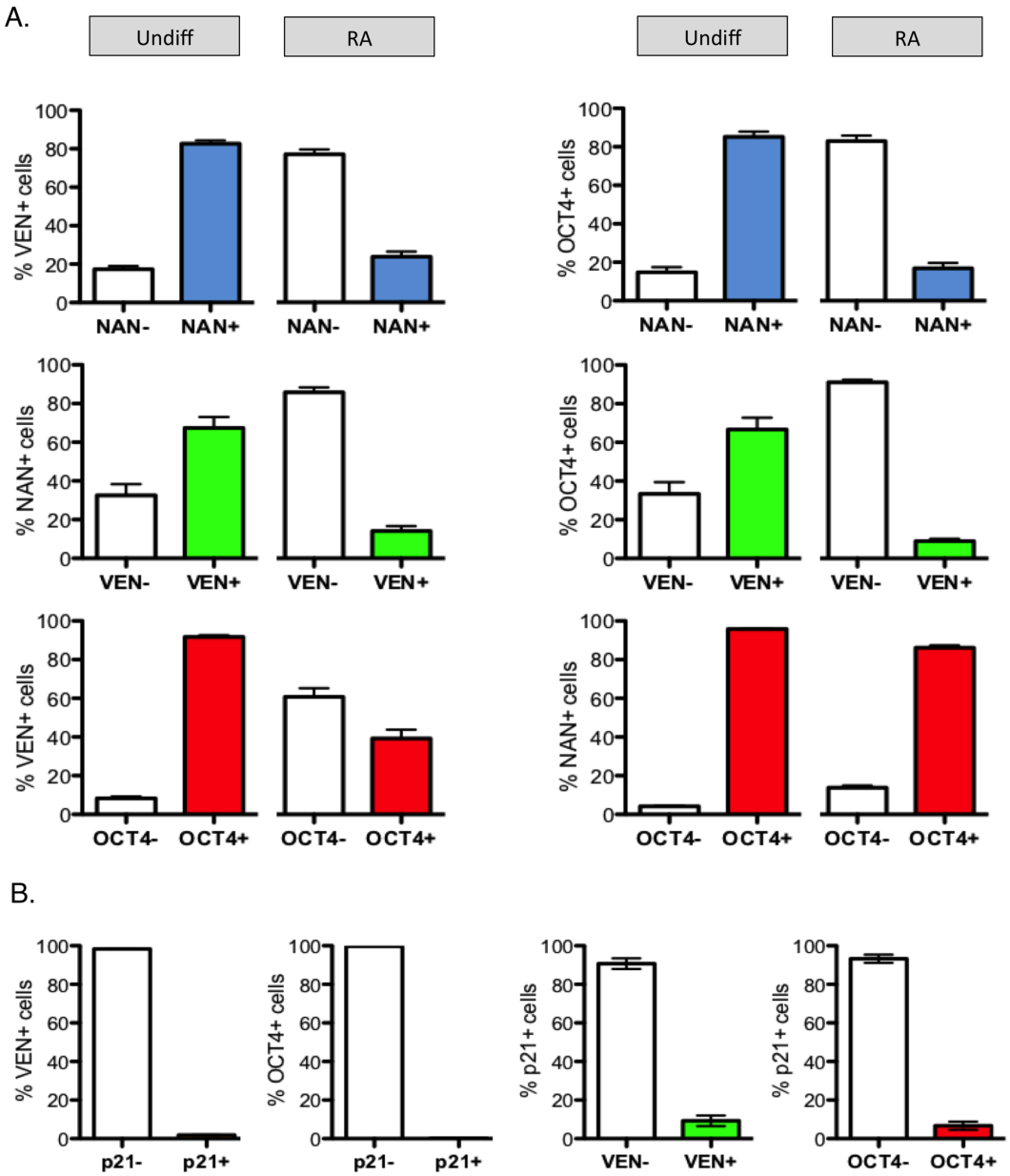
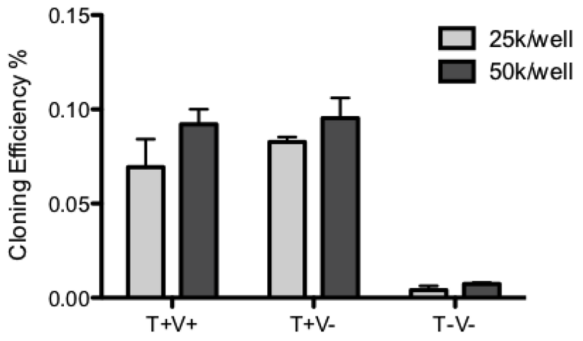

A.
Undiff
RA
Undiff
RA
% VEN+ cells
% OCT4+ cells
% NAN+ cells
% OCT4+ cells
% VEN+ cells
% NAN+ cells
NAN-
NAN+
VEN-
VEN+
OCT4-
OCT4+
B.
% VEN+ cells
% OCT4+ cells
% p21+ cells
% p21+ cells
p21-
p21+
VEN-
VEN+
OCT4-
OCT4+

Figure 4

A.


0.15
0.10
0.05
0.00
Cloning Efficiency %
25k/well
50k/well
T+V+
T+V-
T-V-

B.

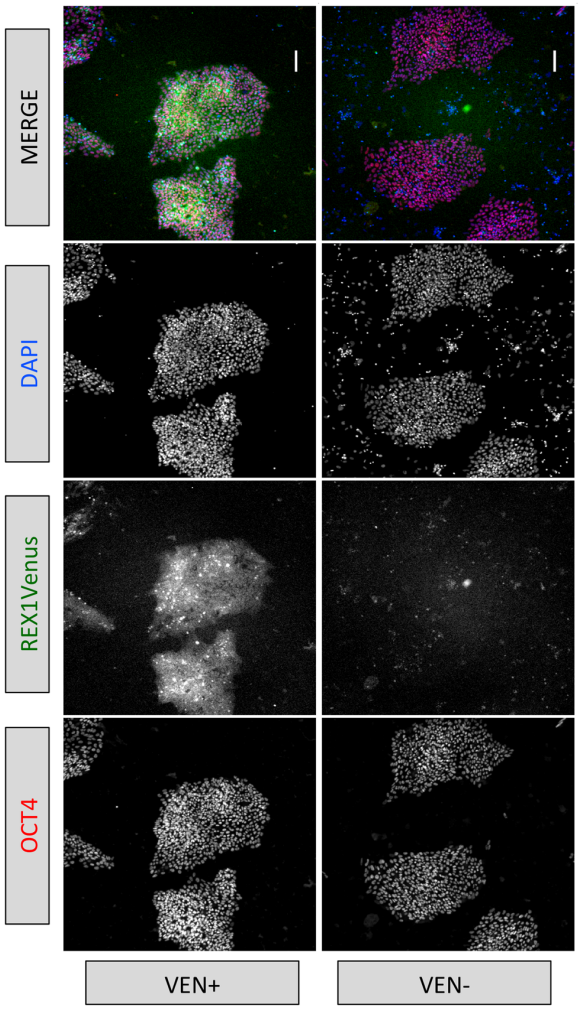

MERGE
DAPI
REX1Venus
OCT4
VEN+
VEN-

Figure 5

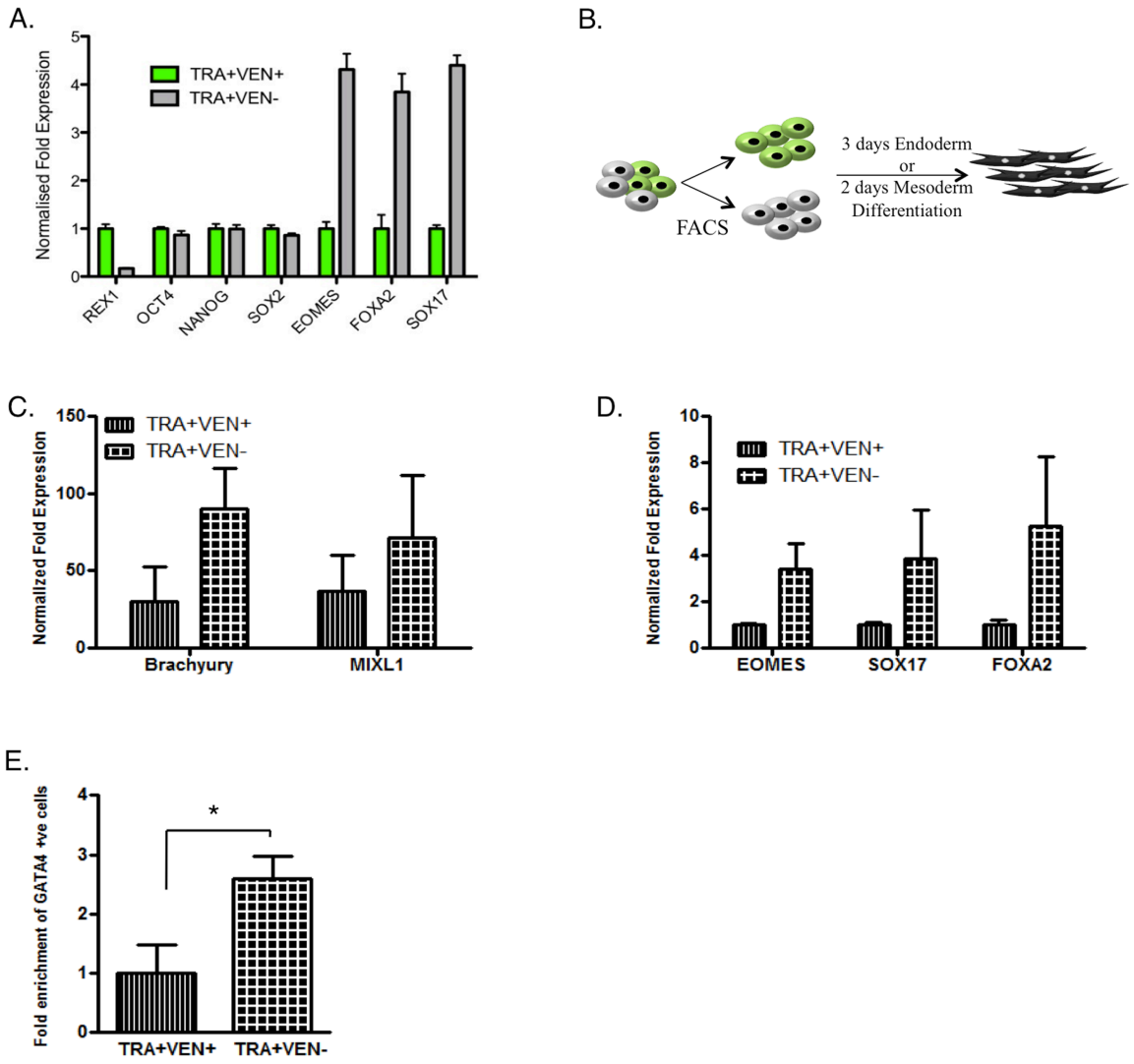
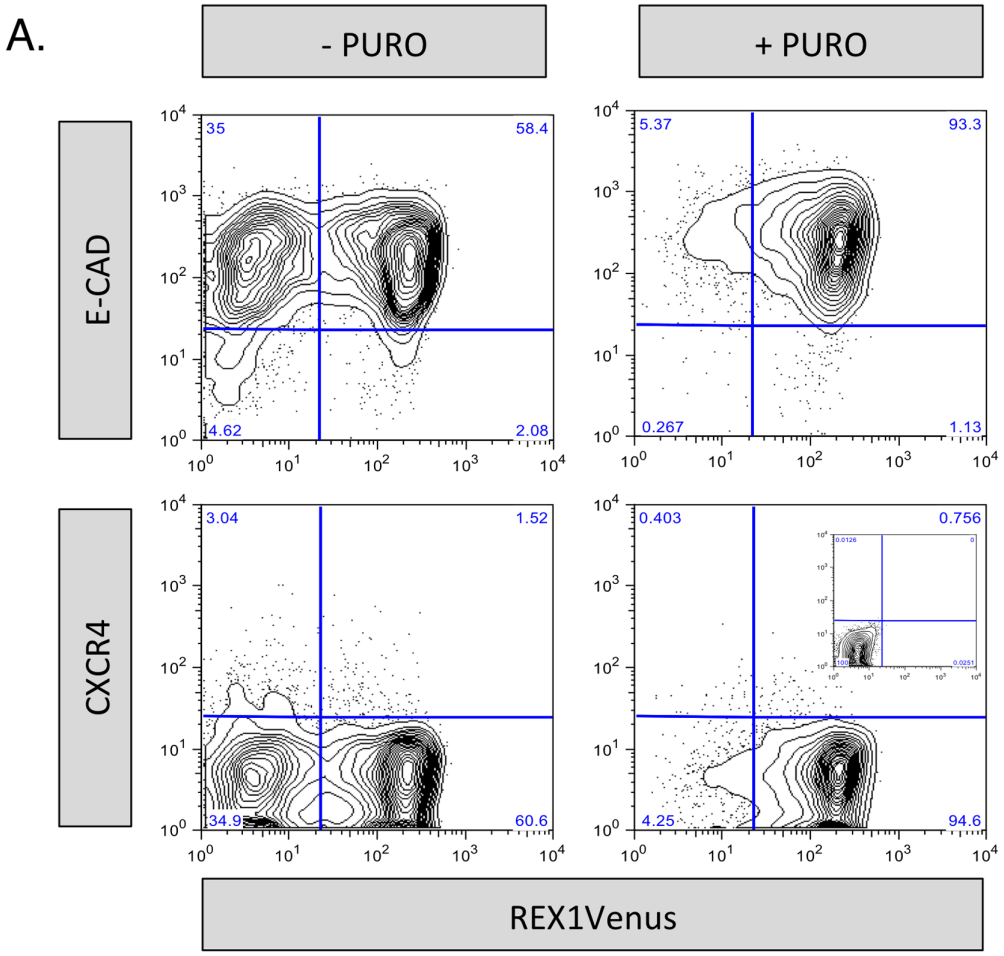
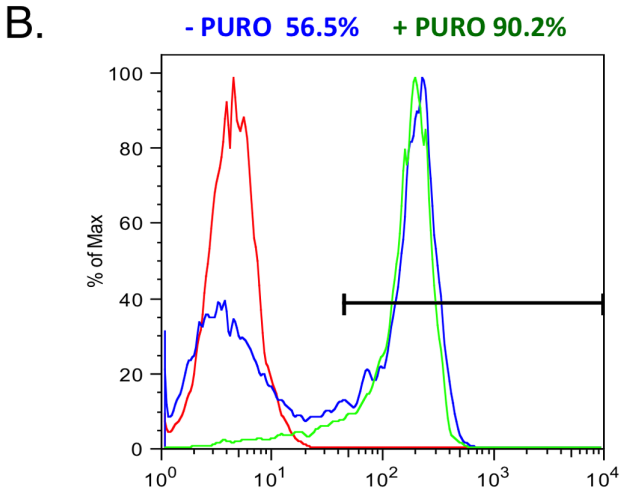

A.
Normalised Fold Expression
TRA+VEN+
TRA+VEN-
REX1
OCT4
NANOG
SOX2
EOMES
FOXA2
SOX17
B.
FACS
3 days Endoderm
or
2 days Mesoderm
Differentiation
C.
Normallzed Fold Expression
TRA+VEN+
TRA+VEN-
Brachyury
MIXL1
D.
Normallzed Fold Expression
TRA+VEN+
TRA+VEN-
EOMES
SOX17
FOXA2
E.
Fold enrichment of GATA4 +ve cells
*
TRA+VEN+
TRA+VEN-

Figure 6

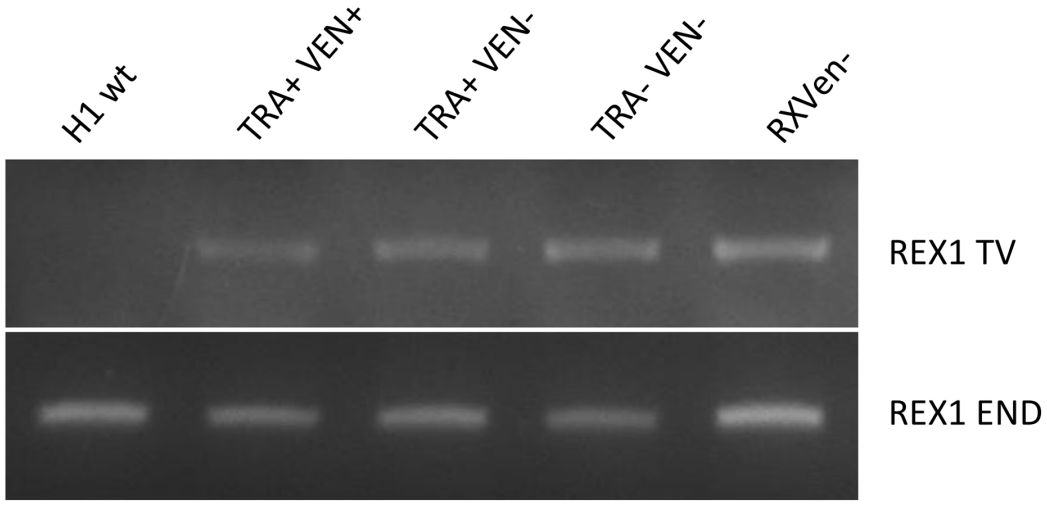
Draper. Supplemental Figure 1. Top

Draper. Supplemental Figure 2. Top

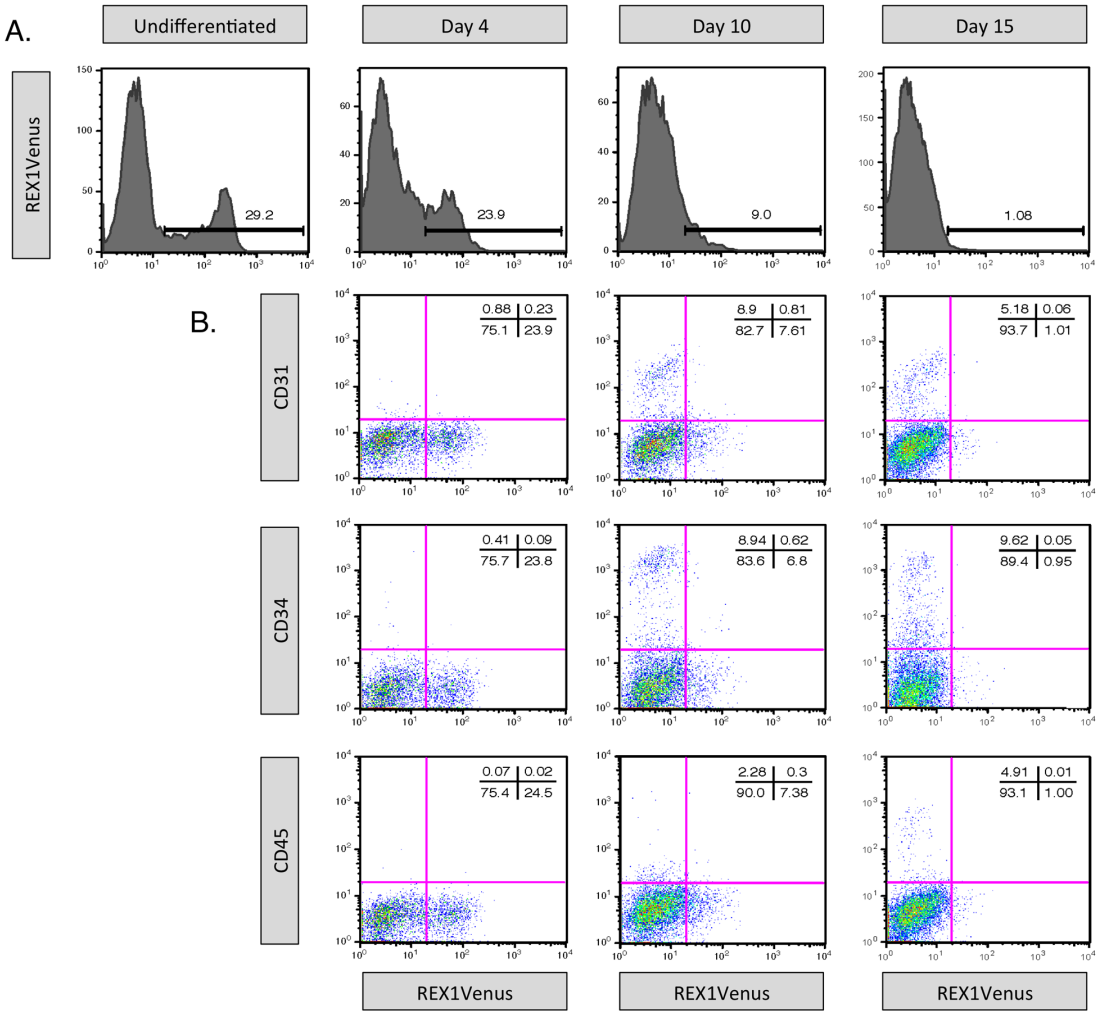




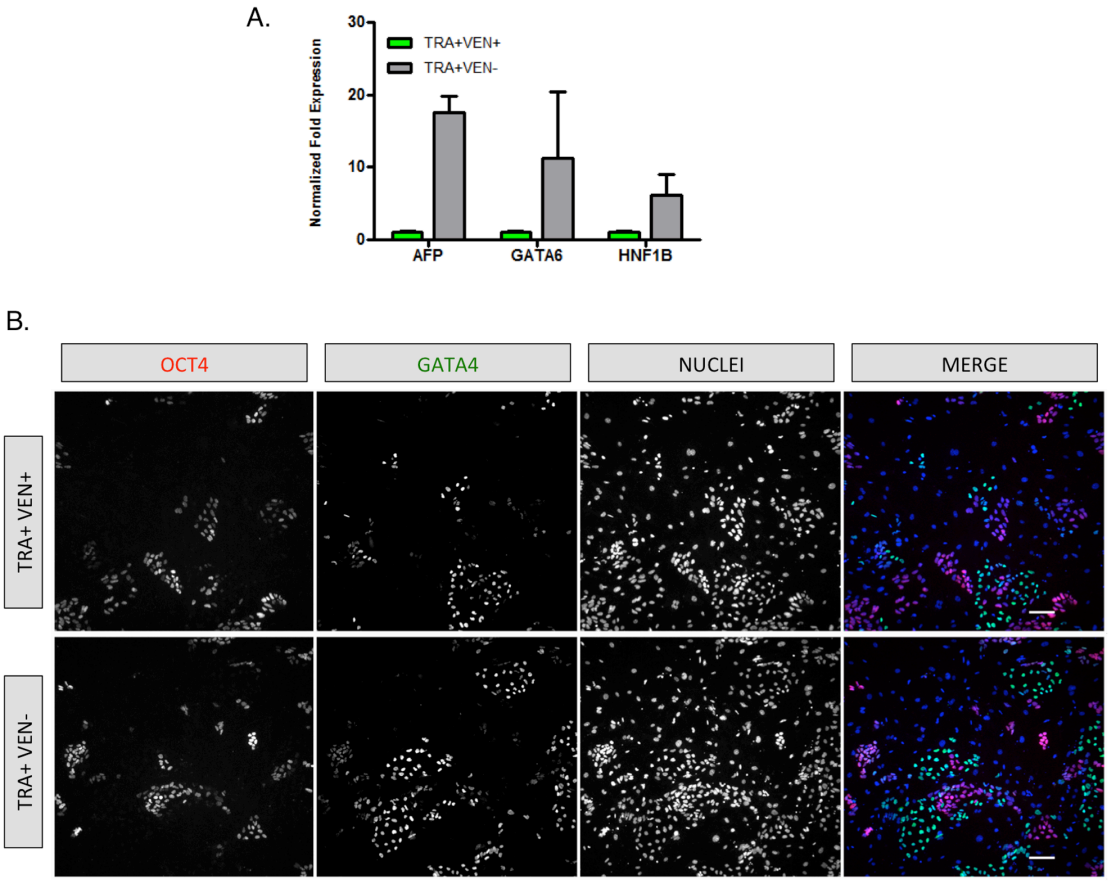
Draper. Supplemental Figure 3. Top



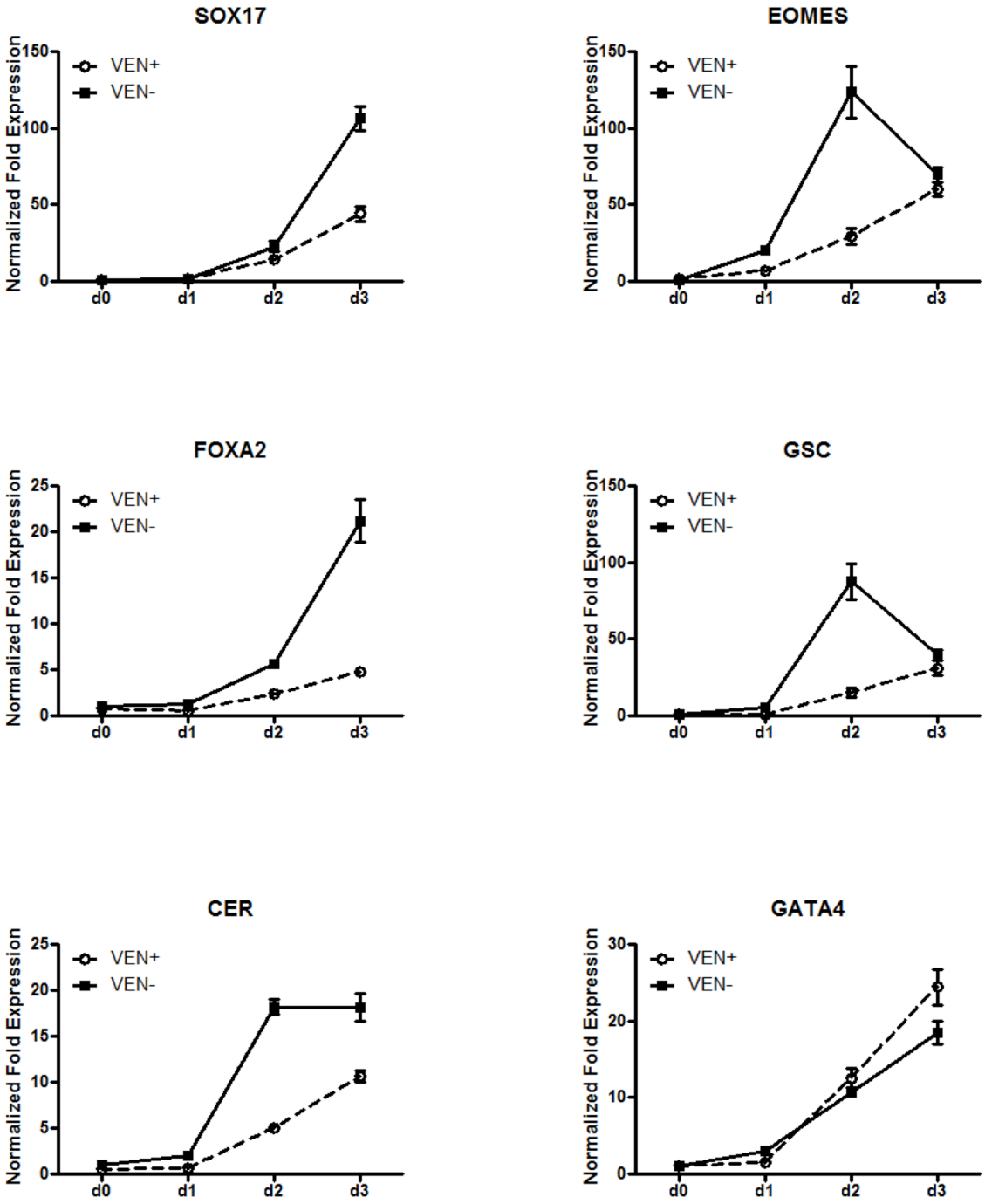
Draper. Supplemental Figure 4. Top



Draper. Supplemental Figure 5. Top



Draper. Supplemental Figure 6. Top



Supplemental Table 1, 2 & 3. Primers used in this study.

Table 1. Recombineering primers

Name	Sequence 5'-3'
REXVenus-F	GGTTGATATATCCTGGTGTAAACCTTCAAGAAGGGCACAGGCAGGAAAACATGGTGAGCAAGGGCGAGGAG
REXVenus-pL451-R	CACTGGGGGCTCTTCCACCCAGGCCTTTCTGGTGTCTTGTCTTTGCCCGTCCGCTCTAGAACTAGTGGAT

Table 2. Q-RT-PCR primers

Gene	Forward primer 5'-3'	Reverse primer 5'-3'
REX1	GCCTTCACTCTAGTAGTGCTCACAGT	GGCAGTAGTGATCTGAGTAAGCTGTCT
OCT4	TGGGCTCGAGAAGGATGTG	GCATAGTCGCTGCTTGATCG
NANOG	TGATTTGTGGGCCTGAAGAAA	GAGGCATCTCAGCAGAAGACA
SOX2	TACAGCATGTCCTACTCGCAG	GAGGAAGAGGTAACCACAGGG
ECAD	AGGAATTCTTGCTTTGCTAATTCTG	CGAAGAAACAGCAAGAGCAGC
NCAD	CCCACACCCTGGAGACATTG	GCCGCTTTAAGGCCCTCA
EOMES	CGGCCTCTGTGGCTCAAA	AAGGAAACATGCGCCTGC
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
CDX2	CTGGAGCTGGAGAAGGAGTTTC	ATTTTAACCTGCCTCTCAGAGAGC
SOX17	GGCGCAGCAGAATCCAGA	CCACGACTTGCCCAGCAT
BRACHYURY	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG
MIXL1	AAGCCCCAGCTGCCTGTT	CCCTCCAACCCCGTTTG
AFP	TGGGACCCGAACTTTCCA	GGCCACATCCAGGACTAGTTTC
HNF1B	TCACAGATACCAGCAGCATCAGT	GGGCATCCCAGGCTTGTA
GATA6	GCGGGCTCTACAGCAAGATG	ACAGTTGGCACAGGACAATCC
GATA4	TCCAAACCAGAAAACGGAAGC	GCCCGTAGTGAGATGACAGG
CER	ACAGTGCCCTTCAGCCAGACT	ACAACTACTTTTTCACAGCCTTCGT
GSC	GAGGAGAAAGTGGAGGTCTGGTT	CTCTGATGAGGACCGCTTCTG

Table 3. Bisulphite DNA methylation sequencing primers

Gene	Forward primer 5'-3'	Reverse primer 5'-3'
REX1	GGTTTAAAAGGGTAAATGTGATTATATTTA	CAAACTACAACCACCCATCAAC
OCT4	ATTTGTTTTTTGGGTAGTTAAAGGT	CCAACTATCTTCATCTTAATAACATCC

Chapter 1.3: Conclusion and Introduction to Section 2

In conclusion, we showed that the TRA-1-60 positive pluripotent compartment in hESCs is divided by the expression of REX1. The TRA+VEN+ and TRA+VEN- pluripotent subpopulations are phenotypically distinct, with the TRA+VEN- cells showing an increased expression of several differentiation genes, suggesting that this subpopulation is primed for differentiation. Indeed, the TRA+VEN- hESCs show more rapid differentiation kinetics compared to the TRA+VEN+ hESCs.

The correlation of loss of REX1 expression in hESCs with a differentiation-primed phenotype suggests that REX1 plays an important role in regulating pluripotency. In mESCs Rex1 is dispensable for maintaining pluripotency, with cells no longer expressing Rex1 still being pluripotent in their capability to contribute to chimeras[23]. However, Rex1 depletion results in some developmental phenotypic defects, and aberrant gene expression profiles[24,25]. It is involved in the regulation of X-inactivation[26] and maintenance of genomic imprinting[27] in the mouse system. Additionally, Rex1 might be associated with various components of the chromatin remodeling machinery[28,29], and the interacts with the epigenetic regulator, lysine demethylase, LSD1[30,31]. Some evidence suggest that in early mouse development Rex1 is involved in the repression of endogenous retroviral gene expression[30].

To explore the role of REX1 in hESCs and pluripotency, I generated hESCs expressing various REX1 fusion constructs. REX1 fusion constructs were made by fusing 3XFLAG to both the C- and N-terminus of REX1, and by fusing the fluorescent protein mKO2 (monomeric Kusabira Orange 2) to the C-terminus. To further explore the role of REX1 protein we generated various 3X FLAG tagged truncation mutants based on

evolutionary sequence conservation (data not shown). Upon analysis of different truncation mutants alongside C- and N-terminally tagged REX1, I observed and later confirmed the existence of three previously unknown translation isoforms of the human REX1 protein. These isoforms were a result of translation products from down-stream start codons in the REX1 gene transcript (Data not shown). Currently, there is no evidence of REX1 isoforms in either human or murine systems, and this observation, therefore, presented us with an opportunity to explore the role of REX1 isoforms in pluripotency. We generated similar fusion constructs in the mouse system to test for the presence and evolutionary conservation of these translation isoforms. The project was passed onto the M.Sc. student, Amanda Hrenzuck, in the Draper lab.

In addition to identifying REX1 translation isoforms, we also observed that the REX1-mKO2 fusion construct was tightly associated with the mitotic chromatin. This observation is previously unreported and is not a characteristic of other pluripotency associated transcription factors, such as OCT4, NANOG and SOX2, as observed by immunostaining. This prompted us to explore the possible explanation for this association, and introduced us to the concept of mitotic bookmarking. During mitosis of the cell cycle, several transcription factors decouple from their DNA binding sites to allow for faithful cell division; a few of the transcription factors, however, remain bound to selected target genes. This phenomenon is referred to as mitotic bookmarking[32-35]. Mitotic bookmarking is a relatively new concept, and is thought to play an important role in protecting cellular identity during cell division[33-35]. At the time, there were no reports of mitotic bookmarking as a mechanism for fate maintenance in ES cells, and therefore we decided to conduct an in-depth analysis on this topic of research. The next section

(Section 2) is a detailed study aimed at characterizing the role of mitotic bookmarking in pluripotent cells.

References (for Chapters 1.1 and 1.3):

1. Thompson, J., Itskovitz, J., Shapiro, S.S., et al. Embryonic stem cell lines derived from human blastocysts. - *Science.1998 Nov 6;282(5391):1145-7.*
2. Evans, M., & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. - *Nature.1981 Jul 9;292(5819):154-6.*
3. Hoffman, L.M, & Carpenter, M.K. Characterization and culture of human embryonic stem cells. - *Nat Biotechnol.2005 Jun;23(6):699-708.*
4. Takahashi, K., Tanabe, K., Ohunki, M., et al. - Induction of pluripotent stem cells from adult human fibroblasts by defined factors. - *Cell.2007 Nov 30;131(5):861-72.*
5. Takahashi, K., & Yamanaka, S. - Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. - *Cell.2006 Aug 25;126(4):663-76.Epub 2006 Aug 10.*
6. Boyer, L.A., Lee, T., Cole, M.F., et al. - Core transcriptional regulatory circuitry in human embryonic stem cells. - *Cell.2005 Sep 23;122(6):947-56.*
7. Kim, J., Chu, J., Shen, X., Wang, J., & Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. - *Cell.2008 Mar 21;132(6):1049-61.doi: 10.1016/j.cell.2008.02.039.*
8. Wang, J., Rao, S., Chu, J., et al. - A protein interaction network for pluripotency of embryonic stem cells. - *Nature.2006 Nov 16;444(7117):364-8.Epub 2006 Nov 8.*
9. Tesar, P.J., Chenoweth, J.G., Brook, F.A., et al. - New cell lines from mouse epiblast share defining features with human embryonic stem cells. - *Nature.2007 Jul 12;448(7150):196-9.Epub 2007 Jun 27.*
10. Han, D.W., Tapia, N., Joo, J.Y., et al. - Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. - *Cell.2010 Nov 12;143(4):617-27.doi: 10.1016/j.cell.2010.10.015.*
11. Dick, E., Matsa, E., Bispham, J., et al. - Two new protocols to enhance the production and isolation of human induced pluripotent stem cell lines. - *Stem Cell Res.2011 Mar;6(2):158-67.doi: 10.1016/j.scr.2010.10.002.Epub 2010 Nov 20.*
12. Chambers, I., Silva, J., Colby, D., et al. - Nanog safeguards pluripotency and mediates germline development. - *Nature.2007 Dec 20;450(7173):1230-4.*
13. Hayashi, K., Lopes, S.M., Tang, F., Surani A.M. - Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. - *Cell Stem Cell.2008 Oct 9;3(4):391-401.doi: 10.1016/j.stem.2008.07.027.*
14. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., Niwa, H. - Identification and characterization of subpopulations in undifferentiated ES cell culture. - *Development.2008 Mar;135(5):909-18.doi: 10.1242/dev.017400.*
15. Furusawa, T., Ohkoshi, K., Honda, C., Takahashi, S., Tonkunaga, T. - Embryonic stem cells expressing both platelet endothelial cell adhesion molecule-1 and stage-specific embryonic antigen-1 differentiate predominantly into epiblast cells in a chimeric embryo. - *Biol Reprod.2004 May;70(5):1452-7.Epub 2004 Jan 21.*
16. Furusawa, T., Ikeda, M., Inoue, F., Ohkoshi, K, Hamano, T., Tokunaga, T. - Gene expression profiling of mouse embryonic stem cell subpopulations. - *Biol Reprod.2006 Oct;75(4):555-61.Epub 2006 May 10.*

17. Chambers, I., Colby, D., Robertson, M., et al. - Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. - *Cell.2003 May 30;113(5):643-55.*
18. Singh, A.M., Hamazaki, T., Hankowski, K., Terada, N. - A heterogeneous expression pattern for nanog in embryonic stem cells. - *Stem Cells.2007 Oct;25(10):2534-42.Epub 2007 Jul 5.*
19. Sato, M., Kimura, T., Kurokawa, K., et al. - Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. - *Mech Dev.2002 Apr;113(1):91-4.*
20. Casanova, J. - Stemness as a cell default state. - *EMBO Rep.2012 May 1;13(5):396-7.doi: 10.1038/embor.2012.47.*
21. Wray, J, Kalkan T., Smith, A. - The ground state of pluripotency. - *Biochem Soc Trans.2010 Aug;38(4):1027-32.doi: 10.1042/BST0381027.*
22. Bao, S., Tang, F., Li, X., et al. - Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. - *Nature.2009 Oct 29;461(7268):1292-5.doi: 10.1038/nature08534.*
23. Masui S, Ohtsuka S, Yagi R, Takahashi K, Ko MS, et al. (2008) Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. BMC developmental biology 8: 45.
24. Rezende NC, Lee MY, Monette S, Mark W, Lu A, et al. (2011) Rex1 (Zfp42) null mice show impaired testicular function, abnormal testis morphology, and aberrant gene expression. Developmental biology 356: 370-382.
25. Scotland KB, Chen S, Sylvester R, Gudas LJ (2009) Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. Developmental dynamics : an official publication of the American Association of Anatomists 238: 1863-1877
26. Navarro P, Oldfield A, Legoupi J, Festuccia N, Dubois A, et al. (2010) Molecular coupling of Tsix regulation and pluripotency. Nature 468: 457-460.
27. Kim JD, Kim H, Ekram MB, Yu S, Faulk C, et al. (2011) Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. Human molecular genetics 20: 1353-1362.
28. Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH. A protein interaction network for pluripotency of embryonic stem cells. Nature 2006 Nov 16;444(7117):364-8.
29. Ho L, Ronan JL, Wu J, Staahl BT, Chen L, Kuo A, Lessard J, Nesvizhskii AI, Ranish J, Crabtree GR. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A 2009 Mar 31;106(13):5181-6.
30. Guallar D, Perez-Palacios R, Climent M, Martinez-Abadia I, Larraga A, Fernandez-Juan M, Vallejo C, Muniesa P, Schoorlemmer J. Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. Nucleic Acids Res 2012 Oct;40(18):8993-9007.31.
31. Garcia-Tunon I, Guallar D, Alonso-Martin S, Benito AA, Benitez-Lazaro A, Perez-Palacios R, Muniesa P, Climent M, Sanchez M, Vidal M, et al. Association of rex-1 to target genes supports its interaction with polycomb function. Stem Cell Res 2011 Jul;7(1):1-16.
32. John, S., & Workman, J. L. (1998). Bookmarking genes for activation in condensed mitotic chromosomes. *Bioessays, 20*(4), 275-279. doi: 10.1002/(sici)1521-1878(199804)20:4<275::aid-bies1>3.0.co;2-p

33. Zaidi SK, Young DW, Montecino M, Lian JB, Stein JL, van Wijnen AJ, Stein GS. Architectural epigenetics: Mitotic retention of mammalian transcriptional regulatory information. Mol Cell Biol 2010 Oct;30(20):4758-66.
34. Zaret, K. S. (2014). Genome reactivation after the silence in mitosis: recapitulating mechanisms of development? *Dev Cell, 29*(2), 132-134. doi: 10.1016/j.devcel.2014.04.019
35. Kadauke S, Udugama MI, Pawlicki JM, Achtman JC, Jain DP, Cheng Y, Hardison RC, Blobel GA. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. Cell 2012 Aug 17;150(4):725-37.

SECTION 2: Characterizing mitotic bookmarking as a mechanism for fate maintenance in pluripotent stem cells

SECTION 2: Characterizing mitotic bookmarking as a mechanism for fate maintenance in pluripotent stem cells

Preface

While I conceived, designed and primarily executed this study with input from Dr. Jonathan Draper, several people contributed towards various portions of the study. I entirely designed, planned, executed and analyzed the experiments outlined in Chapter 2.2. Mehdi Hamzeh assisted with quantitative image analysis. We collaborated with Dr. Aki Minoda (RIKEN, Japan) for ATAC-seq profiling in Chapter 2.3. Dr. Minoda's graduate student Ye Liu performed the ATAC-seq for us. I performed the bioinformatics analysis with help from Drs. Jen-Chien Chang and Chung-Chau Hon. Daisy Deng (Dr. Draper's lab) generated Parp1 knock out lines, performed some characterization assays on these lines and assisted in the collection of material for mitotic release experiments.

This section comprises four subchapters: Chapter 2.1 is an introductory chapter providing a background on cell cycle and the changes that accompany different phases of the cell cycle, focusing greatly on mitosis. Here, I also introduce the concept of mitotic bookmarking and the research goals outlined for this study. Chapters 2.2, and 2.3 are research chapters focusing on the outcomes of the study, and Chapter 2.4 is an overview of our contribution to the field. The references for the entire section are placed at the end.

Chapter 2.1: Introduction

2.1.1 The Cell cycle

The cell cycle is an essential characteristic of proliferating cells aimed at transmitting genetic information from parent to daughter cells. A eukaryotic cell cycle is composed of four main phases: G1, S, G2 and M. G1 is the first gap phase where the decision and preparation to undergo replication is made, S is the synthesis phase where DNA undergoes replication, G2 is the second gap phase where DNA repair occurs and the cell prepares for division, and M is mitosis where cell division occurs. G1, S and G2 phase together comprise the interphase of the cell cycle. Transition between different cell cycle phases is carefully orchestrated by expression and phosphorylation kinetics of different cyclins and cyclin dependent kinases (Cdks) (Amon et al., 1993; Nigg, 2001). In differentiated somatic cells, the fluctuating expression of various cyclins governs cell cycle progression. G1 phase is typically marked by increased Cdk6-Cyclin D3 activity (Quelle et al., 1993; Resnitzky et al., 1994), S phase is associated with high Cyclin E activity (Coverley et al., 2002; Furstenthal et al., 2001; Ohtsubo et al., 1995), G2 is associated with increased Cdk2-cyclin A activity (Geley et al., 2001), and M-phase is associated with Cdk1-cyclin B activity (White and Dalton, 2005). The timing and levels of the various cyclins is critical for normal cell cycle progression, and dysregulation results in an impaired cell cycle, typically altering the length of different phases (Ohtsubo et al., 1995; Quelle et al., 1993; Resnitzky et al., 1994).

Pluripotent Cell Cycle

Pluripotent stem cells exhibit an altered cell cycle profile compared to their differentiated counterparts (Ballabeni et al., 2011; Calder et al., 2013; Stead et al., 2002).

The generation time of pluripotent mouse embryonic stem (ES) cells is about 10 hours compared to >16 hours in differentiated cells (Stead et al., 2002). A shortened cell cycle is a common characteristic of early development. Determination of cell numbers during 4.5dpc to 7.5dpc in a mouse gastrulating embryo showed that cells undergo very rapid rates of division during this time period, with the generation time of ~9hrs from 4.5-6dpc, and ~5hrs from 6.5-7.5dpc (Snow, 1977). Rapid cell divisions are also observed during early embryo development in rats (Mac Auley et al., 1993), zebra fish (Yarden and Geiger, 1996), and drosophila (Edgar and Lehner, 1996). The rapid cell cycle found in early development is often the result of lack of or shortening of G1 and G2 gap phases of the cell cycle.

The difference in cell cycle profiles between ES and differentiated cell types is, in part, regulated by differences in cyclin and Cdk expression and activity. With the exception of mitotic-specific Cdk1-cyclin B, other Cdk-cyclin complexes do not show the characteristic fluctuation during cell cycle progression in ES cells (Stead et al., 2002; White and Dalton, 2005). Additionally, ES cells exhibit atypically high Cdk activity throughout the cell cycle (Stead et al., 2002). In somatic cells Cdk activity is markedly reduced in G1, and this reduction is needed for the establishment of pre-replicative complexes (pre-RCs) at the origins of replication, a requirement for progression into S-phase (Maiorano et al., 2000; Prasanth et al., 2004). The mechanisms underlying the seamless progression of ES cells into S-phase despite a lack of control in Cdk activity are yet to be understood.

The lack of periodicity in Cdk activity also is thought to result in the inactive nature of the pRb-E2F (Retinnoblastoma protein and E2F transcription factors) pathway

in ES cells (Savatier et al., 1994; Stead et al., 2002; White and Dalton, 2005). In somatic cells, G1 phase is normally divided into early and late G1, differentiated by the phosphorylation status of pRb (Savatier et al., 1994). Active pRb is hypophorylated in early G1 and is deactivated in late G1 by phosphorylation. Active pRb interacts with E2F transcription factors, which results in cell cycle dependent gene expression of various cell cycle related genes. In ES cells, pRb is hyperphosphorylated throughout G1, and does not interact with E2F transcription factors, resulting in no cell cycle dependent gene expression of E2F targets (Stead et al., 2002).

Currently, majority of the differences in cell cycle profiles of ES cells are attributed to the G1 phase of the cell cycle. The unique characteristics of the pluripotent cell cycle underlie the ability of ES cells to self-renew indefinitely, while maintaining their ability to differentiate into all three germ layers, the latter two properties providing ES cells with their immense potential for regenerative medicine applications. However, our ability to completely exploit the clinical potential of these cells relies two factors: 1.) the ability to maintain the integrity of an ES cell as a pluripotent cell in culture, and 2.) being able to efficiently alter the fate of ES cells to the desired cell type. Both of these require an in-depth understanding of the intrinsic processes that control cellular identity. Several decades of investigation have revealed the cell cycle to function akin to a multi-tiered decision making process that presents multiple windows of opportunity to decide the fate of the daughter cells. Below, I will discuss the dynamic nature of various phases of the cell cycle, and the processes that come into play to ensure a faithful division to retain cell fate.

2.1.2 Chromatin dynamics throughout the cell cycle

Cell cycle and chromatin structure are highly interlinked. Chromatin structure and organization changes considerably during different phases of the cell cycle, while at the same time gene expression of cell cycle regulators is governed by the underlying chromatin structure. Chromatin structure, composition and organization constitute a cell's epigenetic memory. Epigenetic memory is defined as a mechanism that does not alter the DNA, but governs stable inheritance of a phenotype through mitosis or meiosis (Berger et al., 2009; Waddington, 1953). The epigenetic features of a cell govern its gene and, ultimately, protein expression profile, which is defined as the cell fate or identity.

Chromatin is the nucleoprotein complex that the genome is organized into, and that underlies the chromosome structure (Kornberg and Thomas, 1974). The repeating unit of chromatin is the nucleosome, which comprises of four core histone proteins, H2A, H2B, H3 and H4, which are assembled as pairs into an octamer wrapped by 146bp of DNA (Kornberg and Thomas, 1974; Luger et al., 1997). Chromatin can be modified covalently in a variety of ways leading to multiple variations of chromatin architecture, resulting in a plethora of possible phenotypes for cells with the same underlying genetic information. DNA is commonly modified by methylation of the cytosine residues around CpG islands by *de novo* methyl transferases, and nucleosomes can be modified either at histone tails or the core histone structure by a variety of histone modifying enzymes (Fig 1, **Table 1**). These are categorized as histone mark readers, writers and erasers. Readers will recognize an epigenetic mark, and either maintain it or recruit other catalytic enzymes, the writers or erasers, to alter it. Chromatin remodelers are recruited to sites of activity by an exogenous cell signal, and subsequent interactions with cell type specific transcription factors (Berger et al., 2009; Voss and Hager, 2014).

Chromatin structure is highly dynamic and varies throughout the cell cycle. Despite this, the epigenetic marks deposited onto the chromatin of are faithfully transmitted to the daughter cells in absence of external signals maintaining the cells identity, but the mechanisms underlying the inheritance of these marks are not fully understood. Notwithstanding, two key phases of the cell cycle when active chromatin remodeling occurs are the S-phase and mitosis:

Chromatin remodeling during S-phase: The conservation of epigenetic memory during S-phase is an ongoing topic of study. During S-phase the DNA is replicated to transmit the parental genetic information to the progeny. Newly synthesized histone proteins are deposited simultaneously to the replicated DNA strand (Franco and Kaufman, 2004; Jackson, 1987; Worcel et al., 1978). These nascent histones, however, are devoid of any epigenetic marks, and studies have shown that the deposition of epigenetic marks occurs based on a lateral spreading of chromatin state (Reviewed in (Richards and Elgin, 2002). During DNA strand synthesis, the parental histones along with their covalent modifications, dissociate into dimers of H2A/H2B and tetramers of $(H3/H4)_2$, which then randomly associate with nascent histones, resulting in nucleosomes containing a hybrid of old and new histones (Gruss et al., 1993; Jackson, 1987; Ramachandran and Henikoff, 2015). Epigenetic memory on nascent histones is deposited by chromatin modification complexes based on the epigenetic marks of the parental histones. Similar to parental histone modifications, DNA methylation patterns are deposited to the newly synthesized DNA strand by DNA methyltransferase 1 (Dnmt1), preserving these epigenetic marks through S-phase (Leonhardt et al., 1992; Zhang et al., 2011).

Chromatin remodeling during mitosis: The second phase of the cell cycle where active chromatin remodeling occurs is mitosis. At the onset of mitosis, chromatin undergoes compaction resulting in a visible condensation of the chromosomes (Fig 2.). Several models have been put forth to provide mechanistic insights into metaphase chromatin compaction (Reviewed in Olins and Olins, 2003; Piskadlo and Oliveira, 2016). One of the earliest views proposed was of a random folding of chromatin fiber (DuPraw, 1966). However, this was disputed by findings indicating some order to the chromatin organization. Metaphase chromosomes when stained with specific dyes, such as Giemsa, showed a characteristic banding-pattern; chromosomes always folded into the same lengths; and specific DNA sequences occupied the same regions when tested by *in-situ* hybridizations (Baumgartner et al., 1991; Reviewed in Piskadlo and Oliveira, 2016). These data suggested some reproducible order to chromatin condensation.

One of the widely accepted patterns of chromatin folding is the hierarchical helical-coiling model (Sedat and Manuelidis, 1978). According to this model, each nucleosome is connected to the neighboring nucleosome particles by a strand of linker DNA forming an 11nm 'beads-on-a-string' array (Olins and Olins, 1974; Sedat and Manuelidis, 1978; Woodcock et al., 1976). These arrays are folded into a 30 nm secondary chromatin structure by association with linker histones, H1 and H5. This model, however, has recently been scrutinized in light of some new data, questioning the existence of the 30nm fiber *in vivo* (Joti et al., 2012; Maeshima et al., 2010). More recent studies suggest the existence of a 6 nm stacked layer model of mitotic chromatin condensation. According to this model, chromatin folds into 6 nm layers that are perpendicular to the chromosome axis and each contain about 1Mb of DNA (Daban,

2015). A variety of the chromatin structural components such as condensing proteins, topoisomerase II, chromatin remodelers, and histone modifications play a major role in chromatin compaction (Piskadlo and Oliveira, 2016).

Condensation is, primarily, triggered by a wave of histone modifications during late G2 and early mitosis. Phosphorylation of serine 10 of histone 3 (H3S10P) is a hallmark of mitosis (Crosio et al., 2002; Juan et al., 1998; Wei et al., 1999). H3S10P is initiated by the recruitment of Aurora kinases to phosphorylated histone H3 threonine 3 (H3T3P) at the pericentromeric regions of the chromosomes during late G2 (Kelly et al., 2010). The phosphorylation of H3S10 subsequently spreads throughout the entire chromosome during mitosis (Crosio et al., 2002; Hendzel et al., 1997). H3S10P recruits histone de-acetylase Hst2p, which removes the acetylation group at lysine 16 histone H4 (Wilkins et al., 2014). This deacetylation event results in an interaction between the histones H4 and H2A/B of neighbouring nucleosomes leading to compaction of the chromatin (Wilkins et al., 2014).

Epigenetic marks during mitosis

In addition to physical compaction of the chromatin, many epigenetic marks have an altered profile during mitosis. Phosphorylation of nucleosomes is ubiquitously increased during mitosis, and is typically associated with faithful cell cycle progression during mitosis and into G1. Other gene regulatory histone marks show a variety of profiles during mitosis compared to interphase (**Table1**), but H3K4me3 and H3K79me2 marks generally display an increase in their mitotic occupancy, while others such as H4K5Ac and ubH2B are reduced. H3K27me3 and H3K9me3 appear to be maintained during mitosis, while the active enhancer mark H3K27Ac is restructured and occupies a

large subset of its interphase sites. The mitotic-specific profiling of many other histone marks is currently unknown.

3D landscape of the mitotic chromatin

Condensation of the chromatin structure results in a 3D organization of the mitotic chromatin that is distinct from that in interphase (Naumova et al., 2013). During interphase the chromosomes are spatially folded by long-range chromatin interactions between regions of the genome into topologically associated domains (TADs) (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Markaki et al., 2010). These domains are organized in a pattern that is locus dependent and distinct between different cell types, and is correlative with the expression patterns of genes (Nora et al., 2012). Interestingly, a study of high order chromatin in different phases of the cell cycle showed that cell type-specific spatial organization of chromatin is unique to interphase, and is lost during mitosis (Naumova et al., 2013). In contrast to interphase chromatin, mitotic chromatin has a generic folding pattern that is common between different cell types, and is independent of the gene loci and transcriptional phenotype of the cell (Naumova et al., 2013).

In addition to intra-chromatin interactions, chromosomes are also spatially organized into lamina-associated domains (LADs) through interactions with the lamin proteins that form the nuclear lamina (Chubb et al., 2002; Guelen et al., 2008). These interactions are mediated by the repressive mark H3K9me2 during interphase (Kind et al., 2013). When the nuclear lamina disintegrates during mitosis, the spatial orientation of LADs is lost, but the chromosome lamin interactions persist in a banded pattern alternating with active chromatin marks H3K27Ac and H3K4me2 (Kind et al., 2013).

Upon mitotic exit, LADs are reformed in a stochastic manner in the two daughter cells and do not occupy the same chromosome-nuclear lamina territories that they did in the parental cell (Kind et al., 2013). Unlike the spatial organization of the TADs, which is reset upon mitotic exit (Naumova et al., 2013), LADs are one of the unique molecular memory mechanisms that are not faithfully inherited upon mitotic exit.

Chromatin accessibility

Despite the condensation and compaction of chromatin during mitosis, recent studies have shown that DNA accessibility during mitosis is largely similar to that in interphase (Hsiung et al., 2015; Teves et al., 2016). DNase I sensitivity assays comparing mouse erythroid interphase or mitotic cells showed that chromatin accessibility is largely unchanged during mitosis (Hsiung et al., 2015). The DNase sensitive sites were divided into two groups: “hotspots” (broad, moderately accessible domains, mostly along gene bodies) and “peaks” (narrow sites of hypersensitivity, generally located around regulatory elements) (Hsiung et al., 2015). Both the hotspots and peaks largely retain their accessibility profiles during mitosis, however locus specific reduction of DNase I sensitivity is observed at some peak regions during mitosis (Hsiung et al., 2015). The small fraction of reduced accessibility during mitosis is mostly at distal regulatory elements, while most proximal promoters maintain their accessible nature. Similarly, the ATAC-seq (Assay for Transposase-Accessible Chromatin by sequencing) (Buenrostro et al., 2013) profiles of mitotic and interphase chromatin in mouse ES cells showed no significant differences in chromatin accessibility (Teves et al., 2016). These data suggest that there are several mechanisms in place that are responsible for transmitting locus specific chromatin accessibility information through mitosis.

2.1.3 Transcription during the cell cycle

In eukaryotes, transcription is mediated by three different RNA polymerase (RNA pol) enzymes: i.) RNA pol I, which transcribes ribosomal (rRNA) from the ribosomal DNA clusters that are located at the nucleolus organizing regions (NORs) on specific chromosome sites; ii.) RNA pol III, which transcribes small rRNAs and transfer RNAs; and iii.) RNA pol II, which is largely responsible for transcription of protein coding genes (Bjorklund and Kim, 1996; Kornberg, 1996; Reines et al., 1996; Roeder, 1996). Transcription occurs in three main steps: initiation, elongation and termination. During initiation, RNA polymerase and its basal transcriptional machinery consisting of general transcription factors such as TATA binding protein (TBP) and TFIIA, B, D, E, F and H (Kornberg, 1996; Roeder, 1996) in case of RNA pol II is assembled at the regulatory promoter elements of a gene (Kornberg, 1996). This recruitment in mediated by interactions with cell type specific transcription factors either at gene promoters or enhancer elements resulting a context specific gene expression (Kornberg, 1996).

Transcription ceases from all three RNA polymerase units at the onset of mitosis, albeit via different mechanisms, and several hypotheses have been put forth to explain the underlying mechanisms of transcriptional repression (Gottesfeld and Forbes; Johnson and Holland, 1965; Taylor, 1960). Some early studies had hypothesized that RNA synthesis in mitotic cells is repressed partly due to the condensed nature of the chromatin. *In vitro* transcription of DNA protein complexes from mitotic extracts was much less efficient compared to that from interphase cells (Johnson and Holland, 1965) leading to the hypothesis that chromatin is less accessible during mitosis. However, recent studies have

shown that mitotic chromatin is widely accessible (Hsiung et al., 2015; Martinez-Balbas et al., 1995; Teves et al., 2016).

In addition to chromatin condensation, the repression of core transcriptional machinery and displacement of sequence specific transcription factors have been explored as potential mechanism for transcriptional silencing.

Mitotic phosphorylation of mitotic-specific DNA-protein interaction is now understood to be a mechanism that can elicit transcriptional repression. The core transcription machinery of both RNA polymerase II and III can be inactivated *in vitro* by the addition of cdk1-cyclinB (Hartl et al., 1993; Leresche et al., 1996). At the onset of mitosis, activation of cdk1-cyclinB complex results in a cascade of phosphorylation of the basal transcription machinery resulting in a temporary hiatus in transcription (Hartl et al., 1993). RNA pol II is also shown to dissociate from the mitotic chromatin by both immunocytochemistry and live cell imaging (Parsons and Spencer, 1997; Zhao et al., 2011). In addition to phosphorylation dependent inactivation of basal transcription machinery, several transcription factors are also phosphorylated at the onset of mitosis resulting in their decoupling from the mitotic chromatin. Sequence specific transcription factor Sp1 is bound less efficiently to hsp90 gene promoter compared to other general transcription regulators during mitosis (Martinez-Balbas et al., 1995). Several other transcription factors, Oct1, Oct2, Ets-1, and Bcl6 were shown to be de-coupled from the mitotic chromatin by immunocytochemistry (Martinez-Balbas et al., 1995). A phosphorylation dependent loss of DNA binding during mitosis was also observed in case of zinc finger transcription factor YY1 (Rizkallah and Hurt, 2009), octamer binding

protein Oct-1 (Ohtsubo et al., 1995; Segil et al., 1991) and transcription regulators Myb and Myc (Luscher and Eisenman, 1992).

Unlike RNA pol II and III, where mitotic chromatin dissociation of the polymerase and/or its binding factors blocks transcription initiation, RNA pol I remains bound to the rDNA during mitosis (Weisenberger and Scheer, 1995). In the case of RNA pol I, transcription is largely ceased at the elongation step (Dundr and Olson, 1998; Weisenberger and Scheer, 1995). Similarly, a fraction of the RNA pol II mediated transcription is also repressed by the dissociation of transcription elongation machinery between late prophase and telophase of mitosis (Parsons and Spencer, 1997). Interestingly, however, during early prophase there is a wave of transcription activation that is mediated by the elongation factor p-TEFb, and is required for proper cell division (Liang et al., 2015). In this case, p-TEFb is recruited to genes with an already engaged RNA pol II and is key to completion of transcription and release of RNA pol II from the gene body (Liang et al., 2015).

The phosphorylation status of RNA pol II governs interactions with various components of the RNA processing machinery and, therefore, can be used to detect RNA pol II subunits that are involved in different functions (Ho and Shuman, 1999). RNA pol II phosphorylated at serine 5 in the carboxyl-terminal domain is associated with transcription initiation (Bregman et al., 1994; Komarnitsky et al., 2000), while when phosphorylated at serine 2 it is associated with transcriptional elongation (Bregman et al., 1995; Ho and Shuman, 1999). Using mitotic arrest and release experiments, and by exploiting the various forms of RNA pol II and incorporation of bromo-UTP (Br-UTP) into nascent RNAs, Prashanth and colleagues elegantly demonstrated the kinetics of RNA

transcription and processing (Prasanth et al., 2003). They showed that in early telophase, as lamin B1 and nuclear pore protein p62 form the nuclear lamina, RNA pol II in both hypo- and hyper-phosphorylated forms is entirely cytosolic, as are other members of the RNA processing machinery including the general transcription factor (TFIIE) and splicing factors (Prasanth et al., 2003). At this stage, the cells are transcriptionally silent, as demonstrated by the lack of Br-UTP incorporation. As division progresses and enters late telophase, Serine 5-phosphosphorylated RNA pol II (the initiation form) is detected in the daughter nuclei, while the splicing factors and other processing components are still cytoplasmic (Prasanth et al., 2003). At this stage, Br-UTP was detected, suggesting that RNA transcription occurs immediately upon recruitment of RNA pol II onto the chromatin. Similarly, they showed that further into division, the splicing machinery, along with the elongation form of RNA pol II (Serine 2 phosphorylated), is found in the daughter nuclei and is associated with a strong increase in Br-UTP incorporation (Prasanth et al., 2003).

If RNA is transcribed immediately upon recruitment of RNA pol II to transcription start sites, and this recruitment occurs at mid-telophase, the transcription program must be set before exit from mitosis. And given that core components of general and context specific transcription machinery are dissociated form the mitotic chromatin, the question arises as to how are cells able to retain their transcriptional memory during or after division. This process is even more intriguing when the significant differences in the epigenetic profiles of mitotic and interphase chromatin are taken into account. As cells undergo mitotic division, chromatin is highly remodeled: 1.) long range chromatin interactions are altered but reset upon mitotic exit 2.) interactions with nuclear lamina are

disrupted 3.) some epigenetic memory marks are not retained at their interphase loci 4.) transcription factors and transcriptional machinery are largely dissociated from their normal binding sites. Despite these large changes, once a cell exits mitosis, the daughter cells acquire the parental cell identity and the transcriptional program is reset. Currently, there are significant gaps in our understanding of the mechanisms that play a role in this faithful transmission of cell identity. Over the last few years, the phenomenon of mitotic bookmarking has been proposed and tested to explain the retention of cell fate upon mitotic exit.

2.1.4 Mitotic bookmarking

The concept of mitotic bookmarking stemmed from two main observations in the late 1990s. In the first, it was shown that a greater proportion of mitotic chromatin had single stranded DNA compared to interphase chromatin, by using single strand specific nuclease analysis (Juan et al., 1996). Subsequently, potassium permanganate ($KMnO_4$) DNA foot-printing of several transcription start sites (TSS) revealed that the single stranded nature of chromatin was correlative with the expression profiles of active genes, with repressed genes showing no hypersensitivity to $KMnO_4$ (Michelotti et al., 1997). These observations raise the question that if transcription factors and transcriptional machinery are dissociated from the mitotic chromatin, what causes the preservation of the transcriptional state by DNA confirmation during mitosis? At the time, it proposed that as cells enter G2-phase and are preparing for division, several “bookmarks” are deposited onto the TSS of active genes, which are then retained throughout mitosis while the transcriptional machinery de-couples (John and Workman, 1998; Michelotti et al., 1997).

In the period following these initial studies, further observations have assisted in refining the current definition of mitotic bookmarks to be 'memory signatures' that are preserved throughout mitosis at key fate maintaining/determining genomic sites, and passed on to the daughter cells providing them a blueprint of what genes to turn on or off to maintain the parental fate (Hsiung and Blobel, 2016; Hsiung et al., 2015; Sarge and Park-Sarge, 2009). It is also now understood that the memory signatures that underlie mitotic bookmarks can belong to one of the following categories: DNA methylation profiles, mitotically retained transcription factors, mitotically stable histone variants, or architectural components of the chromatin (Hsiung et al., 2015).

Over the past decade, several studies (summarized in Table 2) have revealed that some transcription factors and chromatin regulators are retained on the mitotic chromatin. MLL (Mixed Lineage Leukemia), a histone methyl transferase, was one of the first components of the chromatin remodeling complex that was shown to be associated with the mitotic chromatin (Blobel et al., 2009) in HeLa cells. The mitotic occupancy of MLL was correlated with rapid transcriptional reactivation of the bookmarked genes (Blobel et al., 2009). Subsequently, other members of chromatin remodeling complexes, including Ring1A (Arora et al., 2012) and Brd4 (Zhao et al., 2011), have been found to be retained on the mitotic chromatin. This behavior is not limited to chromatin-modifying enzymes: transcription factors such as Gata1 (Kadauke et al., 2012), FoxA1 (Caravaca et al., 2013), Parp1 (Lodhi et al., 2014), Esrrb (Festuccia et al., 2016), Hnf1b (Lerner et al., 2016), Sox2 (Liu and Kraus, 2017; Teves et al., 2016) and pluripotency factors Oct4, and Klf4 (Liu et al., 2017b), have been implicated to act as mitotic bookmarks that facilitate a rapid transcriptional program of bookmarked genes. It has also been proposed that

bookmarking factors could act in concert with mitotically retained epigenetic marks to facilitate transcription. Evidence to support this hypothesis exists in case of Brd4, which bookmarks sites that present H4K5Ac mark (Zhao et al., 2011) and H3K27Ac which bookmarks active gene loci in conjunction with transcription factors Oct4, Sox2 and Klf4(Liu et al., 2017b).

In addition to affecting transcriptional kinetics, the bookmarking factor Uhrf1 (Rothbart et al., 2012) is involved in the maintenance of DNA methylation during mitosis, while Rbpj, a transcription regulator, is proposed to affect long-range chromatin interactions during mitosis (Lake et al., 2014).

2.1.5 Summary of Intent

Mitotic bookmarking is a recently defined and underexplored phenomenon, and up until last year, there were no studies detailing the conservation of this mechanism in pluripotent cells. It has now become clear that more factors (Table 2) (Liu et al., 2017b; Teves et al., 2016) are associated with the mitotic chromatin than previously appreciated. To date, most studies have used a biased target-based approach to identify the transcription and chromatin regulatory factors that could be associated with the mitotic chromatin. There are currently no datasets or tools available to discover the mitotic chromatin association of factors in an unbiased manner, frustrating the exploration of mitotic bookmarking at a global proteome level. My research project aims to develop reliable and unbiased methods for identifying and studying proteins that are associated with mitotic chromatin, and that may perform mitotic bookmarking in ES cells. Based on evidence in other cell types, we *hypothesized that in ES cells putative gene bookmarking factors are retained on the mitotic chromatin, and are involved in fate maintenance of*

daughter cells after cell division. Following research aims were put forth to test the hypothesis:

Aim 1: To identify, in an unbiased manner, putative mitotic bookmarking factors (MBFs) which associate with the mitotic chromatin of ES cells.

Aim 2: To identify key loci bookmarked during mitosis, and assay their transcriptional profile upon mitotic exit

Aim 3: To assay the effects of perturbation of a candidate MBF on maintenance of stem cell identity

Aim 1 was addressed by developing a mitotic chromatin immunoprecipitation followed by mass spectrometry (ChIP-MS) assay to identify and characterize putative mitotic bookmarking factors in an unbiased manner. We utilized a mitosis specific histone modification; phosphorylation of serine 10 on histone 3 (H3S10P), to pull down chromatin associated proteins. Putative MBFs were validated for their mitotic association by live cell imaging, and were characterized *in-silico* and by literature search to rank them as putative fate maintaining mitotic bookmarking factors.

Aim 2 was addressed by performing ATAC-seq (Assay for Transposase-Accessible Chromatin followed by sequencing) to study the chromatin dynamics of ES cells in mitosis and upon mitotic exit. We identified key sites that show a more open chromatin structure during mitosis and characterize these as putatively bookmarked sites.

In Aim 3, the mitotic bookmarking capacity of the transcriptional regulator and putative MBF poly(ADP-ribose)polymerase (Parp1), which was identified in Aim 1, was assessed by assaying transcriptional kinetics of Parp1-bookmarked sites. The role of

Parp1 as a putative MBF was then explored by functional evaluation of Parp1 knock out lines to assay effects of its perturbation on cell fate.

All together, the study was designed as an unbiased survey of mitotic bookmarking in ES cells. It encompassed addressing the identity of bookmarked genes, the factors that could potentially bookmark these genes, and the phenotypic outcomes of perturbation of these MBFs.

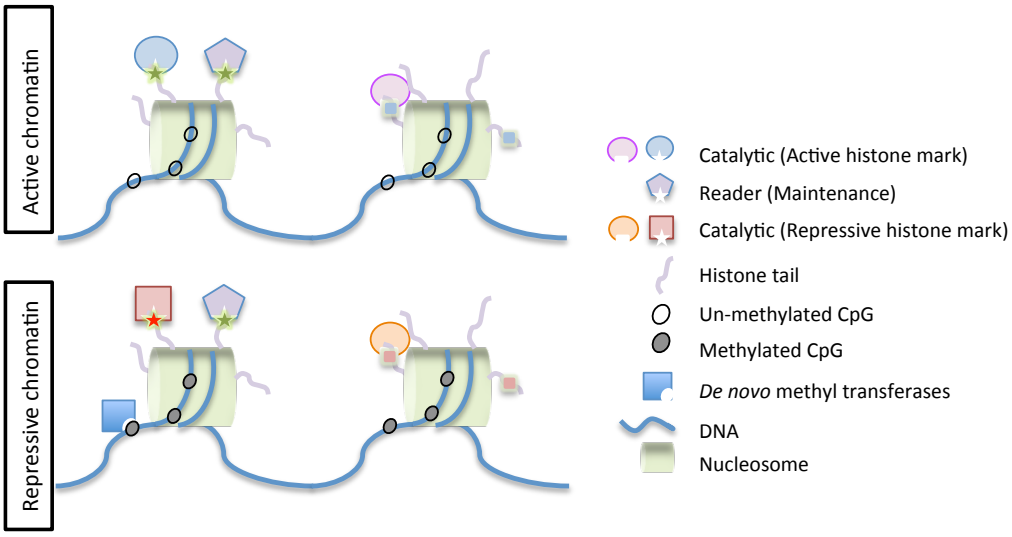


Figure 1: An overview of regulation of epigenetic memory. The deposition and maintenance of epigenetic marks is regulated by different catalytic enzymes. These epigenetic marks could be active, in that they result in a euchromatin chromatin structure resulting in gene expression. Repressive epigenetic marks result in a hetercrhomatic chromatin structure and therefore repress gene expression.

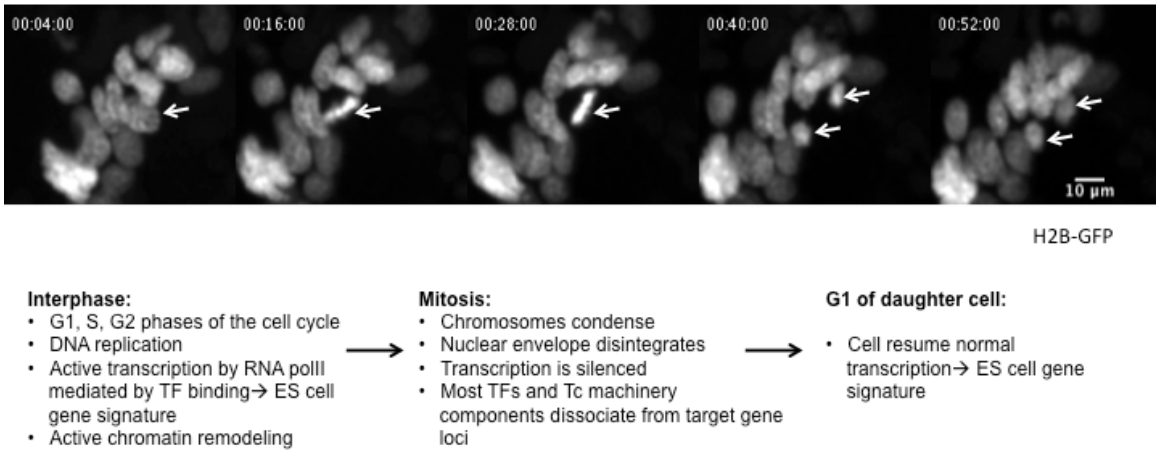


Figure 2: An overview of mitosis. Mouse embryonic stem cells were transfected with H2B-GFP and followed over a round of mitotic division.

Table 1: A summary of different histone post-translational modifications that are a part of the cell epigenetic memory.

Histone marks	Details	Reference	Readers	Writers	Erasers	Change in mitosis
H3K4me3	Generally at TSS of active promoters	(Calo and Wysocka, 2013; Lawrence et al., 2016; Santos-Rosa et al., 2002)	SET1 complex, TFIID, HAT, SIN3A, de novo methyltransferases, NuRD complex (Eberl et al., 2013)	MLL3/4, SET1, ASH1L, SMYD3	LSD1/NuRD complex	increased, (Grandy et al., 2016)maintained (Valls et al., 2005)
H3K27Ac	Active enhancers	(Calo and Wysocka, 2013; Creyghton et al., 2010)	SMRT/HDAC3	p300	TAF1, BRD4, BRG1	maintained(Hsiung et al., 2016), increased (Liu et al., 2017b)
H3K36me3	Transcription elongation mark on active gene bodies	(Chantalat et al., 2011)	PWWP domain (Qin and Min, 2014)	SETD2		maintained at pericentromeric regions during mitosis by IF (Chantalat et al., 2011)
H3K79me3	Active promoters	(Vakoc et al., 2006)		DOT1L (Feng et al., 2002)		maintained in yeast (Kim et al., 2014; Schulze et al., 2009)
H3K79me2	Active transcription, generally at transition sites between H3K4me3 and H3K36me3			DOT1L (Feng et al., 2002)		increased in mitosis in mouse preimplantation embryos (Ooga et al., 2008)and in yeast (Schulze et al., 2009) maintained in human K562s (Fu et al., 2013; Kim et al., 2014)
H3K4me1	Active enhancers	(Calo and Wysocka, 2013)		MLL3/4, SET1, ASH1L	LSD1/NuRD complex	
H3K4me2	Active enhancers and promoters	(Calo and Wysocka, 2013)		MLL3/4, SET1, SMYD3	LSD1/NuRD complex	increased (Liu et al., 2017b)
H4K5ac	Active			HAT1, p300		decrease (except at sites bookmarked by Brd4) (Zhao et al., 2011)
ubH2BK120	Transcribed regions of active genes		Zhu 2005	RNF20		restructured to some promoters (Arora et al., 2012)
H3K20me3	Heterchromatic mark (at telomeres)	(Lawrence et al., 2016; Nishioka et al., 2002)		SUV420H1/2 (Schotta 2004)		
H3K9me3	Transcription repression, imprinting		HP1 proteins (Eberl et al., 2013)	Suv39h1 (Rea et al., 2000)SETDB1 (Schultz et al., 2002)		maintained
H3K27me3	Transcription repression, X-inactivation, bivalent genes along with H3K4me3	(Bernstein et al., 2006)		Ezh2 ((Kuzmichev et al., 2002)		maintained (Grandy et al., 2016)
H2AK119Ub	Transcription repression	(Wang et al., 2004)		PRC1		decrease

Table 2: A summary of recent studies that have identified putative mitotic bookmarking factors in various systems

Mitotic chromatin bound protein	Type	Conservation during mitosis	Epigenetic mark	Mitotic specific role	Reference
MLL	chromatin remodeller (writer, histone methyl transferase)	subset of interphase sites, gene promoters of active genes	H3K4me (associated but dispensible for the mitotic preservation of the mark)	recruitment of other epigenetic machinery	Blobel et al 2009
Brd4	chromatin remodeller (reader)	differential recruitment kinetics with RNA polII upon mitotic exit	H4K5Ac	facilitates transcriptional reactivation	Zhao et al 2011
Ring1A/BMI1	chromatin remodeller (PRC1 complex)	some gene promoters	ubiquitination (UbH2B)		Arora et al 2012
Gata1	transcription factor	retained at a subset of interphase binding sites	NA	facilitates transcriptional reactivation	Kaduke et al 2012
Uhrf1	E3 ubiquitin ligase	retained via binding to H3K9me3	H3K9me3	preservation of DNA methylation during mitosis	Rothbart et al 2012
Foxa1	transcription factor	retained at 15% of interphase sites; also bound non-specifically to the mitotic chromatin	NA	facilitates transcriptional reactivation	Caravaca et al 2013
Rbpj	transcription regulator	retained at 60% of interphase sites	co-localizes with Ctcf sites	proposed role in maintining long range chromatin structure upon mitotic exit	Lake et al 2014
Parp1	catalytic enzyme	retained on promoters genome wide	NA	facilitates transcriptional reactivation	Lodhi et al 2014
Esrrb	transcription factor	retained at a subset of interphase binding sites	NA	facilitates transcriptional reactivation	Festuccia et al 2016
Hnf1b	transcription factor	reversible, thermosensitive binding to the mitotic chromatin	NA	imparied mitotic association of naturally occuring HNF1B mutants in diabetic patients	Lerner et al 2016
Sox2	transcription factor	interaction with mitotic chromatin is more dynamic compared to interphase chromatin	NA	co-related with maintenance of chromatin accessiblity	Teves et al 2016
Sox2, Oct4, Klf4	transcription factors	subset of interphase sites	H3K27Ac	important for efficient induction of pluripotency upon mitotic exit	Liu et al 2017

Chapter 2.2 Identification of pluripotency associated putative mitotic bookmarking factors

Chapter 2.2 Identification of pluripotency associated putative mitotic bookmarking factors

Preface

This chapter addresses aim 1 of the research goals, and was designed entirely by me with assistance from Dr. Jon Draper. I carried out the experiments, performed the analysis and wrote the chapter. Mehdi Hamzeh, a previous undergraduate student, assisted with quantitative data analysis of live cell images (Fig. 7C).

Abstract

Mitotic bookmarking is a mechanism that attempts to explain the preservation of cellular identity during mitotic division. As cells undergo mitosis, several transcription and chromatin regulators decouple from the mitotic chromatin to allow faithful segregation of DNA. Some candidate-based approaches have identified mitotic bookmarking factors (MBFs) that are retained on the mitotic chromatin and preserve the transcriptional memory of the cell. We performed a mitotic chromatin immunoprecipitation followed by mass spectrometry assay (ChIP-MS) to identify putative MBFs in an unbiased manner. We identified a total of 143 MBFs, and validated their mitotic association with a success rate of 70%. The majority of the MBFs identified were involved in chromatin regulation, of both heterochromatin and euchromatin marks, and were bound to varying degrees with the mitotic chromatin. We also identified MBFs that are very highly expressed in ES cells (Utf1, Dnmt3b, Dnmt3L, Msh6, and Parp1) suggesting their potential roles in fate maintenance via mitotic bookmarking in ES cells. Our data has identified proteins, not previously known to be associated with the mitotic chromatin, and therefore, can be used as a starting point for studying mitotic bookmarking.

Background:

Embryonic stem cells are pluripotent in nature, i.e., they can indefinitely self-renew and are capable of giving rise to all three adult lineages. Recently, several research groups have exploited the differentiation capacity of ES cells to give rise to therapeutically relevant cell types from various organ systems (Cantz et al., 2008; Mauritz et al., 2008). The efficient therapeutic usage of ES cells relies on our knowledge

of how these cells maintain their stem cell fate over subsequent divisions, and what fate determining mechanisms can be exploited to differentiate them better towards a desired lineage. Mitotic bookmarking is proposed as a mechanism of fate maintenance during cell division (Sarge and Park-Sarge, 2009; Zaidi et al., 2010).

Mitotic bookmarks are 'memory signatures' retained during mitosis at fate maintaining loci and passed on to the daughter cells providing them the transcriptional program of the parental cell (Hsiung et al., 2015). Stably inherited epigenetic marks, or mitotically retained transcription and chromatin regulators could act as mitotic bookmarks (Hsiung et al., 2015; Sarge and Park-Sarge, 2009). Recent studies suggest that more transcriptional regulators might be associated with the mitotic chromatin than previously known and could play a role in propagation of cellular identity during division. Transcription factors such as Oct4, Sox2, Esrrb have recently been shown to be associated with the mitotic chromatin in ES cells (Festuccia et al., 2016; Liu et al., 2017b; Teves et al., 2016). These studies provide a candidate-based approach for studying bookmarking factors, however, we currently do not know the global proteomic profile of the pluripotent mitotic chromatin.

We used published studies that take a ChIP-based approach to study chromatin-associated transcription factors (Mitchell et al., 2013; Soldi and Bonaldi, 2014) as a starting point, and developed a strategy to look for mitotically-associated transcription factors. We hypothesized that by pulling down the mitotic chromatin, we will be able to identify key pluripotency-associated mitotic bookmarking factors. Histone, H3, is phosphorylated at serine 10 (H3S10P) at the onset of mitosis (Crosio et al., 2002; Hendzel et al., 1997). The phosphorylation event is initiated at the pericentromeric

regions in late G2 (Kelly et al., 2010), and spreads along the body of the chromosomes during mitosis (Crosio et al., 2002; Hendzel et al., 1997). We used an antibody against H3S10P to immunoprecipitate the mitotic chromatin and identified putative bookmarking factors by mass spectrometry.

Materials and methods:

Cell culture: E14TG2A mouse embryonic stem cells were cultured on 0.1% gelatin coated culture dishes in mESC media: DMEM (Sigma Aldrich, D5796), 15% FBS, 1x non-essential amino acids (Life technologies: 11140-050), 1x glutamax (Life technologies: 35050-061), 1x sodium pyruvate (Life technologies: 11360-070), 1x beta-mercaptoethanol (Gibco: 21985-023), mouse recombinant LIF (Amsbio, AMS-263-100). Cells were passage every three days using accutase® (Sigma Aldrich: A6964). ParpKO cells were routinely maintained on a layer of x-ray irradiated mouse embryonic fibroblasts (xMEFs) seeded at a density of $1x10^6$ cells/$60cm^2$. MESCs were pre-plated on gelatin coated dishes for 20minutes to deplete xMEFs.

Mitotic Enrichment: For ChIP-MS and ATAC-seq experiments mESCs were mitotically enriched using 50ng/ml nocodazole (Sigma Aldrich: M1404) for 7hrs. Flow cytometry using Hoechst 33342 (Life technologies: H1399), MPM2 (05-368, Millipore) and H3S10P (05-1336, Millipore) was used to characterize the percentage of cells in mitosis.

Chromatin Immunoprecipitation followed by Mass Spectrometry (ChIP-MS): $1.35x10^8$ M-enriched mESCs were used per IP with 20ug of antibody. Based on our pilot mass spectrometry runs and a 30% enrichment of mitotic cells, the input cell number for

H3S10P (05-1336, Millipore) pull down was doubled while keeping all other parameters the same. Cells were cross-linked in 1% paraformaldehyde (PFA) for 10 minutes at room temperature with shaking; un-used PFA was quenched using final 1X glycine (stock 10X Glycine, Cell Signaling) and the cross-linked pellet was washed 2X in large volumes of PBS. Cross-linked cells were partially lysed in RIPA buffer (50mM Tris-Cl pH7.45, 150mM NaCl, 0.1%SDS, 2%NP-40, 1%Sodium deoxycholate) supplemented with protease inhibitors (11836153001 cOmplete^{TM} mini-tablets Roche, Sigma) followed by mechanical lysis by passing the cells 50x through an 18G needle on a 1 ml syringe. Cells were then checked for partial lysis under the microscope looking for released nuclei. Nuclear and chromatin pellet was collected by gentle centrifugation at 2500g for 5 min at 4C. The pellet was resuspended in RIPA dilution buffer (RDB, 50mM Tris-Cl pH7.45, 150mM NaCl) and supplemented with a final of protease inhibitors, 1mM $CaCl_2$, and micrococcal nuclease (MNase, LS004797 Worthington, used at 0.88U/million cells). Reactions were incubated at 37C for 20mins with shaking at 300 rpm, tubes were inverted to mix intermittently. MNase was inactivated by adding EDTA to a final concentration of 50mM, mixed by inverting and incubated on ice for 2minutes. The reactions were centrifuged at 13000rpm for 1 minute at 4C. The supernatant and the nuclear pellet were collected. The nuclear pellet was lysed in RIPA buffer, incubated on ice for 10mins and passed through an 18G needle 10x to break open the nuclei. The lysate was gently sonicated to release shearing chromatin (6 pulses each with 5s ON, 10s OFF at 30% amplitude). 1ul of the lysate was inspected under the microscope to ensure lack of complete nuclei. Sheared chromatin supernatant was collected by spinning at 10,000g for 10 min at 4C, and combined with the MNase treated supernatant collected

previously. Chromatin was diluted in RDB to get a final SDS concentration of 0.025% to assist in IP. Respective antibodies were added to the chromatin and IPed overnight at 4C (12hrs). 100ul of Dynabeads were added per IP and incubated for 2hrs at 4C. In keeping with data that showed that the supernatant collected still had a lot of unbound H3 and H3S10P (not shown), the IP supernatant was collected and re-IPed with 15ug of antibody for 5 hrs and 100ul of beads for 2 hrs at 4C. The beads from both IPs were washed 2X in low salt buffer (50mM Tris-Cl, pH 7.45, 250mM NaCl, 2mM EDTA, 0.1%SDS w/v, 1% Triton X-100 v/v), and 3X in high salt buffer (same as low salt buffer except use 500mM NaCl) and eluted in reverse crosslinking buffer (1:1 mix of 2X reverse crosslink buffer (4%SDS, 1M BME, 500mM Tris-Cl pH 8.8) and 2X LDS buffer (2X LDS (NuPage), 10% Bond-Breaker^{TM} (Cat# 77720, ThermoFisher).

The samples were reverse cross-linked at 95C for 20 mins, and run on a precast 4-12% bis-tris gel (Cat# NP0322, ThermoFisher) for 53 minutes at 165V. The gel was washed 3x in water, fixed (40%methanol, 10%acetic acid and 50% water) for 60 minutes at RT, rinsed with water 3x and then washed in a 50% ethanol solution overnight at 4C to reduce background. The fixed gel was washed 3x with water and stained with Bio-Safe^{TM} coomassie stain for 1 hr with shaking and washed with water for 1 hr (Fig. 2A). The gel was placed on a thoroughly cleaned glass plate and IP lanes were cut out using clean razor blades in a biosafety cabinet. Each IP lane was split into two samples, one containing the prominent IgG bands and the histones (brackets in Fig.2A), and second containing the rest of the bands. Protein bands were processed and run on a Q-Exactive Orbitrap MS, by the IRIC proteomics facility, Montreal QC.

Chromatin Immunoprecipitation followed by sequencing (ChIP-seq): ChIP-seq was performed on the H3S10P ChIP samples immunoprecipitated as mentioned above. The starting material was reduced to 1/3rd and the DNA was eluted in TE buffer (10mM Tris-Cl pH=8, and 1mM EDTA pH=8) for 30 mins at 37C with shaking at 1200 rpm. The eluted DNA and 2% input were RNAse treated for 2 hrs at 42C and reversed crosslinked overnight at 65C in TE buffer with 200mM NaCl, and ProteinaseK Sequence libraries were prepared using DNA SMART ChIP-kit (Clontech 634865) with the final library size of 300bps as observed by bioanalyzer data. ChIP-seq libraries were sequenced as 50bp single end reads on Illumina HiSeq 2000.

ChIP-seq data processing: Trimmed fastq files were mapped onto mm9 genome using bowtie 2. Following command line function was used:

```
bowtie2 –x [index] –U chip.fastq |samtools view –bhS – | samtools sort – aligned_chip.bam
```

Duplicate reads and blacklisted regions were removed from the aligned files using:

```
samtools rmdup –s aligned_chip.bam – | samtools view –h – | grep –v chrM | grep –v chrY | grep –v chrUn | grep –v random | samtools view –bS –q 10 – | intersectBed –v –abam stdin –b mm9blacklist.bed | samtools sort – aligned_chip_de–dup_q10.bam
```

Macs2 (Zhang et al., 2008) was used to call peaks:

```
macs2 callpeak –t aligned_chip_de–dup_q10.bam –c aligned_input_de–dup_q10.bam –m 3 30 ––bw 200 –n mm9_macs2 –f BAM –B –outdir
```

Bin Counts: bedtools makewindows was used to create 1kb bins across the mm9 genome. The aligned bam files were converted to bed using bedtools bamtobed option. The ChIP bed files were counted against the 1kb bin file using bedtools intersect –c option. Furhter analysis was performed using R packages (Team, 2013). The read counts were used as input values in EdgeR and normalized counts per million (cpm) were computed per bin for each of the libraries. A heat-map was plotted using normalized cpm for a 1000 randomly selected bins.

Mass spectrometry data processing and analysis: Mass spectrometry (MS) data was obtained and analyzed using PEAKS Studio 7.5 software. An overview of the analysis pipeline is showing in Figure 2C. Protein IDs with at least 2 unique peptides with a peptide threshold of -10logP>=1% FDR were kept for down stream analysis. An overlap of these proteins IDs was made between the three replicates and only the ones common to all three were kept for down stream analysis. For identifying putative bookmarking factors The proteins IDs were characterized using ClueGO app (Bindea et al., 2009) for cytoscape. Following selection criteria was used: Statistical Test Used = Enrichment/Depletion (Two-sided hypergeometric test), Correction Method Used = Bonferroni step down, Only show enrichments with $p<0.005$, Min GO Level = 3, Max GO Level = 8, Minimum number of genes per cluster=3.

For the heat-map of putative MBFs association with other epigenetic marks, the MS data was acquired from Ji et al 2015(Ji et al., 2015) (Associated R-script 170524_MBF_overlap_with_other_histone_ChIPMS.R in R scripts). The list of 143 putative MBFs was overlapped with ChIP-MS data for other epigenetic marks, and the ones with at least one overlapping dataset were used for the heat-map. The heat-map

values are represented as log10(MS score+1). For identification of pluripotency related factors, RNAseq was obtained from (Terranova et al., 2015)

Validation of mitotic association of putative MBFs: cDNAs for several putative MBFs were received from TCAG cDNA library (clone information in Table S5) and fused with mKO2 fusion protein by a linker L3 (Cadinanos and Bradley, 2007). pCAG-eGFP-IRES-Puromycin-polyA (pCAG-eGFP-IPpA) vector was used to replace the eGFP with L3-mKO2 using restriction digest at *XhoI* and *NotI* sites to generate the master backbone vector pCAG-L3-mKO2-IPpA. In-fusion® cloning (Clontech 638909) was used to insert all cDNA sequences upstream and in-frame of L3-mKO2 using sites *XhoI* restriction site. cDNAs not available from TCAG were generated by amplification of cDNA converted from mRNA using iSCRIPT cDNA synthesis kit (Biorad: 1708890). The final pCAG-cDNA-L3-mKO2-IPpA vectors were co-transfected to E14TG2A wild type mESCs along with pCAG-H2B-GFP-IPpA using Lipofectamine LTX (ThermoFisher: 15338100) and seeded into 96 well plate imaging plates. Live cell imaging was performed using Perkin Elmer High-Content Imaging system Operetta for 2-3hrs acquiring images at 20X every 4 or 5minutes. Images were processed using FIJI and ImageJ (Schindelin et al., 2012; Schneider et al., 2012).

Results

Development of mitotic ChIP-MS assay

We developed a ChIP-MS based strategy to identify mitotic bookmarking factors (Figure 1A). Briefly, ES cells were enriched in mitosis, cross-linked using PFA, lysed, treated with MNase, immunoprecipitated and mass spectrometry analysis was performed

on the eluted chromatin bound proteins. In normal ES cell culture only a small fraction of cells (~1%) are in mitosis (Figure 1B) and therefore studying proteins associated with the mitotic chromatin requires enrichment of cells in metaphase. We tried various conditions to achieve mitotic enrichment and utilized MPM2 antibody to calculate the mitotic index of ES cells (Figure 1B, S1C). The MPM2 antibody recognizes phosphorylated version of a peptide sequence that is present in over 40 different eukaryotic proteins (Westendorf et al., 1994). The peptide sequence is commonly phosphorylated at the onset of mitosis therefore MPM2 can be used as a marker for mitotic cells (Campbell et al., 2014; Westendorf et al., 1994). The co-occurrence of MPM2+ cells in G2-M population (dot plots in Fig 1B), and IF images with DAPI (Fig S1C), confirm the specificity of the antibody to mitotic cells. A maximal mitotic enrichment of ES cells (30-40%, Figure 1B) was observed with when treated with 50ng/ml of nocodazole for 7 hours.

Phosphorylation of serine 10 of histone 3 (H3S10P) is an epigenetic mark that is considered to be important for the onset of mitosis and is associated with a condensed chromatin structure (Crosio et al., 2002; Hendzel et al., 1997). Aurora kinases phosphorylate H3 at serine 10 at the peri-centromeric foci in late G2 of the cell cycle from where it spreads throughout the metaphase chromosomes (Crosio et al., 2002; Hendzel et al., 1997). There are some reports of expression of the H3S10P mark at some active gene loci in G1 phase of the cell cycle and also during apoptotic condensation of chromatin (Park and Kim, 2012; Perez-Cadahia et al., 2009). To test the specificity of the epigenetic mark, we tested two antibodies specific to H3S10P for IF, and western blotting. The expression correlates with condensed mitotic cells based on Hoechst staining (Figs. 1C for Abcam 05-1336, S1A for Abcam 06-570). In western blot analysis

of untreated and mitotically enriched cells, the antibodies show a much greater signal in the latter population (Figs. 1D, S1B). Based on the mitotic specific recognition in our system, we chose H3S10P as a mark to immunoprecipitate the mitotic chromatin.

After establishing a mitotic cell enrichment protocol and optimal mark for pull down, we optimized a chromatin immunoprecipitation assay to identify proteins present on the mitotic chromatin (Fig. 1A). Briefly, PFA fixed mitotically enriched cells were lysed to isolate the chromatin/nuclear pellet, which was then treated with MNase to result in a wide range of fragments to ensure efficient protein pull-down (Fig. 1E). In normal MNase based ChIP assays, only the nuclear pellet is retained after enzymatic shearing, however, since mitotic cells lack the nuclear membrane we tested for the presence of sheared chromatin in the supernatant (Fig. 1E) and combined the two fractions before immunoprecipitation. MNase treated lysate was immunoprecipitated (IPed) for histone H3, as a global chromatin mark, and H3S10P as a mitotic specific chromatin mark.

H3S10P is an unbiased marker for pulling down mitotic chromatin

We performed ChIP-seq on a fraction of the immunoprecipitated sample to ensure an unbiased chromatin pull-down that equally represented all genomic sites. Macs2 based peak calling analysis of S10 data compared to input did not call any peaks, so we divided the genome into 1kb bins and counted the number of reads in both IP and input samples. A comparative heat-map of normalized reads counts from 1000 randomly selected bins showed similar profiles in both S10 IP and input samples (Fig. 2A). Additionally, a differential analysis performed across the all genome wide bins showed minimal significant differences between the S10 IP and input samples (Fig 2B). This suggests that

S10 pull down is equivalent to the genome wide signal, and hence H3S10P can be used to pull down chromatin in an unbiased manner.

The majority of the immunoprecipitated sample was used for protein identification, and was run on a gel and isolated for mass spectrometry (Fig. 3A). Due to predominance of IgG and histone bands in the IP lanes (bracketed regions in Figure 3A), each sample was run as two separate mass spectrometry runs, one with the bracketed region in Figure 3A, and the other consisting of all other bands. IP efficiency was assessed by western blotting $1/10^{th}$ of each sample with anti-H3, anti-H3S10P and Oct4 (Fig. 2B). Oct4, was used as a negative control for H3S10P as it was known to be displaced from the mitotic chromatin by immunostaining.

Mitotic ChIP followed by mass spectrometry reveals putative MBFs.

To ensure greater fidelity in identification of putative mitotic bookmarking factors, the IP was performed in triplicates and only significant protein IDs present in all three replicas were used for further analysis (Fig. 3C and 4A). The MS/MS spectra were processed and assembled into peptide amino acid sequences using a round of *de novo* sequencing. The peptides were then put through a database search against the cRAP database (http://www.thegpm.org/crap/), to map and identify common mass spectrometry contaminants (e.g. Keratins, albumins etc). The unmapped peptides were then searched against the Uniprot-Swissprot database. Protein IDs were filtered with a peptide threshold set to significance -10lgP≥1% FDR (false discovery rate), and protein threshold significance set to -10lgP≥20. Only protein IDs with at least two unique peptides were chosen for further analysis (Table 1).

When we compared the protein IDs between S10 and H3 IPs, most S10 (143/162) proteins were present in the global H3 IP, further confirming the chromatin bound nature of these proteins. 68 proteins IDs were unique to the H3 IP sample, and were considered to be chromatin bound non-mitotic bookmarking factors.

Characterization of the putative MBFs

We used ClueGO (Bindea et al., 2009) statistical enrichment analysis to characterize the phenotypic groups that are enriched ($p<0.005$) in our list of 143 putative MBFs. With the cellular component analysis we found that the most enriched categories were that of chromosome, ribosomal or nuclear localization (Fig 5C). The most significantly enriched biological processes associated with these proteins were chromatin assembly or disassembly, DNA metabolism or RNA regulation, and majority of other categories enriched were involved in related processes (Fig 5B), and had either RNA or DNA binding as a primary function (Fig.5A, molecular function).

We also compared protein classes enriched in previous analyses of mitotic proteins isolated from metaphase chromosomes (Ohta et al., 2010). This study was the largest multifactorial analysis of metaphase chromosomes, and had identified a list of proteins that have chromosome and nuclear localization as the most significantly enriched category for cellular component, however it also showed significant enrichment of proteins in various other cellular compartments (Fig 5C). Additionally, most of the proteins found were involved in cell cycle regulation and various metabolic processes (Fig 5B), and had either RNA or DNA binding roles (Fig. 5A).

We explored a chromatin regulation role of the putative MBFs, as it was the most highly significant bioprocess associated with this set of proteins (Fig 5B), and found the

presence of various histone variants, high-mobility group proteins (Hmga1, Hmgb1), other nucleosome organization components (Mcm2, Npm1, Set, Smarca5, Supt16), all DNA methyl-transferase enzymes (Dnmt3L, Dnmt1, Dnmt3a, Dnmt3b), and other gene expression regulators (Uhrf1, Parp1, Rbbp4, Pabpc1, Trim28, Chd4) (Table 2).

To further profile the chromatin binding characteristic of these putative MBFs, we compared our dataset to ChIP-MS data for other epigenetic marks (Ji et al., 2015), and found that 38/143 putative MBFs, and 2/8 S10 only proteins (Fig 4A) overlapped with at least one of the epigenetic mark tested (Fig 6). The dataset consisted of proteins associated with heterochromatic epigenetic marks, H4K20me3 and H3K9me3, and euchromatic marks H3K27Ac, H3K4me3, H3K36me3, H3K79me2 (also, see Intro Table 2). The inner centromere protein (Incenp) is exclusively associated with heterochromatic marks, and DNA methyltransferases (Dnmt3a, Dnmt3b), centromere protein V (Cenpv) and chromobox protein 5 (Cbx5) show strong tendency towards closed chromatin marks (Fig. 6). Other proteins show a wide range of association with both open and closed chromatin. Some proteins such as Smarca5, Ssrp1, Uhrf1, Psip1, Hdgf, Rcc2, Rcc1, and Set show the strongest association with the enhancer mark H3K27Ac, while lamina associated proteins (Lmna, Lmnb1), and the chromatin remodeler Rbbp4 strongly associate with active promoter mark H3K4me3 (Fig 6). Proteins such as Parp1, Top2a, Trim28 and Hmgb2 show strong binding towards all histone marks (Fig 6).

Identification of pluripotency associated MBFs

To identify pluripotency specific MBFs from the 143 putative MBFs (Table 3), the filtered MS data was overlaid onto normalized expression data comparing E14T mESCs in LIF and Retinoic acid (RA) (Terranova et al., 2015), pluripotent and

differentiation conditions respectively (Fig. 7A). The data were split into three bins based on normalized expression values, where the -2 to -0.5 expression values correspond to RA specific genes, -0.5 to +0.5 correspond to genes equivalently expressed in RA and LIF, and +0.5 to +2 corresponding to genes more highly expressed in pluripotent state. Each of the categories enabled the classification of putative MBFs into RA-specific MBFs (1 protein ID), generic MBFs (124 protein IDs) and pluripotency-specific MBFs (10 protein IDs), respectively. The list of putative regulatory MBFs was filtered based on GO terms for nucleus/nuclear, transcription and chromatin for further analysis, leaving a total of 51 putative generic and pluripotency-associated MBFs (Fig 7B). A similar analysis was performed to select for negative targets, except that the 68 H3 only proteins were first screened to check for identification in any of the three S10 replicates. If the protein was identified in at least one of the S10 replicate (with ≥2 unique peptides) it was excluded for analysis, narrowing down the list of putative negatives to 12 candidates. The list was then filtered as in Figure 7A, resulting in five stringently selected negative hits (Fig 7C).

In keeping with our original hypothesis that these putative bookmarking factors are retained on the mitotic chromatin to maintain the parental cellular identity, we performed a literature search of the selected putative 51 MBFs, looking for phenotypes upon knockout or knock-down of the factor (Tables S1-S3). For further investigation, we narrowed down the list to 31 candidates (Table S3) by filtering out factors that result in mitotic defects or DNA damage upon knockdown.

Validation of mitotic association of the putative MBFs.

After having identified and characterized putative bookmarking factors, we went on to validate the mitotic association of 20 of the chosen MBFs (Table S3) by either fusing their cDNAs to the fluorescent protein, mKO2 (Fig 8A) or by immunofluorescence, where cDNAs were not available (Table S4). Out of the 20 MBFs tested, a majority of them (70%, 14/20) were associated with the mitotic chromatin (Fig 8B, C, 9), while all H3-only proteins tested (Table S4) were excluded from the mitotic chromatin (Fig 9). The fluorescence intensity analysis of some of the target MBFs shows profiles similar to that of H2B-GFP (Fig 8C), albeit to varying degrees: strong association as with Hmgb2, Psip, and Dnmt3a to weak association as with Parp1 (Fig 8C).

The phenotype of mitotic chromatin association for some candidates differed from others. For instance, most factors, such as Utf1 and Uhrf1, showed clear nuclear localization during interphase followed by mitotic chromatin association, but Dnmt3L was dispersed in the cytoplasm for most of the cell cycle, but upon entry into mitosis was associated closely with the mitotic chromatin (Fig 9). Another factor, Ddx21, an RNA binding protein, showed a granular distribution throughout interphase but was strongly associated with the chromatin during mitosis (Fig 9). The chromatin remodelers, Rbbp4 and Khdr1 (Sam68), showed strong correlation with condensed chromatin at the onset of mitosis. However, during metaphase, they were diffused and not restricted only to the metaphase plate (Fig 9). All three of the heterogeneous nuclear ribonucleoproteins tested (HnrpC, HnrpK, and HnrpF) were excluded from the metaphase plate (Fig 9) indicating that this group represents false positives. Both of the non-mitotic bookmarking factors tested (Mta2 and Ssbp1) were not associated with the mitotic chromatin (Fig 9)

Discussion and Conclusion

We used a modified and targeted ChIP-MS approach to pull-down global and mitotic chromatin in an unbiased manner, and have identified putative pluripotency specific and generic mitotic bookmarking factors (Fig. 7B). Several pieces of data suggest that the pluripotency-specific proteins we identified are very highly expressed in ES cells as opposed to their differentiated counterparts, at both the RNA (Fig 7B) and protein level (Chaerkady et al., 2011; Van Hoof et al., 2006) (Terranova et al., 2015), suggesting that any putative bookmarking capabilities may be unique to ES cells. The 'generic' MBFs identified showed similar expression in RA conditions that promote ES cell differentiation and LIF conditions that maintain self-renewal, suggesting that these factors might play a role in both pluripotent and differentiated cell types. This raises the possibility that even though their bookmarking mechanism might be similar in a variety of cell types, for example binding a particular histone modification, their genetic targets would vary in a context-dependent manner, and that these factors might still play an important role in a pluripotency context.

Additionally, the ChIP-MS screen has identified chromatin bound proteins that are involved in important chromatin related regulatory cellular processes (Fig. 5 & Table 2), highlighting the success of the methodology. Current information on mitotically associated proteins is largely derived from metaphase chromosomes that are isolated on a sucrose gradient (Ohta et al., 2010; Uchiyama et al., 2005). These approaches have methodological disadvantages, including a requirement for large number of cells (between $3x10^8$ and $7.5x10^9$) and require ~90% mitotic enrichment, so are not feasible for a context specific survey of the mitotic chromatin in cells that struggle to meet these

requirements. Most problematic, current approaches result in a lot of contaminating proteins from the mitochondrial (Uchiyama et al., 2005) or other cytosolic compartments (Ohta et al., 2010) (Fig 5). While we were able to get the largest list of proteins putatively associated with metaphase chromosomes from Ohta et al. 2010, their methodology is less ideal for identification of putative bookmarking factors due to the potential dilution and overall under-representation of chromatin and transcriptional regulators amongst proteins involved in other regulatory processes such as cell cycle or metabolism (Fig 5).

Our approach was to immunoprecipitate chromatin, and this specifically enriched for proteins involved in chromatin remodeling and epigenetic mechanisms (Fig 5, 6, Table 2). However, our study yielded poor representation for transcription factors in the mass spectrometry data. This may be because sequence-specific transcription factors are relatively rare on the global chromatin scale, and bind a very small fraction of genomic loci. However, in the ChIP-MS technique we employed, the chance of discovery is directly correlated to the relative abundance, and transcription factors were underrepresented. It is important to note that the number of proteins identified in individual mass spectrometry runs was much greater (Table 1), but due to rigorous filtering out of proteins IDs unique to one or two replica we excluded some information. Scaling up the assay, and performing more replicas would result in a robust and greater detection of transcription factors. Additionally, a recent study suggested that fixation can impact the association and detection of chromatin bound factors in ChIP based and immunostaining assays (Teves et al., 2016). Since our methodology required a fixation step, proteins sensitive to this artifact would have gone undetected.

Out of the 143 putative MBFs identified, we narrowed down our list and validated a subset of proteins that are not involved in regulating cell cycle. Of the ones tested ~70% were associated with the mitotic chromatin during division (Fig 8, 9). However, it is worth mentioning that cell cycle-regulating proteins could also act as bookmarks, specifically those involved in chromatin remodeling such as Chd4, Sspr1, Dnmt3b, Set and Smarca5 (Table S1, S2).

Mitotic ChIP-MS also identified proteins that have previously been described to be retained on the mitotic chromatin, such as Utf1 (van den Boom et al., 2007), Dnmt3a (Easwaran et al., 2004), Uhrf1 (Rothbart et al., 2012), and Parp1 (Lodhi et al., 2014) (Table S3). One of the proteins, Parp1, was recently shown to be a mitotic bookmark that is associated with rapid reactivation of genes upon mitotic exit (Lodhi et al., 2014). Additionally, our dataset provides the opportunity to study various proteins in a novel mitotic specific context. The mitotic retention of splicing factors, RNA binding proteins and post-transcriptional regulators is novel and intriguing, and an investigation could result in novel regulatory mechanisms of gene regulation (Fig 5). We validated the mitotic association of RNA binding proteins Ddx21 and Litd1, both of which strongly co-localize with the mitotic chromatin (Fig 9), providing avenues for future exploration of roles in mitosis. Of note, Ddx21 has recently been shown to be highly involved in regulating RNA pol I and pol II transcription via small chromatin-bound RNAs (AJ et al., 2016; Calo et al., 2015), and the mitotic association of Ddx21 provides a unique opportunity to investigate an RNA mediated mechanism of transcription reactivation upon mitotic exit.

Importantly, our data presents the prospect to study mitosis by investigating the role of various chromatin regulators throughout cell cycle. While many of the identified MBFs associate with gene silencing, and heterochromatic epigenetic marks (Table 2, Fig 6), a significant number, such as Lmna, Lmnb1, Rbbp4, Smarca5, Ssrp1, and Parp1 also associate with various euchromatic marks (Fig 6). This association is not tested in embryonic stem cells, and importantly, has not yet been established during mitosis. Recent studies have shown that open chromatin marks, such as H3K27Ac (Liu et al., 2017b) and H3K4me3 (Grandy et al., 2016; Valls et al., 2005) are abundantly present during mitosis, and perhaps some of these MBFs are involved in the mitotic propagation of these marks. The current school of thought is that bookmarking factors are retained on the mitotic chromatin to mark active gene loci (Zhao et al., 2011, Kaduke et al., 2012, Caravaca et al., 2013, Lodhi et al., 2014, Festuccia et al., 2016, Liu et al., 2017), however the abundance of heterochromatin related proteins suggest that perhaps repressed gene loci are also bookmarked during mitosis. Mitotic bookmarks could, therefore, be important for retaining both the repressive and the active transcriptional memory of the cell during mitosis.

In summary, in this study we developed and deployed an unbiased approach to identifying mitotic chromatin associated factors in ES cells, which has led to the identification of novel mitotic-specific interactions of chromatin related proteins. The pool of candidates that we have identified using our ChIP-MS screen contains putative MBFs that require further validation, but the data provides many opportunities for investigating chromatin remodeling, epigenetic inheritance, and gene expression during mitosis in ES cells, by specific MBFs. Currently, there are significant gaps in the mitotic

bookmarking field of study, of which the paucity of MBF candidates is a major roadblock. The work presented helps surmount this roadblock, and provides a new methodology for MBF identification and a fresh pool of candidate MBFs for future validation. There still remain numerous outstanding questions that surround mitotic bookmarking, such as: what is the role and mechanism of histone mark inheritance during mitosis; is there a bias to retaining active or repressive transcription memory during division; the specificity of MBFs for specific epigenetic marks; the role of DNA methylation retention during division given that all four *de novo methyltransferases* were found to retained on the mitotic chromatin in our screen. The nature of the gene loci that are bookmarked during mitosis by putative MBFs we have identified could provide a useful entry point to understanding if these bookmarked sites control cell fate and phenotypic identity. In the next chapter, we will address the question of what gene loci are potentially bookmarked during mitosis, and explore potential bookmarking capabilities of the putative MBF, Parp1, at these loci.

Figures

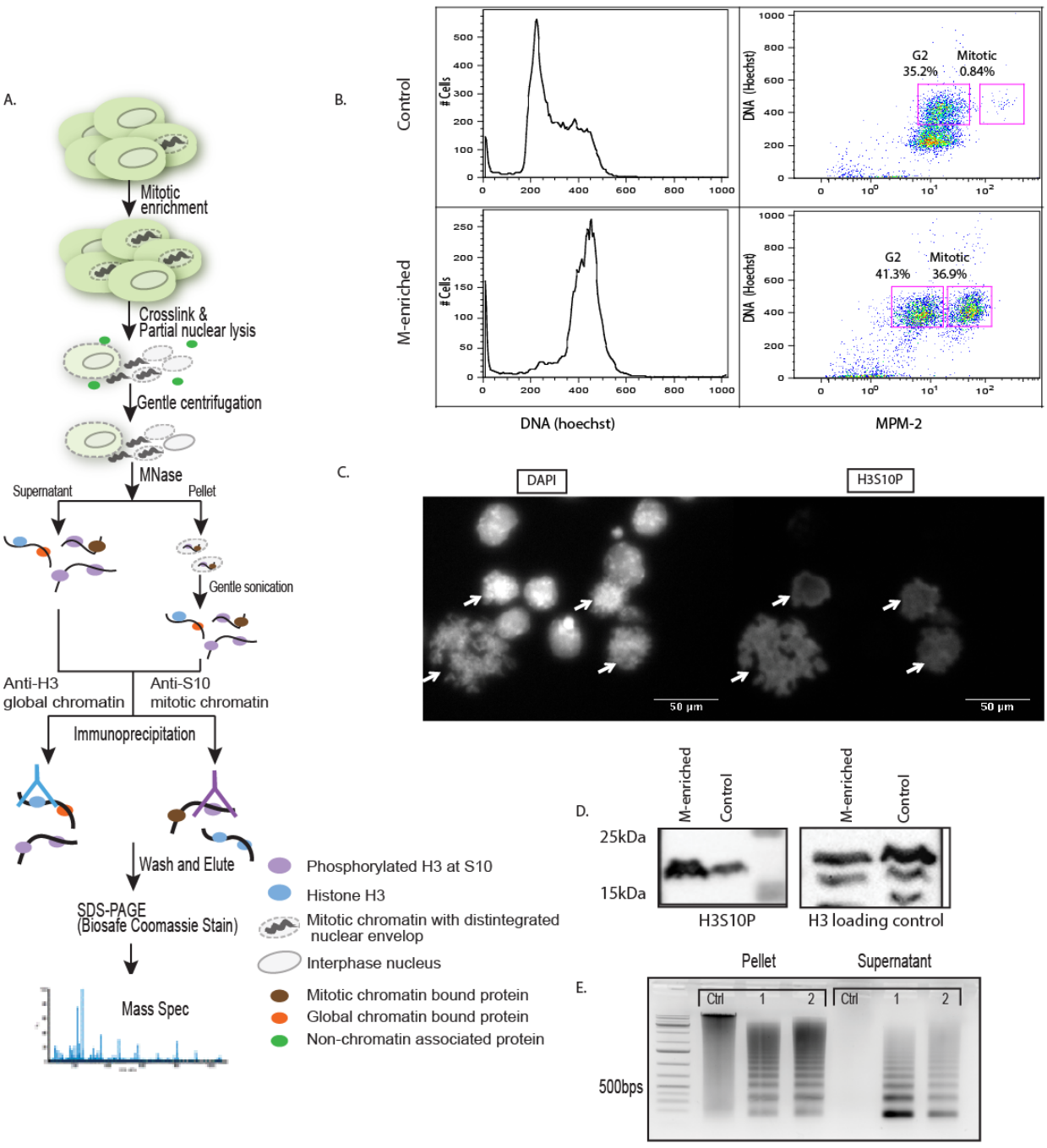

A.
Mitotic enrichment
Crosslink & Partial nuclear lysis
Gentle centrifugation
MNase
Supernatant
Pellet
Gentle sonication
Anti-H3 global chromatin
Anti-S10 mitotic chromatin
Immunoprecipitation
Wash and Elute
SDS-PAGE (Biosafe Coomassie Stain)
Mass Spec
Phosphorylated H3 at S10
Histone H3
Mitotic chromatin with distintegrated nuclear envelop
Interphase nucleus
Mitotic chromatin bound protein
Global chromatin bound protein
Non-chromatin associated protein
B.
Control
M-enriched
Cells
DNA (Hoechst)
G2 35.2%
Mitotic 0.84%
G2 41.3%
Mitotic 36.9%
DNA (hoechst)
MPM-2
C.
DAPI
H3S10P
50 µm
D.
M-enriched
Control
25kDa
15kDa
H3S10P
H3 loading control
E.
Pellet
Supernatant
Ctrl
1
2
500bps

Figure 1: Mitotic ChIP-MS strategy and approach. A.) Schematic showing the protocol used to identify global and mitotic chromatin associated proteins in ES cells. B.) Flow cytometry plots of the cell cycle profiles of control and mitotically enriched (M-enriched) ES cells, stained by Hoechst for DNA content and anti-MPM2 antibody (AF647 secondary) for staining mitotic cells. C.) Immunofluorescence of mitotically enriched ES cells cytospun onto slides and stained with Hoechst and anti-H3S10P. White arrows highlight mitotic cells, scale bar=50μm D.) Western blot of control and M-enriched ES cells showing the specificity of anti-H3S10P antibody compared to the anti-H3 loading control. E.) Agarose gel showing the MNase shearing of MNase treated supernatant and pellet after sonication; Ctrl= no MNase control, 1 & 2= two separate replicates

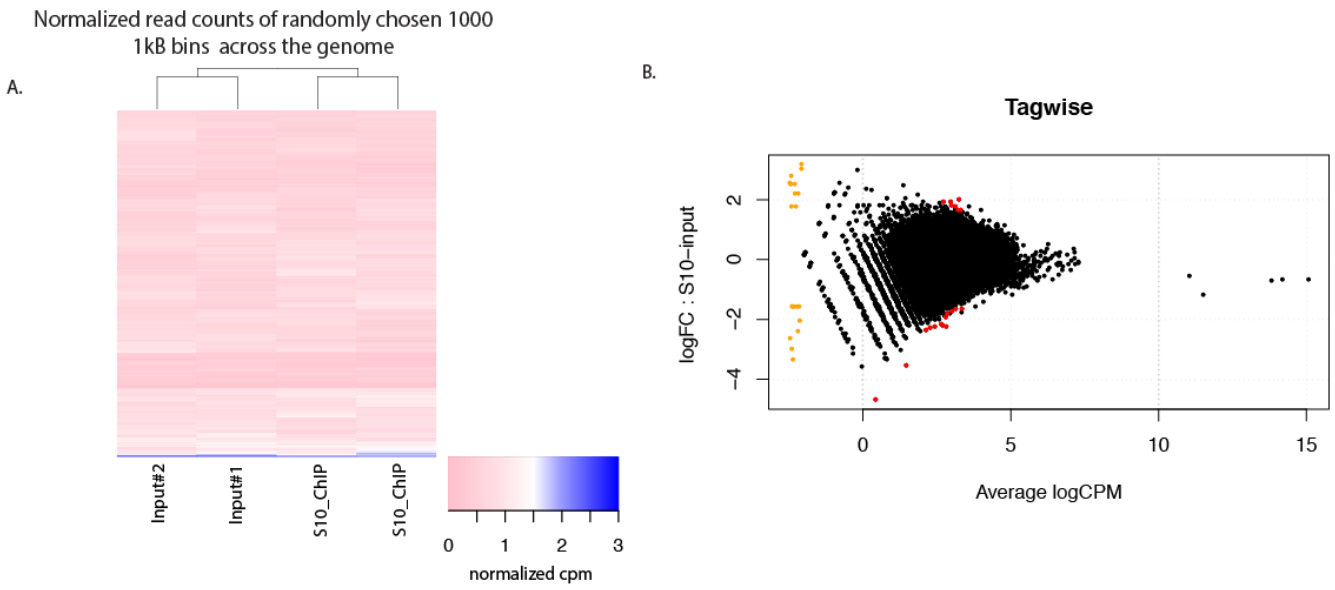


Figure 2. H3S10P is an unbiased marker for pulling down mitotic chromatin bound proteins A.) Heat-map showing the normalized read counts (cpm=counts per million) across 1000 randomly selected 1kb bins across the genome B.) Volcano plot showing differential read counts (log fold change) between H3S10P ChIP and background input controls versus the average read abundance (log cpm) (red: gene loci with significant fold change, orange: gene loci with zero counts in all samples, black: gene loci with no difference in read counts between H3S10 ChIP and input

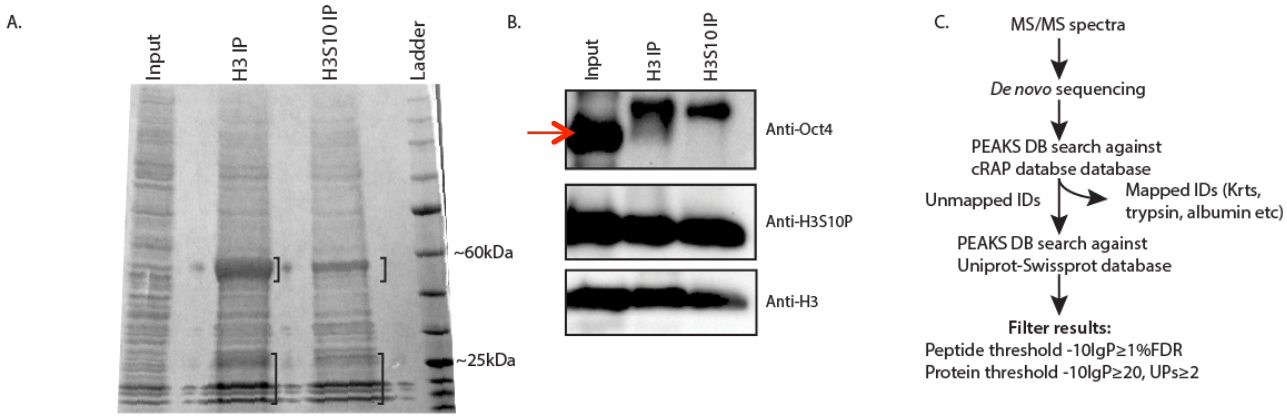


Figure 3. ChIP-MS protein analysis for identification of mitotic chromatin bound proteins. A.) Coomassie biosafe stained SDS PAGE protein gel, for one of the three immunoprecipitation (IP) replicates. H3 IP= IP with anti-H3 (global chromatin) and S10-IP= IP with anti-H3S10P (mitotic chromatin). Each lane was split into two MS runs, one with the bracketed regions and the other consisting of all other bands in the lane B.) Western blot from the IPs probed for Oct4 (non-mitotically associated), histone H3 and histone H3S10P. C.) Strategy to identify protein IDs from mass spectrometry data.

Table 1: Total number of protein IDs identified in each mass spectrometry sample after filtering out contaminants

Samples	Protein IDs>=1 peptide	Proteins IDs>=2 peptides
H3-1	659	459
H3-2	516	316
H3-3	561	313
S10-1	401	243
S10-2	463	269
S10-3	470	292

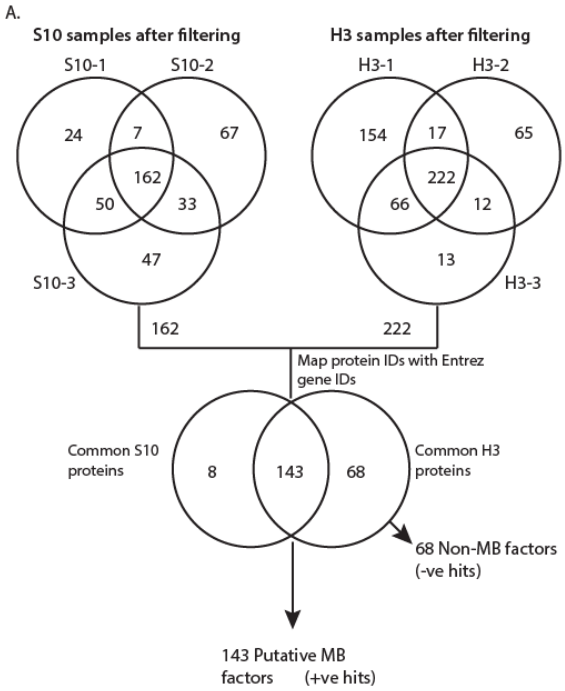


Figure 4. Identification of associated mitotic bookmarking factors (MBFs). A.) Filtration strategy used to identify transcription and chromatin regulating mitotic bookmarking factors (MBFs). We identified a total of 143 putative mitotic bookmarking (MB) factors, and 68 non-bookmarking factors (non MB factors).

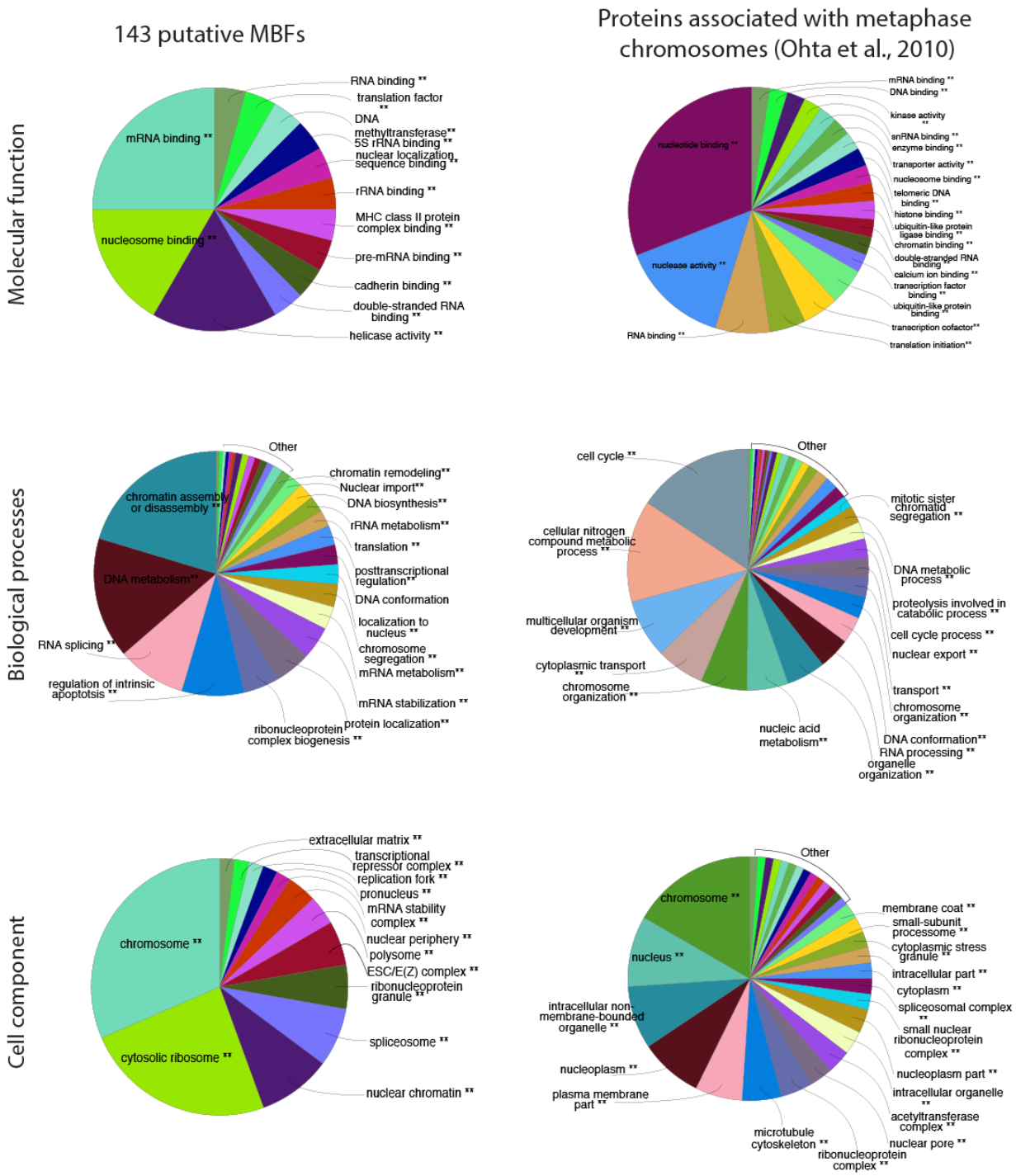

143 putative MBFs
Proteins associated with metaphase chromosomes (Ohta et al., 2010)
Molecular function
Biological processes
Cell component
mRNA binding **
translation factor **
DNA methyltransferase**
5S rRNA binding **
nuclear localization sequence binding **
rRNA binding **
MHC class II protein complex binding **
pre-mRNA binding **
cadherin binding **
double-stranded RNA binding **
helicase activity **
nucleosome binding **
nucleotide binding **
nuclease activity **
RNA binding **
mRNA binding **
DNA binding **
kinase activity **
snRNA binding **
enzyme binding **
transporter activity **
nucleosome binding **
telomeric DNA binding **
histone binding **
ubiquitin-like protein ligase binding **
chromatin binding **
double-stranded RNA binding **
calcium ion binding **
tranecription factor binding **
ubiquitin-like protein binding **
transcription cofactor**
translation initiation**
Other
chromatin remodeling**
Nuclear import**
DNA biosynthesis**
rRNA metabolism**
translation **
posttranscriptional regulation**
DNA conformation
localization to nucleus **
chromosome segregation **
mRNA metabolism**
mRNA stabilization **
protein localization**
ribonucleoprotein complex biogenesis **
chromatin assembly or disassembly **
DNA metabolism**
RNA splicing **
regulation of intrinsic apoptotsis **
cell cycle **
cellular nitrogen compound metabolic process **
multicellular organism development **
cytoplasmic transport **
chromosome organization **
mitotic sister chromatid segregation **
DNA metabolic process **
proteolysis involved in catabolic process **
cell cycle process **
nuclear export **
transport **
chromosome organization **
DNA conformation**
RNA processing **
organelle organization **
nucleic acid metabolism**
extracellular matrix **
transcriptional repressor complex **
replication fork **
pronucleus **
mRNA stability complex **
nuclear periphery **
polysome **
ESC/E(Z) complex **
ribonucleoprotein granule **
spliceosome **
nuclear chromatin **
chromosome **
cytosolic ribosome **
chromosome **
nucleus **
intracellular non-membrane-bounded organelle **
nucleoplasm **
plasma membrane part **
microtubule cytoskeleton **
ribonucleoprotein complex **
membrane coat **
small-subunit processome **
cytoplasmic stress granule **
intracellular part **
cytoplasm **
spliceosomal complex **
small nuclear ribonucleoprotein complex **
nucleoplasm part **
intracellular organelle **
acetyltransferase complex **
nuclear pore **

Figure 5. Characterization of the 143 S10 and H3 overlapping proteins IDs identified by mass-spectrometry. ClueGO (Bindea et al., 2009) analysis of the molecular functions, biological processes and cellular component terms significant over-represented in dataset. Similar enrichment analysis was performed for previously identified proteins associated with metaphase chromosome spreads by Ohta et al. 2010. ** significantly enriched categories p<0.005

Table 2: Putative MBFs that are associated with significantly enriched chromatin regulating bioprocesses based on ClueGO enrichment analysis (corrected Pvalue<0.001)

GOID	GOTerm	Term PValue Corrected with Bonferroni step down	Group PValue Corrected with Bonferroni step down	% Associated Genes	Nr. Genes	Associated Genes Found
GO:0006333	chromatin assembly or disassembly	29.0E-15	35.0E-24	12.50	16.00	[Cenpv, H2afy, Hist1h1a, Hist1h1b, Hist1h1c, Hist1h1d, Hist1h1e, Hist2h2aa1, Hmga1, Hmgb1, Mcm2, Npm1, Rbbp4, Set, Smarca5, Supt16]
GO:0034728	nucleosome organization	68.0E-12	35.0E-24	11.40	13.00	[H2afy, Hist1h1a, Hist1h1b, Hist1h1c, Hist1h1d, Hist1h1e, Hist2h2aa1, Mcm2, Npm1, Rbbp4, Set, Smarca5, Supt16]
GO:0016458	gene silencing	790.0E-9	35.0E-24	6.06	12.00	[Dnmt1, Dnmt3a, Dnmt3b, H2afv, H2afy, H2afz, Hist2h2aa1, Pabpc1, Rbbp4, Smarca5, Trim28, Uhrf1]
GO:2001251	negative regulation of chromosome organization	720.0E-6	35.0E-24	6.09	7.00	[Dnmt1, Dnmt3b, H2afy, Hnrnpa1, Hnrnpc, Hnrnpu, Parp1]
GO:0000183	chromatin silencing at rDNA	4.3E-3	35.0E-24	11.11	4.00	[Dnmt1, Dnmt3b, Rbbp4, Smarca5]
GO:0043044	ATP-dependent chromatin remodeling	8.0E-3	35.0E-24	9.09	4.00	[Chd4, Hnrnpc, Rbbp4, Smarca5]
GO:0090116	C-5 methylation of cytosine	220.0E-6	35.0E-24	60.00	3.00	[Dnmt1, Dnmt3a, Dnmt3b]
GO:0006304	DNA modification	210.0E-6	35.0E-24	7.53	7.00	[Dnmt1, Dnmt3a, Dnmt3b, Dnmt3l, Parp1, Trim28, Uhrf1]
GO:0010608	posttranscriptional regulation of gene expression	9.7E-12	8.1E-15	5.47	21.00	[Ddx3x, Dhx9, Eef2, Eif4a3, Eif5a, Hnrnpa0, Hnrnpc, Hnrnpl, Hnrnpu, Igf2bp1, Khdrbs1, Khsrp, Ncl, Npm1, Rpl13a, Rpl26, Rpl30, Rpl5, Rps3, Rps4x, Rps9]
GO:0051568	histone H3-K4 methylation	180.0E-6	14.0E-15	10.34	6.00	[Dnmt1, Dnmt3b, H2afy, Hist1h1c, Hist1h1d, Hist1h1e]
GO:0048255	mRNA stabilization	190.0E-6	14.0E-15	15.15	5.00	[Dhx9, Hnrnpa0, Hnrnpc, Hnrnpu, Igf2bp1]
GO:0043484	regulation of RNA splicing	1.7E-9	11.0E-12	10.34	12.00	[Ddx5, Eif4a3, Hnrnpa1, Hnrnpf, Hnrnph1, Hnrnpk, Hnrnpl, Hspa8, Khdrbs1, Npm1, Ptbp1, Srsf3]
GO:0006338	chromatin remodeling	140.0E-6	920.0E-9	6.40	8.00	[Cenpv, Chd4, Hmga1, Hnrnpc, Rbbp4, Smarca5, Tcp1, Top1]

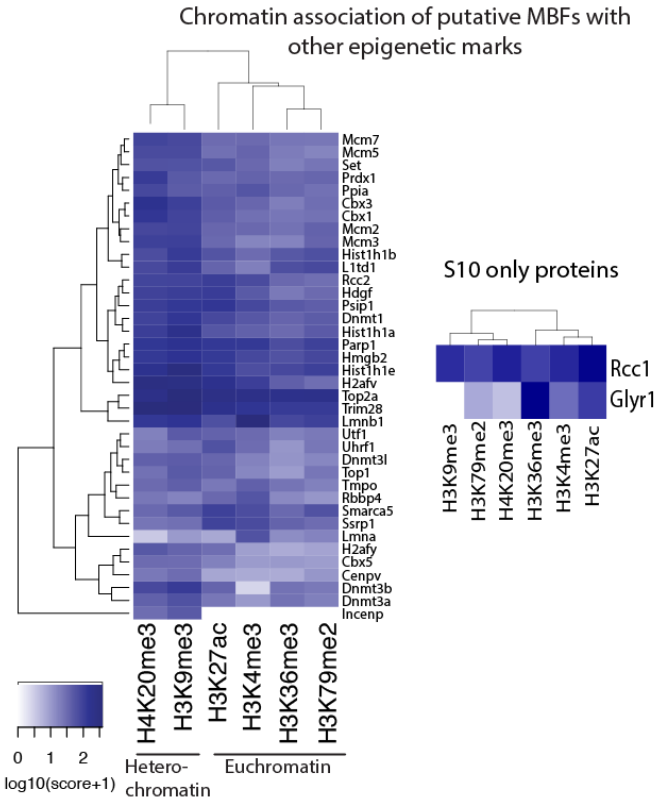


Figure 6. Epigenetic marks associated with putative MBFs. Heat-map showing the relative abundance of putative bookmarking factors when pulled down by other epigenetic marks. Abundance is represented as log10 of MS_score+1. Epigenetic mark data is from Ji et al., 2015, and only the proteins from our dataset that overlapped with theirs are shown. Left panel represents proteins characterized as putative MBFs and on the right (S10 only proteins) are the proteins that were unique to the S10 datasets are not present in H3 pulldown.

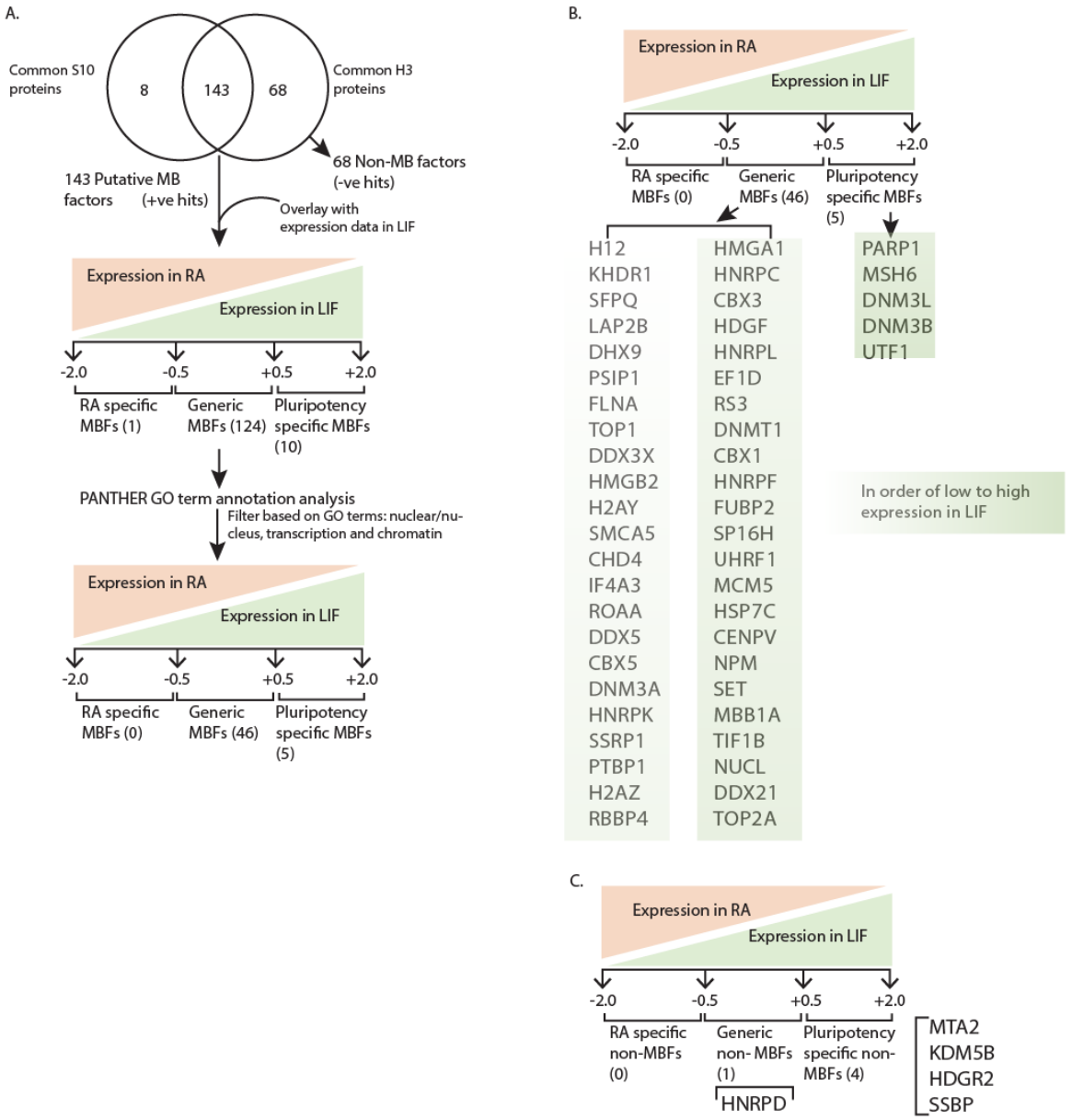

A.
Common S10 proteins
8
143
68
Common H3 proteins
68 Non-MB factors (-ve hits)
143 Putative MB factors (+ve hits)
Overlay with expression data in LIF
Expression in RA
Expression in LIF
-2.0
-0.5
+0.5
+2.0
RA specific MBFs (1)
Generic MBFs (124)
Pluripotency specific MBFs (10)
PANTHER GO term annotation analysis
Filter based on GO terms: nuclear/nucleus, transcription and chromatin
Expression in RA
Expression in LIF
-2.0
-0.5
+0.5
+2.0
RA specific MBFs (0)
Generic MBFs (46)
Pluripotency specific MBFs (5)
B.
Expression in RA
Expression in LIF
-2.0
-0.5
+0.5
+2.0
RA specific MBFs (0)
Generic MBFs (46)
Pluripotency specific MBFs (5)
H12
KHDR1
SFPQ
LAP2B
DHX9
PSIP1
FLNA
TOP1
DDX3X
HMGB2
H2AY
SMCA5
CHD4
IF4A3
ROAA
DDX5
CBX5
DNM3A
HNRPK
SSRP1
PTBP1
H2AZ
RBBP4
HMGA1
HNRPC
CBX3
HDGF
HNRPL
EF1D
RS3
DNMT1
CBX1
HNRPF
FUBP2
SP16H
UHRF1
MCM5
HSP7C
CENPV
NPM
SET
MBB1A
TIF1B
NUCL
DDX21
TOP2A
PARP1
MSH6
DNM3L
DNM3B
UTF1
In order of low to high expression in LIF
C.
Expression in RA
Expression in LIF
-2.0
-0.5
+0.5
+2.0
RA specific non-MBFs (0)
Generic non- MBFs (1)
Pluripotency specific non-MBFs (4)
HNRPD
MTA2
KDM5B
HDGR2
SSBP

Figure 7. Identification of pluripotency associated mitotic bookmarking factors. A.) Filtration strategy used to identify pluripotency associated MBFs. The 143 putative MBFs were overlapped with RNA-seq data and categorized into bins from -2 to -0.5, -0.5 to +0.5 and +0.5 to +2 based on fold changed in LIF vs RNA. Based on GO analysis Nuclear, chromatin and transcription related factors were selected for further characterization. B& C.) List of putative (B.) MBFs and (C.) non-MBFs sorted on their relative expression in pluripotent (LIF) conditions.

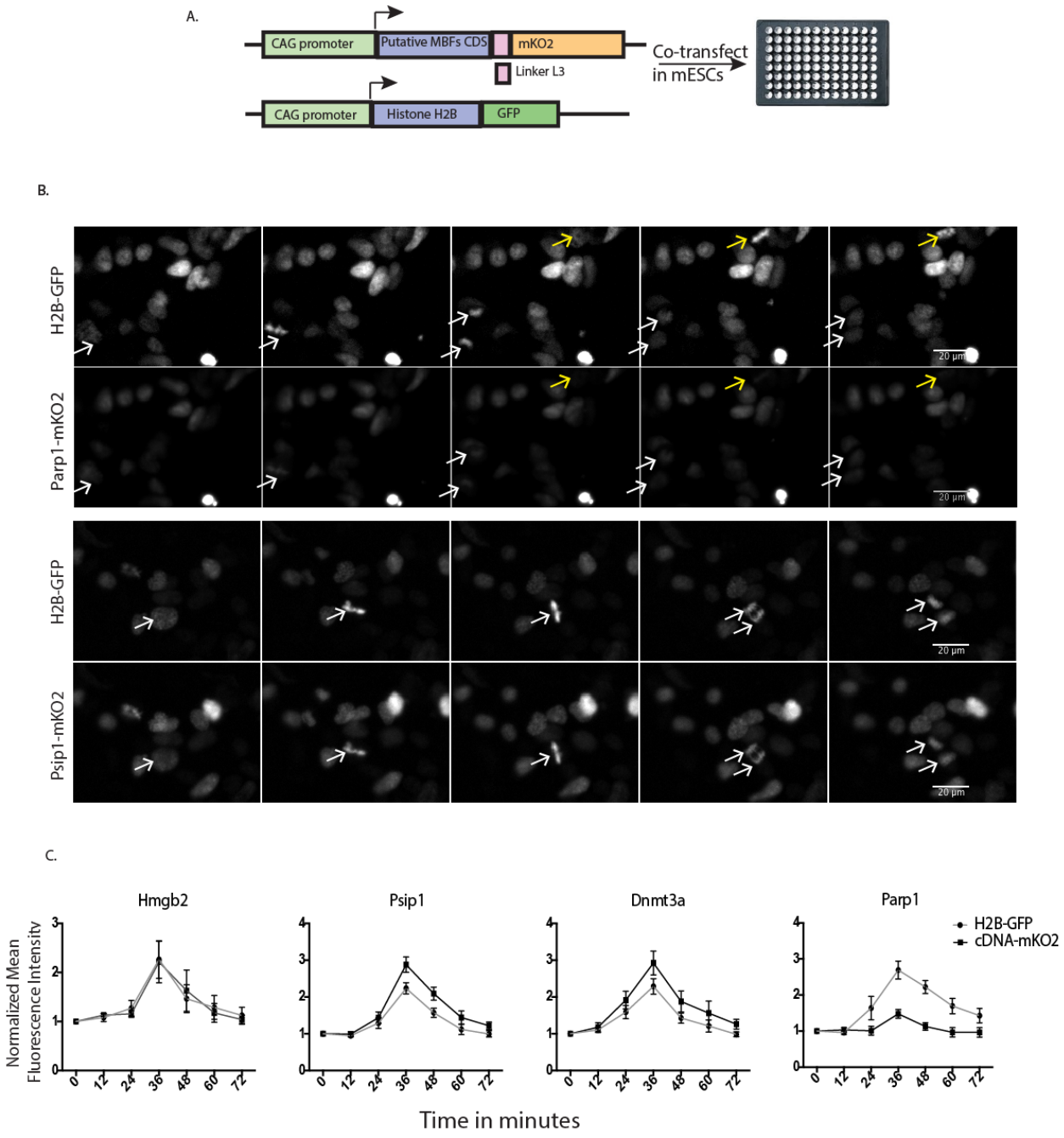
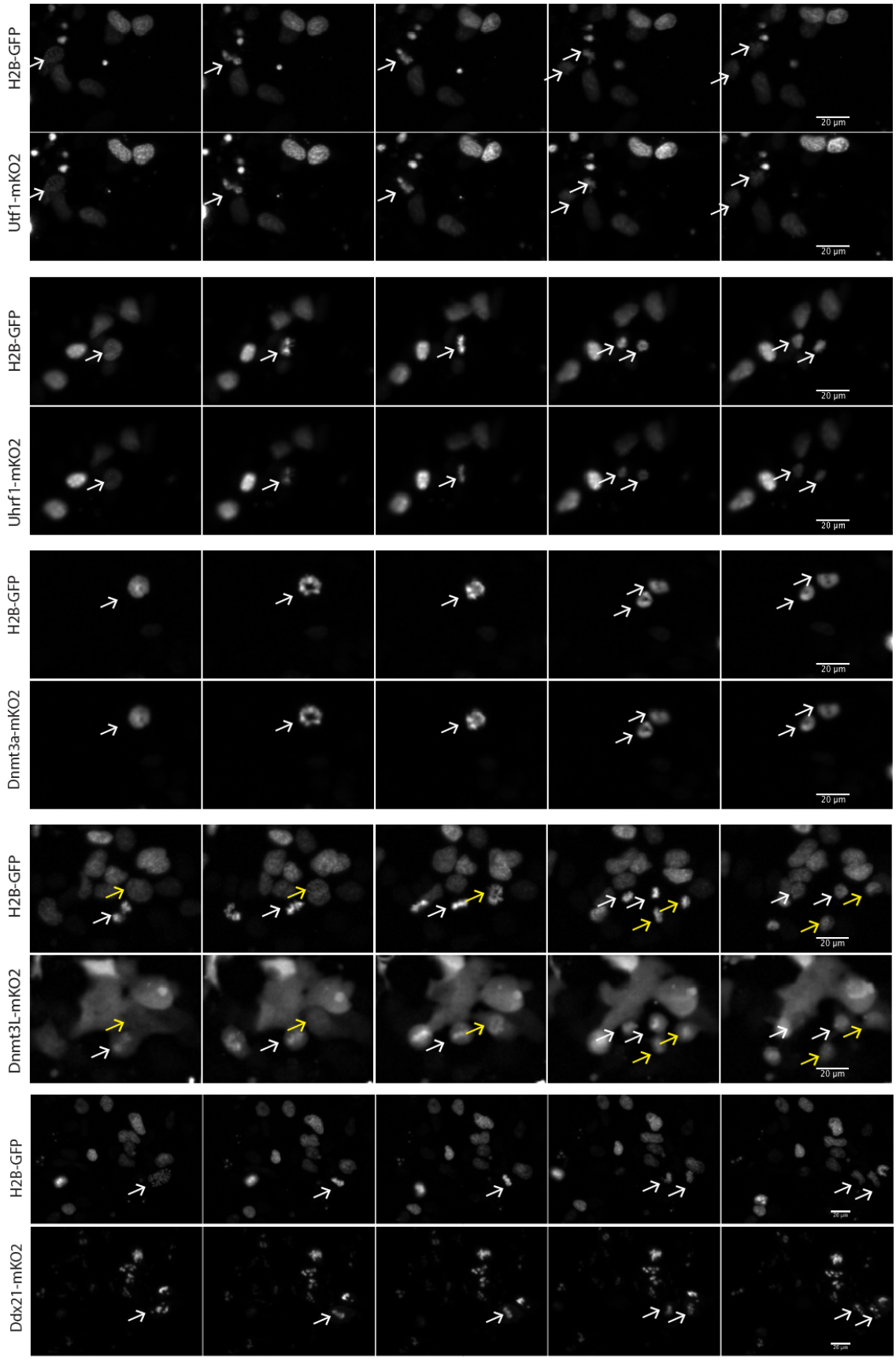
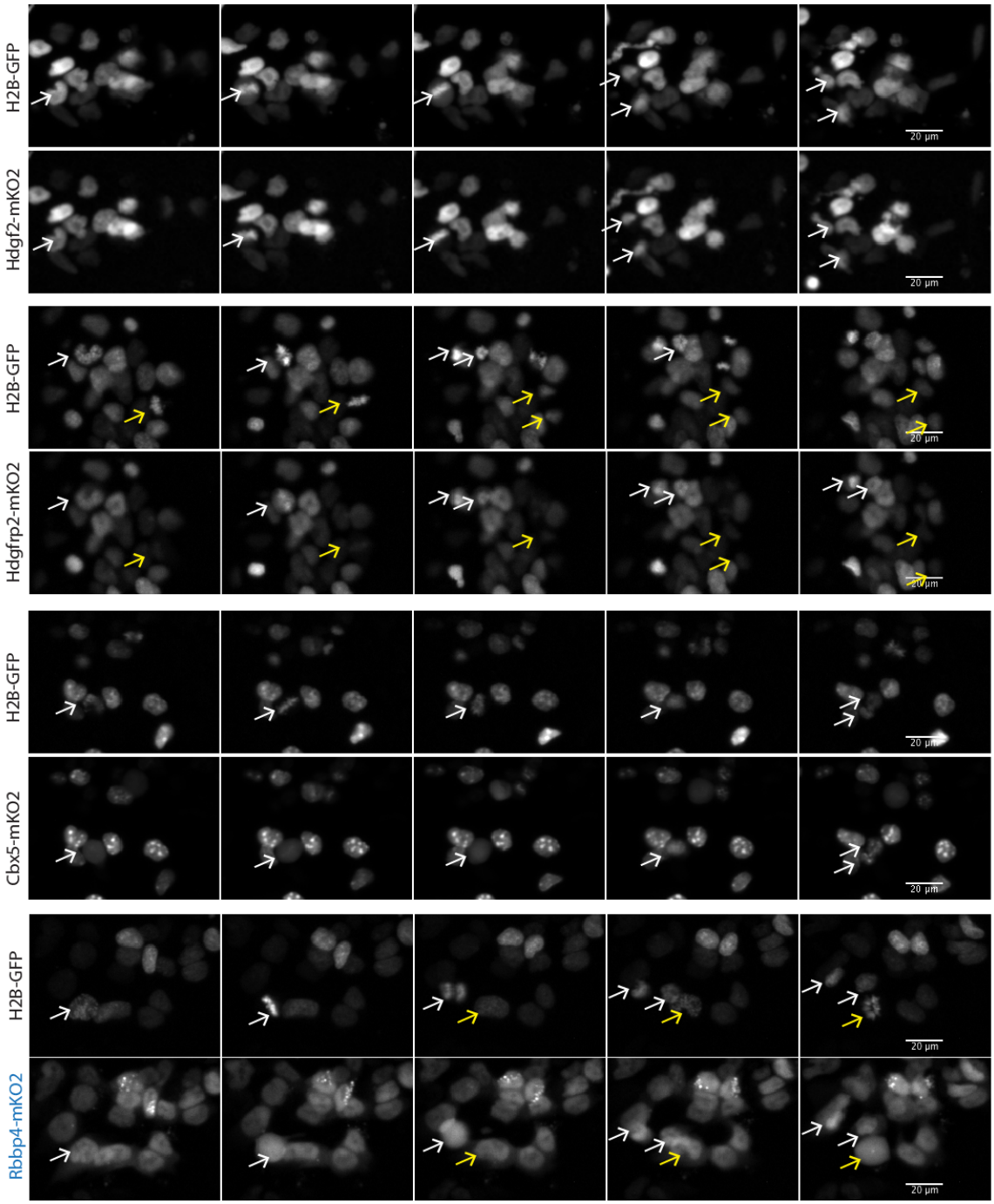
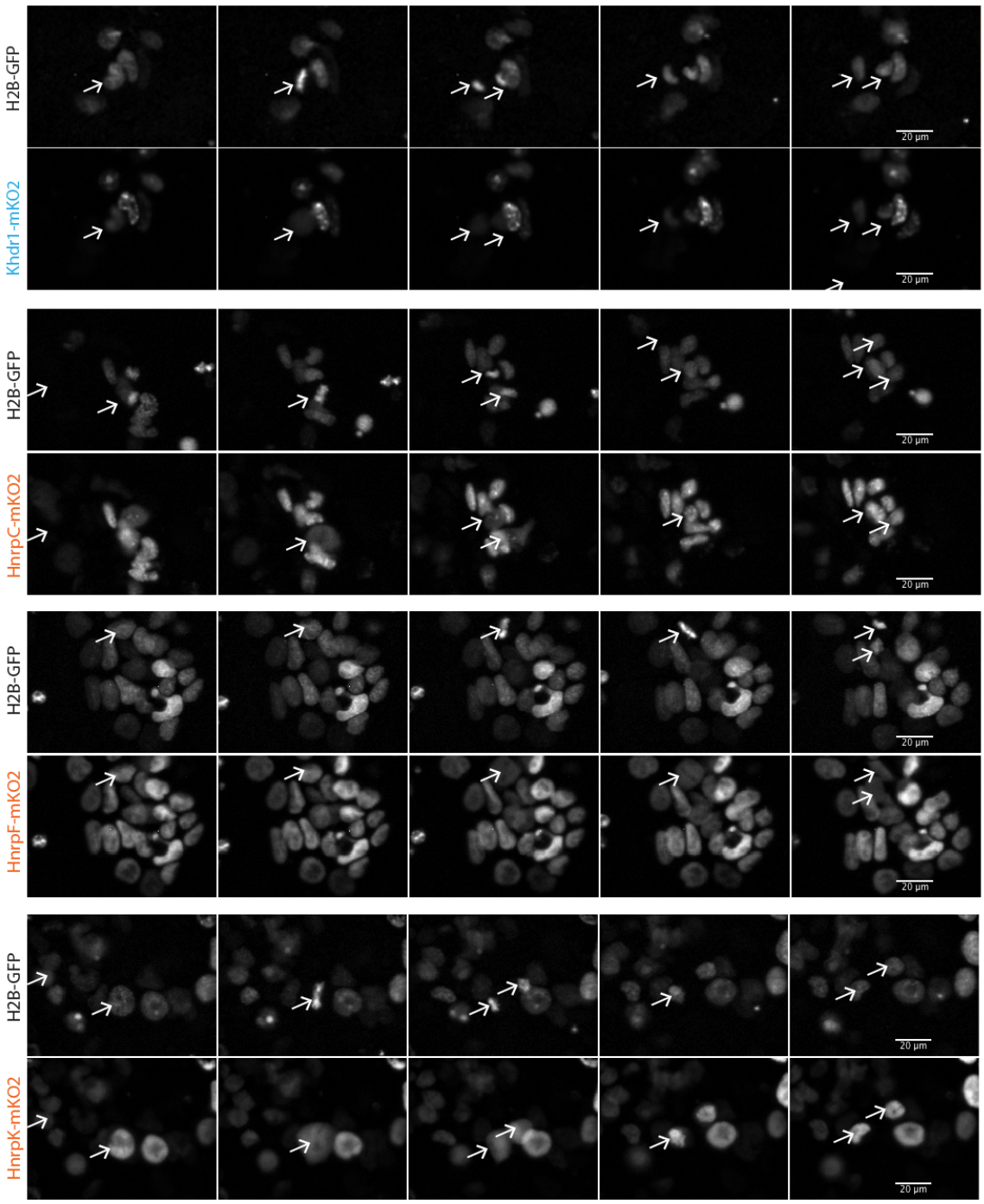

A.
CAG promoter
Putative MBFs CDS
mKO2
Linker L3
CAG promoter
Histone H2B
GFP
Co-transfect
in mESCs
B.
H2B-GFP
Parp1-mKO2
H2B-GFP
Psip1-mKO2
20 µm
C.
Hmgb2
Psip1
Dnmt3a
Parp1
H2B-GFP
cDNA-mKO2
Normalized Mean
Fluorescence Intensity
Time in minutes

Figure 8. Validation of the mitotic association of putative MBFs A.) Experimental scheme to validate the mitotic association of candidate MBFs by live cell imaging. Cells were co-transfected with cDNA-mKO2 fusion and H2B-GFP fusion constructs and imaged for 2hrs at 4minute intervals. B.) Representative images for putative MBFs (Parp1 and Psip1). Arrows point towards a dividing cell. Scale =20um C.) Normalized mean fluorescence intensity measurements for candidate MBFs as cells undergo mitosis. n=4 (Hmgb2), n= 8 (Psip1), n=8 (Dnmt3a), n=6 (Parp1). To capture the complete mitotic event graphs were generated from t-9 frames, where t was the frame with the highest intensity representing a metaphase plate.

Figure 9. (Below, four pages) Validation of the mitotic association of putative MBFs A.) Representative live cell images for other putative MBFs tested, along with H2B-GFP. Arrows point towards a dividing cell. Yellow and white arrows represent multiple mitotic events in the same frame. Scale =20um. cDNA= MBFs diffused during mitosis but were not excluded from the mitotic chromatin, cDNA= excluded from the mitotic chromatin, cDNA= associated strongly with the mitotic chromatin. Mta2, and Ssbp1 are the non-MBFs (S10 negatives) tested


H2B-GFP
Utf1-mKO2
H2B-GFP
Uhrf1-mKO2
H2B-GFP
Dnmt3a-mKO2
H2B-GFP
Dnmt3L-mKO2
H2B-GFP
Ddx21-mKO2
20 µm


H2B-GFP
Hdgf2-mKO2
H2B-GFP
Hdgfrp2-mKO2
H2B-GFP
Cbx5-mKO2
H2B-GFP
Rbbp4-mKO2
20 µm


H2B-GFP
Khdr1-mKO2
H2B-GFP
HnrpC-mKO2
H2B-GFP
HnrpF-mKO2
H2B-GFP
HnrpK-mKO2
20 µm

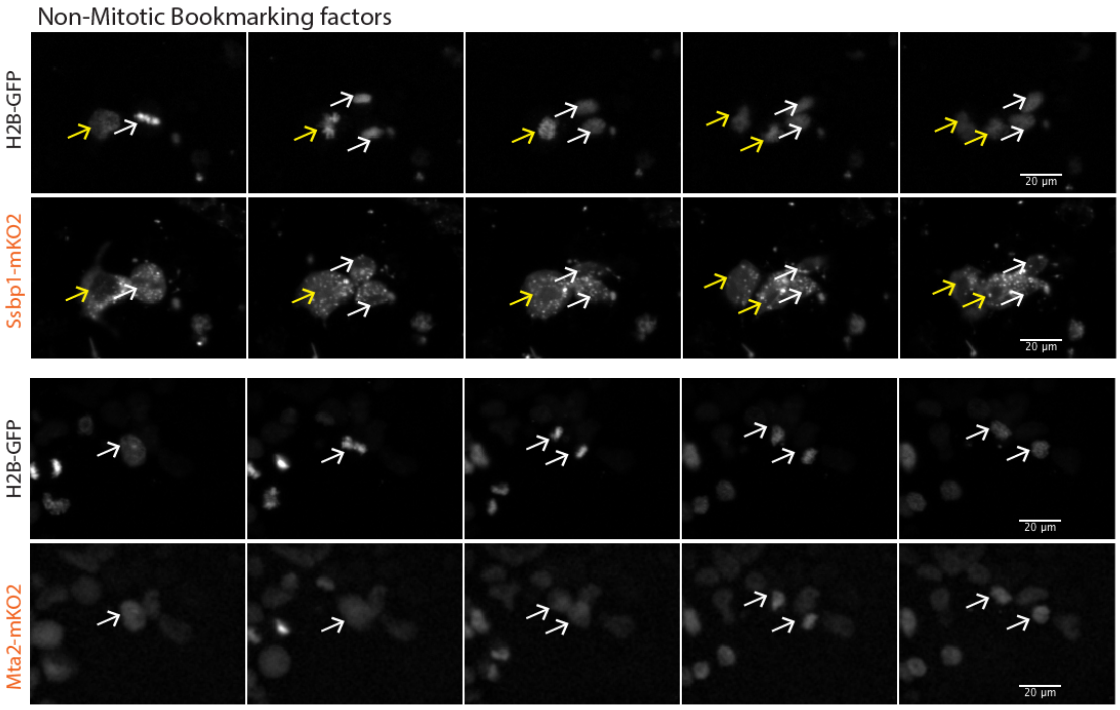

Non-Mitotic Bookmarking factors
H2B-GFP
Ssbp1-mKO2
H2B-GFP
Mta2-mKO2
20 µm

Table 3: Complete list of the putative MBFs identified, along with the mass-spectrometry peak scores, and fold change values (normalized expression) in LIF compared to RA conditions (2 pages)

		MS Peak Score						
Putative MBFs	Protein_ID	S10-1	S10-2	S10-3	H3-1	H3-2	H3-3	Normalized expression in LIF compared to RA
UTF1_MOUSE	Q6J1H4	109.67	91.2	143.67	154.31	140.43	176.59	2.0156212
LITD1_MOUSE	Q587J6	113.02	170.5	199.76	169.06	186.62	235.85	2.0411978
DNM3B_MOUSE	O88509	120.57	114.74	166.54	179.17	158.53	178.66	1.4405861
DNM3L_MOUSE	Q9CWR8	132.58	125.21	174.17	165.54	142.99	149.21	0.8377609
MSH6_MOUSE	P54276	101.79	87.49	140.22	164.53	127.52	171.59	0.7905941
HSPB1_MOUSE	P14602	80.78	154.79	105.67	101.84	184.51	112.91	0.79298973
PGK1_MOUSE	P09411	53.39	125.35	132.65	79.8	136.31	146.08	0.6789799
HS90A_MOUSE	P07901	139.48	203.56	192.74	187.55	237.52	249.45	0.7170563
PARP1_MOUSE	P11103	126.7	191.84	161.38	202.53	136.82	163.01	0.61285305
LMNA_MOUSE	P48678	129.89	210.3	227.78	162.95	188.69	251.68	0.5516634
KPYM_MOUSE	P52480	83.47	176.09	121.96	108.93	168.89	128.73	0.539999
ENOA_MOUSE	P17182	135.06	192.64	189.6	158.69	294.58	264.82	0.33584976
TOP2A_MOUSE	Q01320	197.13	211.05	263.72	280.33	222.74	321.28	0.26179695
DDX21_MOUSE	Q9JIK5	166.29	173.67	208.49	234.97	205.89	259.95	0.49538136
NUCL_MOUSE	P09405	180.01	226.76	248.06	257.28	284.19	326.16	0.610404
ROA3_MOUSE	Q8BG05	150.5	115.56	156.32	197.82	147.75	231.39	0.4477396
TIF1B_MOUSE	Q62318	191.07	220.68	271.33	235.15	298.78	351.32	0.44473267
MCM3_MOUSE	P25206	68.53	75.59	116.59	138.54	114.82	197.14	0.39921284
EF2_MOUSE	P58252	91.67	195	170.8	194.15	219.92	228.72	0.31895447
MBB1A_MOUSE	Q7TPV4	159.76	161.83	206.06	214.2	158.3	221.92	0.40689945
RL29_MOUSE	P47915	83.39	98.77	128.99	101.42	113.19	133.06	-0.003119469
SET_MOUSE	Q9EQU5	62.33	84.33	122.17	114.54	87.41	117.74	0.6707649
MYH9_MOUSE	Q8VDD5	112.35	446.56	198.9	151.98	460.71	334.94	0.09022236
NPM_MOUSE	Q61937	195.86	249.52	222.28	227.64	269.19	339.76	0.49975777
PAIRB_MOUSE	Q9CY58	110.39	105.58	113.77	191.62	112.1	106.34	0.4524355
PABP1_MOUSE	P29341	59.12	87.76	130.96	88.09	69.84	94.03	0.41602135
RL8_MOUSE	P62918	131.74	151.44	139.36	148.67	139.74	210.1	0.35197067
CENPV_MOUSE	Q9CXS4	48.79	139.21	82.12	132.44	90.6	94.21	0.39480972
HSP7C_MOUSE	P63017	169.8	273.98	256.99	222.68	294.65	330.17	0.3806858
MCM6_MOUSE	P97311	83.12	45.56	125.34	126.44	114.19	176.82	0.42368317
RL3_MOUSE	P27659	128.31	111.64	131.47	187.48	182.91	196.78	0.31113815
RL28_MOUSE	P41105	100.96	129.71	118.2	161.15	174.42	193.84	0.302186
RL4_MOUSE	Q9D8E6	180.73	168.1	233.21	240.64	247.13	281.25	0.40075874
RL30_MOUSE	P62889	63.24	57.88	62.73	137.67	100.15	131.65	0.37894917
MCM5_MOUSE	P49718	98.4	70.1	138.39	120.54	101.11	177.67	0.36904716
RL14_MOUSE	Q9CR57	108.94	118.88	124.3	143.6	176.83	179.47	0.27376938
RL23A_MOUSE	P62751	100.2	109.79	114.61	165.51	175.83	186.59	0.29267406
RS4X_MOUSE	P62702	136.04	190.81	150.75	178.09	169.45	228.05	0.3810463
RCC2_MOUSE	Q8BK67	160.81	192.59	214.3	171.21	164.52	190.04	0.36276245
IPO5_MOUSE	Q8BKC5	49.41	77.84	83.23	57.93	49.61	55.44	0.3485403
RS2_MOUSE	P25444	113.59	121.15	130.28	121.72	101.73	160.97	0.35983467
RL12_MOUSE	P35979	88.73	156.65	139.34	134.02	161.05	194.24	0.18281364
RL10_MOUSE	Q6ZWV3	127.11	151.46	153.63	176.82	179.52	195.03	0.2928629
UHRF1_MOUSE	Q8VDF2	74.9	61.34	96.16	137.28	88.06	106.2	0.29533577
MCM2_MOUSE	P97310	84.92	80.85	92.08	121.5	72.45	136.27	0.31054497
SP16H_MOUSE	Q920B9	101.09	136.83	136.9	210.19	149.65	210.83	0.35726357
IMB1_MOUSE	P70168	79.71	94.45	113.41	116.1	122.49	145.81	0.27930164
IF4A1_MOUSE	P60843	115.5	167.72	165.21	121.79	167.66	174.62	0.4453802
RL26_MOUSE	P61255	116.03	144.14	120.66	169.65	174.79	174.9	0.30155563
RL13_MOUSE	P47963	127.63	133.05	143.6	173.44	237.81	218.44	0.3137684
RL27_MOUSE	P61358	89.38	68.02	101.31	162.44	134.75	136.92	0.2983637
HS90B_MOUSE	P11499	158.28	196.85	206.31	216.32	257.27	272.55	0.31961632
PRDX1_MOUSE	P35700	88.09	94.01	93.88	112.77	128.38	148.02	0.49136925
FUBP2_MOUSE	Q3U0V1	85.54	77.07	111.8	147.91	126.06	209.29	0.32692146
RL6_MOUSE	P47911	150.35	163.19	186.7	164.13	201.98	256.47	0.2596035
HNRPF_MOUSE	Q9Z2X1	144.86	102.74	161.63	171.62	93.98	180.36	0.36139965
IF5A1_MOUSE	P63242	81.01	117.73	73.54	134.89	129.06	121.9	0.5202484
CBX1_MOUSE	P83917	94.63	95.73	103.46	144.22	165.18	170.09	0.3218851
RL21_MOUSE	O09167	62.51	96.36	105.32	133.16	124.43	146.36	0.28993797
NOP2_MOUSE	Q922K7	82.23	80.33	122.46	126.38	72.99	120.11	0.31312466
RL7A_MOUSE	P12970	163.56	147.12	154.09	216.29	150.58	242.6	0.24067497
RS20_MOUSE	P60867	104.34	153.54	119.69	147.47	157.89	162.61	0.30631924
DNMT1_MOUSE	P13864	57.84	35.01	91.1	126.19	68.29	114.74	0.0652504
RL10A_MOUSE	P53026	81.56	74.43	132.1	147	154.97	175.48	0.25081158
MCM7_MOUSE	Q61881	92.38	132.03	173	146.99	196.08	248.93	0.23360062
RRS1_MOUSE	Q9CYH6	53.37	71.41	117.52	113.66	132	152.65	0.25920963
RL18_MOUSE	P35980	120.72	154.09	166.04	178.38	183.75	185.5	0.17068577
RL7_MOUSE	P14148	122.62	93.21	185.64	175.18	140.26	223.77	0.32143784

RS3_MOUSE	P62908	114.4	142.68	108.04	147.28	103.75	149.32	0.18379402
RS16_MOUSE	P14131	71.06	132.28	76.11	124.11	164.28	137.91	0.06770897
RS3A_MOUSE	P97351	157.92	130.83	134.38	158.07	116.26	204.95	0.2780018
RL24_MOUSE	Q8BP67	104.13	124.15	145.96	140.33	162.69	172.18	0.19452477
PPIA_MOUSE	P17742	68.52	164.02	158.02	133.14	170.66	179.35	0.31924915
EF1G_MOUSE	Q9D8N0	112.27	93.87	133.11	139.63	192.85	192.68	0.28319263
EF1D_MOUSE	P57776	47.22	115.12	136.99	101.19	97.26	157.5	0.21950531
EF1A1_MOUSE	P10126	188.64	216.55	190.32	235.63	280.48	258.69	0.19044876
RL11_MOUSE	Q9CXW4	102.73	152.53	139.26	137.09	166.2	156.23	0.22263908
RS15A_MOUSE	P62245	94.23	167.95	68.18	170.95	191.31	171.72	0.1836977
RS7_MOUSE	P62082	101.91	167.49	127.49	197.91	197.21	177.68	0.20681381
RS12_MOUSE	P63323	95.09	122.22	101.46	122.15	104.38	126.97	0.06772995
RL5_MOUSE	P47962	102.82	90.57	150.72	136.27	139.44	190.35	0.12741756
H13_MOUSE	P43277	224.88	310.75	287.75	244.45	295.14	343	0.9284706
HNRPL_MOUSE	Q8R081	108.62	60.76	132.34	147.72	126.2	160.1	0.27976418
RL18A_MOUSE	P62717	116.35	91.26	78.66	146.34	128.88	160.51	0.24042511
RL13A_MOUSE	P19253	122.92	76.78	95.51	144.73	149.22	174.68	0.066967964
INCE_MOUSE	Q9WU62	81.49	102.94	106.49	159.36	96.83	114.4	0.083109856
RS9_MOUSE	Q6ZWN5	110.87	168.33	142.8	155.01	166.4	157.83	0.19768906
HDGF_MOUSE	P51859	86.88	132.79	138.02	147.33	124	152.28	0.2697115
CBX3_MOUSE	P23198	121.22	152.9	181.85	161.08	186.21	193.46	0.3217106
H15_MOUSE	P43276	210.36	298.38	283.67	229.81	274.89	311.37	1.395544
RL15_MOUSE	Q9CZM2	116.36	89.17	101.5	160.98	178.68	180.8	0.28746605
HNRPC_MOUSE	Q9Z204	85.47	108.11	108.06	140.49	149.64	197.15	0.19274712
HMGA1_MOUSE	P17095	45.98	122.13	144.86	145.82	124.88	149.65	0.2569666
RL27A_MOUSE	P14115	71.94	100.76	122.81	177.89	165.79	166.81	0.24435997
IMA1_MOUSE	P52293	88.78	122.24	144.32	151.51	126.82	188.35	0.26095772
RS8_MOUSE	P62242	136.33	117.28	131.22	176.29	195.01	185.42	0.12450695
RL9_MOUSE	P51410	55.19	79.98	85.47	74.93	162.22	121.55	0.24909782
RS11_MOUSE	P62281	120.76	135.67	140.1	147.52	171.16	169.7	0.07344341
RBBP4_MOUSE	Q60972	124.47	97.61	122.34	194.99	113.17	148.61	0.2260046
H2AZ_MOUSE	P0C0S6	98.52	199.15	192.32	134.23	168.27	211.4	0.22065353
ROA1_MOUSE	P49312	132.74	130.61	154.84	167.49	141.96	185.55	0.11215401
TBB5_MOUSE	P99024	215.96	298.69	247.89	260.74	302.96	300.87	0.21878624
SRSF3_MOUSE	P84104	126.9	126.48	138.1	139.05	155.89	179.7	0.29718685
PTBP1_MOUSE	P17225	101.36	126.32	100.59	137.91	100.59	151	0.11719513
TCPA_MOUSE	P11983	57.22	39.03	103.43	97.52	78.98	153.67	0.12405872
SSRP1_MOUSE	Q08943	56.72	59.33	141.08	133.26	112.76	185.36	0.15133286
HNRPK_MOUSE	P61979	133.43	87.89	180.87	156.69	135.99	204.07	0.23636055
RS23_MOUSE	P62267	89.56	129.44	80.39	101.93	82.84	139.61	0.06407547
DNM3A_MOUSE	O88508	76.08	68.61	139.6	160.08	116.46	116.82	0.14112854
RL17_MOUSE	Q9CPR4	98.96	132.08	142.67	210.14	172.66	184.73	0.0790844
H2AV_MOUSE	Q3THW5	98.52	199.15	192.32	134.23	168.27	211.4	0.21310043
CLH1_MOUSE	Q68FD5	91.79	217.17	137.92	102.22	222.3	182.15	-0.011450768
HMGB1_MOUSE	P63158	134.74	167.81	160.6	183.06	152.18	184.05	0
H2A2A_MOUSE	Q6GSS7	177.47	283.78	271.47	259.38	272.46	274.6	0
RL34_MOUSE	Q9D1R9	55.77	70.92	40.21	101.57	97.55	108.74	0
G3P_MOUSE	P16858	152.07	254.68	224.38	217.72	237.49	289.17	0
CBX5_MOUSE	Q61686	74.22	119.66	114.63	166.16	131.4	151.19	-0.03255844
DDX5_MOUSE	Q61656	133.47	154.13	210.29	232.73	150.33	187.38	0.05992031
ROAA_MOUSE	Q99020	91.11	82.35	118.05	100.7	115.46	165.14	0.22601414
IF4A3_MOUSE	Q91VC3	109.08	114.82	138.4	125.31	130.41	145.42	0.15531635
CHD4_MOUSE	Q6PDQ2	91.2	55.21	137.32	179.09	116.7	193.55	-0.16739082
SMCA5_MOUSE	Q91ZW3	148.79	162.21	209.75	195.58	166.5	238.49	0.11253452
H2AY_MOUSE	Q9QZQ8	145.59	132.81	189.82	184.31	142.2	219.46	0.028535843
HMGB2_MOUSE	P30681	171.04	195.33	189.87	184.73	183.05	227.57	0.12150288
LMNB1_MOUSE	P14733	172.98	234.2	277.4	225.38	288.26	327.24	0.009764671
DDX3X_MOUSE	Q62167	100.58	146.72	158.95	150.37	151.73	116.98	0.009571075
TOP1_MOUSE	Q04750	114.38	130.46	156.41	170.13	110.12	123.53	-0.015608788
FLNA_MOUSE	Q8BTM8	84.23	280.03	58.32	119.72	307.28	222.7	-0.25331593
PSIP1_MOUSE	Q99JF8	167.64	233.87	204.44	245.99	212.05	235.47	0.058997154
HNRPU_MOUSE	Q8VEK3	187.72	210.96	217.21	218.26	238.21	275.92	-0.091340065
DHX9_MOUSE	O70133	99.69	69.87	135.1	179.65	92.6	138.49	-0.007488251
IF2B1_MOUSE	O88477	85.65	140.35	148.39	143.19	156.03	185.62	-0.04246235
LAP2B_MOUSE	Q61029	111.19	149.36	173.85	194.4	204.83	196.84	-0.090922356
SFPQ_MOUSE	Q8VIJ6	147.24	144.35	169.47	186.45	148.63	227.42	-0.4206934
HNRPM_MOUSE	Q9D0E1	179.46	162.1	214.93	175.11	177.61	276.73	-0.08760834
HNRH1_MOUSE	O35737	141.87	132.4	139.32	171.16	99.68	184.69	-0.12410545
ROA0_MOUSE	Q9CX86	82.14	78.67	102.91	97.02	92.23	131.09	-0.011597633
KHDR1_MOUSE	Q60749	63.37	65.49	78.9	108.75	91.28	115.62	-0.24306297
H11_MOUSE	P43275	211.5	309.11	291.99	244.92	252.56	314.97	-0.16702795
H12_MOUSE	P15864	216.01	308.43	290.85	237.93	289.31	329.52	-0.41924
H14_MOUSE	P43274	219.94	301.99	282.73	241.56	282.07	342.26	0.21973419
VIME_MOUSE	P20152	189.19	358.21	254.22	240.76	347.47	296.03	-0.67206955
H4_MOUSE	P62806	242.81	325.73	303.97	273.81	332.84	327.2	0.12418604

Supplementary material

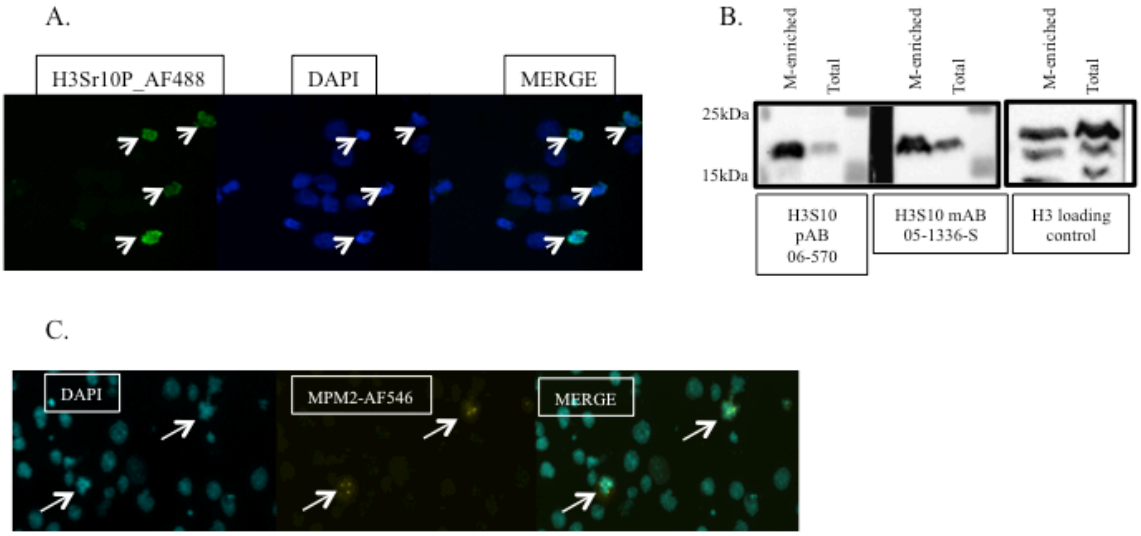


Figure S1: A &B) Immunostaining (A) and western blot (B) showing the specificity of two different H3S10P antibodies. C) Immunostaining showing the specificity of the MPM2 antibody used to check for M-enriched cells

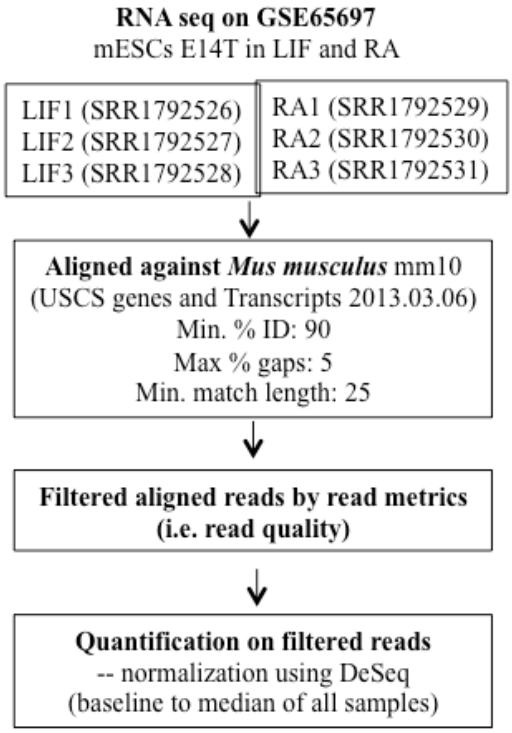


Figure S2: Strand-NGS RNA seq analysis pipeline for GEO dataset GSE65697

Table S1: List of putative MBFs that result in some form of mitotic defect upon depletion

Putative MBFs associated with Mitotic Defects			
Putative MBFs	**Literature evidence for mitotitc association**	**KO phenotype**	**KO/KD Phenotype**
Top2a	Yes	Top2a knockout mouse are not viable, conditional knockout during mitosis results in mitotic failure and cell death PMID: 15456904	mitotic defect
Cbx1	Yes	KO results in impaired neural differentiation and genomic instability via mitotic defects (PMID: 19015315), KO ES results in increased retroviral gene expression (PMID: 21774827)	Differentiation impairment, mitotic defect
Dnmt3b	Yes (using GFP-Dnm3b) PMID: 15148359, and 19482874	siDnm3b results in mitotic defects (PMID: 19482874)	mitotic defect
Ncl		knockdown results in impaired microtubule organization	mitotic defect
Npm1	Associated with centrosomes (PMID: 12214246)	RNAi results in distortion of nucleolar and nuclear structures PMID: 18729828	mitotic defect
Cenpv	Yes, localizes to kinetochores (PMID: 18772885)	Depletion results in mitotic defects PMID: 18772885	mitotic defect
Rps3	Localizes to the mitotic spindle PMID: 23131551	Depletion causes mitotic arrest in metaphase PMID: 23131551	mitotic defect
Ssrp1	Yes (only associated with alpha tubulin during mitosis) PMID: 19995907	Knockdown results in impaired mitosis	mitotic defect
Top1	yes	functional mutant results in mitotic defects (PMID: 8895658)	mitotic defect
Chd4	associated with mitotic spindle PMID: 24268414	Depletion results in spindle assembly defects PMID: 24268414	mitotic defect
Sfpq	not known	knockdown results in impaired chromosome segregation and impairment in DSB repair (PMID: 20813759)	DNA repair, mitotic defect

Table S2 : List of putative MBFs that result in some form of DNA damage upon depletion

Putative MBFs associated with DNA damage			
MBFs	**Literature evidence for mitotitc association**	**KO phenotype**	**KO/KD Phenotype**
Msh6	Not known	Involved in mismatch repair, defects result in increased mutation rate	DNA repair
Set	Not known	siRNA depletion results in increase DNA damage response	DNA repair
Sfpq	not known	knockdown results in impaired chromosome segregation and impairment in DSB repair (PMID: 20813759)	DNA repair, mitotic defect
Smarca5	possibly, is recruited to damaged DNA, colocalizes with H2A gamma PMID: 23264744	Depletion results in senstitivity to DNA damage PMID: 23264744	DNA repair
Dhx9	not known	depletion results in genomic instability PMID: 24049074	DNA repair

Table S3: List of putative pluripotency specific MBFs to follow up on ranked in order of priority. Putative MBFs were prioritized first based on the phenotype upon depletion. Factors in the same phenotypic category were prioritized based on high to low expression in LIF conditions.

Putative MBFs ranked in order of priorty			
MBFs	**Literature evidence for mitotitc association**	**KO phenotype**	**KO/KD Phenotype**
Trim28	Not associated by IF beyond prophase (PMID: 19442252)	Trim28-/- ES cells gradual loose self-renewal+aberrant gene expression PMID: 20075919	Stem cell maintenance
Rbbp4	Not known	siRNA depletion results in a more differentiation characteristic of hESCs (PMID: 21689726)	Stem cell maintenance
Utf1	Yes (using GFP-UTF1) PMID 17785516	knockdown results in delayed differentiation (PMID: 20715181)	Differentiation impairment
Dnmt3l	Not known	Dnmt3L-/- ESCs are morphologically normal, but show impairment of PGC markers and differentiation to PGCs (PMID: 24074865)	Differentiation impairment
Parp1	Yes (using IF) PMID: 24861619	Parp1-/- ES cells have high differentiation propensity towards trophoblast PMID: 12729565	Differentiation bias
Hnrnpl	not known	Conditional KO results in impaired T-cell differentiation	Differentiation impairment
Hmga1-rs1	Yes (using GFP-Hmga1) PMID: 15213251	Hmga1-/- ES show impaired hematopoetitic differentiation (PMID: 12824305)	Differentiation impairment
Cbx5	Yes PMID: 10460410	knockdown impairs neural differentiation (PMID: 17627279), KO ES results in increased retroviral gene expression (PMID: 21774827)	Differentiation impairment
Hmgb2	yes PMID: 12925773	hmgb2-/- MSCs have increase osteogenic potential PMID: 21890638	Differentiation impairment
Psip1	yes, showed by gfp fusion	psip1 -/- mice show skeletal defects	Differentiation impairment
Khdrbs1	not known, localizes to chromatin upon genotoxic stress PMID: 21355037)	Sam68-/- cells have differentiation bias towards osteogeneic differentiation (PMID: 16362077)	Differentiation impairment
Uhrf1	Yes (using a peptide array binding) PMID: 23022729	depletion affects DNA snythesis, Uhrf1-/- ESCs show embryonic lethality but ESC cultures are normally maintainable (PMID: 17673620, 12084726), methylation defects PMID: 23463006	DNA synthesis, Epigenetic
Dnmt1	Yes (using GFP-Dnmt1) PMID: 15550930	DNMT1 -/- hESCs in global demethylation and cell death PMID: 25822089, only demethylation in mESCs PMID: 16824199	epigenetic
Supt16h	Not known	in HeLas depletion results in decrease nucleosome occupancy on active ene bodies PMID: 23325844	chromatin remodelling, epigenetic
Dnmt3a	Yes PMID:	methylation defects PMID: 16824199	epigenetic
Khsrp	No direct evidence, is associated with permangate sensitive DNA of closed chromatin PMID: 8940189	knockdown affects RNA processing PMID: 19458619	Transcription regulation
Hnrnpf	not known	not known	Transcription regulation
Hnrnpc	not known	knockdown results in increased exonization of Alu repeats (PMID: 23374342)	Transcription regulation
Hnrnpk	No direct evidence, linked to FBP2 PMID: 8940189	knockdown results in increased pre-mRNA levels PMID: 23857582	Transcription regulation
Ddx21	Indirectly shown by IP of mitotic phosphoprotein phosphatases PMID: 22761809	knockdown impairs the production of nascent RNA transcripts from bound promoters PMID: 25470060	RNA regulation
Mybbp1a	Localizes with phospho-H3 at anaphase at parachromosomal region (PMID: 23056166)	Cell cycle and growth rate defects upon downregulation in HeLa PMID: 23056166	Cell cycle

Putative MBFs ranked in order of priority (contd.)			
MBFs	**Literature evidence for mitotic association**	**KO phenotype**	**KO/KD Phenotype**
Hdgf	Not associated by IF of phosp-Hdgf (PMID: 21489262)	knockdown affects cell growth PMID: 20043047	cell proliferation
H2afz	Yes PMID: 20864037	Depletion impairs proliferation (PMID: 24240188)	cell proliferation
Ddx5	Yes, by IF (PMID: 22034099)	KO mice are lethal after E11.5	developmental defects
Eif4a3	not known	RNA binding protein, knockdown results in embryo development defects PMID: 20549732	developmental defects
Ddx3x	not known	Knockdown in zygotes results in impaired embryo development PMID: 25050112	developmental defects
Hspa8	not known (localizes to nucleus upon stress)	not known	not known
Eef1d	No, localizes in a peri-metphase plate ring during mitosis PMID: 14618264	not known	not known
Flna	not known	flna-/- mice are lethal with multiple organ defects	survival
Hnrnpab	not known	not known	not known
H2afy		KO has little affects in mice	not known

Table S4: Genes corresponding to the bioprocesses enriched in S10 and H3 overlap dataset

Function	Group Genes
DNA conformation change	Cenpv|Ddx3x|H2afy|Hist1h1a|Hist1h1b|Hist1h1c|Hist1h1d|Hist1h1e|Hist2h2aa1|Hmga1|Hmgb1|Hmgb2|Mcm2|Mcm3|Mcm5|Mcm6|Mcm7|Npm1|Parp1|Rbbp4|Set|Smarca5|Top1|Top2a
DNA topological change	Hmgb2|Top1|Top2a
RNA splicing	Ddx21|Ddx3x|Ddx5|Eif4a1|Eif4a3|Hnrnpa1|Hnrnpa3|Hnrnpc|Hnrnpf|Hnrnph1|Hnrnpk|Hnrnpm|Hnrnpu|Hspa8|Khsrp|Mybbp1a|Npm1|Pabpc1|Psip1|Ptbp1|Sfpq|Srsf3|Tcp1
cellular response to interleukin-4	Hsp90ab1|Mcm2|Rpl3|Rps2
chromatin assembly	Cenpv|Ddx3x|Dnmt1|Dnmt3b|H2afy|Hist1h1a|Hist1h1b|Hist1h1c|Hist1h1d|Hist1h1e|Hist2h2aa1|Hmga1|Hmgb1|Hmgb2|Mcm2|Mcm3|Mcm5|Mcm6|Mcm7|Npm1|Rbbp4|Set|Smarca5|Top1|Top2a
chromatin remodeling	Cenpv|Chd4|Dnmt1|Dnmt3b|Hmga1|Hnrnpc|Rbbp4|Smarca5|Tcp1|Top1
cytoplasmic translation	Eef2|Rpl15|Rpl26|Rpl6|Rpl7|Rpl8|Rpl9
mRNA stabilization	Dhx9|Dnmt1|Dnmt3b|H2afy|Hmgb1|Hnrnpa0|Hnrnpc|Hnrnpu|Igf2bp1|Msh6|Parp1|Rps3|Tcp1|Trim28
mRNA transcription	Ddx5|Eef1d|Flna
mesenchymal cell proliferation	Hmga1|Hmgb1|Lmna|Tcp1
nucleosome assembly	Chd4|Dnmt1|Dnmt3a|Dnmt3b|Dnmt3l|H2afv|H2afy|H2afz|Hist1h1a|Hist1h1b|Hist1h1c|Hist1h1d|Hist1h1e|Hist2h2aa1|Hnrnpc|Hnrnpk|Hnrnpu|Ipo5|Mcm2|Msh6|Npm1|Pabpc1|Parp1|Rbbp4|Rps3|Set|Smarca5|Tcp1|Trim28|Uhrf1
platelet aggregation	Flna|Hspb1|Myh9
positive regulation of cellular amide metabolic process	Ddx3x|Eef2|Eif4a1|Eif4a3|Eif5a|Hnrnpa1|Igf2bp1|Khdrbs1|Khsrp|Npm1|Rpl13a|Rps4x|Rps9|Srsf3|Tcp1
positive regulation of protein localization to nucleus	Flna|Hmgb1|Hmgb2|Hnrnpm|Hsp90ab1|Ipo5|Khdrbs1|Kpna2|Kpnb1|Npm1|Parp1|Rps3|Tcp1|Trim28
positive regulation of translation	Ddx3x|Eef2|Eif4a1|Eif4a3|Eif5a|Khdrbs1|Npm1|Rcc2|Rpl13a|Rps4x|Rps9|Tcp1
regulation of nuclease activity	Ddx3x|Ddx5|Flna|Hmga1|Hmgb1|Hmgb2|Hnrnpc|Hnrnpu|Hsp90ab1|Ipo5|Msh6|Npm1|Parp1|Ppia|Rps3|Top2a|Trim28
regulation of transcription from RNA polymerase I promoter	Flna|H2afy|Ncl
regulation of translational initiation	Ddx3x|Eif4a1|H2afy|Khdrbs1|Lmnb1|Npm1|Rcc2|Rpl13a|Sfpq|Tcp1
regulation of viral genome replication	Ddx3x|Ddx5|Hmgb1|Ppia|Rps3|Top2a|Trim28
ribonucleoprotein complex biogenesis	Ddx21|Ddx3x|Eif4a3|Ncl|Nop2|Npm1|Psip1|Rpl10|Rpl11|Rpl12|Rpl13a|Rpl14|Rpl23a|Rpl24|Rpl26|Rpl3|Rpl34|Rpl5|Rpl6|Rpl7|Rpl7a|Rps16|Rps7|Rps8|Rrs1
ribosome biogenesis	Ddx21|Ddx3x|Eif4a3|Ncl|Nop2|Npm1|Psip1|Rpl10|Rpl11|Rpl12|Rpl13a|Rpl14|Rpl23a|Rpl24|Rpl26|Rpl3|Rpl34|Rpl5|Rpl6|Rpl7|Rpl7a|Rps16|Rps7|Rps8|Rrs1
spindle assembly	Chd4|Cltc|Flna|Kpnb1|Myh9|Rps3|Tubb5
translation	Ddx3x|Eef1a1|Eef1d|Eef1g|Eef2|Eif4a1|Eif4a3|Eif5a|Igf2bp1|Khdrbs1|Npm1|Rpl10|Rpl10a|Rpl11|Rpl12|Rpl13|Rpl13a|Rpl14|Rpl15|Rpl17|Rpl18|Rpl18a|Rpl21|Rpl23a|Rpl24|Rpl26|Rpl27|Rpl27a|Rpl28|Rpl3|Rpl30|Rpl34|Rpl4|Rpl5|Rpl6|Rpl7|Rpl8|Rpl9|Rps11|Rps12|Rps15a|Rps16|Rps2|Rps20|Rps23|Rps3|Rps3a1|Rps4x|Rps7|Rps8|Rps9
translational elongation	Eef1a1|Eef1d|Eef1g|Eef2|Eif5a|Rps9

Table S5: Validation of the mitotic association of the putative MBFs pulled down in S10P ChIP. Putative MBFs in red did not show localization with the mitotic chromatin.

Putative MBF	pCAG-cDNAL 3mko2	Mitotic association	cDNA origin	cDNA-F	cDNA-R	Infusion cloning F primer	Infusion R primer
Uhrf1	✓	+	TCAG			caaagaattcctcgaggccaccatgtggatccaggttcgaactat	ccagcttcatCTCGAgccggccgctgccatagcc
Parp1	✓	+	TCAG			caaagaattccTCGAGgccaccatggcggaggcctcg	ccagcttcatCTCGAgccacagggatgtcttaaaattgaacttgag
Utf1	✓	+	E14T cDNA	ATGCTGCTTCGTCCCCGG	TTATTGGCGCAAGTCCCCAAG	caaagaattccTCGAGgccaccATGCTGCTTCGTCCCCGG	ccagcttcatCTCGAgTTGGCGCAAGTCCCCAAGGA
Dnmt3l	✓	+	TCAG			caaagaattccTCGAGgccaccatgggttcccgggagac	ccagcttcatCTCGAgaagaggaagtgagttttgggaaaaatact
Rbbp4	✓	not excluded	E14T cDNA	atggctgacaaggaagcgg	ctaggacccttgtccctctgga	caaagaattccTCGAGgccaccatggctgacaaggaagcgg	ccagcttcatCTCGAgggaccctttgtccctctggat
Trim28	incorrect	not excluded by IF				caaagaattccTCGAGgccaccatggcggcctcggcg	ccagcttcatCTCGAgggggccatcaccaggg
Dnmt1	incorrect	+ by IF				caaagaattccTCGAGgccaccatgccagcgcgaaca	ccagcttcatCTCGAggtccttggtagcagc
Dnmt3a	✓	+	TCAG			caaagaattccTCGAGgccaccatgccctccagcggc	ccagcttcatCTCGAgcacacaagcaaaatattccttcagcg
Hmga1-rs1	✓	+	E14T cDNA	atgagcgagtcggggctcaaa	tcactgctcctcctcagaggac	caaagaattccTCGAGgccaccatgagcgagtcgggc	ccagcttcatCTCGAgctgctcctcctcaga
Hmgb2	✓	+	TCAG			caaagaattccTCGAGgccaccatgggcaagggtgaccc	ccagcttcatCTCGAgttcttcatcctcctcttcttcctcttcatcttcc
Psip1	✓	+	TCAG			caaagaattccTCGAGgccaccatgactcgcgatttcaaacctg	ccagcttcatCTCGAggttatctagtgtagactctttcagagatatttcagcc
Khdrbs1	✓	not excluded	TCAG			caaagaattccTCGAGgccaccatgcagcgccgggac	ccagcttcatCTCGAgataacgtccatatggatgctctctgt
Hnrnpl	✓	-	TCAG			caaagaattccTCGAGgccaccatgtcgcggaggctgc	ccagcttcatCTCGAgggaggcgtgctgtgca
Cbx5	✓	+	TCAG			caaagaattccTCGAGgccaccatgggaaagaagaccaagaggacag	ccagcttcatCTCGAggctcttcgcgctttctttttctttg
Hnrnpf	✓	-	TCAG			caaagaattccTCGAGgccaccatgatgctgggccctgag	ccagcttcatCTCGAgatcatatccgcccatgctgt
Litd1	✓	+	E14T cDNA	atgtcgggcgtgcagtc	ctaatgtatattgttcaacagatctttcaaatagggtatatcaga	caaagaattccTCGAGgccaccatgtcgggcgtgcag	ccagcttcatCTCGAgatgtatattgttcaacagat
Hnrnpc	✓	-	TCAG			caaagaattccTCGAGgccaccatggctagcaatgttaccaacaag	ccagcttcatCTCGAgagagtcatcctccccattgg
Hnrnpk	✓	-	TCAG			caaagaattccTCGAGgccaccatggagaccgaacagccag	ccagcttcatCTCGAggaatccttcaacatctgcatactgctt
Ddx21	✓	+	TCAG			caaagaattccTCGAGgccaccatgccgggaaaactccg	ccagcttcatCTCGAgctgaccaaacgctttactaaaactccg
Hdgf	✓	+	TCAG			caaagaattccTCGAGgccaccatgtcgcgatccaaccgg	ccagcttcatCTCGAgcaggctctcatgatctctgacg
H3 only							
Mta2	✓	-	TCAG			caaagaattccTCGAGgccaccatggcggctaacatgtaccg	ccagcttcatCTCGAgatcctccaggacaatgggc
Ssbp	✓	-	TCAG			caaagaattccTCGAGgccaccatgtttcgaagacctgtgttacagg	ccagcttcatCTCGAgtgccttttcttttgtctggtcactcagaaatataatg

Chapter 2.3

Assaying chromatin accessibility through mitosis reveals bookmarked sites at proximal gene promoters

Chapter 2.3

Assaying chromatin accessibility through mitosis reveals bookmarked sites at proximal gene promoters

Preface

This chapter addresses aims 2 and 3 of the research goal. I designed the study along with input from Dr. Jon Draper and Dr. Aki Minoda, and carried out majority of the experiments outlined in this study. Ye Liu, a graduate student in Dr. Minoda's lab, performed the ATAC-seq. I performed the data analysis with assistance from Drs. Jen-Chien Chang and Chung-Chau Hon (RIKEN institute, Japan). Daisy Deng, a graduate student in Dr. Draper's lab, generated Parp1 knockout lines, performed characterization assays on them and assisted in collection of material for mitotic release qRT-PCR experiments, contributing to figures 7 & 8.

Abstract

Mitotic bookmarks are memory signatures that allow the faithful inheritance of cellular identity to daughter cells after division. One of the key outstanding questions surrounding mitotic bookmarking is to identify those genomic loci in which bookmarking facilitates an altered pattern of transcription. To begin the process of addressing this key question, we profiled the accessible regions of the chromatin using the Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq), on interphase, mitotic (G2M) and G1 populations. We defined bookmarked loci as genomic sites whose accessibility is preserved throughout interphase, mitosis and back into G1 phase. Our data demonstrates that a large portion of the interphase sites (~31%) fit this bookmarking model of constant accessibility through mitosis, and these sites most frequently occupied promoter regions within 3kb of the transcription start site. Significantly, these putative bookmarked sites strongly co-related with the occupancy of H3K27Ac modifications, a mark previously characterized as being associated with mitotically bookmarked sites (Liu et al., 2017b). In contrast, the non-bookmarked sites were largely present in distal inter-genic regions and did not co-relate with H3K27Ac mark. We also identified subsets of pluripotency-associated accessible gene regions that appear to be bookmarked by a variety of transcription factors such as Oct4, Sox2 and Klf4, as well as Parp1, a hit identified in our ChIP-MS screen. Additionally, we show that there is a subset of bookmarked sites that are more accessible in G2M compared to G1 (G2M-enriched). These G2M-enriched sites are ubiquitously marked by H3K27Ac, and show a strong bias towards a pluripotent gene signature. Genes associated with G2M-enriched loci are rapidly transcribed upon mitotic exit. We tested the role of Parp1 as a potential bookmark

at these sites, but the limited nature of our study did not provide conclusive evidence to establish the role that Parp1 plays in mitotic bookmarking of ES cells. This study provides a new perspective and fresh avenues for exploring the mitotic bookmarking capability of key pluripotency proteins.

Background

The proposed role of mitotic bookmarking is the preservation of transcription states through cell division (Kaduke et al, 2012, Zaidi et al., 2010, Hsiung et al., 2015, Zaret, 2014). Transcription via all three RNA polymerase enzymes is stalled during mitosis. This occurs as a result of dissociation of core transcription machinery form the mitotic chromatin, as with RNA pol II (Parsons and Spencer, 1997, Zhao et al., 2011) and RNA pol III (Leresche et al., 1996) or by inactivation of RNA pol I during transcriptional elongation (Weisenberger and Scheer, 1995). Phosphorylation dependent uncoupling of cell type-specific transcription factors is also involved in the cessation of transcription during mitosis (Martinez-Balbas et al., 1995, Roberts et al., 1991, Segil et al., 1991, Luscher and Eisenman, 1992). The transcription machinery is the executioner, and requires a blueprint to be pre-set to carry out RNA transcription. The chromatin state of the genome is the blueprint for transcription, and therefore, it must be established before the reactivation of transcription upon mitotic exit. Transcription is initiated during late telophase of mitotic division, and proceeds through to G1 in a sequential and orchestrated manner (Prashanth et al., 2003).

We hypothesize that in ES cells, mitotic bookmarking factors (MBFs) play a role in preserving the chromatin state of key pluripotency related genes, resulting in a rapid reactivation of the pluripotent transcriptional program. To test this we first establish the chromatin accessibility profiles of ES cells in interphase, during mitosis and upon mitotic exit into G1 using the Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq) (Buenrostro et al., 2013). The ATAC-seq assay involves treating cells with a ‘transposome’, which contains a hyperactive Tn5 transposase and DNA

adapters (Buenrostro et al., 2013). Upon encountering genomic sites that are unprotected, Tn5 cleaves the DNA and results in insertion of DNA adapters, which can later be sequenced to identify regions of high accessibility. DNaseI profiling of mitotic and asynchronous erythroid cells had previously revealed that the chromatin was largely accessible during mitosis, except at certain distal transcription factor binding sites (Hsiung et al., 2015). We aimed to use ATAC-seq in a similar manner to profile the changing chromatin dynamic during mitosis and into G1, and propose that it would be modulated by binding of one or more MBFs.

We chose to focus on Parp1 as a potential mitotic bookmarking candidate in ES cells, as this factor had been identified in our ChIP-MS screen for mitotic chromatin-association in Chapter 2.2. Parp1 (poly [ADP ribose] polymerase I) catalyzes the addition of ADP-ribosyl group from NAD+ to various protein including histones, and Parp1 itself. Poly-ADP-ribosylation of a variety of proteins serves as an important post-translation regulatory mechanism. As a result, Parp1 is involved in a variety of processes including DNA damage response (Shall and de Murcia, 2000; Wang et al., 1995), regulation of chromatin structure (Krishnakumar and Kraus, 2010; Liu and Kraus, 2017), differentiation (Hemberger et al., 2003; Nozaki et al., 1999; Roper et al., 2014) and transcription regulation (Ogino et al., 2007; Zhang et al., 2016). Additionally, it was shown to bookmark certain gene promoters in HEK293 cells (Lodhi et al., 2014) resulting in rapid transcriptional reactivation of certain genes. It was also shown to be important for the maintenance of pluripotency in ES cells cultures (Jiang et al., 2015; Roper et al., 2014), and is highly expressed in ES cells (Chapter 2.2 Fig 7). Hence, Parp1

represented a suitable pilot study for assaying mitotic bookmarking within the context of the new bookmarking data we present in this chapter.

Materials and methods

Cell culture: E14TG2A wild type mouse ES cells were cultured on 0.1% gelatin coated dished with or without x-ray irradiated mouse embryonic fibroblasts (x-MEFs). X-Mefs were seeded at a density of $1x10^6$ per full 6 well plate. Cells were cultured in DMEM (Sigma: D5796), supplemented with 15% FBS (Gibco), 1X non-essential amino acids (11140-050, Thermo Fisher), 1X glutamax (35050-061, Thermo Fisher), 1x sodium pyruvate (11360-070, Thermo Fisher) and 1x BME (21985-023, Thermo Fisher), and murine recombinant LIF (AMS-263-100, Amsbio). Cells were passaged every 2-3 days with Accutase® (Sigma: A6964).

Mitotic enrichment: Mitotic enrichment for ATAC-seq experiments was performed by treating mouse ES cells with 50ng/ml of nocodazole for 7hours at 37C. For later, mitotic release qPCRs with wild-type and knockout cells a double thymidine and nocodazole block was used (Teves et al., 2016). For this, cells were seeded at a density of 2.5M per T75 0.1% gelatin coated tissue culture flasks. Each flask was supplemented with 0.5M x-ray irradiated MEFs (mouse embryonic fibroblasts).

Intracellular flow cytometry: Cell cycle profiles were established by fixing cells at various time points using the BD fixation and permeabilization kit (Cat # 554714). The fixation solution was diluted with 3 parts PBS to achieve a final paraformaldehyde (PFA) concentration of 1%, and cells were fixed for 10minutes at room temperature. Cells were Hoechst (Life technologies: H1399) for DNA and MPM2 (05-368, Millipore) as a mitotic marker. Donkey anti-mouse Alexa Fluor 647 (Thermo Fisher A-31571) was used as a

secondary antibody. Samples were acquired on MACSQuant® analyzer (Milteny Biotech), and analyzed using FlowJo.

Mitotic release experiments for ATAC-seq: mESCs were mitotically enriched for 7hrs with 50ng/ml nocodazole. At 6.5hrs with nocodazole Hoechst33342 was added to the media at 10ug/ml to stain the nuclei at 37C for 30minutes. Mitotic cells were collected by mitotic shake off, and were released (M-released) into regular mESC media after washing off nocodazole (1X with PBS) for 20mins at 37C. Cells were gently dissociated and passed through a 40um cell strainer. M-released cells (G1t20) were kept cold from this point on, and were resuspended in PBS with 2%BSA, 5mM EDTA and 7AAD (BD 559925 at 1:100); live cells were sorted using Beckman Coulter's MoFlo based on DNA content. A fraction of G1t20 cells were washed, and resuspended gently in mESC media and release backed into G1 at 37C for 15mins. Cells were then collected for G1t35. 20K and 40K cells from each fraction (Interphase, G2M, G1t20 and G1t35) were cryopreserved in 15% FBS and 10%DMSO until processing for ATAC-seq.

ATAC-seq: ATAC-seq samples were prepared from 20,000 cells. The transposase reactions were carried as previously described (Buenrostro et al., 2015) with 10 to 13 total PCR cycles. Amplified DNA fragments were purified with QIAGEN MinElute PCR Purification Kit and size selected twice with Agencourt AMPure XP (1:1.4 and 1: 0.5 sample to beads, Beckman Coulter). Libraries were quantified with KAPA Library Quantification Kit for Illumina Sequencing Platforms (KAPA Biosystems), and size distribution was checked on a Bioanalyzer (Agilent High Sensitivity DNA chip, Agilent Technologies). Libraries were sequenced on a HiSeq2500 Paired-end 50 base (Illumina),

and were mapped using HISATII (Kim et al., 2015) using Gencode GRCm38 genome assembly. Following command was used for alignment:

```
hisat2-2.0.4/hisat2 -p 16 --fr -x \
hisat2_index/mm10_with_gencodeM10/index/index \
-1 sample1_R1 \
-2 sample1_R2 \
| /samtools-1.3.1/samtools sort - \
-o sample1.sort.bam
```

PCR duplicates were removed from the alignments as follows:

```
samtools-1.3.1/samtools rmdup -S sample1.sort.bam
sample1.sort.rmdup.bam
```

ATAC-seq data analysis:

Global peak analysis: Macs2 peak caller (Zhang et al., 2008) was used to identify total ATAC-seq accessible peaks. Following parameters were used:

```
macs2 -t $bam_file -f BAM -s 145 -n sample_ID_narrow -g mm -p 0.01 --
nomodel -s 145 -B --outdir $subdir_narrow
```

Bedtools intersect (`bedtools intersect -u`) (Quinlan and Hall, 2010) was used to find common peaks between the two replicates for G2M, G1t20 and G1t35 samples. Bedtools was also used to find overlaps with other datasets and between samples.

PePr differential Peak Calling (final parameters): Peak-calling and Prioritization pipeline (PePr) (Zhang et al., 2014) was used to identify differentially accessible regions between different cell cycle populations. Following parameters were used:

```
python PePr.py \
-c $sample_1_rep1,$sample_1_rep2 \
--chip2 $sample_1_rep1,$sample_1_rep2 \
-f bam --diff -n G1t35vsG2M --peaktype sharp --normalization intra-
group \
--output-directory
```

Post-processing data analysis was performed using R (http://www.R-project.org/) (Team, 2013), individual scripts are deposited here (See R scripts). Briefly, ChIPseeker (Yu et al., 2015) was used to perform feature distribution, and peak annotation analysis for both global and differential ATAC-seq peaks. R package venneuler was used to generate venn diagrams, and UpSet (Lex et al., 2014)was used to generate UpSet plot for showing relationships between different overlapping datasets.

Post-processing (see R-scripts):

170413_ATAC_ChIPseeker.R (for peak annotation and profiling)

170705_ATACseq_macs2_venndiagrams (for venn-diagrams and Upset Plots)

170712_pluripotent_bookmared_RNAexpr (comparison of RNA expression with Parp1, klf4 and H3K27Ac bookmarked sits)

GEO datasets used: RNA seq analysis for RA vs LIF conditions (GSE65697 (Terranova et al., 2015), analyzed as in Chapter 2.2). H3K27Ac mitotic and interphase ChIP-seq, mitotic Oct4, mitotic Klf4 and mitotic Sox2 ChIP-seq (GSE92846) (Liu et al., 2017b) and Parp1 Interphase ChIPseq (GSE81168) (Liu and Kraus, 2017).

q-RT-PCR: RNA was isolated using Trizol™ LS (Thermo Fisher 10296028) according to manufacture's protocols. For nascent RNA q-RT-PCR, RNA was DNaseI treated in solution and purified using Qiagen RNeasy Micro Kit (Qiagen Cat # 74004). cDNA was prepared for 1ug of RNA using SensiFAST cDNA Synthesis Kit (Froggabio BIO-65054), and q-PCR was performed using SensiFAST SYBR No-ROX Kit (Froggabio CSA-01194). CFX Manager™ was used to analyze the data (Biorad, Software #1845000).

Generation of Parp1 knockout mouse ES cells: Parp1 KO mouse ES lines were generated by using guide RNAs again exon 2 of mouse Parp1 gene. Guide RNAs were

designed using Benchling and cloned into CRISPR/Cas9 backbone, pSpCas9(BB)-2A-Puro (PX459) V2.0 (a gift from Feng Zhang (Addgene plasmid # 62988)) (Ran et al., 2013). 10 ug of the cloned CRISPR/Cas9 plasmid was used to transfect a well of 6-well plate of wild type ES cells (p6). Cells with selected with 2ug/ml of Puromycin for 72hours, and colonies were allowed to form. Individual colonies were hand picked and screen by western blot with anti-Parp1 antibody (Abcam ab194586). Positive clones were amplified around the targeted region and sequenced using Sanger sequencing.

Colony forming assay: Cells were seeded at a density of 250cells/well onto a well of a 12 well plate with or without x-MEFs. For each experiment cells were seeded in a technical triplicate and cultured for 5days. At day 5, colonies were fixed in the dish with 250ul of 4% PFA (Electron microscopy sciences, Cat # 15710) for 1-2minutes and washed with water. Alkaline phosphatase (AP) staining was performed as described in Sigma AP staining kit (86R-1KT, Sigma). Plates were scanned on EPSON Scanner with 3200dpi and 24-bit colour and analyzed on ImageJ (Schneider et al., 2012). Dense colonies with intense AP staining were characterized as AP positive (AP+ve) while the less dense ones with dispersed pink staining around the edges were characterized as mixed colonies.

Statistical Analysis: All statistical analyses were performed using GraphPad Prism (version 6), except ATAC-seq related analyses. Unless otherwise stated error bars represent standard error of the mean and alpha of 0.05 was used as a cut-off for statistical significance.

Results

Cell cycle kinetics of ES cells

To assay the transcription profiles of mouse ES cells upon mitotic exit we first studied the kinetics of mitotic exit in ES cells. ES cells were treated with 50ng/ml of nocodazole for 7hrs, collected by mitotic shake off (t0) and released into G1, with media without nocodazole at 37C. Cell cycle profiles of mitotic-release (m-release) were established by PFA fixation and staining with mitotic marker MPM2 (Campbell et al., 2014, Westendorf et al, 1994), and the DNA dye, hoechst. At t0, majority (~80%) had a 2N DNA content and about half of those were in mitosis as observed by MPM2-APC staining (Fig 1A-B). When released into media without nocodazole, they very rapidly exited out of mitosis and within 30mins of release about 10% of cells are in G1 phase of cell cycle, with numbers increasing rapidly until about 75minutes post-release when cells appear to establish a momentary equilibrium (Fig. 1A-B). Of note, from 75minutes to 120 minutes post-release, 15% of the cells still remain in G2, and a negligible percentage is in mitosis, suggesting that most of the G2 cells observed at t0 were arrested in late G2 or early metaphase, and only about 15% of the total t0 population was in early G2.

We aimed to assay the chromatin and transcription profile of the earliest stages of G1. To this end, we released nocodazole-arrested cells into G1 for 20mins (G1t20) at 37C, and FACS sorted live cells based on DNA content (Fig 1C). G1t20 cells were placed into culture and released for another 15mins at 37C, collecting a later G1t35 cell population (Fig 1C).

Assay for Transposase Accessible Chromatin (ATAC-seq) reveals putatively bookmarked gene loci

ATAC-seq was performed on interphase (asynchronous) control, G2M, early G1 (G1t20) and late G1 (G1t35) cells (Fig. 1C & Table S1). Samples were sequenced to a depth of over a 100 million raw reads per sample, and filtered out for quality and PCR duplicates (Table S1). Macs2 (Zhang et al., 2008) was used to call peaks for individual samples (Table S2). The number of peaks between different samples was very similar, except in interphase, which had slightly higher number of peaks (Table S2). All samples also showed a similar peak distribution centered around the transcription start site (TSS) (Fig 2A), however the interphase sample showed slightly higher number of peaks in distal-intergenic regions compared to G2M, G1t20 and G1t35 samples (Fig 2B). For further identification of putatively bookmarked gene loci, we retained peaks that were common between the two replicates for G2M, G1t20 and G1t35 samples (Table S2).

We defined mitotically bookmarked sites as genomic regions that maintain their chromatin accessibility state throughout mitosis into G1. To identify genes that were putatively bookmarked during mitosis, we compared the overlap of G2M accessible sites with other cell populations. 42% (45486/107869) of sites that were accessible in interphase retained accessibility during G2M (Fig 2C), and 29% (32127/107869) of interphase sites also open in both G1t20 and G1t35 populations (Fig 2C and E). We found that majority of the sites identified in G2M were also accessible in early and late G1 (Fig 2D), but only 9.4% (5167/54963) were accessible uniquely during mitosis (Fig. 2E). We categorize the 32127 sites that retain their chromatin state throughout interphase, mitosis and into G1 as being putatively mitotically bookmarked sites, and the remaining interphase sites (74365) as not being mitotically bookmarked (Fig 2F). Comparing the feature distribution of these two classes, we observed that a majority (47.8%: 43.5%

within 1kb of TSS, 2.49% between 1-2kb and 1.85% between 2-3kb of the TSS) of the bookmarked sites are located within gene promoters, and 28% are distal intergenic (Fig 3A). Conversely, the non-mitotically bookmarked sites are mostly present in distal intergenic regions (48.16%), and only 9.96% within up to 3kb of the TSS (Fig 3A). It is worth noting that in our analysis the non-mitotically bookmarked sites were classified as shared between interphase and 2 or less of the 3 cell cycle populations tested.

Recently, Liu et al., 2017b had shown that the H3K27Ac epigenetic mark is retained on mitotic chromatin as are Oct4, Sox2, and Klf4, and could bookmark some loci. We examined the overlap of putative bookmarked sites identified using our ATAC-seq data with the mitotic specific sites of H3K27Ac, Oct4, Klf4, and Sox2 (Fig 3A-B, data from (Liu et al., 2017b)), as well as for the total known binding sites of the putative bookmarking factor Parp1 (Fig 3B, data from (Liu and Kraus, 2017)). Since, H3K27Ac marks enhancer elements, we also explored the putative overlap of pluripotency related super-enhancers (SE) (Fig 3B, data from (Whyte et al., 2013)). We found that mitotic H3K27Ac mark occupied 33.8% of the total putative bookmarked sites identified using our ATAC-seq data (Fig 3A-B), and only co-localized with 6.8% of the non-mitotically bookmarked sites (Fig 3A). Various combinations of the pluripotency factors occupied only small subsets of our bookmarked sites (Fig. 3B). Additionally, 41.6% (1110 out of the total 2563) super enhancer sites (Whyte et al., 2013) are bookmarked during mitosis based on our data. Majority of the H3K27Ac bookmarked sites (78%, 8504/10863), did not co-localize with any of the other marks tested, while 8% (875/10863) were occupied by Klf4 only and 6% (622/10863) putatively by Parp1 only (inset, Fig 3B). A comparison of expression of the genes bookmarked by H3K27Ac, Klf4 and Parp1 in MEF

(differentiation) vs ESC condition identifies some genes that show higher expression in pluripotency (Fig 3C). It is important to note, though, that some of the gene loci bookmarked by this group are equivalently expressed in ESC and MEF (a differentiated cell type) (Fig 3C).

Assaying the changing chromatin landscape upon mitotic exit

Differential peak analysis reveals regions of variable accessibility upon mitotic exit

To get a better resolution of the changing chromatin landscape of the bookmarked sites upon mitotic exit, we performed a pairwise differential peak analysis using Peak-calling and Prioritization Pipeline (PePr) (Zhang et al., 2014). Differential analysis showed distinct accessibility patterns between different cell cycle fractions (Fig. 4A-B). Pairwise comparisons were made between G2M and G1t20, G2M and G1t35, as well as between and G1t20 and G1t35 populations (Fig 4 & Table 1). This analysis revealed that the number of differentially accessible peaks was highest in G1t20 compared to G2M (G1t20overG2M), although is should be noted that the fold change differences were minimal. To better understand the nature of potentially bookmarked sites, we compared the differentially accessible sites to the bookmarked sites identified in Fig 2F, and found that a majority of the sites that were enriched in G2M (both G2MoverG1t20 and G2MoverG1t35; 82% and 63%, respectively) were identified as being bookmarked, while only a small subset of the G1 enriched sites were classified as being bookmarked (Table 1). Furthermore, regions that are more accessible in G2M compared to both G1 fractions (G2M over G1t20 and G2M over G1t35) are mostly present at gene promoters and flank the transcription start site (TSS) (Fig. 4B, 4C (red-box)). G2M enriched peaks also

exhibit a distinct feature distribution pattern (Fig 4B) compared to global profiles (Fig 2A-B) and regions more accessible in G1 (Fig 4A). Additionally, genomic sites more accessible in G2M (G2MoverG1t20 or G2MoverG1t35) have a broad region of accessibility around promoter sites, unlike a sharp peak normally observed at the TSS (Fig 4C, red box). We identified the sites that were enriched in G2M compared to both early and late G1 time points, 'G2M enriched common' (Table 1) and hypothesized that these sites are associated for rapid transcriptional reactivation of pluripotency genes upon mitotic exit.

G2M-enriched loci are ubiquitously occupied by H3K27Ac mark, and are associated with pluripotency-related gene expression

To characterize the genes associated with regions of differential accessibility, we overlaid the differentially accessible genomic sites with RNA expression profiles in pluripotent and RA differentiation conditions (Fig 5A). Genes associated with G2M-enriched peaks had significantly higher expression profiles in pluripotent cell conditions compared to genes associated with regions more accessible in G1 (G1t20 over G2M and G1t35 over G2M) (Fig. 5A). Common G2M-enriched regions ('G2M_enriched_common') (Table S3) that are present in both 'G2M over G1t20' and 'G2M over G1t35', show a definite bias towards a pluripotency-associated gene expression signature (Fig. 5A). Although these genes (Table S3) appear to have a more pluripotent expression signature, they are involved in generic cellular pathways such as cell-cycle phase transition, DNA biosynthesis, and RNA regulation (Fig 5B).

Next, we performed a survey of the mitotic occupancy of H3K27Ac, and revealed that the majority of the G2M enriched sites (77% of G2MoverG1t20 (674/895) and 66%

of G2MoverG1t35 (199/301)) are bookmarked during mitosis and interphase by H3K27Ac, while a small subset of the G1t20 and G1t35 enriched sites are occupied by this mark (Fig 6). A recent study had shown that some genomic sites undergo a spike in transcriptional activity when transitioning from mitosis to G1 (Hsiung et al., 2016), and that these sites are marked by H3K27ac during mitosis, and could play an important role in establishing the gene expression profile of the daughter cells, ergo the cell fate. To test whether the G2M-enriched sites follow this pattern, we assayed the gene expression levels of the gene associated with these sites upon mitotic exit.

Transcriptional reactivation profiles of differentially accessible sites upon mitotic exit

To assay gene transcription profiles upon mitotic exit we used a double thymidine (Fig 7A) followed by nocodazole block (Teves et al., 2016) to get mitotic populations that are slightly purer (Fig. 7B) than the ones collected after a 7hr nocodazole block (Fig 1A-B). Mitotic cells (t0) were collect by mitotic shake off (mso) and released into G1 for varying time points and collected for RNA. Mouse ES cells that were mitotically enriched by this method still followed the same kinetics of cell cycle progression upon release as for the previous enrichment protocol, with a significant number of cells entering G1 within 35minutes of release (Fig 7B and 1B). We also collected a primarily G2 fraction (postmso; post mitotic shake-off) that remained adherent after collecting cells by shake-off (Fig 7B). We assayed mature RNA transcript levels at different points upon mitotic release for genes that were enriched in G2M or genes that were more accessible in G1 compared to G2M (G1t35overG2M). Despite obvious differences in the cell cycle

profile of these populations (Fig 7B) no significant gene expression differences were observed at the transcript level for these genes (Fig 7C) between different time points.

We then sought to check for transcription re-activation by measuring nascent, unspliced RNA transcripts. We designed primers (Table S4) that bind both intronic and exonic regions of genes to quantify the newly synthesized un-spliced transcripts. We compared different fractions using t0 as a control. For G2M enriched genes Eda2r and Kdm6a, nascent transcript level remained fairly constant throughout the cell cycle, while Atf5 showed greater levels of transcription in t0 and t20, and Huwe1 showed a small steady increase in expression through G1 (Fig 7D). Interestingly, Elavl2, a genomic site more accessible in G1t35 shows a spike in expression of nascent RNA levels in G2 (postmso) and interphase cells, the late stages of the cell cycle (Fig. 7D).

Assaying the mitotic bookmarking capabilities of Parp1

Knocking out Parp1 affects the self-renewal capability of ES cells without affecting their cell cycle profile

In order to ascertain the role of Parp1 in mitotic bookmarking we generated Parp1 knockout (Parp1KO) lines by using CRISPR-cas9 (Cong et al., 2013) to disrupt exon2 of Parp1 (Fig 8A). We generated two independent clones of Parp1KO mESC lines (#24, and #8), where clone Parp1KO24 had a homozygous mutation and Parp1KO8 had independent mutations in both the alleles or was a mixed clone with two independent homozygous mutations. In both cases the open reading frame of Parp1 was disrupted and lead to a complete loss of detection of the protein product (Fig 8B-C).

To test the self-renewal capability of Parp1KO cells, colony-forming unit assays were performed by looking at the formation of pluripotent alkaline phosphatase (AP)

(Palmqvist et al., 2005; Pease et al., 1990) colonies. Knocking out Parp1 significantly reduces the number AP+ve pluripotent colonies formed compared to wild-type cells (Fig. 8D), suggesting a role in maintaining pluripotency. Additionally, the cell cycle profile of asynchronous Parp1KO cells is unchanged (Fig. 8E) compared to wild-type ES cells. Similarly, when arrested in mitosis and released into G1, Parp1KO cells showed similar kinetics of cell cycle progression compared to wild-types (Fig. 8E), suggesting that the association of Parp1 with mitotic chromatin is not involved in cell cycle progression.

Transcriptional reactivation of Parp1 targets upon mitotic exit

We tested nascent RNA expression of Parp1 target genes (Liu and Kraus, 2017) that were also differentially accessible by ATAC-seq in a mitotic release experiment. Since the cell cycle profile upon mitotic exit was similar between the two cell types (Fig 8F), we compared the differences in expression at each time point to its wild-type counterpart. For the G2M enriched Parp1 targets Eda2r and Kdm6a, knocking out Parp1 results in a slightly increased transcript expression in mid- to late- G1, while no significant changes were observed for Huwe1 and Stella (Fig 9). The G1t35 enriched locus, Elavl2, showed slightly increased expression in mid- to late- G1, but was significantly higher only in asynchronous control samples (Fig 9).

Discussion and conclusion

The aim of this research study was to characterize the extent of mitotic bookmarking in ES cells. To that end, we first identified mitotically bookmarked genomic sites by ATAC-seq. Mitotic bookmarking is a phenomenon that attempts to explain the preservation of cellular identity from the parent cell through mitosis and into G1 of the daughter cells (Kadauke and Blobel, 2012; Sarge and Park-Sarge, 2009; Zaret,

2014). To satisfy this criterion, we defined mitotically bookmarked sites as genomic sites that retain their ATAC-seq accessible nature from interphase, through mitosis and into G1. We designed a mitotic release assay (Fig 1C), and FACS isolated G2M, and early and mid- G1 populations. Interestingly, we found that majority of the sites were shared between these three populations, however, there were significant differences when compared to the number of accessible sites profiled in asynchronous interphase cell populations (Fig 2C-D).

We identified 32127 genomic sites that were shared in all four (interphase, G2M, early and mid G1) samples as being mitotically bookmarked and a larger number of interphase sites that were considered to be non-bookmarked (Fig 2F). Majority of the bookmarked sites were present in promoter regions within 3kb of the TSS and were in part occupied by H3K27Ac while the non-bookmarked sites were largely intergenic and devoid of H3K27Ac (Fig 3A-B). This is in concert with data from DNAse I hypersensitive assay performed on asynchronous and mitotic erythroid cells, where it was observed that chromatin is largely accessible at promoter regions bookmarked by H3K27Ac during mitosis, while distal regions are generally closed off (Hsiung et al., 2015). Our data, however, is somewhat in contrast with a recent study where ATAC-seq was perform in asynchronous and nocodazole arrested mouse ES cells, and it was suggested that chromatin accessibility is unchanged during mitosis (Teves et al., 2016). However, when we carried out a genomic loci comparison (as performed in our study) of their dataset, it revealed 54% of the asynchronous sites are bookmarked while the remaining 46% are specific to interphase (Table S5). Additionally, 85% of the sites

common to asynchronous and mitotic datasets in their study, overlapped with our bookmarked sites (Fig 2F, Table S5).

These data suggest that while a significant portion of the genome retains the same mitotic chromatin accessibility signature, as during interphase, there is a large percentage that loses accessibility during mitosis and early G1 phases. This raises the possibility that these non-bookmarked sites are dispensable for establishing pluripotent identity, and their altered accessibility is a consequence, but not a cause, of cell fate decisions. In contrast, bookmarked sites would contribute in multiple ways to maintain the various processes that make up a pluripotent cell. It is possible that majority of these genomic loci are important for generic cellular process, while only a subset is specific to pluripotent cells. In order to identify pluripotency associated bookmarked sites, we looked for occupancy of pluripotency related factors, Oct4, Sox2 and Klf4 at these loci, and identified various sets of genes that were bookmarked by a combination of these factors (Fig 3B). We also found that about half of the total pluripotency associated super-enhancers (Whyte et al., 2013) are bookmarked during mitosis, and some of these sites are co-occupied by other pluripotency transcription factors (Fig 3B). These analyses revealed that a combination of Klf4 and Parp1 binding, in addition to H3K27Ac, identified a small subset of pluripotency genes (Fig 3C). Further experimentation, and comparison of bookmarked sites in other cell types is needed to reveal the cell type specific subset of the total bookmarked sites.

To understand the relationship of the ATAC-seq peaks with G1-phase transcription reactivation, we performed a differential peak analysis between the accessibility data obtained from the G2M, G1t20 and G1t35 cycle fractions. It has

previously been shown that in late M-phase some H3K27Ac bookmarked genes are occupied by RNA polII, resulting in a spike in transcriptional activity (Hsiung et al., 2016). In our differential peak analysis we identified a subset of bookmarked sites that were more accessible in G2M when compared to G1t20 and G1t35 (Fig. 4). These sites are ubiquitously occupied by H3K27Ac during mitosis, and seem to correlate with genes that have higher expression in pluripotency conditions (Fig 5). Quantitative RT-PCR for the nascent transcripts of these genes (Fig 7) shows that they are expressed at similar levels in G2M, early G1 and interphase controls. However, Elavl2, which was shown to be most accessible in G1t35, displayed a spike in transcriptional activity in interphase and G2 cells, suggesting that such loci may be primed for transcription during late M and into G1.

We checked the transcriptional profiles of Parp1 target genes (Liu and Kraus, 2017) upon mitotic release of Parp1KO cells, however minimal differences were observed. Eda2r and Kdm6a seemed to be slightly up-regulated in mid- to late-G1 (Fig 9) suggesting that perhaps Parp1 acts as a repressor of some genes upon mitotic exit into G1 for maintaining normal ES cell expression profiles.

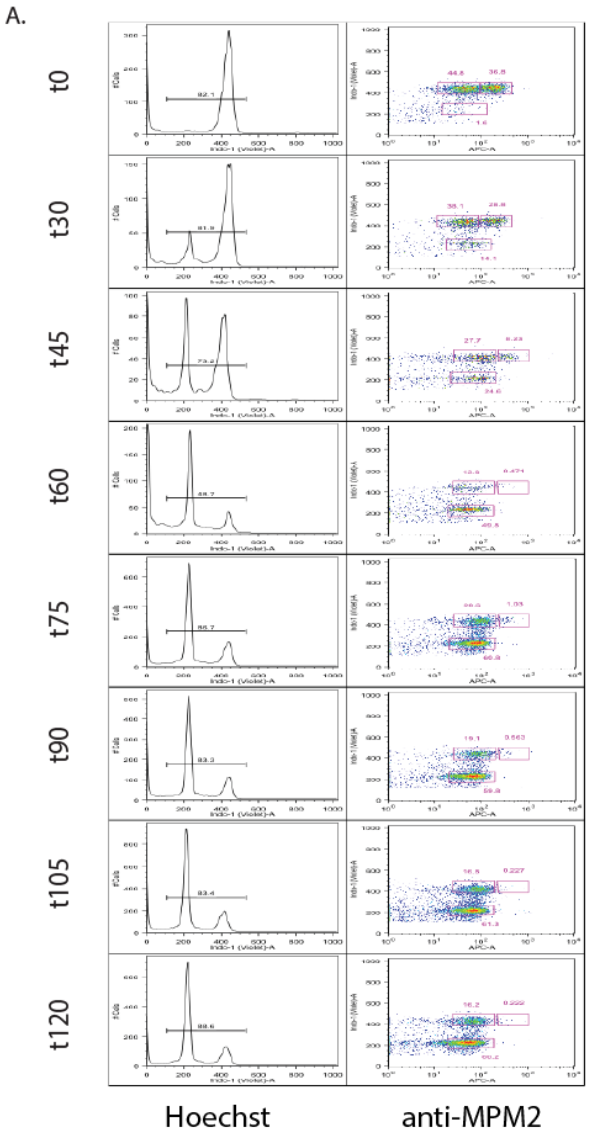
It is worth mentioning that these genes represent a very small fraction of the total or pluripotency specific sites bookmarked, and that further characterization of other gene loci are required before ruling out the role of Parp1 as a mitotic bookmarking factor. As an attempt to address this in the near future, we are currently in the process of assaying the affects of expressing Parp1 protein fused to a mitotic degron domain (MD) that would decouple it from the chromatin during M-G1 phase transition (Kadauke et al., 2012). We will use an inactive degron unit, MD* as a control. This approach has been used by recent

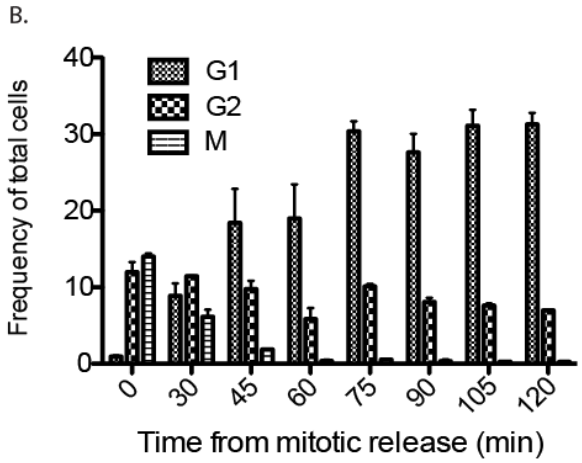
studies to assess the role of mitotic occupancy of various bookmarking factors(Deluz et al., 2016; Kadauke et al., 2012; Liu et al., 2017b). We will assay the phenotypic affects of perturbation of Parp1 during late mitosis by Parp1-MD, and Parp1-MD* in wild-type and Parp1KO cells in a colony forming unit assay. We hypothesize that Parp1-MD* and not Parp1-MD will be able to rescue the colony forming defects observed in Parp1KO cells (Fig. 7D).

Additionally, some recent studies have shown that very minimal differences are observed in global gene expression profiles upon mitotic exit by RNA-seq (Festuccia et al., 2016). Sox2, a pluripotency factor, is now established as a mitotic bookmarking factor and its abrogation during M-G1 phase transition impairs pluripotency, however, it does not seem to disrupt gene transcription upon mitotic exit (Deluz et al., 2016; Deluz et al., 2017). We attempted to use RNAseq to profile the global RNA expression profile of ES cells in the same cell cycle phase stages that we profiled with the ATAC-seq, but no differences were observed in the mature RNA transcriptome upon mitotic exit (data not shown) and the sequencing depth limited the identification of nascent transcripts. Analysis of the replicates suggested that there were probably technical issues that confounded this experiment, but more sensitive technologies, such as GRO-seq (Core et al., 2008), could be employed for establishing the nascent RNA transcription activation profiles upon mitotic exit.

In sum, our data reveals that accessibility of majority of proximal promoters is preserved during mitosis, and suggests that global transcription is likely reset very soon in late mitosis and early G1. However, there are still a significant number of genomic sites whose chromatin state is apparently dispensable for early G1 phase reactivation, and

therefore not preserved during mitosis. Various bookmarks occupy small subsets of these promoter sites, including H3K27Ac, Klf4, Parp1, Oct4 and Sox2, and are potentially involved in propagating the open chromatin state of these sites into G1. The analysis also revealed a large number of bookmarked sites that do not co-relate with occupancy of any of these factors, warranting an investigation into the preservation of these sites by other bookmarking factors, such as those identified in Chapter 2.2.


A.
t0
t30
t45
t60
t75
t90
t105
t120
Hoechst
anti-MPM2


B.
Frequency of total cells
40
30
20
10
0
G1
G2
M
0
30
45
60
75
90
105
120
Time from mitotic release (min)

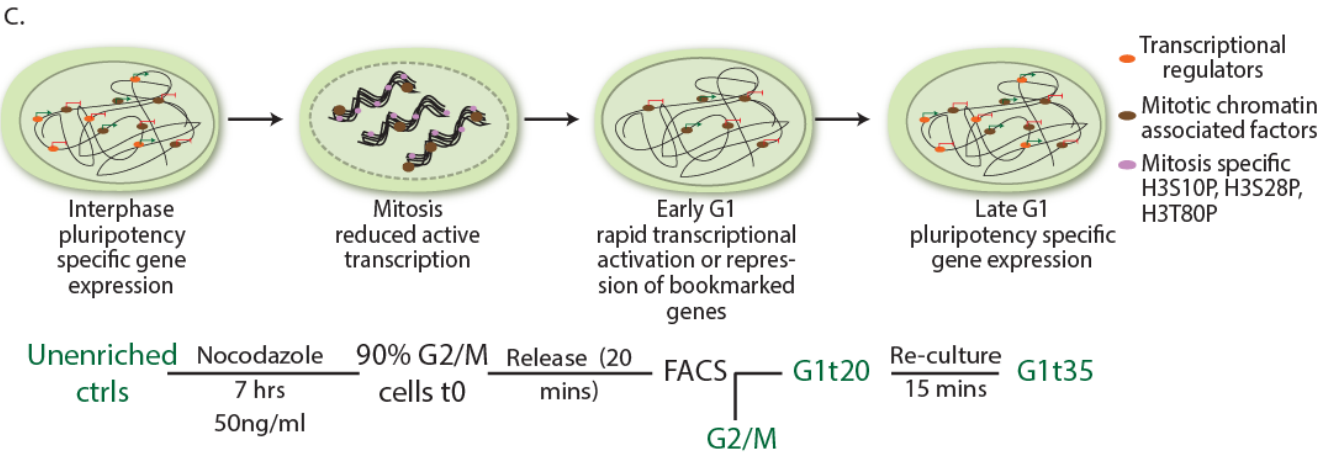

C.
Interphase pluripotency specific gene expression
Mitosis reduced active transcription
Early G1 rapid transcriptional activation or repression of bookmarked genes
Late G1 pluripotency specific gene expression
Transcriptional regulators
Mitotic chromatin associated factors
Mitosis specific H3S10P, H3S28P, H3T80P
Unenriched ctrls
Nocodazole 7 hrs 50ng/ml
90% G2/M cells t0
Release (20 mins)
FACS
G1t20
G2/M
Re-culture 15 mins
G1t35

Figure 1. A.) Mitotic release time course experiment. Mitotic cells were collected with mitotic shake-off at t0 and released into G1 for 30, 45, 60, 75, 105 and 120minutes (t30, t45, t60, t75, t105 and t120 respectively). The first panel shows cell cycle profiles with DNA dye Hoechst, the line gate represents the population used in the subsequent panel; the second panel is Hoechst, on the y-axis and mitotic marker MPM2 on the x-axis. B.) Relative frequency of different cell cycle phases identified in second panel in A (n=2). C.) Schematic showing the experimental plan for the mitotic release experiment to identify mitotically bookmarked sites. Populations highlighted in green were processed for ATAC-seq.

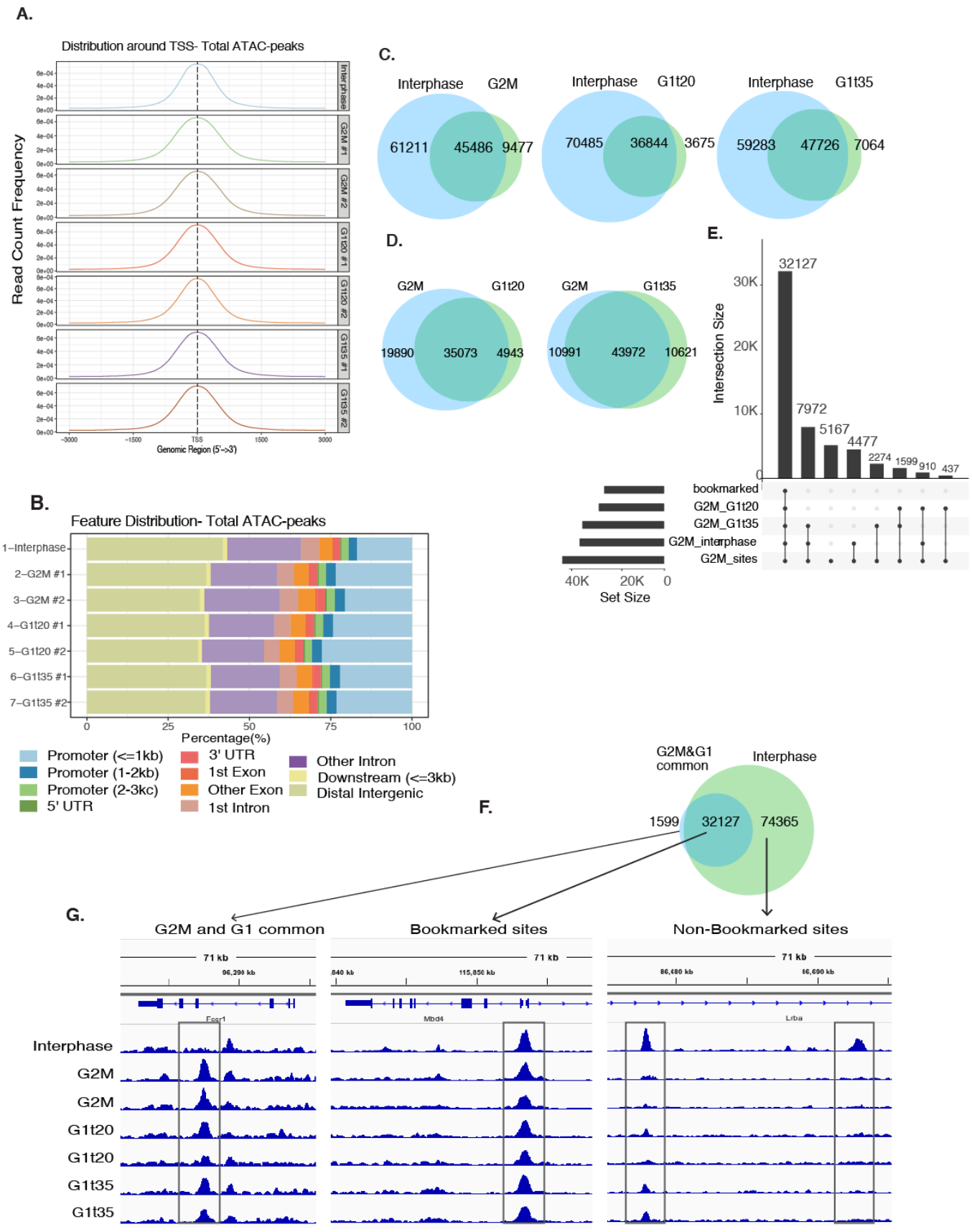
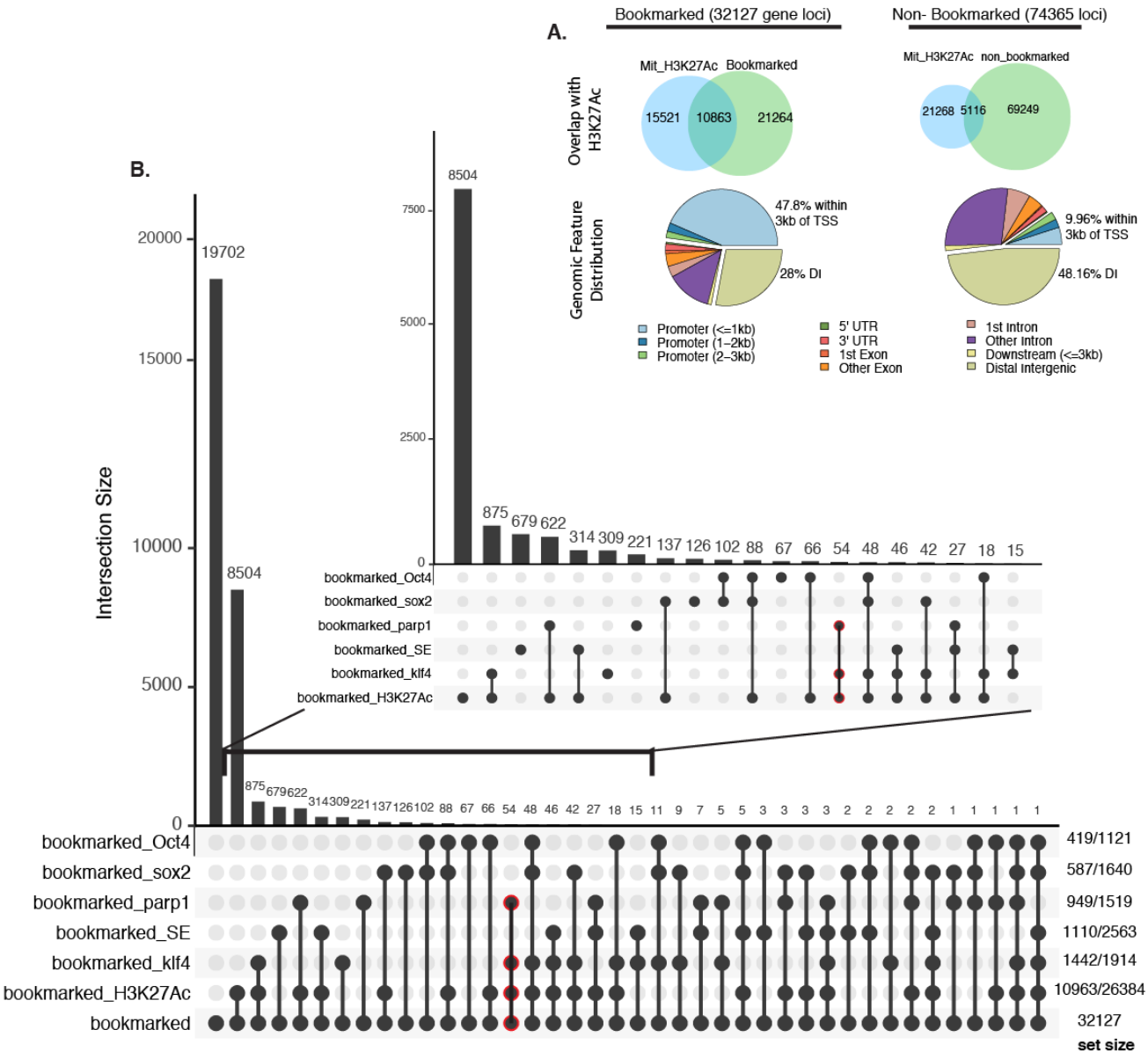

A.
Distribution around TSS- Total ATAC-peaks
Read Count Frequency
Interphase
G2M #1
G2M #2
G1t20 #1
G1t20 #2
G1t35 #1
G1t35 #2
-3000
-1500
TSS
1500
3000
Genomic Region (5'->3')
B.
Feature Distribution- Total ATAC-peaks
1-Interphase
2-G2M #1
3-G2M #2
4-G1t20 #1
5-G1t20 #2
6-G1t35 #1
7-G1t35 #2
0
25
50
75
100
Percentage(%)
Promoter (<=1kb)
Promoter (1-2kb)
Promoter (2-3kc)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=3kb)
Distal Intergenic
C.
Interphase
G2M
61211
45486
9477
Interphase
G1t20
70485
36844
3675
Interphase
G1t35
59283
47726
7064
D.
G2M
G1t20
19890
35073
4943
G2M
G1t35
10991
43972
10621
E.
32127
7972
5167
4477
2274
1599
910
437
30K
20K
10K
0
Intersection Size
bookmarked
G2M_G1t20
G2M_G1t35
G2M_interphase
G2M_sites
40K
20K
0
Set Size
F.
G2M&G1 common
Interphase
1599
32127
74365
G.
G2M and G1 common
Bookmarked sites
Non-Bookmarked sites
71 kb
Fgr1
Mbd4
Lrba
Interphase
G2M
G2M
G1t20
G1t20
G1t35
G1t35

Figure 2. Assay for Transposase Accessible Chromatin (ATAC-seq) reveals putatively bookmarked gene loci. A.) Peak distribution for different ATAC-seq sample +/-3kB from the transcription start site (TSS). B.) Feature distribution of all the peaks across the whole genome for the different samples. C.) Comparison between interphase peaks and peaks common between the two replicates of G2M, G1t20 and G1t35 samples. D.) Peak comparison between G2M and early and late G1 populations E.) Plot showing the relationship of G2M (G2M_sites) sites shared between different overlapping samples. The overlaps compared are as follows: G2M sites shared with interphase (G2M_interphase), G2M sites shared with G1t20 (G2M_G1t20), and G2M sites shared with G1t35 (G2M_G1t35) and are shown at the bottom; set size refers to the number of gene loci within each corresponding overlapping sample. Filled circles represent sites shared with the respective sets, intersection size= number of overlaps represented by the filled circles. For example, 7972 sites are shared between G2M, Interphase and G1t35 but not G1t20. F.) Venn diagram of sites common between G2M, G1t20 and G1t35 (G2M&G1 common) and interphase, identifying the bookmarked and non-bookmarked sites G.) Representative signal tracks showing the three fractions identified in F.)


A.
Bookmarked (32127 gene loci)
Non- Bookmarked (74365 loci)
Overlap with H3K27Ac
Mit_H3K27Ac Bookmarked
15521 10863 21264
Mit_H3K27Ac non_bookmarked
21268 5116 69249
Genomic Feature Distribution
47.8% within 3kb of TSS
28% DI
9.96% within 3kb of TSS
48.16% DI
Promoter (<=1kb)
Promoter (1–2kb)
Promoter (2–3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=3kb)
Distal Intergenic
B.
Intersection Size
bookmarked_Oct4
bookmarked_sox2
bookmarked_parp1
bookmarked_SE
bookmarked_klf4
bookmarked_H3K27Ac
bookmarked
419/1121
587/1640
949/1519
1110/2563
1442/1914
10963/26384
32127
set size

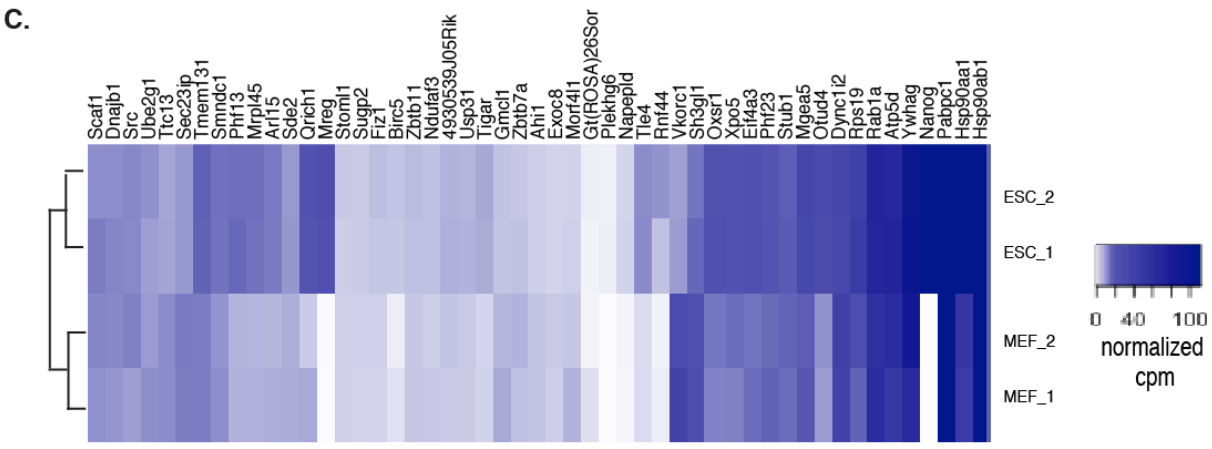

C.
ESC_2
ESC_1
MEF_2
MEF_1
0 40 100
normalized cpm

Figure 3. Transcription factor and epigenetic occupancy of bookmarked sites. A.) Characterization of the bookmarked and non-bookmarked sites. (top) Venn diagram showing the overlap with mitotic H3K27Ac mark and (bottom) the genomic feature distribution of genomic sites that are either bookmarked or non-bookmarked. B.) Occupancy of the bookmarked sites. UpSet plot showing the relationship between occupancy of bookmarked sites with mitotic specific binding of Oct4, Sox2, Klf4, and H3K27Ac and total binding sites for Parp1, and pluripotency specific super-enhancers (SE). Filled circles represent overlap between the different datasets. Intersection size= number of sites corresponding to each overlap. Set size= (total number of gene loci in the overlap (an aggregate of the filled circles) over total number of genes in the original dataset eg. for bookmarked_oct4: 419 is the total number of our bookmarked sites occupied by Oct4, 1121 is the total number of genomic sites present in Oct4 mitotic ChIP). The inset shows the magnified version of the bookmarked sites highlighting the most predominant overlaps. C.) Relative abundance of expression of Parp1, Klf4, and H3K27Ac bookmarked genes (red circled in 3B) in MEF vs ESC conditions, cpm= counts per million

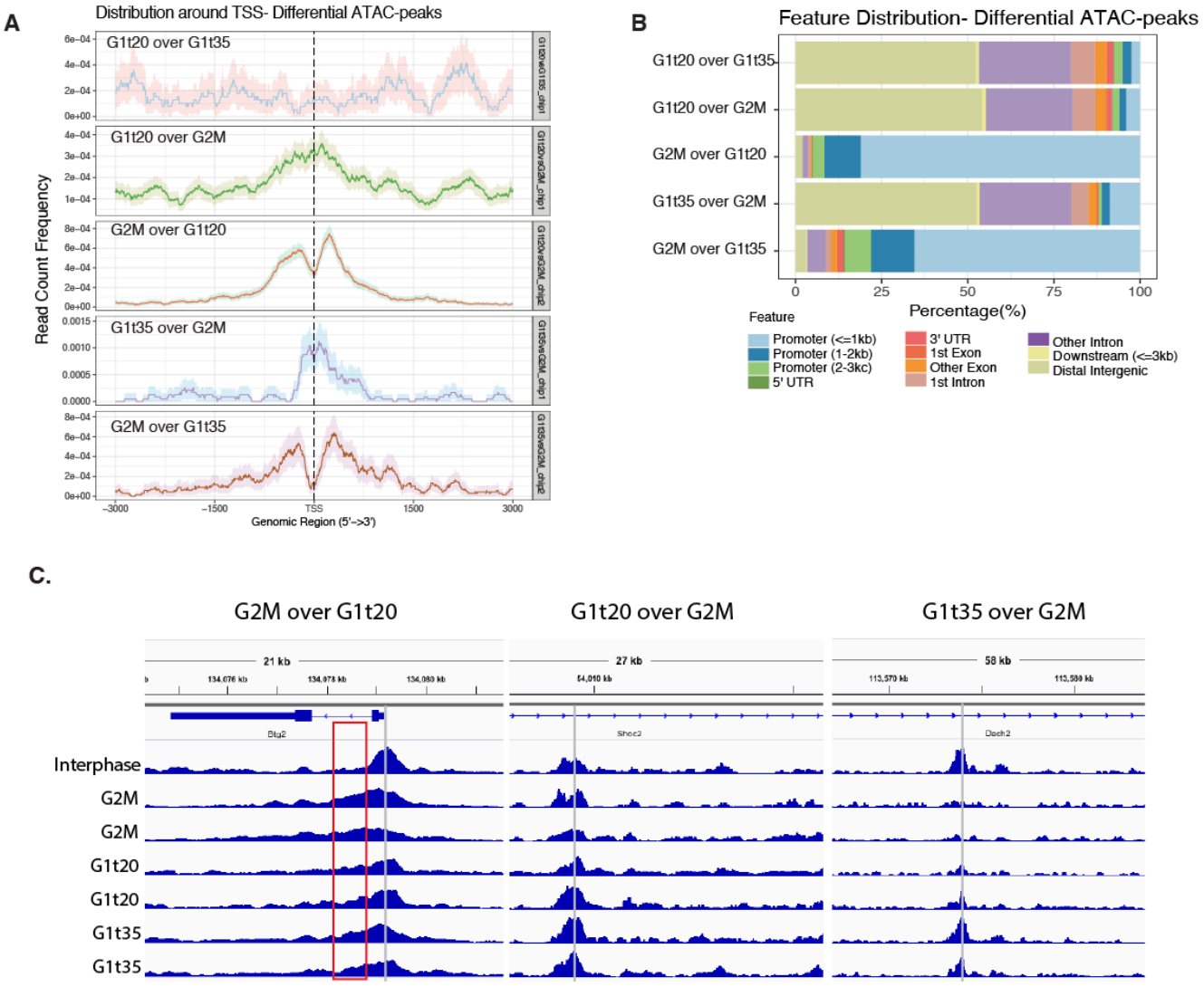


Figure 4. Differential peak analysis of accessible regions upon mitotic exit. A.) Peak distribution of differentially accessible ATAC-seq peaks within +/- 3kB of TSS. B.) Genoimc feature distribution of differentially accessible ATAC-seq peaks across the whole genome C.) Representative ATAC-seq signal tracks of regions identified as being differentially accessible. The vertical grey line marks the center of the peak during interphase, the red box marks the broad region of increased accessibility often observed at G2M enriched loci.

Table 1. Summary of differentially accessible ATAC-seq peaks between different cell cycle fractions. G2M over G1t20/G1t35= peaks enriched in G2M compared to G1t20/G1t35, G1t20/G1t35 over G2M= peaks enriched in G1t20/G1t35 compared to G2M, G2M_enriched_common= common peaks that are enriched in both G2M over G1t20 and G2M over G1t35. Highlighted row represents the % of bookmarked site identified in each sample

	G2M over G1t20	G1t20 over G2M	G2M over G1t35	G1t35 over G2M	G1t20 over G1t35	G1t35 over G1t20	G2M_enriched _common
Total Peaks	835	18055	301	932	2868	15	289
Total Genes Annotated	835	18022	301	932	2868	NA	289
Total Peaks (+/-3kB from TSS)	794	1440	258	113	218	NA	279
% in promoters (+/-3kB)	95.1	8.0	85.7	12.1	7.6	NA	96.5
% of bookmarked sites identified (Fig 2F)	82%	4%	63%	31%			78%
Total Genes Annotated (unique)	666	6292	259	759	1863	NA	123
Total unique genes (RNA overlap)	359	2056	138	233	643	NA	65
Total unique genes (up in ESC)	119	575	51	73	206	NA	32
% up in ES cells	33.15	27.97	36.96	31.33	32.04	NA	49.23

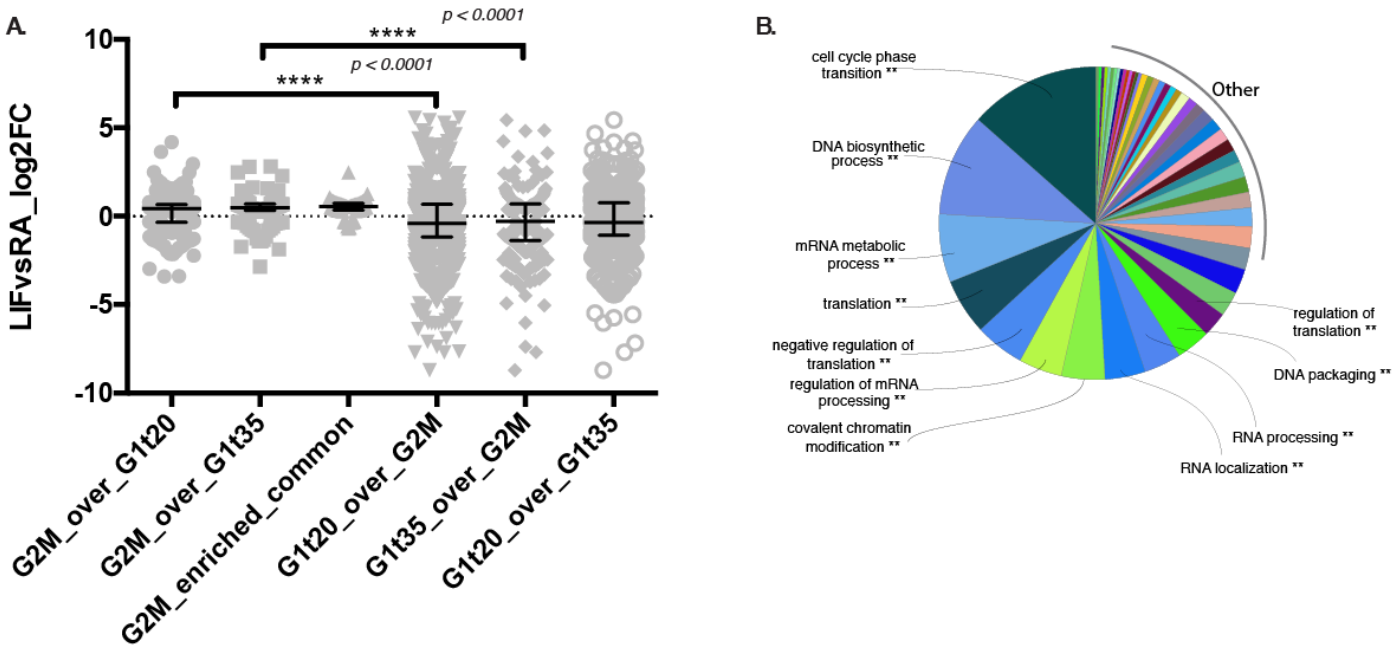


Figure 5. G2M-enriched loci show a bias towards a pluripotency gene expression signature. A.) Gene expression profile of the genes associated with differentially accessible ATAC-seq peaks, LIFvsRA_log2FC= log 2 fold change between ESCs grown in LIF (leukemia inhibitory factor, pluripotency conditions) and RA (retinoic acid, differentiation condition). pvalues were computed using two tailed unpaired t-test. B.) Biological processes up-regulated in G2M enriched dataset using ClueGO overrepresentation test ** pvalue<0.005

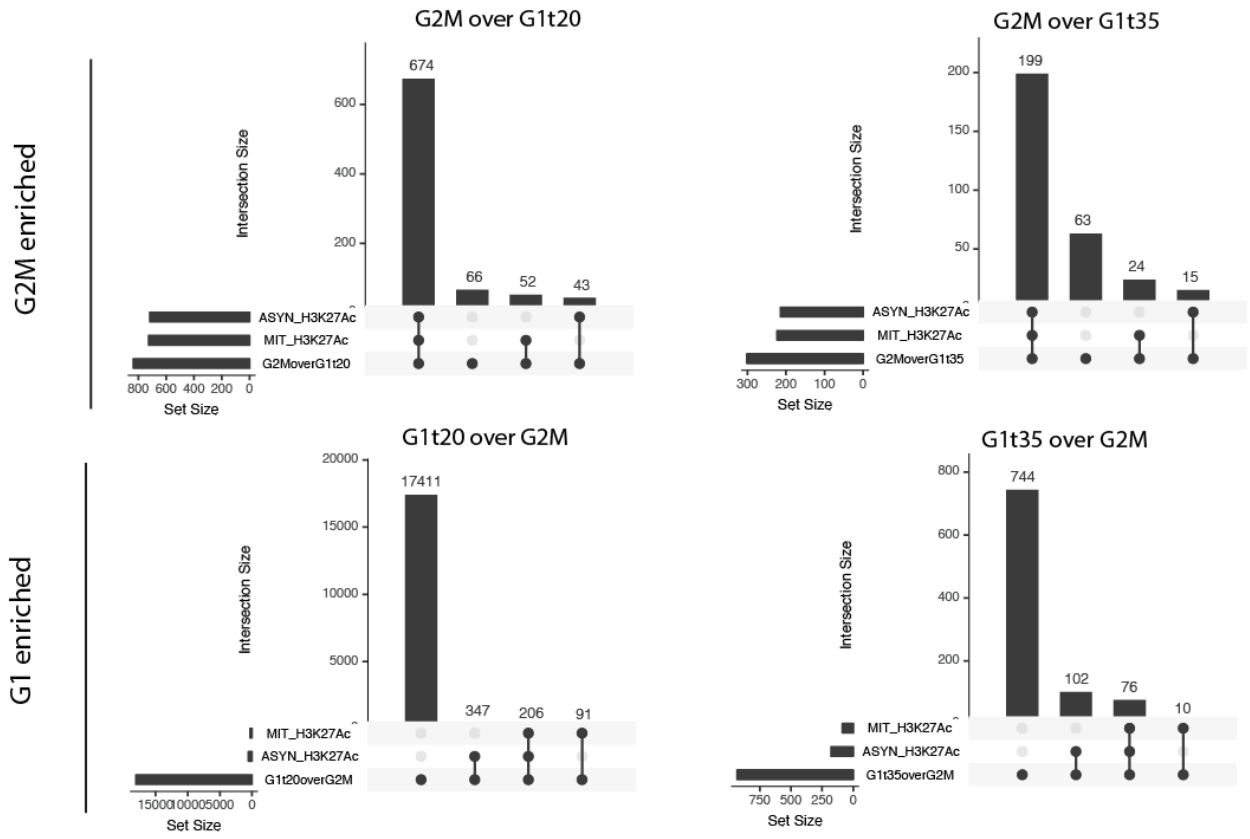


Figure 6. G2M-enriched loci are ubiquitously bookmarked by H3K27Ac during mitosis. A comparison of mitotic specific (MIT_H3K27Ac) or asynchronously (ASYN_H3K27Ac) associated H3K27Ac epigenetic mark with genomic sites that are differentially accessible between various populations.

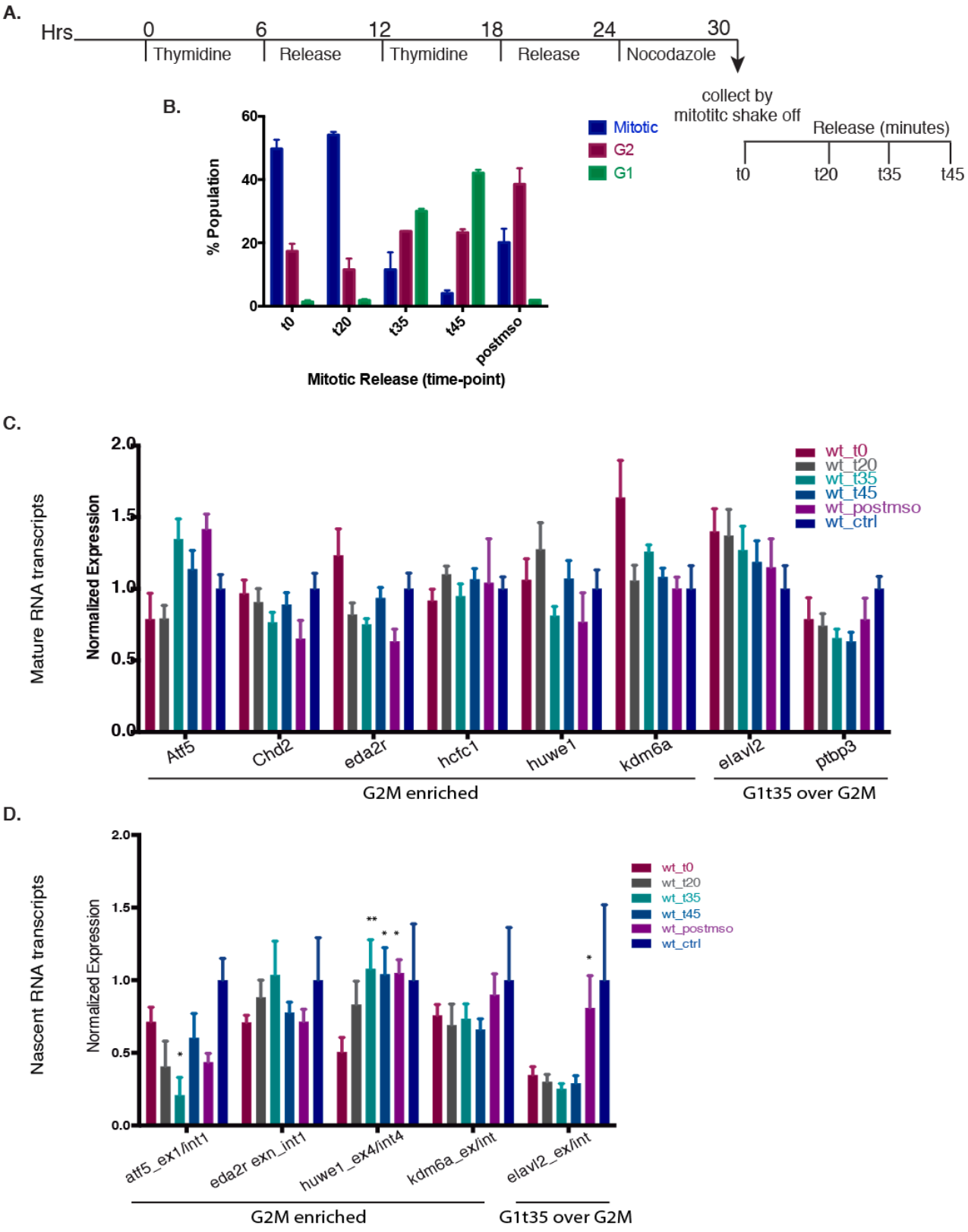

A.
Hrs
0
6
12
18
24
30
Thymidine
Release
Thymidine
Release
Nocodazole
collect by
mititc shake off
Release (minutes)
t0
t20
t35
t45
B.
% Population
Mitotic
G2
G1
Mitotic Release (time-point)
postmso
C.
Mature RNA transcripts
Normalized Expression
wt_t0
wt_t20
wt_t35
wt_t45
wt_postmso
wt_ctrl
Atf5
Chd2
eda2r
hcfc1
huwe1
kdm6a
elavl2
ptbp3
G2M enriched
G1t35 over G2M
D.
Nascent RNA transcripts
Normalized Expression
atf5_ex1/int1
eda2r exn_int1
huwe1_ex4/int4
kdm6a_ex/int
elavl2_ex/int
G2M enriched
G1t35 over G2M

Figure 7. Transcriptional profiles of selected differentially accessible gene loci upon mitotic exit: A.) Schematic of the mitotic release experiment B.) Frequency of cell cycle populations of wild-type (wt) cells profiled at each time point. Error bars= standard error of the mean (SEM) (n=2). t=0 mitotic cells collected by mitotic shake-off, t20, t35, t45 cells collected upon 20, 35 and 45 minute release into G1, post-mso= G2 cells that remain attached to the plate after collecting mitotic cells by shake-off. C-D.) qRT-PCR of C) mature and D.) nascent RNA transcripts of G2M enriched and G1t35 over G2M gene loci. Gene expression was normalized using Rpl13a and Tbp as reference genes. Statistical significance was assessed by multiple comparisons in a 2way ANOVA test using t0 samples as control *pval<0.05, **pval<0.01 (n=2 for mature transcripts and n=3 for nascent RNA qRT-PCR), error bars= SEM.

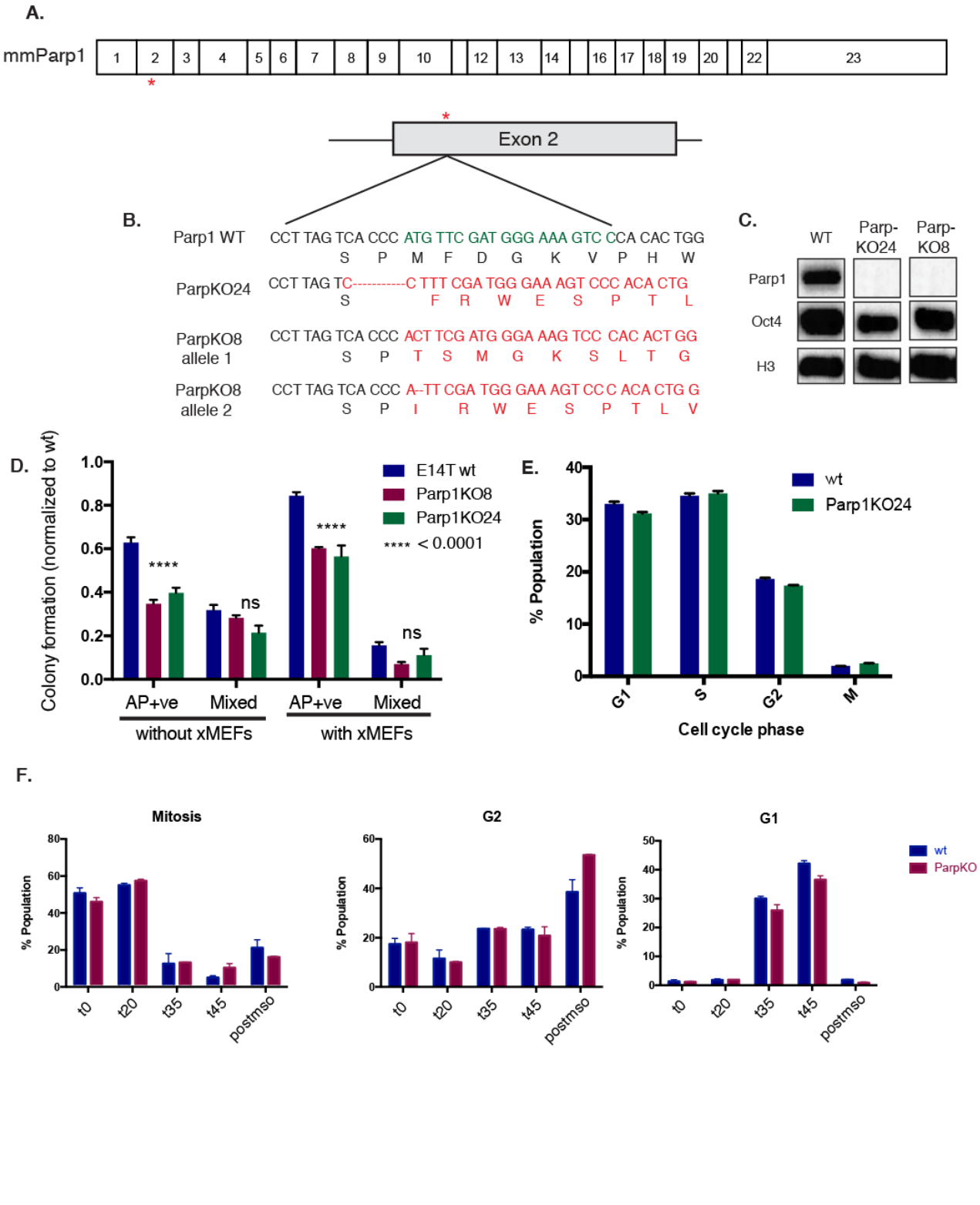

A.
mmParp1
1 2 3 4 5 6 7 8 9 10 12 13 14 16 17 18 19 20 22 23
Exon 2
B.
Parp1 WT CCT TAG TCA CCC ATG TTC GAT GGG AAA GTC CCA CAC TGG
S P M F D G K V P H W
ParpKO24 CCT TAG TC----------C TTT CGA TGG GAA AGT CCC ACA CTG
S F R W E S P T L
ParpKO8 allele 1 CCT TAG TCA CCC ACT TCG ATG GGA AAG TCC CAC ACT GG
S P T S M G K S L T G
ParpKO8 allele 2 CCT TAG TCA CCC A--TT CGA TGG GAA AGT CC C ACA CTG G
S P I R W E S P T L V
C.
WT
Parp-KO24
Parp-KO8
Parp1
Oct4
H3
D.
Colony formation (normalized to wt)
E14T wt
Parp1KO8
Parp1KO24
**** < 0.0001
****
ns
AP+ve
Mixed
without xMEFs
with xMEFs
E.
% Population
wt
Parp1KO24
G1
S
G2
M
Cell cycle phase
F.
Mitosis
G2
G1
% Population
t0
t20
t35
t45
postmso
wt
ParpKO

Figure 8. Generation and characterization of Parp1 knock-out (Parp1KO) mouse ES lines. A.) Protein (top) and genomic structure (bottom, numbers represent the number of exons) of mmParp1 (not to scale). Red * diagrammatically represents the site targeting by CRISPR/Cas9. B.) genomic sequence of the wild-type (wt) and two Parp1KO clones generated. C.) Western blot showing the protein expression of Parp1, Oct4 and loading control histone H3 in wt, and Parp1KO lines. D.) Colony forming assay for wt, Parp1KO8 and Parp1KO24, normalized to total number of wt colonies formed. AP+ve= alkaline phosphatase positive pluripotent colonies, mixed= heterogeneous clusters of cells. ****pvalue<0.0001 using multiple comparisons by 2-way ANOVA (n=3, alpha =0.05). E.) Cell cycle profile of asynchronous wt and Parp1KO24 using Hoechst and anti-MPM2 staining (n=2) F.) Cell cycle profile of Parp1KO cells upon mitotic arrest and release compared to wt, t=0 mitotic cells collected by mitotic shake-off, t20, t35, t45 cells collected upon 20, 35 and 45 minute release into G1, postmso= G2 cells that remain attached to the plate after collecting mitotic cells by shake-off (n=2).

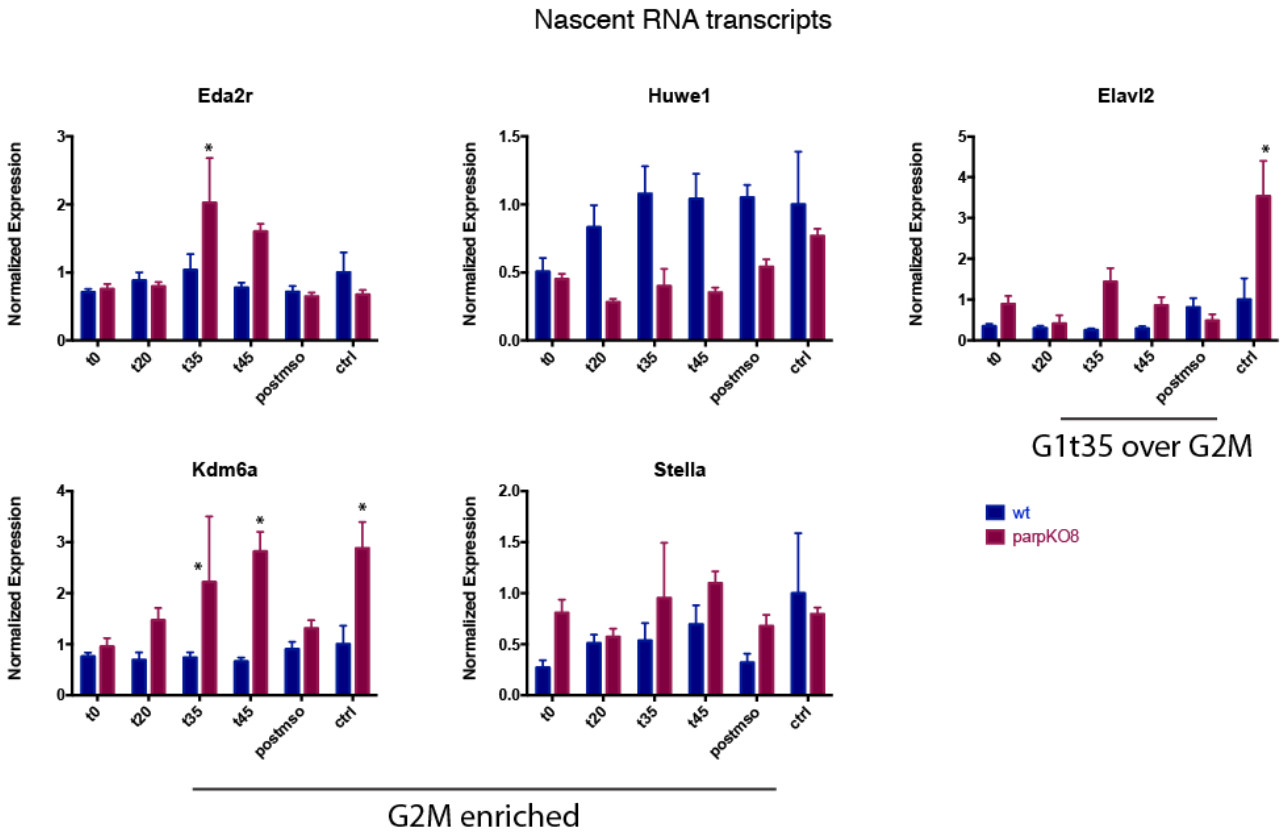


Figure 9. Transcriptional profiles of Parp1 targets upon mitotic exit in wild-type (wt) and Parp1 knock-out (parpKO8) cells. qRT-PCR of nascent RNA transcripts of G2M enriched and G1t35 over G2M gene loci. Gene expression was normalized using Rpl13a and Tbp as reference genes. t=0 mitotic cells collected by mitotic shake-off, t20, t35, t45 cells collected upon 20, 35 and 45 minute release into G1, post-mso= G2 cells that remain attached to the plate after collecting mitotic cells by shake-off. * Statistically significant as assessed by multiple t-tests (alpha=0.05) (n=3 for wt and n=2 for ParpKO8).

Table S1: Summary of ATAC-seq sequencing reads for all samples. Reads were mapped to mm10 genome and filtered out for PCR duplicates

Sample	sample_ID	raw reads	mapped reads	filtered
Interphase	AThi10016_GGACTCCT.sort.rmdup.bam	111141464	73791447	66279067
G2/M_rep2	AThi10016_CGAGGCTG.sort.rmdup.bam	124464846	82377956	74511908
G2/M_rep1	AThi10016_TAGGCATG.sort.rmdup.bam	118699507	57893404	52092076
G1T35_rep2	AThi10016_GTAGAGGA.sort.rmdup.bam	132001333	90744570	81191379
G1T35_rep1	AThi10016_CAGAGAGG.sort.rmdup.bam	135058762	65176238	58464246
G1T20_rep2	AThi10016_AAGAGGCA.sort.rmdup.bam	200771846	131914840	117240931
G1T20_rep1	AThi10016_CTCTCTAC.sort.rmdup.bam	128125779	80135766	71409670

Table S2: Summary of ATAC-seq peak calling using macs2. Peaks that were common between replicas were combined for further analysis

	Interphase	G2/M_rep1	G2/M_rep2	G1T20_rep1	G1T20_rep2	G1T35_rep1	G1T35_rep2
Total Peaks	107869	72027	85358	70265	60407	78562	75045
Total Genes Annotated	107797	71966	85294	70204	60342	78511	74983
Total Peaks (+/-3kB from TSS)	23436	19798	22389	20835	19903	21726	20964
% in promoters (+/-3kB)	21.7	27.5	26.2	29.7	33.0	27.7	28.0
Total peaks common between replicates	NA	54963		40519		54790	

Table S3: Top 110 ATAC-seq peaks for G2M_enriched_common sites (sorted based on significance)

chr	start	end	width	FC	qval	annotation	distanceToTSS	SYMBOL	GENENAME
chr1	134078122	134079110	989	2.002	1.02E-14	Promoter (<=1kb)	45	Btg2	B cell translocation gene 2, anti-proliferative
chr6	86525232	86525890	659	1.992	3.01E-13	Promoter (<=1kb)	275	Pcbp1	poly(rC) binding protein 1
chr14	50924832	50925270	439	2.173	4.34E-13	Promoter (<=1kb)	0	Apex1	apurinic/apyrimidinic endonuclease 1
chr11	6526302	6527070	769	2.558	2.59E-12	Promoter (1-2kb)	1690	Snhg15	small nucleolar RNA host gene 15
chr4	11156422	11157190	769	2.144	4.77E-12	Promoter (<=1kb)	0	Trp53inp1	transformation related protein 53 inducible nuclear protein 1
chr17	25272832	25273380	549	2.336	1.01E-11	Promoter (<=1kb)	534	Ube2i	ubiquitin-conjugating enzyme E2I
chr7	16098612	16099160	549	2.188	1.59E-11	Promoter (<=1kb)	0	Napa	N-ethylmaleimide sensitive fusion protein attachment protein alpha
chr1	178335192	178335960	769	2.938	1.70E-11	Promoter (1-2kb)	1824	Hnrnpu	heterogeneous nuclear ribonucleoprotein U
chr12	110694962	110695680	719	2.057	2.22E-11	Promoter (<=1kb)	92	Hsp90aa1	heat shock protein 90, alpha (cytosolic), class A member 1
chr16	11203062	11203390	329	2.235	2.39E-11	Promoter (<=1kb)	0	Rsl1d1	ribosomal L1 domain containing 1
chrX	162828712	162829370	659	2.022	4.43E-11	Promoter (<=1kb)	84	Txlng	taxilin gamma
chr2	34772102	34773310	1209	2.167	7.22E-11	Promoter (<=1kb)	12	Hspa5	heat shock protein 5
chr7	78896072	78897060	989	2.092	8.05E-11	Promoter (<=1kb)	14	Aen	apoptosis enhancing nuclease
chr12	110694542	110695640	1099	2.437	8.96E-11	Promoter (<=1kb)	132	Hsp90aa1	heat shock protein 90, alpha (cytosolic), class A member 1
chr1	178336952	178337610	659	1.725	1.06E-10	Promoter (<=1kb)	174	Hnrnpu	heterogeneous nuclear ribonucleoprotein U
chr2	132252012	132252560	549	2.816	1.07E-10	Promoter (<=1kb)	620	Pcna	proliferating cell nuclear antigen
chr11	40754232	40755110	879	2.116	1.08E-10	Promoter (<=1kb)	176	Ccng1	cyclin G1
chr10	117709242	117710120	879	2.120	1.14E-10	Promoter (<=1kb)	0	Mdm2	transformed mouse 3T3 cell double minute 2
chr11	6526172	6526710	539	2.494	3.88E-10	Promoter (2-3kb)	2050	Snhg15	small nucleolar RNA host gene 15
chrX	73966422	73966970	549	2.508	4.94E-10	Promoter (<=1kb)	-65	Hcfc1	host cell factor C1
chr7	24884642	24885520	879	2.008	6.39E-10	Promoter (<=1kb)	0	Rps19	ribosomal protein S19
chr19	6364602	6364930	329	3.048	6.86E-10	Promoter (<=1kb)	912	Sf1	splicing factor 1
chr11	69757712	69758260	549	2.410	7.35E-10	Promoter (<=1kb)	373	Polr2a	polymerase (RNA) II (DNA directed) polypeptide A
chr2	34773422	34774190	769	3.445	7.35E-10	Promoter (1-2kb)	1332	Hspa5	heat shock protein 5
chr7	4811952	4812390	439	2.043	7.65E-10	Promoter (<=1kb)	0	Ube2s	ubiquitin-conjugating enzyme E2S
chr3	96558442	96558880	439	2.634	1.01E-09	Promoter (<=1kb)	485	Txnip	thioredoxin interacting protein
chr1	135767282	135767940	659	2.942	1.03E-09	Promoter (1-2kb)	1197	Phlda3	pleckstrin homology like domain, family A, member 3
chr17	29093792	29094890	1099	1.835	1.12E-09	Promoter (<=1kb)	23	Cdkn1a	cyclin-dependent kinase inhibitor 1A (P21)
chr11	106787122	106787560	439	2.136	1.16E-09	Promoter (<=1kb)	934	Ddx5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
chr14	7243472	7244190	719	7.208	1.24E-09	Intron (uc011zfr.1/100038847, intron 5 of 6)	70847	Gm10406	predicted gene 10406
chr16	11253662	11253990	329	2.175	1.36E-09	Promoter (<=1kb)	335	Gspt1	G1 to S phase transition 1
chr14	7243502	7244050	549	5.353	1.47E-09	Intron (uc011zfr.1/100038847, intron 5 of 6)	70987	Gm10406	predicted gene 10406
chrX	97375632	97376510	879	2.744	1.49E-09	Promoter (<=1kb)	699	Eda2r	ectodysplasin A2 receptor

Table S3: Top 110 ATAC-seq peaks for G2M_enriched_common sites (sorted based on significance) (contd.)

chr	start	end	width	FC	qval	annotation	distanceTo TSS	SYMBOL	GENENAME
chr17	33955242	33955680	439	1.865	1.50E-09	Promoter (<=1kb)	0	Rps18	ribosomal protein S18
chr8	13159192	13160180	989	1.876	1.73E-09	Promoter (<=1kb)	31	Lamp1	lysosomal-associated membrane protein 1
chr9	106891382	106891650	269	2.314	2.77E-09	Promoter (<=1kb)	0	Manf	mesencephalic astrocyte-derived neurotrophic factor
chr11	62552052	62552710	659	1.829	2.78E-09	Promoter (<=1kb)	321	Ubb	ubiquitin B
chr11	6475482	6476360	879	2.010	3.49E-09	Promoter (<=1kb)	0	Purb	purine rich element binding protein B
chr7	45525152	45525700	549	3.032	3.49E-09	Promoter (<=1kb)	446	Ppp1r15a	protein phosphatase 1, regulatory (inhibitor) subunit 15A
chr10	91122842	91123110	269	3.327	3.53E-09	Promoter (<=1kb)	853	Slc25a3	solute carrier family 25 (mitochondrial carrier, phosphate carrier), member 3
chr7	3217832	3218270	439	2.558	3.96E-09	Promoter (<=1kb)	-356	Mir290a	microRNA 290a
chr9	106891292	106891840	549	1.953	4.55E-09	Promoter (<=1kb)	0	Manf	mesencephalic astrocyte-derived neurotrophic factor
chr7	28392982	28393200	219	2.128	4.86E-09	Promoter (<=1kb)	0	Paf1	Paf1, RNA polymerase II complex component
chr6	86525552	86526090	539	1.863	5.30E-09	Promoter (<=1kb)	75	Pcbp1	poly(rC) binding protein 1
chr11	69578822	69579090	269	2.117	6.33E-09	Promoter (<=1kb)	234	Wrap53	WD repeat containing, antisense to Trp53
chr1	55087122	55087560	439	2.057	7.97E-09	Promoter (<=1kb)	372	Hspd1	heat shock protein 1 (chaperonin)
chr11	6626182	6626730	549	2.080	8.10E-09	Promoter (<=1kb)	-115	Tbrg4	transforming growth factor beta regulated gene 4
chr7	45525812	45526250	439	2.345	8.73E-09	Promoter (<=1kb)	0	Ppp1r15a	protein phosphatase 1, regulatory (inhibitor) subunit 15A
chr10	117709112	117709560	449	2.458	8.90E-09	Promoter (<=1kb)	468	Mdm2	transformed mouse 3T3 cell double minute 2
chr5	125388892	125389440	549	1.954	1.08E-08	Promoter (<=1kb)	314	Ubc	ubiquitin C
chr10	117707592	117708140	549	2.281	1.16E-08	Promoter (1-2kb)	1888	Mdm2	transformed mouse 3T3 cell double minute 2
chr14	20929482	20929700	219	2.091	1.29E-08	Promoter (<=1kb)	49	Vcl	vinculin
chr19	5845182	5845400	219	2.313	1.39E-08	Promoter (<=1kb)	80	Neat1	nuclear paraspeckle assembly transcript 1 (non-protein coding)
chr7	16309042	16310140	1099	1.777	1.86E-08	Promoter (<=1kb)	0	Bbc3	BCL2 binding component 3
chr7	16310362	16310690	329	2.112	2.00E-08	Promoter (<=1kb)	779	Bbc3	BCL2 binding component 3
chr7	24885812	24886080	269	3.467	2.10E-08	Promoter (1-2kb)	1098	Rps19	ribosomal protein S19
chr2	34870662	34870990	329	1.996	2.18E-08	Promoter (<=1kb)	0	Psmd5	proteasome (prosome, macropain) 26S subunit, non-ATPase, 5
chr2	130274912	130275270	359	1.935	2.32E-08	Promoter (<=1kb)	-245	Snord110	small nucleolar RNA, C/D box 110
chr15	11995612	11995940	329	1.925	2.41E-08	Promoter (<=1kb)	67	Sub1	SUB1 homolog (S. cerevisiae)
chr1	134078492	134078850	359	1.677	2.42E-08	Promoter (<=1kb)	305	Btg2	B cell translocation gene 2, anti-proliferative
chr7	127968392	127969050	659	2.077	2.65E-08	Promoter (<=1kb)	913	Fus	fused in sarcoma
chr10	117709832	117710100	269	1.752	2.83E-08	Promoter (<=1kb)	0	Mdm2	transformed mouse 3T3 cell double minute 2
chr3	32365192	32365520	329	2.035	2.85E-08	Promoter (<=1kb)	0	4930429B21Rik	RIKEN cDNA 4930429B21 gene
chr17	25272992	25273170	179	2.324	2.97E-08	Promoter (<=1kb)	744	Ube2i	ubiquitin-conjugating enzyme E2I
chr10	34282272	34282600	329	2.115	3.19E-08	Promoter (<=1kb)	82	Tspyl1	testis-specific protein, Y-encoded-like 1
chr9	20652062	20652500	439	1.754	3.23E-08	Promoter (<=1kb)	0	Pin1	protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting 1
chr7	34389302	34389520	219	2.292	3.50E-08	Promoter (<=1kb)	20	Lsm14a	LSM14A mRNA processing body assembly factor
chr5	24576972	24577410	439	1.980	3.64E-08	Promoter (<=1kb)	57	Abcf2	ATP-binding cassette, sub-family F (GCN20), member 2
chr11	40733112	40733440	329	1.913	3.67E-08	Promoter (<=1kb)	0	Hmmr	hyaluronan mediated motility receptor (RHAMM)
chr10	81177692	81178020	329	3.125	3.94E-08	Promoter (<=1kb)	36	Eef2	eukaryotic translation elongation factor 2

Table S3: Top 110 ATAC-seq peaks for G2M_enriched_common sites (sorted based on significance) (contd.)

chr	start	end	width	FC	qval	annotation	distanceToTSS	SYMBOL	GENENAME
chr12	3426302	3426660	359	2.097	4.12E-08	Promoter (<=1kb)	87	1110002L01Rik	RIKEN cDNA 1110002L01 gene
chr11	51606392	51606830	439	1.762	4.34E-08	Promoter (<=1kb)	51	Hnrnpab	heterogeneous nuclear ribonucleoprotein A/B
chr15	31601132	31601790	659	1.939	4.49E-08	Promoter (<=1kb)	14	Cct5	chaperonin containing Tcp1, subunit 5 (epsilon)
chr17	29033292	29033840	549	2.808	4.52E-08	Promoter (<=1kb)	632	Srsf3	serine/arginine-rich splicing factor 3
chr17	34118482	34118920	439	1.821	4.60E-08	Promoter (<=1kb)	980	Brd2	bromodomain containing 2
chr7	44816422	44816750	329	1.849	4.65E-08	Promoter (<=1kb)	0	Atf5	activating transcription factor 5
chr6	124711652	124711920	269	2.235	4.69E-08	Promoter (<=1kb)	258	Emg1	EMG1 N1-specific pseudouridine methyltransferase
chr7	24885852	24886620	769	2.609	4.90E-08	Promoter (1-2kb)	1138	Rps19	ribosomal protein S19
chr5	32459352	32459790	439	2.025	5.08E-08	Promoter (<=1kb)	382	Ppp1cb	protein phosphatase 1, catalytic subunit, beta isoform
chr7	4811852	4812210	359	2.358	5.94E-08	Promoter (<=1kb)	130	Ube2s	ubiquitin-conjugating enzyme E2S
chr17	24415052	24415380	329	2.197	5.94E-08	Promoter (<=1kb)	377	Rnps1	ribonucleic acid binding protein S1
chr7	28393022	28393290	269	2.066	6.27E-08	Promoter (<=1kb)	26	Paf1	Paf1, RNA polymerase II complex component
chr3	95659632	95660070	439	2.545	6.38E-08	Promoter (<=1kb)	911	Mcl1	myeloid cell leukemia sequence 1
chr17	17178702	17179140	439	8.260	6.65E-08	5' UTR	57319	Zfp97	zinc finger protein 97
chr5	31054872	31055310	439	2.290	8.05E-08	Promoter (<=1kb)	92	Cad	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase
chr11	3648372	3648920	549	1.830	8.71E-08	Promoter (<=1kb)	0	Tug1	taurine upregulated gene 1
chr17	29034832	29035160	329	3.231	9.06E-08	Promoter (2-3kb)	2172	Srsf3	serine/arginine-rich splicing factor 3
chr11	51606362	51606630	269	1.852	9.59E-08	Promoter (<=1kb)	251	Hnrnpab	heterogeneous nuclear ribonucleoprotein A/B
chr11	80383382	80383600	219	2.007	1.01E-07	Promoter (<=1kb)	103	Zfp207	zinc finger protein 207
chr15	31601072	31601520	449	2.245	1.07E-07	Promoter (<=1kb)	284	Cct5	chaperonin containing Tcp1, subunit 5 (epsilon)
chr8	70510112	70510330	219	1.793	1.07E-07	Promoter (<=1kb)	37	Uba52	ubiquitin A-52 residue ribosomal protein fusion product 1
chr14	46084172	46084610	439	1.938	1.08E-07	Promoter (<=1kb)	145	Gm1821	ubiquitin pseudogene
chr14	31206452	31206780	329	2.230	1.24E-07	Promoter (<=1kb)	46	Nisch	nischarin
chr8	85024832	85025160	329	1.852	1.31E-07	Promoter (<=1kb)	118	Asna1	arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)
chr17	29094482	29094750	269	1.977	1.33E-07	Promoter (<=1kb)	713	Cdkn1a	cyclin-dependent kinase inhibitor 1A (P21)
chr10	42275092	42275420	329	2.060	1.38E-07	Promoter (<=1kb)	178	Foxo3	forkhead box O3
chr3	108384762	108384980	219	2.358	1.38E-07	Promoter (<=1kb)	924	Psrc1	proline/serine-rich coiled-coil 1
chr13	23531752	23532080	329	2.100	1.43E-07	Promoter (<=1kb)	708	Hist1h4h	histone cluster 1, H4h
chr10	117708582	117708800	219	3.626	1.49E-07	Promoter (1-2kb)	1228	Mdm2	transformed mouse 3T3 cell double minute 2
chr17	33940832	33941160	329	1.976	1.57E-07	Promoter (<=1kb)	109	Wdr46	WD repeat domain 46
chr7	45525152	45525420	269	2.654	1.57E-07	Promoter (<=1kb)	726	Ppp1r15a	protein phosphatase 1, regulatory (inhibitor) subunit 15A
chr11	5098722	5099050	329	1.744	1.57E-07	Promoter (<=1kb)	38	Ewsr1	Ewing sarcoma breakpoint region 1
chr12	17545442	17545770	329	2.781	1.60E-07	Promoter (<=1kb)	569	Odc1	ornithine decarboxylase, structural 1
chr11	69578742	69579180	439	1.960	1.67E-07	Promoter (<=1kb)	144	Wrap53	WD repeat containing, antisense to Trp53
chr4	126200792	126201060	269	2.860	1.71E-07	Promoter (1-2kb)	1650	Thrap3	thyroid hormone receptor associated protein 3
chr2	132251942	132252480	539	2.439	1.73E-07	Promoter (<=1kb)	700	Pcna	proliferating cell nuclear antigen
chr13	23531852	23532120	269	2.429	1.81E-07	Promoter (<=1kb)	808	Hist1h4h	histone cluster 1, H4h
chr8	3500532	3500750	219	2.046	1.96E-07	Promoter (<=1kb)	13	Mcoln1	mucolipin 1
chr11	106787252	106787520	269	1.934	2.03E-07	Promoter (<=1kb)	974	Ddx5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
chr19	5963872	5964200	329	1.870	2.09E-07	Promoter (<=1kb)	6	Pola2	polymerase (DNA directed), alpha 2

Table S4: Primers used in the study

CrsprParp1_gRNA-F	CACCGGGACTTTCCCATCGAACAT	sgRNA for CRISPR-Cas9 mediated KO
CrsprParp1_gRNA-R	aaacATGTTCGATGGGAAAGTCCC	sgRNA for CRISPR-Cas9 mediated KO
CrsChk_Parp1_Int1_F2	CCA GGA TGA GAA GCC AGA AG	Primer for checking genomic Parp1 DNA after Ko
CrsChk_Parp1_Exn2_R2	CAG AAG CAA CTC AGC AGA TAG A	Primer for checking genomic Parp1 DNA after Ko
Qmm_Huwe1_Ex4F	ATGATGAGCAACTCCTCTTGG	Nascent transcripts
Qmm_Huwe1_Int4R	GCATGTTCCCTATCCTCTGTTAT	Nascent transcripts
Qmm_Eda2r_Ex1F	CACCTATTGTGAGAGCGGTATG	Nascent transcripts
Qmm_Eda2r_Int1R	CATTCGAGTACAGAGCAGACAC	Nascent transcripts
Qmm_Atf5_ExnInt1F	GGAATAAGATGAGGTGGGTAGG	Nascent transcripts
Qmm_Atf5_Int1R	CACACATTCCCAGGGACATTA	Nascent transcripts
Qmm_Stella_Ex2F	CTTTGTTGTCGGTGCTGAAAG	Nascent transcripts
Qmm_Stella_Int2R	GCTGGAGTTGCTCTTAGGTC	Nascent transcripts
Qmm_Elavl2_ExnF	CACAGTATGGGCGCATCATTA	Nascent transcripts
Qmm_Elavl2_IntR	TCAGTCAGGGAGCACAAGA	Nascent transcripts
Qmm_Kdm6a_ExnF	AGACCTAGTCCTCAGATCATACC	Nascent transcripts
Qmm_Kdm6a_IntR	ATCGTCAAACACTTCACTCTGT	Nascent transcripts
Qmm_Kdm6a_F	GGCCTTGCTGGAGCTCTTAA	Mature transcripts
Qmm_Kdm6a_R	TGGTTCCAGTAGGGTCCCAA	Mature transcripts
Qmm_Atf5_F	TGGGCTGGCTCGTAGACTAT	Mature transcripts
Qmm_Atf5_R	CCCGCTCAGTCATCCAATCA	Mature transcripts
Qmm_Eda2r_F	GCAGACTCCTTCTTCCGAGG	Mature transcripts
Qmm_Eda2r_R	TTCCCACCAGTGCAACAAGT	Mature transcripts
Qmm_Hcfc1_F	CCCAAGATTGCTACTGGCCA	Mature transcripts
Qmm_Hcfc1_R	TTGACAGCAGAGACGGTGAC	Mature transcripts
QMM_Huwe1_1468_F	GAACTCCCAAGCTCAGCAGT	Mature transcripts
QMM_Huwe1_1594_R	TGATGGGGGTATGGGTCCAT	Mature transcripts
QMM_Chd2_4921_F	AAGTGAGCCTGTTCCCATCG	Mature transcripts
QMM_Chd2_5016_R	GCCTTCTTCACAGGCCTCAT	Mature transcripts
QMM_Tbp_F	AAGAGAGCCACGGACAACTG	Mature transcripts
QMM_Tbp_R	AGCCCAACTTCTGCACAACT	Mature transcripts
QmmRpl13a-AS	GTC ACT GCC TGG TAC TTC C	Mature transcripts
QmmRpl13a-S	TCC CTC CAC CCT ATG ACA AG	Mature transcripts

Table S5: Bedtools comparison between the genomic intersection of asynchronous and mitotic ATAC-seq peaks identified by Teves et al., 2016

Teves et al 2016	Total number of peaks	Peaks common between replicates	Async only	Mitotic only	Bookmarked	
GSM2259901_Arep1peaks_formatted.bed	34044	23583	10896			
GSM2259902_Arep2peaks_formatted.bed	34495				12687	
GSM2259903_Mrep1peaks_formatted.bed	27450	15503		2816		
GSM2259904_Mrep2peaks_formatted.bed	20654					
Shared with our 'bookmarked' sites					10824	85%

Chapter 2.4 Conclusions and Future Directions

Section 2 of this thesis encompassed a research goal to understand and characterize the role of mitotic bookmarking in maintaining pluripotent cell fate. When we commenced this project there were no reports of mitotic bookmarking in pluripotent cells. Over the course of last year a significant development has been made in this area of research with several publications addressing the concept of mitotic bookmarking in ES cells by various pluripotency transcription factors (Deluz et al., 2016; Festuccia et al., 2016; Liu et al., 2017b; Teves et al., 2016). All these studies involve a thorough candidate based approach and have contributed to the field in multiple ways. Over the course of this chapter I will discuss how the landscape of understanding around this subject has changed and the ways in which our study contributes towards it.

The changing mitotic bookmarking landscape

Mitosis is the most visibly distinguishable phase of the cell cycle, characterized by chromosome condensation. Due to the intense condensation of chromatin during mitosis, and various early studies showing the de-coupling of transcription regulatory components from the mitotic chromatin (Luscher and Eisenman, 1992; Martinez-Balbas et al., 1995; Parsons and Spencer, 1997; Roberts et al., 1991) it was thought that: a) DNA loses accessibility during mitosis and b) it was the norm for most chromatin bound components to decouple during mitosis.

However, various subsequent studies that provided proof against those assumptions lead to the conception of the idea of mitotic bookmarking. The concept was first conceived as a mechanism to explain the co-relation of DNA sites that were single stranded during mitosis, with genes that were actively transcribed in G1 (Juan et al.,

1996; Michelotti et al., 1997). It was known at that the time that various transcription factors and members of the transcription machinery decouple from the mitotic chromatin (Luscher and Eisenman, 1992; Martinez-Balbas et al., 1995; Parsons and Spencer, 1997; Roberts et al., 1991) resulting in a transcriptionally silent state (Johnson and Holland, 1965; Taylor, 1960). So, how was it then, that during this transcriptionally dead state and amidst the loss of transcription machinery, the cells knew to keep gene loci of the genes that would be transcribed in later G1 phase, in a single stranded state during mitosis? It was proposed that certain "bookmarks" were deposited onto the chromatin in G2 to mark active genes, and were propagating through mitotic division into G1 of the daughter cells to re-establish their transcriptional identity (John and Workman, 1998; Michelotti et al., 1997).

Since then several studies have identified a wide variety of 'memory signatures' (Hsiung et al., 2015) that could act as mitotic bookmarks including, several mitotically retained transcription factors, and various epigenetic marks (Summarized in Chapter 2.1, Table 2). In contrast to these studies, we assayed the global protein occupancy of mitotic chromatin with the aim to identify putative mitotic bookmarking factors. Surprisingly, we found that a majority of the global chromatin bound factors are retained on the mitotic chromatin (Chapter 2.2). This observation, in light of the ATAC-seq data, however, is not surprising since a large portion of the genomic loci retain their chromatin accessibility during mitosis (Chapter 2.3).

Initial research that focused on mitotic bookmarking based on the mitotic retention of transcription and chromatin regulatory factors suggested that: a.) this was a rare phenomenon, b.) only a select few proteins would be capable of it, and c.) a small

subset of genomic sites would be bookmarked (Blobel et al., 2009, Zhao et al., 2011, Kaduke et al, 2012, Arora et al., 2012). Recent data including this study, however, suggest that mitotic bookmarking by chromatin bound factors is likely to be more prevalent than was previously appreciated and that a large portion of the genome is bookmarked during mitosis than what we had previously known (Teves et al., 2016, Hsuing et al., 2015, Liu et al., 2017b, Caravaca et al., 2013, Chapters 2.2-2.3). Even though mitotic bookmarking by chromatin bound factors is not considered to be rare anymore, it by no means is a characteristic of all chromatin-binding proteins. There is live cell imaging data clearly showing that not all interphase chromatin bound proteins are retained on the mitotic chromatin (Caravaca et al., 2013, Teves et al., 2016, Chapter 2.2). Similarly, a large portion but not all genomic sites retain their accessible nature during mitosis (Chapter 2.3).

Figure 1 represents the conclusion of our study. Bookmarked sites retain their accessible nature from interphase, through mitosis and into G1, and tend to be at proximal gene promoters. Subsets of these sites are occupied by H3K27Ac, Parp1 and various other factors such as Oct4, Sox2, and Klf4. We speculate that these factors, and potentially other MBFs identified here, act as bookmarks of these sites preserving their accessibility status into the G1 of daughter sites. However, the relationship between the occupancy of these MBFs during interphase, mitosis and the subsequent G1 still needs to be established on a candidate basis (Fig 1).

The genome sites that we identified as being non-bookmarked lose their accessibility from interphase to G1, and tend to be more distal inter-genic. We speculate that in ES cells these sites are dispensable for the acquisition of a pluripotent identity but

are opened up during later cell cycle phases as a result of a pluripotent state. These sites could lose their accessibility status either by loss of a euchromatic mark or by deposition of a heterochromatic mark. It is possible that a bookmark is deposited at these sites during late G2/M to actively close them off (Fig 1). A number of MBFs identified (Chapter 2.2) tend to associate with heterochromatic marks and could, therefore, be involved in maintenance of a repressive chromatin state. Mitotic bookmarking, is defined as mechanism for maintaining the transcriptional identity of a cell (Kadauke and Blobel, 2012; Sarge and Park-Sarge, 2009; Zaidi et al., 2010), however, historically has only been studied in context of preservation of active transcriptional states (Blobel et al., 2009; Caravaca et al., 2013; Festuccia et al., 2016; Lodhi et al., 2016). Considering that a significant portion of the chromatin retains accessibility during mitosis, while the rest is actively shut off perhaps maintenance of a repressive state of certain gene loci is equally as important for preserving cellular identity.

In conclusion, this study has contributed towards understanding the extent of involvement of mitotic bookmarking as a mechanism for maintaining cellular identity. We provide evidence that challenges some of the basic assumptions that have set the basis for this field of study. These data are in concert with some other recently published studies (Teves et al., 2016, Hsuing et al., 2015), and offer a means to further investigate this field of research. We provide many avenues to explore that could eventually lead to a better understanding of the mitotic bookmarking phenomenon. I'll further discuss some of the questions that can be addressed using this study as a basis and the data generated from it.

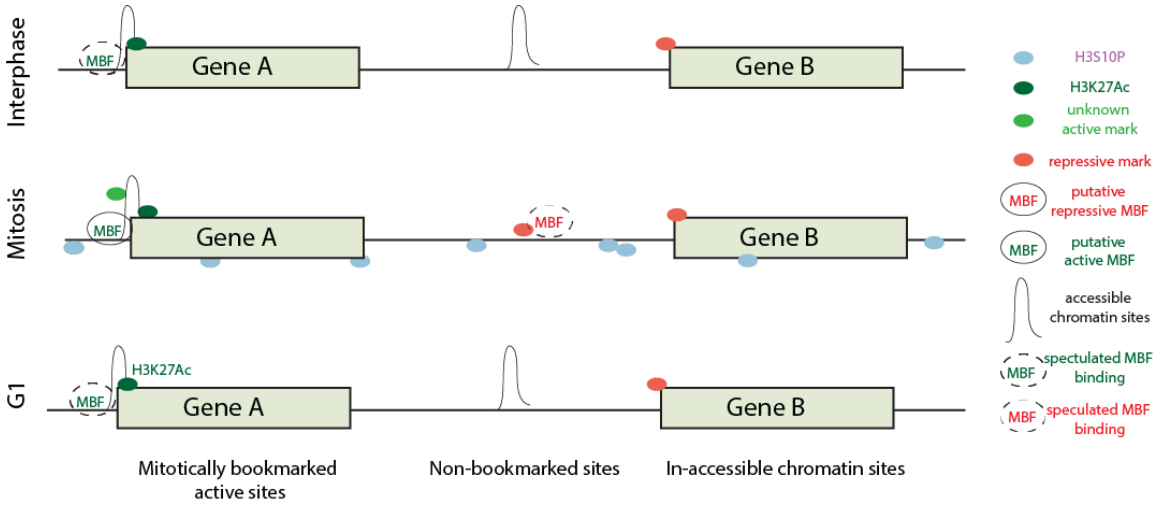


Figure 1. A summary of this study. Mitotic chromatin is ubiquitously coated with H3S10P mark. The schematic represents the different accessibility profiles observed in our study. Gene A represents a gene whose TSS is bookmarked through mitosis by a putative MBF, while the accessibility of the non-bookmarked site is not preserved through mitosis. Regions not accessible in interphase remain unchanged during mitosis.

Future Directions

In keeping with the shifting idea of a more involved role of mitotic bookmarking in maintaining cellular fate, our ChIP-MS screen (Chapter 2.2) provides a list of putative mitotic bookmarking factors (MBFs). These MBFs have not yet been studied in context of their mitosis dependent role, and an investigation can provide important gene regulatory information. We identified several categories of putative bookmarks that interact with both heterochromatin and euchromatin and therefore can act be involved in preservation of chromatin state of active and/or repressed genes during mitosis (Fig 1). We also identify a number of chromatin remodeling proteins, epigenetic regulators, and

RNA binding proteins that could be involved in maintaining the accessible status of various genes during mitosis. The ATAC-seq identified potentially bookmarked and non-bookmarked sites (Chapter 2.3). It would be informative to assay the affects of perturbation of one or more MBFs, specifically those involved in chromatin remodeling, on the accessibility of bookmarked and non-bookmarked sites.

There are several questions that arise out of this data. What happens to the balance between bookmarked and non-bookmarked sites when altering a cell's fate, such as during reprogramming and differentiation? Do certain non-bookmarked sites become bookmarked upon a transition of fates? The hypothesis would be that bookmarked genomic loci would vary between different cell types, the extent to which they would vary, however, is unknown. Could we establish and use this information to better understand cellular identity at a chromatin level, and employ it for purposes of differentiation and reprogramming. Such a study would also provide concrete evidence associating mitotic bookmarking of genomic sites with cell fate.

Currently, there is a gap in understanding the relationship between bookmarked sites, the transcriptional profiles of bookmarked genes, and cellular identity. It is a challenging problem to address because of the small window for profiling transcription between early mitosis (where some transcription is still ongoing), metaphase (which is transcriptionally silent), late telophase (where transcription is reactivated) and G1 (where transcription proceeds normally). Traditional RNA-seq based assays are not sensitive enough to pick up differences in nascent RNA transcription over short periods of time, and perhaps techniques like GRO-seq can be used to address this (Core et al., 2008; Liu et al., 2017a).

Additionally, there is a lack of concrete evidence linking the mitotic association of bookmarking factors with cell fate. Several groups have provided evidence for the importance of bookmarking factors during phase transition from mitosis to G1 (Kadauke et al., 2012, Liu et al., 2017, Festucia et al., 2016, Deluz et al., 2017), but were unable to robustly conclude the importance of mitotic association of these factors. This warrants a study into the development of putative degron domains that could de-couple a protein from chromatin at the onset of mitosis until G1. Additionally, putative protein mutants for various MBFs can be identified and synthesized (such as in Rothbart et al., 2012) with the aim to specifically degrade them during mitosis.

Concluding remarks

Overall, we performed an unbiased and thorough investigation into the study of mitotic bookmarking in ES cells. Our study has furthered the understanding of the field by providing evidence that suggest a more prominent role of mitotic bookmarking in maintenance of cell identity. It has also opened up a series of questions that need to be addressed to establish a causal relationship between mitotic bookmarking and fate maintenance. The data provided here could be used as a basis to start answering those questions, and robustly classify mitotic bookmarking as a fate-determining phenomenon.

R-Scripts

170404_ATAC_ChIPseeker.R 2017-07-14, 11:13 AM

```
#170327_ATAC_PePr_bam_sharp_chipseeker analysis

library(GenomicFeatures)
library(GenomicRanges)
library(TxDb.Mmusculus.UCSC.mm10.knownGene)
library(org.Mm.eg.db)
library(ChIPseeker)

## ------------------------------------------------------------------------
## loading packages
library(ChIPseeker)
library(TxDb.Mmusculus.UCSC.mm10.knownGene)
txdb <- TxDb.Mmusculus.UCSC.mm10.knownGene
library(magrittr)
library(UpSetR)
library(dplyr)

setwd("/Users/draperlabair1/Documents/McMaster/Data folder_Backup/
    Bioinformatics/PePr_diff_peak_calling_R/PePr_results_bam_sharp/
    PePr_processed")

# get parp_chip targets
parp_beds= list.files(path="/Users/draperlabair1/Documents/McMaster/Data
    folder_Backup/Bioinformatics/Parp1_ChipSeq_Processed/Parp1_GSE81168",
    pattern=".bed", all.files = TRUE, full.names = TRUE)
parp_beds=parp_beds[-3]


for (i in 1:length(parp_beds)){
  new_file_name= paste0(unlist(strsplit(parp_beds[i], split="/"))[10])
  data=read.table(parp_beds[i])
  data= filter(data, V5>=200)
  write.table(data, file= paste0(new_file_name,"_filtered_signal_200"),
      sep="\t", col.names = FALSE, row.names = FALSE)
  assign(paste(new_file_name), data)
}

parp_beds_filtered= list.files(path="/Users/draperlabair1/Documents/McMaster/
    Data folder_Backup/Bioinformatics/PePr_diff_peak_calling_R/
    PePr_results_bam_sharp/PePr_processed/",
                               pattern="*.bed", full.names=T, recursive=FALSE)

parp_beds_combined= merge(GSM1910638_ESC_WT_Parp1_new_Rep1_rm_blacklist.bed,
    GSM1910639_ESC_WT_PARP1_new_Rep2_rm_blacklist.bed, all=TRUE)
write.table(parp_beds_combined, file=
    paste0("combined_all_parp1_filtered_signal_200.bed"), sep="\t", col.names =
    FALSE, row.names = FALSE)

beds = list.files(path="/Users/draperlabair1/Documents/McMaster/Data
    folder_Backup/Bioinformatics/PePr_diff_peak_calling_R/
    PePr_results_bam_sharp/PePr_bed_files/",
                  pattern="*peaks.bed", full.names=T, recursive=FALSE)

fileNumbers = seq(beds)
df_list= list()
for (i in fileNumbers) {
```

```
  sample= read.table(beds[i],comment="#", header=FALSE)
  new_file_name = paste0(unlist(strsplit(beds[i],"/"))[12])
  assign(paste(new_file_name), sample)
  df_list= c(df_list, new_file_name)
}

dfs <- ls()[sapply(mget(ls(), .GlobalEnv), is.data.frame)]
dfs_parp= ls()[sapply(mget(ls(), .GlobalEnv), is.data.frame)]
dfs_parp= dfs_parp[7:8]

#sample#2 (g1t20vsg1t23_chip2) in the dfs has too low number of peaks to run
    through, so adjusted accordingly
for (i in (3:length(dfs))) {
  data= get(dfs[i])
  graph_title= unlist(strsplit(dfs[i],".bed"))[1]
  dir.create(paste0("./",graph_title))
  print("now reading file:")
  print(graph_title)
  setwd(paste0("./",graph_title))
  colnames(data)= c('chr','start', 'end', 'id', 'signal', 'V6',
      'FC','pval','qval')
  data_2 <- with(data, GRanges(chr, IRanges(start+1, end), id=id, V5= signal,
      pval=pval, FC=FC, qval=qval))

  data_2$V5 = as.numeric(data_2$V5)

  #######not printing
  pdf(paste0(graph_title, "covplot%02d.pdf"))
  print("plotting coverage plot")
  print(covplot(data_2, weightCol="V5", title= graph_title))
  dev.off()
  #######not printing

  #######prints
  pdf(paste0(graph_title,"%02d.pdf"))
  #plotting tagHeatmap

  promoter <- getPromoters(TxDb=txdb, upstream=3000, downstream=3000)
  tagMatrix <- getTagMatrix(data_2, windows=promoter)


  #plotting peakHeatmap
  print("plotting peakHeatmap")
  peakHeatmap(data_2, TxDb=txdb, upstream=3000, downstream=3000, color="red",
      title= graph_title)


  #plotting peakAnnos
  print("plotting peakAnnos")
  peakAnno <- annotatePeak(data_2, tssRegion=c(-3000, 3000),
                           TxDb=txdb, annoDb="org.Mm.eg.db")
  assign(paste0(graph_title,"annotated_peaks", sep="_"), peakAnno)

  peakAnno_filter_down= as.data.frame(peakAnno@anno) %>% filter(.,
      distanceToTSS >= 0 & distanceToTSS <= 450)
  peakAnno_filter_up= as.data.frame(peakAnno@anno) %>% filter(., distanceToTSS
```

```
    >= -450 & distanceToTSS <= 0)

write.table(peakAnno_filter_down, file= paste(graph_title,
    "_distToTSS_0_downstream450bps"), sep="\t", row.names=TRUE, qmethod=
    "double", col.names=NA)
write.table(peakAnno_filter_up, file= paste(graph_title,
    "_distToTSS_0_upstream450bps"), sep="\t", row.names=TRUE, qmethod=
    "double", col.names=NA)
write.table(peakAnno@anno, file=(paste0(graph_title,"annotated_peaks")),
    sep="\t", row.names=TRUE, qmethod= "double", col.names=NA)


plotAnnoPie(peakAnno)

plotAnnoBar(peakAnno)

vennpie(peakAnno)


#plotting upset plots
print("plotting upset plots")
upsetplot(peakAnno)

dev.off()

pdf(paste0(graph_title,"AvgProf%02d.pdf"))

#plotting AvgProf

print("plotting plotAvgProf")
print(plotAvgProf(tagMatrix, xlim=c(-3000, 3000),
                  xlab="Genomic Region (5'->3')", ylab = "Read Count
                      Frequency"))

print("plotting plotAvgProf2")
print(plotAvgProf2(data_2, TxDb=txdb, upstream=3000, downstream=3000,
                   xlab="Genomic Region (5'->3')", ylab = "Read Count
                       Frequency"))

print(plotAvgProf(tagMatrix, xlim=c(-3000, 3000), conf = 0.95, resample =
    1000))

dev.off()

pdf(paste0(graph_title,"DistoTSS%02d.pdf"))
#plotting Dis to TSS
print("plotting distance to TSS")
print(plotDistToTSS(peakAnno,
                    title="Distribution of transcription factor-binding loci
                        \nrelative to TSS"))

dev.off()

setwd("/Users/draperlabair1/Documents/McMaster/Data folder_Backup/
    Bioinformatics/PePr_diff_peak_calling_R/PePr_results_bam_sharp/
    PePr_processed")
```

```
}

sessionInfo()
```

```
#170705_ATACseq_macs2_venndiagrams.R

library(plyr)
library(dplyr)
library(UpSetR)
library(tidyr)

pdf("ATACseq_macs2_bedintersects.pdf", width=3, height=4,pointsize=5)

#G2Mvsinterphase

G2MvsInt= venneuler(c(G2M= 9477, Int= 61211, "G2M&Int"= 45486))
plot(G2MvsInt)

#G1t20vsinterphase
G1t20vsInt= venneuler(c(G1t20= 3675, Int= 70485, "G1t20&Int"= 36844))
plot(G1t20vsInt)

#G1t35vsinterphase
G1t35vsInt= venneuler(c(G1t35= 7064, Int= 59283, "G1t35&Int"= 47726))
plot(G1t35vsInt)

#G2MvsG1t20
G2MvsG1t20= venneuler(c(G1t20= 4943,G2M= 19890, "G2M&G1t20"= 35073))
plot(G2MvsG1t20)

G2MvsG1t35= venneuler(c(G1t35= 10621, G2M= 10991, "G2M&G1t35"= 43972))
plot(G2MvsG1t35)
pdf("G2MvsG1t35&G2MvsG1t20.pdf")
common= venneuler(c(G2MvsG1t35= 11846, G2MvsG1t20= 2946,
    "G2MvsG1t35&G2MvsG1t20"= 32127))
plot(common)
dev.off()

#common G2MvsInterphase
G2M_commonvsInt= venneuler(c(Int= 74365, G2M_common= 1599, "G2M_common&Int"=
    32127))
plot(G2M_commonvsInt)

#common G2MvsInt== bookmarked sites

bookmarked= venneuler(c(bookmarked= 21264, Mit_H3K27Ac= 15521,
    "bookmarked&Mit_H3K27Ac"= 10863))
plot(bookmarked)

dev.off()

pdf("parp1_200_bookmarked.pdf")
parp1_bookmarked= venneuler(c(parp1_200= 217, bookmarked= 30828,
    "bookmarked&parp1_200"= 1299))
plot(parp1_bookmarked)
dev.off()

pdf("non-bookmarked_Mit_H3K27Ac.pdf")
non_bookmarked= venneuler(c(non_bookmarked= 69249, Mit_H3K27Ac= 21268,
    "non_bookmarked&Mit_H3K27Ac"= 5116))
```

```
plot(non_bookmarked)
dev.off()



bookmarked= venneuler(c(bookmarked= 32127, Mit_H3K27Ac= 15521,
    "bookmarked&Mit_H3K27Ac"= 10863))


##upset
comparisons= list.files("/Users/draperlabair1/Documents/McMaster/Data
    folder_Backup/Bioinformatics/ATACseq/170124_macs2_peakcall/",
                        pattern= "*_common_u.bed", full.names = TRUE)

comparisons= comparisons[5:9]

for (i in (1:length(comparisons))) {
  sample= read.table(comparisons[i], header=FALSE, sep="\t")
  sample_name= paste0((unlist(strsplit(comparisons[i], "/")))[11]))
  sample_name= paste0((unlist(strsplit(sample_name, ".bed")))[1]))
  sample= select(sample, V4,V5)
  colnames(sample)= c("V4", paste0(sample_name))
  assign(paste(sample_name), sample)
}

dfs= list(G2M_macs2_common_u, G2M_interphase, G2M_G1t20, G2M_G1t35,
    G2M_G1t20_G1t35)
upset_combined_data_2 = join_all(dfs, by= "V4", type ="left", match= "first")
#  # this command will join all datasets based on the value in column V4, and
    for all rows present in left-most data-frame in the list
row_names = upset_combined_data_2$V4


## convert to binary data
upset_combined_data_2[upset_combined_data_2>0]<-1
upset_combined_data_2[is.na(upset_combined_data_2)] <- 0
## convert to numeric
upset_combined_data_2 <- sapply(upset_combined_data_2, as.numeric )
## convert back to data.frame from a matrix
upset_combined_data_2= data.frame(upset_combined_data_2)
upset_combined_data_2$V4= row_names
write.table(upset_combined_data_2, "G2M_comparisons_upsetdata", sep="\t")

G2M_total_peaks= read.table("/Users/draperlabair1/Documents/McMaster/Data\
    folder_Backup/Bioinformatics/ATACseq/170705_macs2_bedintersect/
    170705_parp1_liftedover_ATACmacs2/170705_parp1_GSE81168_chipseq_ATAC/
    G2M_common_peaks_annotated_peaks", sep="\t", header=TRUE, row.names=1)
G2M_total_peaks= rename(G2M_total_peaks, V4=id)

concatenated_data= join(G2M_total_peaks, upset_combined_data_2, by ="V4",
    type="right")
concatenated_data_with_RNAexpr= join(RNAexpr, concatenated_data, by ="SYMBOL",
    type="right")

write.table(concatenated_data, file="G2M_comparisons_upset_concatenated",
    sep="\t", quote= FALSE, row.names = FALSE)
```

```
write.table(concatenated_data_with_RNAexpr,
    file="G2M_comparisons_upset_concatenated_withRNA", sep="\t", quote= FALSE,
    row.names = FALSE)


pdf("G2M_comparisons_upset.pdf", height=3, width=3)
upset(upset_combined_data_2, order.by= "freq")
dev.off()


#bookmarked sites
bedintersects= list.files("/Users/draperlabair1/Documents/McMaster/Data
    folder_Backup/Bioinformatics/ATACseq/170124_macs2_peakcall/
    bedintersect_analysis/",
                          pattern= "bookmarked_*", full.names = TRUE)
bedintersects= bedintersects[-c(5,9)]

for (i in (1:length(bedintersects))) {
  sample= read.table(bedintersects[i], header=FALSE, sep="\t")
  sample_name= paste0((unlist(strsplit(bedintersects[i], "/"))[12]))
  sample_name= paste0((unlist(strsplit(sample_name, ".bed"))[1]))
  sample= select(sample, V4,V5)
  colnames(sample)= c("V4", paste0(sample_name))
  assign(paste(sample_name), sample)
}

dfs= list(G2M_G1t20_G1t35, bookmarked_mit_H3K27Ac, bookmarked_klf4,
    bookmarked_Oct4, bookmarked_sox2, bookmarked_superenchancers,
          bookmarked_parp1, bookmarked_parp1_200)
upset_combined_data = join_all(dfs, by= "V4", type ="left", match= "first")
#  # this command will join all datasets based on the value in column V4, and
    for all rows present in left-most data-frame in the list
row_names = upset_combined_data$V4

## convert to binary data
upset_combined_data[upset_combined_data>0]<-1
upset_combined_data[is.na(upset_combined_data)] <- 0
## convert to numeric
upset_combined_data <- sapply(upset_combined_data, as.numeric )
## convert back to data.frame from a matrix
upset_combined_data= data.frame(upset_combined_data)
upset_combined_data$V4= row_names

write.table(upset_combined_data, "G2M_bookmarked_promoter_elements", sep="\t")

G2M_annotated_peaks= read.table("/Users/draperlabair1/Documents/McMaster/Data\
    folder_Backup/Bioinformatics/ATACseq/170705_macs2_bedintersect/
    170705_parp1_liftedover_ATACmacs2/170705_parp1_GSE81168_chipseq_ATAC/
    bookmarked_sites_annotated_peaks", sep="\t", header=TRUE, row.names=1)
G2M_annotated_peaks= rename(G2M_annotated_peaks, V4=id)
RNAexpr= read.table("/Users/draperlabair1/Documents/McMaster/Data\
    folder_Backup/Bioinformatics/DataMining/Transcriptomics/mouse_ENCODE/
    170706_mouse_ENCODE_ESCvsMEF/mouse_ENCODE_ESCvsMEF_RNAexpr", sep="\t",
    header=TRUE)
RNAexpr= rename(RNAexpr, SYMBOL= gene_symbol)
```

```
concatenated_data= join(G2M_annotated_peaks, upset_combined_data, by ="V4",
    type="right")
concatenated_data_with_RNAexpr= join(RNAexpr, concatenated_data, by ="SYMBOL",
    type="right")
write.table(concatenated_data,
    file="G2M_bookmarked_promoter_upset_concatenated", sep="\t", quote= FALSE,
    row.names = FALSE)
write.table(concatenated_data_with_RNAexpr,
    file="G2M_bookmarked_promoter_upset_concatenated_withRNA", sep="\t", quote=
    FALSE, row.names = FALSE)

pdf("G2M_bookmarked_promoter_elements2.pdf", height=5, width = 8)
upset(upset_combined_data, order.by = "freq")
upset(upset_combined_data, order.by = "freq", nsets=)
upset(upset_combined_data, order.by= "freq", sets= c("bookmarked_klf4",
    "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", sets= c("G2M_G1t20_G1t35",
    "bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", sets= c("G2M_G1t20_G1t35",
    "bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", sets= c("G2M_G1t20_G1t35",
    "bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs", "bookmarked_parp1"))
upset(upset_combined_data, order.by= "freq", nintersects=20, sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", nintersects=20, sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",

                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs", "bookmarked_parp1"))
upset(upset_combined_data, order.by= "freq",sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",

                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
```

```
concatenated_data= join(G2M_annotated_peaks, upset_combined_data, by ="V4",
    type="right")
concatenated_data_with_RNAexpr= join(RNAexpr, concatenated_data, by ="SYMBOL",
    type="right")
write.table(concatenated_data,
    file="G2M_bookmarked_promoter_upset_concatenated", sep="\t", quote= FALSE,
    row.names = FALSE)
write.table(concatenated_data_with_RNAexpr,
    file="G2M_bookmarked_promoter_upset_concatenated_withRNA", sep="\t", quote=
    FALSE, row.names = FALSE)

pdf("G2M_bookmarked_promoter_elements2.pdf", height=5, width = 8)
upset(upset_combined_data, order.by = "freq")
upset(upset_combined_data, order.by = "freq", nsets=)
upset(upset_combined_data, order.by= "freq", sets= c("bookmarked_klf4",
    "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", sets= c("G2M_G1t20_G1t35",
    "bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", sets= c("G2M_G1t20_G1t35",
    "bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", sets= c("G2M_G1t20_G1t35",
    "bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs", "bookmarked_parp1"))
upset(upset_combined_data, order.by= "freq", nintersects=20, sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",
                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
upset(upset_combined_data, order.by= "freq", nintersects=20, sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",

                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs", "bookmarked_parp1"))
upset(upset_combined_data, order.by= "freq",sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",

                                                     "bookmarked_sox2",
                                                      "bookmarked_superenchance
                                                      rs",
                                                      "bookmarked_parp1_200"))
```

```
upset(upset_combined_data, order.by= "freq",sets=
    c("bookmarked_mit_H3K27Ac","bookmarked_klf4", "bookmarked_Oct4",

                                                    "bookmarked_sox2",
                                                    "bookmarked_superenchance
                                                    rs", "bookmarked_parp1"))


dev.off()
```

```
library(plyr)
library(dplyr)
if (!require("gplots")) {
  install.packages("gplots", dependencies = TRUE)
  library(gplots)
}
if (!require("RColorBrewer")) {
  install.packages("RColorBrewer", dependencies = TRUE)
  library(RColorBrewer)
}

pluripotent_bookmarked= bookmarked_klf4_oct4_sox4_k27Ac
pluripotent_bookmarked= bookmarked_klf4_parp1_k27Ac
pluripotent_bookmarked <- na.omit(pluripotent_bookmarked)
pluripotent_bookmarked= pluripotent_bookmarked[!
    duplicated(pluripotent_bookmarked$SYMBOL),]



mat_data= data.matrix(pluripotent_bookmarked[,2:ncol(pluripotent_bookmarked)])
rownames(mat_data)= pluripotent_bookmarked$SYMBOL



write.table(as.data.frame(mat_data_log),
    file="MBF_overlap_with_other_ChipMS_data", sep="\t", quote=FALSE, col.names
    = NA)

my_palette <- colorRampPalette(c("white", "darkblue"))(n = 299)
col_breaks = c(seq(0,2,length=50),  # for red
               seq(2.01,18,length=150),           # for yellow
                           seq(18.01,111,length=100))


png("putative_MBFs_overlap_histone_ChIP-MS.png",    # create PNG for the heat
    map
    width = 6*300,        # 5 x 300 pixels
    height = 8*300,
    res = 300,            # 300 pixels per inch
    pointsize = 8)        # smaller font size

pdf("klf4_parp1_K27_bookmarkedsites.pdf", height=8, width=6)
png("klf4_parp1_K27_bookmarkedsites.png")
heatmap.2(mat_data,
          main = "Pluripotency related (Parp1) bookmarked sites", # heat map
              title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(15,10),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          dendrogram="column",     # only draw a column dendrogram
          hclustfun = hclust,
          breaks= col_breaks, cexRow=0.6, cexCol=0.6, keysize=1)
```

```
dev.off()               # close the PNG device

pdf("klf4_parp1_K27_bookmarkedsites2.pdf")
heatmap.2(mat_data ,col=my_palette, breaks=col_breaks, scale="none",key=TRUE,
          symkey=FALSE, symm=FALSE, density.info="none", trace="none",
          cexRow=0.6,   Rowv=TRUE, Colv=FALSE, dendrogram = c("row"),
          cexCol=0.6, sepcolor="lightgray",
          rowsep=FALSE, sepwidth=c(0.005,0.005),
          margins=c(9,9), main= "Pluripotency related (Parp1) bookmarked
              sites" )
dev.off()
```

```
library(plyr)
library(dplyr)
if (!require("gplots")) {
  install.packages("gplots", dependencies = TRUE)
  library(gplots)
}
if (!require("RColorBrewer")) {
  install.packages("RColorBrewer", dependencies = TRUE)
  library(RColorBrewer)
}

setwd("/Users/draperlabair1/Documents/McMaster/Data folder_Backup/
    Bioinformatics/DataMining/Proteomics")

library(readxl)
chip_MS_for_histone_markes_PMID25755260 <- read_excel("~/Documents/McMaster/
    Data folder_Backup/Bioinformatics/DataMining/Proteomics/
    chip_MS_for_histone_markes_PMID25755260.xls",
                                                      skip = 1)

histone_chip_MS= select(chip_MS_for_histone_markes_PMID25755260, 1, 7:18)

S10_H3 <- read_excel("~/Google Drive/MB_project manuscript/Data/Mitotic
    Bookmarking 290615 110515_H3_S10_IP-MS/290615 110615_PEAKS_combined/S10+H3/
    S10^H3.xlsx",
                     skip = 17)
MBF_IDs= select(S10_H3, 39, 2)

concat_data= left_join(MBF_IDs, histone_chip_MS, by= "Gene Symbol")

heatmap_data= select(concat_data, 1,3:8)
heatmap_data[is.na(heatmap_data)] <- 0

heatmap_matrix= filter(heatmap_data, !(H3K27ac==0 & H3K4me3==0 & H3K79me2==0 &
    H3K36me3==0 & H3K9me3==0 & H4K20me3==0))

mat_data= data.matrix(heatmap_matrix[,2:ncol(heatmap_matrix)])
rownames(mat_data)= heatmap_matrix$`Gene Symbol`

mat_data_log= log10(mat_data+1)

write.table(as.data.frame(mat_data_log),
    file="MBF_overlap_with_other_ChipMS_data", sep="\t", quote=FALSE, col.names
    = NA)

my_palette <- colorRampPalette(c("white", "darkblue"))(n = 299)

png("putative_MBFs_overlap_histone_ChIP-MS.png",    # create PNG for the heat
    map
    width = 6*300,        # 5 x 300 pixels
    height = 8*300,
    res = 300,            # 300 pixels per inch
    pointsize = 8)        # smaller font size

pdf("putative_MBFs_overlap_histone_ChIP-MS.pdf", height=8, width=6)
heatmap.2(mat_data_log,
```

```
            notecol="black",      # change font color of cell labels to black
            density.info="none",  # turns off density plot inside color legend
            trace="none",         # turns off trace lines inside the heat map
            margins =c(15,10),     # widens margins around plot
            col=my_palette,       # use on color palette defined earlier
            dendrogram="both",     # only draw a column dendrogram
            hclustfun = hclust,
            keysize=1.2, key.xlab="log10(score+1)")           # turn off column
                clustering

dev.off()               # close the PNG device

### S10 only heatmap

S10_only= data.frame("Gene Symbol" = character(8), "S10_only_proteins" =
    character(8),stringsAsFactors=FALSE)
S10_genes= c("Rld1",
             "Ddx17",
             "Glyr1",
             "Tadbp",
             "Rcc1",
             "Smc2",
             "Rs25",
             "Dkc1")

S10_only$Gene.Symbol= S10_genes
S10_proteins= c("RL1D1_MOUSE",
                "DDX17_MOUSE",
                "GLYR1_MOUSE",
                "TADBP_MOUSE",
                "RCC1_MOUSE",
                "SMC2_MOUSE",
                "RS25_MOUSE",
                "DKC1_MOUSE")

S10_only$S10_only_proteins= S10_proteins

colnames(S10_only)= c("Gene Symbol", "S10_only_proteins")
S10_only_concat= left_join(S10_only, histone_chip_MS, by= "Gene Symbol")
S10_only_heatmap= select(S10_only_concat, 1,3:8)
S10_only_heatmap[is.na(S10_only_heatmap)] <- 0

S10_only_heatmap= filter(S10_only_heatmap, !(H3K27ac==0 & H3K4me3==0 &
    H3K79me2==0 & H3K36me3==0 & H3K9me3==0 & H4K20me3==0))

S10_only_mat_data= data.matrix(S10_only_heatmap[,2:ncol(S10_only_heatmap)])
rownames(S10_only_mat_data)= S10_only_heatmap$`Gene Symbol`

S10_only_mat_data= log10(S10_only_mat_data+1)

pdf("S10_only_overlap_histone_ChIP-MS.pdf", height=8, width=6)
heatmap.2(S10_only_mat_data,
            notecol="black",      # change font color of cell labels to black
            density.info="none",  # turns off density plot inside color legend
            trace="none",         # turns off trace lines inside the heat map
            margins =c(15,10),     # widens margins around plot
```

```
              col=my_palette,       # use on color palette defined earlier
              dendrogram="both",     # only draw a column dendrogram
              hclustfun = hclust,
              keysize=1.5, key.xlab="log10(score+1)")           # turn off column
                  clustering

dev.off()
```

References

Adachi, K., Suemori, H., Yasuda, S.Y., Nakatsuji, N., and Kawase, E. (2010). Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes to cells : devoted to molecular & cellular mechanisms *15*, 455-470.

AJ, C.Q., Bugai, A., and Barboric, M. (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic acids research *44*, 7527-7539.

Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.P., Harris, C.P., Waknitz, M.A., Itskovitz-Eldor, J., and Thomson, J.A. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. Developmental biology *227*, 271-278.

Amon, A., Tyers, M., Futcher, B., and Nasmyth, K. (1993). Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. Cell *74*, 993-1007.

Arora, M., Zhang, J., Heine, G.F., Ozer, G., Liu, H.W., Huang, K., and Parvin, J.D. (2012). Promoters active in interphase are bookmarked during mitosis by ubiquitination. Nucleic acids research *40*, 10187-10202.

Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M.L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. The EMBO journal *22*, 1347-1358.

Babaie, Y., Herwig, R., Greber, B., Brink, T.C., Wruck, W., Groth, D., Lehrach, H., Burdon, T., and Adjaye, J. (2007). Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells *25*, 500-510.

Ballabeni, A., Park, I.H., Zhao, R., Wang, W., Lerou, P.H., Daley, G.Q., and Kirschner, M.W. (2011). Cell cycle adaptations of embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *108*, 19252-19257.

Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature *461*, 1292-1295.

Baumgartner, M., Dutrillaux, B., Lemieux, N., Lilienbaum, A., Paulin, D., and Viegas-Pequignot, E. (1991). Genes occupy a fixed and symmetrical position on sister chromatids. Cell *64*, 761-766.

Beddington, R.S., and Robertson, E.J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development *105*, 733-737.

Bendall, S.C., Stewart, M.H., Menendez, P., George, D., Vijayaragavan, K., Werbowetski-Ogilvie, T., Ramos-Mejia, V., Rouleau, A., Yang, J., Bosse, M., *et al.*

(2007). IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. Nature *448*, 1015-1021.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes & Development *23*, 781-783.

Bernardo, A.S., Faial, T., Gardner, L., Niakan, K.K., Ortmann, D., Senner, C.E., Callery, E.M., Trotter, M.W., Hemberger, M., Smith, J.C.*, et al.* (2011). BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. Cell Stem Cell *9*, 144-155.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics (Oxford, England) *25*, 1091-1093.

Bjorklund, S., and Kim, Y.J. (1996). Mediator of transcriptional regulation. Trends in biochemical sciences *21*, 335-337.

Blobel, G.A., Kadauke, S., Wang, E., Lau, A.W., Zuber, J., Chou, M.M., and Vakoc, C.R. (2009). A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. Molecular cell *36*, 970-983.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Bregman, D.B., Du, L., Li, Y., Ribisi, S., and Warren, S.L. (1994). Cytostellin distributes to nuclear regions enriched with splicing factors. Journal of cell science *107 (Pt 3)*, 387-396.

Bregman, D.B., Du, L., van der Zee, S., and Warren, S.L. (1995). Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. The Journal of cell biology *129*, 287-298.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A.*, et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods *10*, 1213-1218.

Cadinanos, J., and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. Nucleic acids research *35*, e87.

Calder, A., Roth-Albin, I., Bhatia, S., Pilquil, C., Lee, J.H., Bhatia, M., Levadoux-Martin, M., McNicol, J., Russell, J., Collins, T.*, et al.* (2012). Lengthened G1 Phase Indicates Differentiation Status in Human Embryonic Stem Cells. Stem cells and development.

Calder, A., Roth-Albin, I., Bhatia, S., Pilquil, C., Lee, J.H., Bhatia, M., Levadoux-Martin, M., McNicol, J., Russell, J., Collins, T.*, et al.* (2013). Lengthened G1 phase indicates differentiation status in human embryonic stem cells. Stem cells and development *22*, 279-295.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249-253.

Campbell, A.E., Hsiung, C.C., and Blobel, G.A. (2014). Comparative analysis of mitosis-specific antibodies for bulk purification of mitotic populations by fluorescence-activated cell sorting. BioTechniques *56*, 90-91, 93-94.

Cantz, T., Bleidissel, M., Stehling, M., and Scholer, H.R. (2008). In vitro differentiation of reprogrammed murine somatic cells into hepatic precursor cells. Biological chemistry *389*, 889-896.

Caravaca, J.M., Donahue, G., Becker, J.S., He, X., Vinson, C., and Zaret, K.S. (2013). Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes & Development *27*, 251-260.

Cerdan, C., Hong, S.H., and Bhatia, M. (2007). Formation and hematopoietic differentiation of human embryoid bodies by suspension and hanging drop cultures. Current protocols in stem cell biology *Chapter 1*, Unit 1D 2.

Chaerkady, R., Letzen, B., Renuse, S., Sahasrabuddhe, N.A., Kumar, P., All, A.H., Thakor, N.V., Delanghe, B., Gearhart, J.D., Pandey, A.*, et al.* (2011). Quantitative temporal proteomic analysis of human embryonic stem cell differentiation into oligodendrocyte progenitor cells. Proteomics *11*, 4007-4020.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.

Chan, K.K., Zhang, J., Chia, N.Y., Chan, Y.S., Sim, H.S., Tan, K.S., Oh, S.K., Ng, H.H., and Choo, A.B. (2009). KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells. Stem Cells *27*, 2114-2125.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544-547.

Cherry, A., and Daley, G.Q. (2010). Another horse in the meta-stable state of pluripotency. Cell Stem Cell *7*, 641-642.

Chubb, J.R., Boyle, S., Perry, P., and Bickmore, W.A. (2002). Chromatin motion is constrained by association with nuclear compartments in human cells. Current biology : CB *12*, 439-445.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A*., et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY) *339*, 819-823.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (New York, NY) *322*, 1845-1848.

Costa, M., Dottori, M., Sourris, K., Jamshidi, P., Hatzistavrou, T., Davis, R., Azzola, L., Jackson, S., Lim, S.M., Pera, M*., et al.* (2007). A method for genetic modification of human embryonic stem cells using electroporation. Nature protocols *2*, 792-796.

Coverley, D., Laman, H., and Laskey, R.A. (2002). Distinct roles for cyclins E and A during DNA replication complex assembly and activation. Nature cell biology *4*, 523-528.

Crosio, C., Fimia, G.M., Loury, R., Kimura, M., Okano, Y., Zhou, H., Sen, S., Allis, C.D., and Sassone-Corsi, P. (2002). Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. Molecular and cellular biology *22*, 874-885.

Daban, J.R. (2015). Stacked thin layers of metaphase chromatin explain the geometry of chromosome rearrangements and banding. Scientific reports *5*, 14891.

Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., and Walker, P.R. (2005). Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. The Journal of biological chemistry *280*, 6257-6260.

Deluz, C., Friman, E.T., Strebinger, D., Benke, A., Raccaud, M., Callegari, A., Leleu, M., Manley, S., and Suter, D.M. (2016). A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes & Development *30*, 2538-2550.

Deluz, C., Strebinger, D., Friman, E.T., and Suter, D.M. (2017). The elusive role of mitotic bookmarking in transcriptional regulation: insights from Sox2. Cell cycle (Georgetown, Tex), 0.

Deuve, J.L., and Avner, P. (2010). The Coupling of X-Chromosome Inactivation to Pluripotency. Annual review of cell and developmental biology.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Draper, J.S., Pigott, C., Thomson, J.A., and Andrews, P.W. (2002). Surface antigens of human embryonic stem cells: changes upon differentiation in culture. J Anat *200*, 249-258.

Dundr, M., and Olson, M.O. (1998). Partially processed pre-rRNA is preserved in association with processing components in nucleolus-derived foci during mitosis. Molecular biology of the cell *9*, 2407-2422.

DuPraw, E.J. (1966). Evidence for a 'folded-fibre' organization in human chromosomes. Nature *209*, 577-581.

Easwaran, H.P., Schermelleh, L., Leonhardt, H., and Cardoso, M.C. (2004). Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO reports *5*, 1181-1186.

Edgar, B.A., and Lehner, C.F. (1996). Developmental control of cell cycle regulators: a fly's perspective. Science (New York, NY) *274*, 1646-1652.

Egozi, D., Shapira, M., Paor, G., Ben-Izhak, O., Skorecki, K., and Hershko, D.D. (2007). Regulation of the cell cycle inhibitor p27 and its ubiquitin ligase Skp2 in differentiation of human embryonic stem cells. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology *21*, 2807-2817.

Enver, T., Soneji, S., Joshi, C., Brown, J., Iborra, F., Orntoft, T., Thykjaer, T., Maltby, E., Smith, K., Abu Dawud, R.*, et al.* (2005). Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. Human molecular genetics *14*, 3129-3140.

Festuccia, N., Dubois, A., Vandormael-Pournin, S., Gallego Tejeda, E., Mouren, A., Bessonnard, S., Mueller, F., Proux, C., Cohen-Tannoudji, M., and Navarro, P. (2016). Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. Nature cell biology.

Franco, A.A., and Kaufman, P.D. (2004). Histone deposition proteins: links between the DNA replication machinery and epigenetic gene silencing. Cold Spring Harbor symposia on quantitative biology *69*, 201-208.

Furstenthal, L., Kaiser, B.K., Swanson, C., and Jackson, P.K. (2001). Cyclin E uses Cdc6 as a chromatin-associated receptor required for DNA replication. The Journal of cell biology *152*, 1267-1278.

Garcia, E., Marcos-Gutierrez, C., del Mar Lorente, M., Moreno, J.C., and Vidal, M. (1999). RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. The EMBO journal *18*, 3404-3418.

Garcia-Tunon, I., Guallar, D., Alonso-Martin, S., Benito, A.A., Benitez-Lazaro, A., Perez-Palacios, R., Muniesa, P., Climent, M., Sanchez, M., Vidal, M.*, et al.* (2011). Association of Rex-1 to target genes supports its interaction with Polycomb function. Stem cell research *7*, 1-16.

Geley, S., Kramer, E., Gieffers, C., Gannon, J., Peters, J.M., and Hunt, T. (2001). Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. The Journal of cell biology *153*, 137-148.

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van, I.W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386-390.

Gottesfeld, J.M., and Forbes, D.J.

Graf, T., and Stadtfeld, M. (2008). Heterogeneity of embryonic and adult stem cells. Cell Stem Cell *3*, 480-483.

Grandy, R.A., Whitfield, T.W., Wu, H., Fitzgerald, M.P., VanOudenhove, J.J., Zaidi, S.K., Montecino, M.A., Lian, J.B., van Wijnen, A.J., Stein, J.L.*, et al.* (2016). Genome-Wide Studies Reveal that H3K4me3 Modification in Bivalent Genes Is Dynamically Regulated during the Pluripotent Cell Cycle and Stabilized upon Differentiation. Molecular and cellular biology *36*, 615-627.

Greber, B., Wu, G., Bernemann, C., Joo, J.Y., Han, D.W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P.*, et al.* (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. Cell Stem Cell *6*, 215-226.

Gruss, C., Wu, J., Koller, T., and Sogo, J.M. (1993). Disruption of the nucleosomes at the replication fork. The EMBO journal *12*, 4533-4545.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W.*, et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental cell *18*, 675-685.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc Natl Acad Sci U S A *107*, 9222-9227.

Hartl, P., Gottesfeld, J., and Forbes, D.J. (1993). Mitotic repression of transcription in vitro. The Journal of cell biology *120*, 613-624.

Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell *3*, 391-401.

Hemberger, M., Nozaki, T., Winterhager, E., Yamamoto, H., Nakagama, H., Kamada, N., Suzuki, H., Ohta, T., Ohki, M., Masutani, M.*, et al.* (2003). Parp1-deficiency induces differentiation of ES cells into trophoblast derivatives. Developmental biology *257*, 371-381.

Henderson, J.K., Draper, J.S., Baillie, H.S., Fishel, S., Thomson, J.A., Moore, H., and Andrews, P.W. (2002). Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens. Stem Cells *20*, 329-337.

Hendzel, M.J., Wei, Y., Mancini, M.A., Van Hooser, A., Ranalli, T., Brinkley, B.R., Bazett-Jones, D.P., and Allis, C.D. (1997). Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads

in an ordered fashion coincident with mitotic chromosome condensation. Chromosoma *106*, 348-360.

Ho, C.K., and Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Molecular cell *3*, 405-411.

Hsiung, C.C., Bartman, C.R., Huang, P., Ginart, P., Stonestrom, A.J., Keller, C.A., Face, C., Jahn, K.S., Evans, P., Sankaranarayanan, L.*, et al.* (2016). A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. Genes & Development *30*, 1423-1439.

Hsiung, C.C., and Blobel, G.A. (2016). A new bookmark of the mitotic genome in embryonic stem cells. Nature cell biology *18*, 1124-1125.

Hsiung, C.C., Morrissey, C.S., Udugama, M., Frank, C.L., Keller, C.A., Baek, S., Giardine, B., Crawford, G.E., Sung, M.H., Hardison, R.C.*, et al.* (2015). Genome accessibility is widely preserved and locally modulated during mitosis. Genome research *25*, 213-225.

Huang, S., Guo, Y.P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental biology *305*, 695-713.

Humphrey, R.K., Beattie, G.M., Lopez, A.D., Bucay, N., King, C.C., Firpo, M.T., Rose-John, S., and Hayek, A. (2004). Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells *22*, 522-530.

Hyslop, L., Stojkovic, M., Armstrong, L., Walter, T., Stojkovic, P., Przyborski, S., Herbert, M., Murdoch, A., Strachan, T., and Lako, M. (2005). Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. Stem Cells *23*, 1035-1043.

Jackson, V. (1987). Deposition of newly synthesized histones: new histones H2A and H2B do not deposit in the same nucleosome with new histones H3 and H4. Biochemistry *26*, 2315-2325.

Ji, X., Dadon, D.B., Abraham, B.J., Lee, T.I., Jaenisch, R., Bradner, J.E., and Young, R.A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. Proceedings of the National Academy of Sciences of the United States of America *112*, 3841-3846.

Jiang, B.H., Tseng, W.L., Li, H.Y., Wang, M.L., Chang, Y.L., Sung, Y.J., and Chiou, S.H. (2015). Poly(ADP-Ribose) Polymerase 1: Cellular Pluripotency, Reprogramming, and Tumorogenesis. International journal of molecular sciences *16*, 15531-15545.

John, S., and Workman, J.L. (1998). Bookmarking genes for activation in condensed mitotic chromosomes. BioEssays : news and reviews in molecular, cellular and developmental biology *20*, 275-279.

Johnson, T.C., and Holland, J.J. (1965). Ribonucleic acid and protein synthesis in mitotic HeLa cells. The Journal of cell biology *27*, 565-574.

Joti, Y., Hikima, T., Nishino, Y., Kamada, F., Hihara, S., Takata, H., Ishikawa, T., and Maeshima, K. (2012). Chromosomes without a 30-nm chromatin fiber. Nucleus (Austin, Tex) *3*, 404-410.

Juan, G., Pan, W., and Darzynkiewicz, Z. (1996). DNA segments sensitive to single-strand-specific nucleases are present in chromatin of mitotic cells. Experimental cell research *227*, 197-202.

Juan, G., Traganos, F., James, W.M., Ray, J.M., Roberge, M., Sauve, D.M., Anderson, H., and Darzynkiewicz, Z. (1998). Histone H3 phosphorylation and expression of cyclins A and B1 measured in individual cells during their progression through G2 and mitosis. Cytometry *32*, 71-77.

Kadauke, S., and Blobel, G.A. (2012). "Remembering" tissue-specific transcription patterns through mitosis. Cell cycle (Georgetown, Tex) *11*, 3911-3912.

Kadauke, S., Udugama, M.I., Pawlicki, J.M., Achtman, J.C., Jain, D.P., Cheng, Y., Hardison, R.C., and Blobel, G.A. (2012). Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. Cell *150*, 725-737.

Kelly, A.E., Ghenoiu, C., Xue, J.Z., Zierhut, C., Kimura, H., and Funabiki, H. (2010). Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. Science (New York, NY) *330*, 235-239.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature methods *12*, 357-360.

Kim, J.D., Faulk, C., and Kim, J. (2007). Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. Nucleic acids research *35*, 3442-3452.

Kim, J.D., Kim, H., Ekram, M.B., Yu, S., Faulk, C., and Kim, J. (2011). Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. Human molecular genetics *20*, 1353-1362.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. Cell *153*, 178-192.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development *14*, 2452-2460.

Kornberg, R.D. (1996). RNA polymerase II transcription control. Trends in biochemical sciences *21*, 325-326.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science (New York, NY) *184*, 865-868.

Krishnakumar, R., and Kraus, W.L. (2010). PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. Molecular cell *39*, 736-749.

Kumaki, Y., Oda, M., and Okano, M. (2008). QUMA: quantification tool for methylation analysis. Nucleic acids research *36*, W170-175.

Lake, R.J., Tsai, P.F., Choi, I., Won, K.J., and Fan, H.Y. (2014). RBPJ, the major transcriptional effector of Notch signaling, remains associated with chromatin throughout mitosis, suggesting a role in mitotic bookmarking. PLoS genetics *10*, e1004204.

Lanner, F., and Rossant, J. (2010). The role of FGF/Erk signaling in pluripotent cells. Development *137*, 3351-3360.

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J.*, et al.* (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell *141*, 872-883.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-873.

Leresche, A., Wolf, V.J., and Gottesfeld, J.M. (1996). Repression of RNA polymerase II and III transcription during M phase of the cell cycle. Experimental cell research *229*, 282-288.

Lerner, J., Bagattin, A., Verdeguer, F., Makinistoglu, M.P., Garbay, S., Felix, T., Heidet, L., and Pontoglio, M. (2016). Human mutations affect the epigenetic/bookmarking function of HNF1B. Nucleic acids research.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE transactions on visualization and computer graphics *20*, 1983-1992.

Liang, K., Woodfin, A.R., Slaughter, B.D., Unruh, J.R., Box, A.C., Rickels, R.A., Gao, X., Haug, J.S., Jaspersen, S.L., and Shilatifard, A. (2015). Mitotic Transcriptional Activation: Clearance of Actively Engaged Pol II via Transcriptional Elongation Control in Mitosis. Molecular cell *60*, 435-445.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY) *326*, 289-293.

Liu, Y., Chen, S., Wang, S., Soares, F., Fischer, M., Meng, F., Du, Z., Lin, C., Meyer, C., DeCaprio, J.A.*, et al.* (2017a). Transcriptional landscape of the human cell cycle. Proceedings of the National Academy of Sciences of the United States of America *114*, 3473-3478.

Liu, Y., Pelham-Webb, B., Di Giammartino, D.C., Li, J., Kim, D., Kita, K., Saiz, N., Garg, V., Doane, A., Giannakakou, P.*, et al.* (2017b). Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. Cell reports *19*, 1283-1293.

Liu, Z., and Kraus, W.L. (2017). Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. Molecular cell *65*, 589-603 e589.

Lodhi, N., Ji, Y., and Tulin, A. (2016). Mitotic bookmarking: maintaining post-mitotic reprogramming of transcription reactivation. Current molecular biology reports *2*, 10-16.

Lodhi, N., Kossenkov, A.V., and Tulin, A.V. (2014). Bookmarking promoters in mitotic chromatin: poly(ADP-ribose)polymerase-1 as an epigenetic mark. Nucleic acids research *42*, 7028-7038.

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

Luscher, B., and Eisenman, R.N. (1992). Mitosis-specific phosphorylation of the nuclear oncoproteins Myc and Myb. The Journal of cell biology *118*, 775-784.

Mac Auley, A., Werb, Z., and Mirkes, P.E. (1993). Characterization of the unusually rapid cell cycles during rat gastrulation. Development (Cambridge, England) *117*, 873-883.

Maeshima, K., Hihara, S., and Takata, H. (2010). New insight into the mitotic chromosome structure: irregular folding of nucleosome fibers without 30-nm chromatin structure. Cold Spring Harbor symposia on quantitative biology *75*, 439-444.

Maiorano, D., Moreau, J., and Mechali, M. (2000). XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. Nature *404*, 622-625.

Markaki, Y., Gunkel, M., Schermelleh, L., Beichmanis, S., Neumann, J., Heidemann, M., Leonhardt, H., Eick, D., Cremer, C., and Cremer, T. (2010). Functional nuclear organization of transcription and DNA replication: a topographical marriage between chromatin domains and the interchromatin compartment. Cold Spring Harbor symposia on quantitative biology *75*, 475-492.

Martinez-Balbas, M.A., Dey, A., Rabindran, S.K., Ozato, K., and Wu, C. (1995). Displacement of sequence-specific transcription factors from mitotic chromatin. Cell *83*, 29-38.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A*., et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-635.

Masui, S., Ohtsuka, S., Yagi, R., Takahashi, K., Ko, M.S., and Niwa, H. (2008). Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. BMC developmental biology *8*, 45.

Mauritz, C., Schwanke, K., Reppel, M., Neef, S., Katsirntaki, K., Maier, L.S., Nguemo, F., Menke, S., Haustein, M., Hescheler, J.*, et al.* (2008). Generation of functional murine cardiac myocytes from induced pluripotent stem cells. Circulation *118*, 507-517.

Michelotti, E.F., Sanford, S., and Levens, D. (1997). Marking of active genes on mitotic chromosomes. Nature *388*, 895-899.

Migeon, B.R. (2003). Is Tsix repression of Xist specific to mouse? Nature genetics *33*, 337; author reply 337-338.

Migeon, B.R. (2011). The single active X in human cells: evolutionary tinkering personified. Human genetics *130*, 281-293.

Mitchell, L., Huard, S., Cotrut, M., Pourhanifeh-Lemeri, R., Steunou, A.L., Hamza, A., Lambert, J.P., Zhou, H., Ning, Z., Basu, A.*, et al.* (2013). mChIP-KAT-MS, a method to map protein interactions and acetylation sites for lysine acetyltransferases. Proceedings of the National Academy of Sciences of the United States of America *110*, E1641-1650.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotechnol *20*, 87-90.

Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D., and Tesar, P.J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. Cell Stem Cell *8*, 318-325.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science (New York, NY) *342*, 948-953.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457-460.

Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development *128*, 2333-2339.

Nigg, E.A. (2001). Mitotic kinases as regulators of cell division and its checkpoints. Nature reviews Molecular cell biology *2*, 21-32.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics *24*, 372-376.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Nozaki, T., Masutani, M., Watanabe, M., Ochiya, T., Hasegawa, F., Nakagama, H., Suzuki, H., and Sugimura, T. (1999). Syncytiotrophoblastic giant cells in teratocarcinoma-like tumors derived from Parp-disrupted mouse embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *96*, 13345-13350.

Ogino, H., Nozaki, T., Gunji, A., Maeda, M., Suzuki, H., Ohta, T., Murakami, Y., Nakagama, H., Sugimura, T., and Masutani, M. (2007). Loss of Parp-1 affects gene expression profile in a genome-wide manner in ES cells and liver cells. BMC genomics *8*, 41.

Ohta, S., Bukowski-Wills, J.C., Sanchez-Pulido, L., Alves Fde, L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., *et al.* (2010). The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. Cell *142*, 810-821.

Ohtsubo, M., Theodoras, A.M., Schumacher, J., Roberts, J.M., and Pagano, M. (1995). Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. Molecular and cellular biology *15*, 2612-2624.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science (New York, NY) *183*, 330-332.

Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. Nature reviews Molecular cell biology *4*, 809-814.

Palmqvist, L., Glover, C.H., Hsu, L., Lu, M., Bossen, B., Piret, J.M., Humphries, R.K., and Helgason, C.D. (2005). Correlation of murine embryonic stem cell gene expression profiles with functional measures of pluripotency. Stem cells (Dayton, Ohio) *23*, 663-680.

Park, C.H., and Kim, K.T. (2012). Apoptotic phosphorylation of histone H3 on Ser-10 by protein kinase Cdelta. PloS one *7*, e44307.

Parsons, G.G., and Spencer, C.A. (1997). Mitotic repression of RNA polymerase II transcription is accompanied by release of transcription elongation complexes. Molecular and cellular biology *17*, 5791-5802.

Pease, S., Braghetta, P., Gearing, D., Grail, D., and Williams, R.L. (1990). Isolation of embryonic stem (ES) cells in media supplemented with recombinant leukemia inhibitory factor (LIF). Developmental biology *141*, 344-352.

Pelton, T.A., Sharma, S., Schulz, T.C., Rathjen, J., and Rathjen, P.D. (2002). Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. J Cell Sci *115*, 329-339.

Perez-Cadahia, B., Drobic, B., and Davie, J.R. (2009). H3 phosphorylation: dual role in mitosis and interphase. Biochemistry and cell biology = Biochimie et biologie cellulaire *87*, 695-709.

Piskadlo, E., and Oliveira, R.A. (2016). Novel insights into mitotic chromosome condensation. F1000Research *5*.

Prasanth, K.V., Sacco-Bubulya, P.A., Prasanth, S.G., and Spector, D.L. (2003). Sequential entry of components of the gene expression machinery into daughter nuclei. Molecular biology of the cell *14*, 1043-1057.

Prasanth, S.G., Mendez, J., Prasanth, K.V., and Stillman, B. (2004). Dynamics of pre-replication complex proteins during the cell division cycle. Philosophical transactions of the Royal Society of London Series B, Biological sciences *359*, 7-16.

Quelle, D.E., Ashmun, R.A., Shurtleff, S.A., Kato, J.Y., Bar-Sagi, D., Roussel, M.F., and Sherr, C.J. (1993). Overexpression of mouse D-type cyclins accelerates G1 phase in rodent fibroblasts. Genes & Development *7*, 1559-1571.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) *26*, 841-842.

Ramachandran, S., and Henikoff, S. (2015). Replicating Nucleosomes. Science advances *1*.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

Reines, D., Conaway, J.W., and Conaway, R.C. (1996). The RNA polymerase II general elongation factors. Trends in biochemical sciences *21*, 351-355.

Renfree, M.B., and Shaw, G. (2000). Diapause. Annual review of physiology *62*, 353-375.

Resnitzky, D., Gossen, M., Bujard, H., and Reed, S.I. (1994). Acceleration of the G1/S phase transition by expression of cyclins D1 and E with an inducible system. Molecular and cellular biology *14*, 1669-1679.

Rezende, N.C., Lee, M.Y., Monette, S., Mark, W., Lu, A., and Gudas, L.J. (2011). Rex1 (Zfp42) null mice show impaired testicular function, abnormal testis morphology, and aberrant gene expression. Developmental biology *356*, 370-382.

Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.

Rizkallah, R., and Hurt, M.M. (2009). Regulation of the transcription factor YY1 in mitosis through phosphorylation of its DNA-binding domain. Molecular biology of the cell *20*, 4766-4776.

Roberts, S.B., Segil, N., and Heintz, N. (1991). Differential phosphorylation of the transcription factor Oct1 during the cell cycle. Science (New York, NY) *253*, 1022-1026.

Roeder, R. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends in biochemical sciences *21*.

Roper, S.J., Chrysanthou, S., Senner, C.E., Sienerth, A., Gnan, S., Murray, A., Masutani, M., Latos, P., and Hemberger, M. (2014). ADP-ribosyltransferases Parp1 and Parp7 safeguard pluripotency of ES cells. Nucleic acids research *42*, 8914-8927.

Rossant, J. (2008). Stem cells and early lineage development. Cell *132*, 527-531.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M.*, et al.* (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nature structural & molecular biology *19*, 1155-1160.

Sarge, K.D., and Park-Sarge, O.K. (2009). Mitotic bookmarking of formerly active genes: keeping epigenetic memories from fading. Cell cycle (Georgetown, Tex) *8*, 818-823.

Savatier, P., Huang, S., Szekely, L., Wiman, K.G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene *9*, 809-818.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nature methods *9*, 676-682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods *9*, 671-675.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Scotland, K.B., Chen, S., Sylvester, R., and Gudas, L.J. (2009). Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. Developmental dynamics : an official publication of the American Association of Anatomists *238*, 1863-1877.

Sedat, J., and Manuelidis, L. (1978). A direct approach to the structure of eukaryotic chromosomes. Cold Spring Harbor symposia on quantitative biology *42 Pt 1*, 331-350.

Segil, N., Roberts, S.B., and Heintz, N. (1991). Mitotic phosphorylation of the Oct-1 homeodomain and regulation of Oct-1 DNA binding activity. Science (New York, NY) *254*, 1814-1816.

Shall, S., and de Murcia, G. (2000). Poly(ADP-ribose) polymerase-1: what have we learned from the deficient mouse model? Mutation research *460*, 1-15.

Snow, M. (1977). Gastrulation in the mouse: Growth and regionalization of the epiblast. J Embryol Exp Morph

Soldi, M., and Bonaldi, T. (2014). The ChroP approach combines ChIP and mass spectrometry to dissect locus-specific proteomic landscapes of chromatin. Journal of visualized experiments : JoVE.

Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D., and Dalton, S. (2002). Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. Oncogene *21*, 8320-8333.

Stewart, M.H., Bendall, S.C., Levadoux-Martin, M., and Bhatia, M. (2010). Clonal tracking of hESCs reveals differential contribution to functional assays. Nature methods *7*, 917-922.

Stewart, M.H., Bosse, M., Chadwick, K., Menendez, P., Bendall, S.C., and Bhatia, M. (2006). Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. Nature methods *3*, 807-815.

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

Taylor, J.H. (1960). Nucleic acid synthesis in relation to the cell division cycle. Annals of the New York Academy of Sciences *90*, 409-421.

Team, R.C. (2013). R: A language and environment for statistical

computing. In R Foundation for Statistical Computing, Vienna, Austria.

Terranova, C., Narla, S.T., Lee, Y.W., Bard, J., Parikh, A., Stachowiak, E.K., Tzanakakis, E.S., Buck, M.J., Birkaya, B., and Stachowiak, M.K. (2015). Global Developmental Gene Programing Involves a Nuclear Form of Fibroblast Growth Factor Receptor-1 (FGFR1). PloS one *10*, e0123380.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Teves, S.S., An, L., Hansen, A.S., Xie, L., Darzacq, X., and Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. eLife *5*.

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

Tomoda, K., Takahashi, K., Leung, K., Okada, A., Narita, M., Yamada, N.A., Eilertson, K.E., Tsang, P., Baba, S., White, M.P.*, et al.* (2012). Derivation conditions impact x-inactivation status in female human induced pluripotent stem cells. Cell Stem Cell *11*, 91-99.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909-918.

Tucker, K.L., Wang, Y., Dausman, J., and Jaenisch, R. (1997). A transgenic mouse strain expressing four drug-selectable marker genes. Nucleic acids research *25*, 3745-3746.

Uchiyama, S., Kobayashi, S., Takata, H., Ishihara, T., Hori, N., Higashi, T., Hayashihara, K., Sone, T., Higo, D., Nirasawa, T.*, et al.* (2005). Proteome analysis of human metaphase chromosomes. The Journal of biological chemistry *280*, 16994-17004.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci *118*, 4495-4509.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P.*, et al.* (2009a). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339-1349.

Vallier, L., Touboul, T., Brown, S., Cho, C., Bilican, B., Alexander, M., Cedervall, J., Chandran, S., Ahrlund-Richter, L., Weber, A.*, et al.* (2009b). Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. Stem Cells *27*, 2655-2666.

Valls, E., Sanchez-Molina, S., and Martinez-Balbas, M.A. (2005). Role of histone modifications in marking and activating genes through mitosis. The Journal of biological chemistry *280*, 42592-42600.

van den Boom, V., Kooistra, S.M., Boesjes, M., Geverts, B., Houtsmuller, A.B., Monzen, K., Komuro, I., Essers, J., Drenth-Diephuis, L.J., and Eggen, B.J. (2007). UTF1 is a chromatin-associated protein involved in ES cell differentiation. The Journal of cell biology *178*, 913-924.

Van Hoof, D., Passier, R., Ward-Van Oostwaard, D., Pinkse, M.W., Heck, A.J., Mummery, C.L., and Krijgsveld, J. (2006). A quest for human and mouse embryonic stem cell-specific proteins. Molecular & cellular proteomics : MCP *5*, 1261-1273.

Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nature reviews Genetics *15*, 69-81.

Waddington, C. (1953). Genetic Assimilation of an Acquired Character. Evolution *7*.

Wang, J., Sarov, M., Rientjes, J., Fu, J., Hollak, H., Kranz, H., Xie, W., Stewart, A.F., and Zhang, Y. (2006). An improved recombineering approach by adding RecA to lambda Red recombination. Molecular biotechnology *32*, 43-53.

Wang, Z.Q., Auer, B., Stingl, L., Berghammer, H., Haidacher, D., Schweiger, M., and Wagner, E.F. (1995). Mice lacking ADPRT and poly(ADP-ribosyl)ation develop normally but are susceptible to skin disease. Genes & Development *9*, 509-520.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Muguruma, K., and Sasai, Y. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat Biotechnol *25*, 681-686.

Wei, Y., Yu, L., Bowen, J., Gorovsky, M.A., and Allis, C.D. (1999). Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. Cell *97*, 99-109.

Weisenberger, D., and Scheer, U. (1995). A possible mechanism for the inhibition of ribosomal RNA gene transcription during mitosis. The Journal of cell biology *129*, 561-575.

Westendorf, J.M., Rao, P.N., and Gerace, L. (1994). Cloning of cDNAs for M-phase phosphoproteins recognized by the MPM2 monoclonal antibody and determination of the phosphorylated epitope. Proceedings of the National Academy of Sciences of the United States of America *91*, 714-718.

White, J., and Dalton, S. (2005). Cell cycle control of embryonic stem cells. Stem cell reviews *1*, 131-138.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Wilkins, B.J., Rall, N.A., Ostwal, Y., Kruitwagen, T., Hiragami-Hamada, K., Winkler, M., Barral, Y., Fischle, W., and Neumann, H. (2014). A cascade of histone modifications induces chromatin condensation in mitosis. Science (New York, NY) *343*, 77-80.

Woodcock, C.L., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. Experimental cell research *97*, 101-110.

Worcel, A., Han, S., and Wong, M.L. (1978). Assembly of newly replicated chromatin. Cell *15*, 969-977.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochemical Society transactions *38*, 1027-1032.

Yarden, A., and Geiger, B. (1996). Zebrafish cyclin E regulation during early embryogenesis. Developmental dynamics : an official publication of the American Association of Anatomists *206*, 1-11.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics (Oxford, England) *31*, 2382-2383.

Zaidi, S.K., Young, D.W., Montecino, M., Lian, J.B., Stein, J.L., van Wijnen, A.J., and Stein, G.S. (2010). Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. Molecular and cellular biology *30*, 4758-4766.

Zaret, K.S. (2014). Genome reactivation after the silence in mitosis: recapitulating mechanisms of development? Developmental cell *29*, 132-134.

Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., Li, J., Dong, S., Koseki, H., and Wong, J. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. Cell research *21*, 1723-1739.

Zhang, Y., Lin, Y.H., Johnson, T.D., Rozek, L.S., and Sartor, M.A. (2014). PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. Bioinformatics (Oxford, England) *30*, 2568-2575.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W*., et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Zhang, Y., Zhang, D., Li, Q., Liang, J., Sun, L., Yi, X., Chen, Z., Yan, R., Xie, G., Li, W*., et al.* (2016). Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. Nature genetics *48*, 1003-1013.

Zhao, R., Nakamura, T., Fu, Y., Lazar, Z., and Spector, D.L. (2011). Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. Nature cell biology *13*, 1295-1304.