A DUAL FUNCTION OF PRDM14 IN HUMAN EMBRYONIC STEM CELLS AND REPROGRAMMING

CHAN YUN SHEN

(B. Sc. (Hons.), NUS)

A THESIS SUBMITTED FOR THE DEGREE OF PHILOSOPHY

NUS GRADUATE SCHOOL FOR INTERGRATIVE SCIENCES AND ENGINEERING

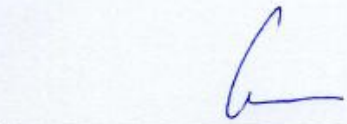
NATIONAL UNIVERSITY OF SINGAPORE

2012

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



CHAN YUN SHEN

12th December 2012

ACKNOWLEDGEMENTS

I would like to thank the NUS Graduate School for Integrative Sciences and Engineering for funding my PhD candidature and Dr Ng Huck Hui for his patient guidance and support throughout the past 4 years. I would like to express my gratitude to all who have helped me make this study possible. I thank Dr Chia Na Yu for sharing her PRDM14 reagents, Dr Jonathan Göke and Dr Yuriy Orlov for their help in the bioinformatics analyses, Dr Feng Bo for her guidance and help in human iPSC reprogramming, Dr Yu Qiang for the EZH2 plasmid, Dr Nandini Venkatesan and Dr Su I-hsin for providing the purified PRC2 complex and help on the GST pull-down assays, Ms Katty Kuay for assisting in generating the PRDM14 antibody and Miss Yang Lin, Miss Lau Mei-sheng, Miss Goh Kim-jee and Mr Terence Tng Jie-lun for their technical support. I thank Ms Yaw Lai Ping for the ordering and purchasing of reagents for the project. I am grateful to Dr Tara Huber and Prof Davor Solter for their support and guidance as my thesis advisory committee. I thank Kevin Andrew Uy Gonzales and Yeo Jia Chi for their critical comments on the thesis. I would also like to thank all the colleagues in the Genome Institute of Singapore for their various support in administration and research, all the thought provoking discussions and kind friendships over the years. Lastly, I would like to thank my family and Miss Chia Shumei for their relentless support and encouragement that carried me through these years.

TABLE OF CONTENTS

SUMMARY

Human embryonic stem cells (ESC) defined by their ability to self-renew indefinitely and to give rise to all adult cell types, are potentially invaluable in the field of regenerative medicine. However, this expanded cell fate potency also proves to be a hurdle for precise control of human ESC differentiation for various applications. Without detailed knowledge on how this cellular plasticity is regulated, their full potential remains to be harnessed. Extensive efforts have been placed in understanding how mouse ESC pluripotency is regulated, however much more needs to be done for human ESC given the mounting evidence that ESC from both species are intrinsically different. PRDM14 was identified in a genome-wide RNAi screen as an important determinant of the human ESC identity. PRDM14 is essential for maintaining the human ESC state and also enhances the efficiency of reprogramming fibroblasts back to a pluripotent state. In this study, I dissected the mechanism in which PRDM14 safeguards and induce pluripotency. PRDM14's global binding profile overlaps with the core ESC regulators OCT4 and NANOG across the genome of human ESC. Importantly, PRDM14 was found to directly regulate the expression of *OCT4* in human ESC. Corresponding to their overlapping binding profiles near pluripotency associated genes, PRDM14 interacts with NANOG and the pair displays synergistic activity at the *OCT4* enhancer and in iPSC reprogramming. On the other hand, PRDM14 also binds many developmental genes across the genome and its binding sites are enriched with the repressive tri-methylation of histone H3 lysine 27 (H3K27me3) mark. Majority of PRDM14 bound genes were upregulated when PRDM14 is depleted in the human ESC. PRDM14 was found to interact with the polycomb repressive complex II (PRC2) and this interaction is important for PRDM14-mediated repressive activity in human ESC and iPSC reprogramming. In summary, PRDM14 was found to play dual functions with different partners in maintaining and inducing the human pluripotent cell state.

LIST OF PUBLICATIONS

1. Chan YS, Jonathan Göke J, Lu XY, Venkatesan N, Feng B, Su IH, Ng HH. (2012). A PRC2 dependent repressive role of PRDM14 in human embryonic stem cells and iPSC reprogramming. STEMCELLS (Accepted. Waiting for publication date).
2. Chan YS, Yang L, Ng HH. (2011). Transcriptional regulatory networks in embryonic stem cells. Prog Drug Res. 67, 239-52.
3. Chia NY*, Chan YS*, Feng B*, Lu X, Orlov YL, Moreau D, Kumar P, Yang L, Jiang J, Lau MS, Huss M, Soh BS, Kraus P, Li P, Lufkin T, Lim B, Clarke ND, Bard F, Ng HH. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature 468, 316-20 (*co-authors)
4. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 42, 631-4.
5. Chan KK, Zhang J, Chia NY, Chan YS, Sim HS, Tan KS, Oh SK, Ng HH, Choo AB (2009). KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells. Stem Cells 27, 2114-25.
6. Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J, Vega VB, Cacheux-Rataboul V, Lim B, Lufkin T, Ng HH. (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol. 11, 197-203.
7. Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, Lim CA, Robson P, Zhong S, Ng HH. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol. 10, 353-60.

LIST OF TABLES

LIST OF FIGURES

LIST OF ILLUSTRATIONS

LIST OF SYMBOLS AND ABBREVIATIONS

Symbol	Definition
˚C	Degree Celsius
%	Percentage
e	Exponential
µg	Microgram
µl	Microlitre
µM	MicroMolar
mg	Milligram
ml	Millilitre
mM	MilliMolar
k	1000 Units
min	Minute
hr	Hour

Abbreviation	
4F	Reprogramming factors OCT4, SOX2, KLF4 and c-MYC
bFGF	Basic Fibroblast growth factor
BMP	Bone morphogenetic protein
bp	Base pairs
cDNA	Complementary DNA
ChIP	chromatin immunoprecipitation
ChIP-chip	ChIP coupled with microarray chips
ChIP-PET	ChIP coupled with Pair-end ditags
ChIP-Seq	ChIP coupled with high throughput sequencing
CR	Conserved region
Ct	Cycle threshold
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DBD	DNA binding domain
EB	Embryoid body
EC	Embryonal carcinoma
EpiSC	Epiblast stem cell
ESC	Embryonic stem cell
FBS	Fetal bovine serum
GFP	Green fluorescent protein
GO	Gene Ontology
GST	Glutathione S-Transferase
H3K4me1	Mono methylated histone 3 lysine 4
H3K27ac	Acetylated histone 3 lysine 27
H3K27me3	Tri-methylated histone 3 lysine 27
iPSC	Induced pluripotent stem cells
kb	Kilo base pair
LIF	Leukemia inhibitory factor
MEF	Mouse embryonic fibroblast
PCR	Polymerase chain reaction

PRC2	Polycomb repressive complex 2
qPCR	Quantitative PCR
RNAi	Ribonucleic acid interference
shRNA	Short hairpin ribonucleic acid
S.e.m	Standard error of the mean
SSEA-1	Stage specific antigen 1
SSEA-3	Stage specific antigen 3
SSEA-4	Stage specific antigen 4
Znf	Zinc finger

Chapter 1: LITERATURE REVIEW

1.1 Stem cells and their potentials

Stem cell and tissue replacement

Stem cells are defined by their capacity to self-renew and give rise to specialized progeny cells. The existence of stem cells originates from the notion that progenitor cells are essential for the constant replenishment of various tissues throughout adult life (Fuchs and Segre, 2000); hair follicles, hepatocytes, skin and blood turnover and even neurons in the brain which was once considered irreplaceable (Temple, 2001). The existence of such cells was supported by the successful isolation of progenitor cells from different adult tissues and these cells are able to differentiate into functional progenies *in vitro* (Rheinwald and Green, 1975; Yang *et al*, 1993; Temple, 1989; Reynold and Weiss, 1992; Pittenger *et al*, 1999; Asahara *et al*, 1997; Beltrami *et al*, 2003). The successful isolation of adult stem cells brings great promises for a new era of regenerative medicine (Körbling and Estrov, 2003); cells and tissues can be generated *in vitro* for replacing worn out and disease ones in the human body. However, there are inherent disadvantages regarding the use of adult stem cells: the invasive methodology in deriving some of these cells (especially the organ specific ones), the rarity of these populations of cells *in vivo* and their limited proliferative capacity especially with the requirement of large quantity of cells for replacement therapy.

Pluripotent stem cells

Embryonic stem cells (ESC), as the name suggests, is a naïve group of cells isolated from early embryos. These cells possess two unique properties that potentially challenge the need

for adult stem cells: the capacity to self-renew indefinitely *in vitro* and the potential to give rise to all adult cell types (also known as pluripotency) (Illustration 1). ESC was first successfully isolated from the inner cell mass of pre-implantation embryos in mouse (Martin, 1981; Evans and Kaufman, 1981). The isolation of ESC was preceded by a long and convoluted history of research on a unique class of tumours known as teratocarcinomas (for review, see Davor, 2006). This tumour contains tissues of all three germ layers, highlighting the expanded cell fate capacity of these cancerous cells. The successful establishment of an embryonal carcinoma (EC) cell line from teratocarcinomas by Martin and Evans set the stage for their subsequent isolation of ESC using similar culture conditions (Martin and Evans, 1974). These EC cells were able to contribute to the formation of chimeric mice when injected into early blastocysts, highlighting the developmental potential of these cells (Brinster, 1974; Mintz and Illmensee, 1975). ESC derived using similar culture methods are also capable of forming chimeras and contributing to the germline (Bradley *et al*, 1984). In addition, mouse ESC have also been shown to display similar developmental potential *in vitro*; these cells have been differentiated *in vitro* into neurons (Strübing *et al*, 1995), glia (Fraichard A *et al*, 1995) and keratinocytes (Bagutti *et al* 1996) of the ectoderm lineage, cardiomyocytes (Doestschman *et al*, 1985), adipocytes (Dani *et al*, 1997) and erythrocytes (Nakano *et al*, 1996) of the mesodermal lineage and, hepatocytes (Jones *et al*, 2002) and pancreatic cells (Lumelsky *et al*, 2001) of the endoderm lineage.

Almost two decades after the isolation of mouse ESC, Thomson *et al* isolated the first human ESC line (Thomson *et al*, 1998). When injected into immuno-compromised mice, the cells similarly form tumours that contain tissues of all three germ lineage, exhibiting similar pluripotent capacity as the mouse ESC. The successful isolation of human ESC represents a milestone for using ESC in regenerative therapy. However, the use of human ESC is plagued

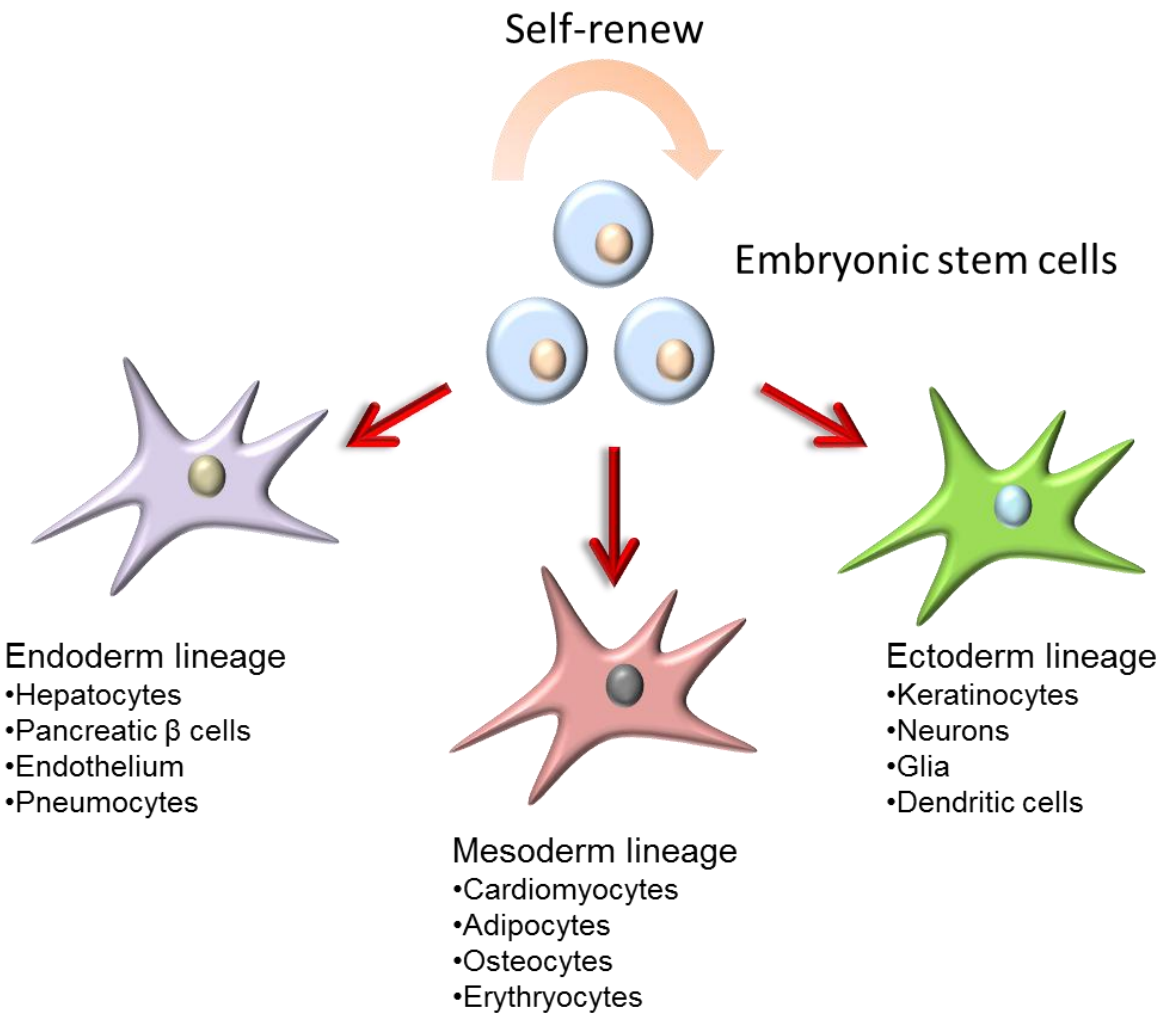


Illustration 1. Characteristic of embryonic stem cells. ESC are defined by their ability to self-renew limitlessly and retain the potential to differentiate into cells of all 3 germ lineage; ectoderm, mesoderm and enododerm. Listed are some of the various somatic tissues found in each lineage

with ethical issues regarding the use of human oocytes and embryos towards deriving these cell lines. With the limited cell lines available due to such complications, clinicians may face an issue of graft rejection as there are not enough human ESC lines to cater for all patients' immuno-compatibility. A major breakthrough in looking for new ways to generate pluripotent stem cells was achieved by Takahashi and Yamanaka, whereby they demonstrated that differentiated mouse fibroblast cells can be reverted back to the pluripotent state by the ectopic expression of key ESC regulators (Takahashi and Yamanaka, 2006). These cells termed as induced pluripotent stem cells (iPSC) exhibited similar pluripotent characteristic as the mouse ESC and could even give rise to the entire organism through tetraploid complementation assays (Zhao *et al*, 2009). Using similar strategy, Takahashi *et al* showed that human iPSC can also be generated from human adult somatic cells (Takahashi *et al*, 2007). The establishment of the iPSC technology overcame two major hurdles faced when using ESC in regenerative medicine: avoid the use of oocytes and embryos to generate pluripotent stem cells, and deriving patient immuno-compatible cells. The advent of iPSC technology brings human pluripotent stem cells one step closer towards translational applications.

1.2 Applications of pluripotent stem cells (PSC)

The infinite proliferative capacity of these PSC (ESC and iPSC) potentially provides limitless amount of cells for therapeutic applications. This strongly favours the use of PSC for regenerative therapy. With increasing knowledge of how these cells function, the use of PSC have expanded beyond generating somatic cells for tissue replacement; PSC have been employed in the field of organogenesis, in the process of drug discovery, for genome editing and in basic research (Table 1).

Table 1. Applications of pluripotent stem cell (PSC)	
Regenerative medicine	A renewable source for producing adult tissues for replacement therapies and for organogenesis.
Drug discovery	Facilitates the creation of transgenic mouse model for *in vivo* system studies. Also for the production of various tissue types for high throughput drug screening *in vitro.*
Genome editing	Gene correction therapy and generating hereditary disease models
A basic research tool	An opportunity to study early cell fate development *in vitro.* Also a mammalian primary cell line for various basic researches.

Organogenesis

One of the ultimate goals in regenerative medicine is the replacement of diseased and worn out organs in the human body. The formation of an organ would require cell types from multiple lineages and building organs with somatic cells may require a *tour de force* assembly of large numbers of different specialized cells in three dimensional cultures. Impressively, this technical feat has been achieved with ESC culture *in vitro*; by carefully manipulating the culture conditions of ESC aggregates in floating cultures, Eiraku *et al* successfully generated an self-organising optic cup structure *in vitro* (Eiraku *et al*, 2011). The aggregates of floating mouse ESC display dynamic self-assembling and self-formation properties to achieve this complex morphogenesis in the petri dish. More significantly, Spence *et al* showed that by using a similar stepwise approach with different culture medium, human ESC can also be directed *in vitro* to generate intestinal-like tissues (Spence *et al*, 2011). These challenging feats may not have been possible without the inherent capacity of ESC to proliferate and give rise to different cell types.

Genetically modified mice and drug discovery

A Nobel-winning application of ESC is the generation of genetically modified mice (GeMM) (Smithies *et al*, 1985; Robertson *et al*, 1986; Thomas and Capecchi, 1987). Robertson *et al* displayed how genetic modifications in mouse ESC could be germline transmitted, hence allowing the generation of transgenic mice with both alleles of the targeted genes altered. This strategy has been extensively utilised to generate transgenic mice, mostly target gene knockout (KO) models, for functional studies and drug discovery (McNeish, 2004). The KO mice provided an opportunity for drug discovery researchers to test their candidates' function, specificity and toxicity *in vivo*. Thus far, the drug's effects seen in the KO mice are well recapitulated when used in human therapeutics (Zambrowicz and Sands, 2003). Nevertheless, moving on from animal to human trials, it would be desirable to first test drugs on human tissues *in vitro*. Previously, drug discovery is dependent on the use of primary tissues, tumorigenic cell lines or immortalized cell lines (McNeish, 2004). However, primary tissues that are often rare or have limited proliferative capacity is not cost effective for large scale drug screening. Results obtained from immortalized and cancerous cell lines on the other hand may be erroneous due to the altered nature of these cells. Human ESC is thus highly favoured in the pipeline for large scale drug screenings as it potentially provides limitless amount of various primary tissues.

Genome editing and genetic disease models

While genome editing in mice provides transgenic model towards the *in vivo* study of genetic diseases, the ability to replicate such studies in human cells would be of higher clinical relevance. Similar to mouse ESC, human ESC can be used in generating transgenic human cell lines (Zwaka and Thomson, 2003). The ability to edit the human ESC genome expands

the potential use of these cells to create genetic disease cell line for drug screening and research, and importantly, correction of mutations in genetic diseases. These potential applications drive researchers to improve the genome editing efficiency in human PSC using various technologies (Hockemeyer *et al*, 2009; Khan *et al*, 2010; Hockemeyer *et al*, 2011; Song *et al*, 2010). While the use of genome editing in human ESC have provided a new avenue for studying genetic diseases (Soldner *et al*, 2011), the iPSC technology have similarly allowed researchers to generate cell lines with hereditary diseases. The ability to convert disease patient's fibroblasts to iPSC potentially allows the establishment of an arsenal of disease model cell lines (Park *et al*, 2008). Importantly, these iPSCs acquire the disease phenotypes when differentiated to the respective cell types (Ebert *et al*, 2009; Lee *et al*, 2009; Raya *et al*, 2009; Ye *et al*, 2009), thus providing valuable opportunity for researchers to understand how the diseases manifest when the specialized cells are formed.

A basic research tool

While disease-patient iPSC provides valuable cell lines for studying genetic diseases, the ESC differentiation process provides unprecedented opportunities for scientists to dissect mechanisms of specification and also early developmental processes *in vitro*. ESC represents one of the most naïve cells *in vivo* and its commitment into the various lineages would reveal insights to the molecular machinery that drives cell fate transition during gastrulation (D'Amour *et al*, 2005; Oldershaw *et al*, 2010; Kamiya *et al*, 2011). Indeed, by using mouse ESC, Hayashi *et al* successfully recapitulated the entire germ cell specification events, capturing the formation of different intermediate states such as the epiblast stem cell (EpiSC) like and primordial germ cell (PGC) like cells *in vitro* (Hayashi *et al*, 2011).

The dissection of cellular processes in early studies in eukaryotes was performed in simple unicellular organisms such as *Saccharomyces cereversiae* (yeast). The ease of handling and expansion of yeast culture and its small genome makes it an ideal model organism for dissecting fundamental principles in cellular proliferation, metabolism and specification. In particular, large scale mapping studies have been carried out in yeast to reveal for the first time, intricate networks of transcriptional regulation that govern different cellular processes (Lee *et al*, 2002). Dissecting transcription regulation in higher multicellular organism such as the sea urchin revealed a much complex gene regulation program (Davidson *et al*, 2002). This prompts for a need to carry out such studies directly in higher mammals. However, such mapping studies require large quantities of homogenous primary cells which prove to be a formidable challenge. ESC proves to be a feasible alternative for such studies. The primary ESC cell line allows limitless expansion to provide the numbers for such mapping studies to be done. In addition, ESC differentiation *in vitro* also allows researchers to perform such studies to dissect the specification process. In summary, PSC proves to be more than an excellent candidate for regenerative therapies; these cells are also invaluable for industrial applications and basic research.

1.3 Hurdles and challenges in utilising PSC

Homogeneity of differentiated cells

The ability of PSC to self-renew indefinitely and differentiate to diverse cell types proves to be a double-edged sword. A major issue with differentiation of PSC to terminally differentiated cells is the homogeneity of the cells derived. The differentiation of PSC to the desired somatic cells involves the transition through progenitor states with multiple developmental potentials; when ESC differentiate towards the endoderm and mesoderm

lineage, the cells pass through a bi-potent intermediate mesoendoderm progenitor (Rodaway and Patient, 2001), and ESC differentiation to the neurons involves the transition to multiple glial progenitors that give rise to both astrocytes and neurons (Kriegstein and Alvarez-Buylla, 2009). Each progenitor state thus serves as a gateway to multiple cell types between two different lineages or within the same lineage (Illustration 2). This amplifies the diversity of different cell types generated during the ESC differentiation process. While the control of culture conditions potentially eliminates cells of diverse lineages from arising, overlapping signalling requirements in different cell types still result in heterogeneous cell cultures. In an effort to produce hormone-secreting endocrine cells which would be beneficial for diabetic treatment, D'Amour *et al* established a 5 step differentiation protocol involving multiple intermediate cell states under different culture conditions (D'Amour *et al*, 2006). The two weeks long differentiation process only generated 12% of insulin-producing cells in the culture. The authors emphasized the importance of generating pure populations of definitive endoderm progenitor in the first stage and highlighted the presence of neural or extra-embryonic lineage cells at this stage. Similar issues plagued the differentiation of human ESC into the anterior foregut endoderm (Green *et al*, 2011) and hepatocytes (Cai *et al*, 2007). The need to constantly enrich for the population of target cells in each stage results in high cost incurred in the process of differentiation and harsh treatment to the cells in the purification processes reduces cell viability. The large starting number of human ESC required and the high cost of enrichment and culture media, make such protocols unsuitable in generating sufficient amount of cells for therapeutic purposes. Hence, there is a need to unravel better strategies to shut down and manipulate the pluripotency network to allow more efficient differentiation of human ESC into specific cell types.

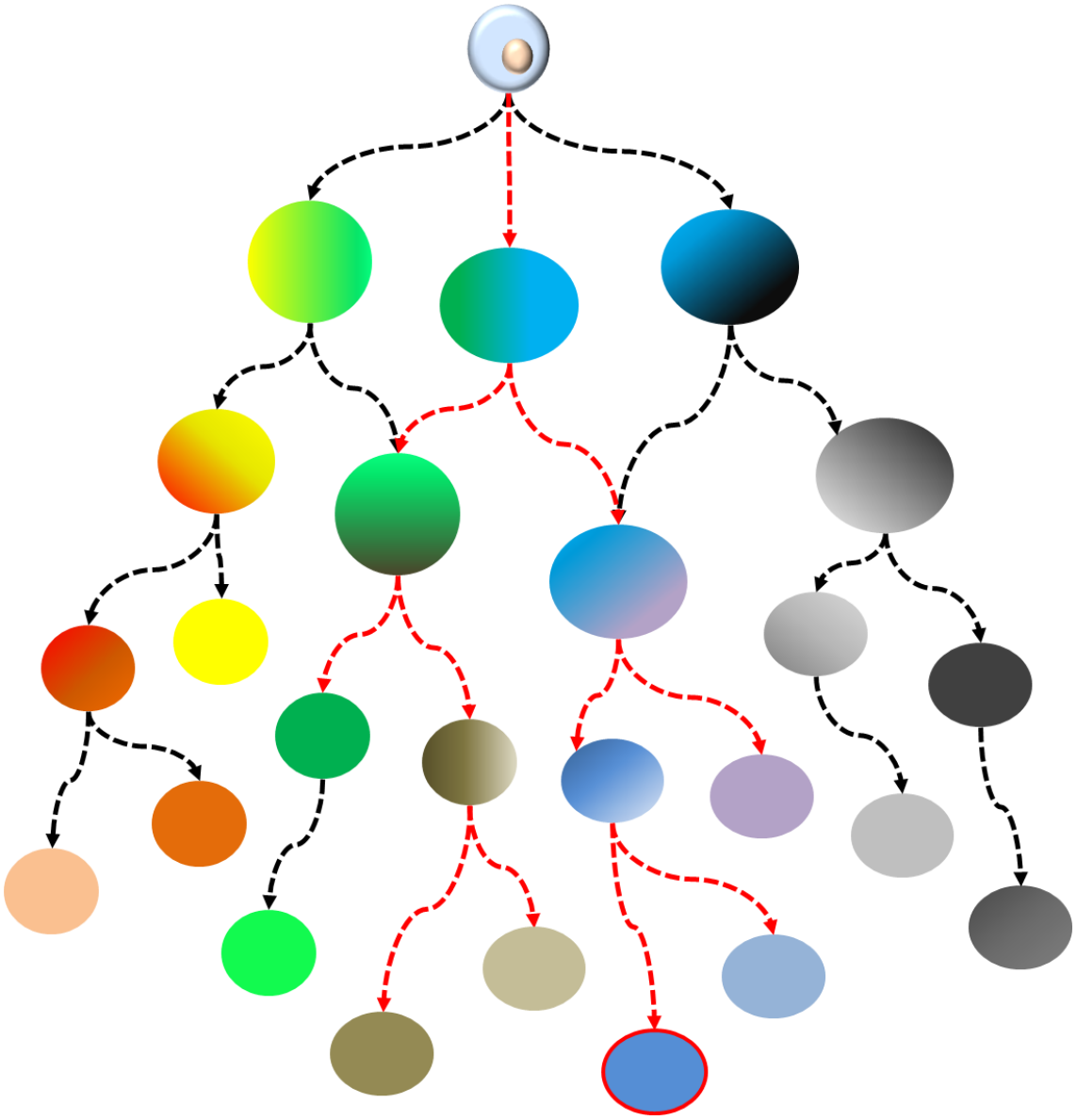


Illustration 2. Transient amplifying progenitor in ESC differentiation. A graphical illustration of ESC differentiation to somatic cells of various lineages. Each circle represents a cell type that can be derived from the ESC. The pluripotent stem cell is at the apex of cell fate potential while terminally differentiated somatic cells are at the base (single coloured). Differentiation of ESC involves the formation of intermediate progenitor that can give rise to more than 1 type of cells (Multi-potent progenitor are highlighted with a gradient of 2 colours). Dotted arrows highlighted in red depicts how the differentiation of ESC to a somatic cell (outlined in red) can potentially generate more than 1 cell type due to the transition into multiple progenitors.

Problem of teratogenicity arising from contaminating pluripotent human ESC

The need to efficiently shutdown the ESC network work is also highlighted by the growth of tumours in the transplanted tissues derived from human ESC (Fujikawa *et al*, 2005; Li *et al*, 2008). The prospect of having tumour growth in the transplanted tissue calls for stringent screening of potential proliferative cells within the differentiated pool of cells. In the study by Fujikawa *et al* to produce insulin-secreting cells for transplantation, the human ESC were differentiated up to a month and after which, the cells were negatively sorted for OCT4 and SSEA-1 (ESC specific markers). Unexpectedly, tumours that contain small populations of OCT4 and SSEA-1 cells were formed three weeks post transplantation. This study highlights the potential existence of progenitor stem cell populations within sorted cells that retain the tumorigenic properties of the parental human ESC. The extension of the course of differentiation reduces the incidence of tumour formation; however, the viability of the cells is greatly compromised in the process (Li *et al*, 2008). Thus, there is a need to understand the underlying mechanisms that maintain the ESC regulatory network, to devise strategies that ensure a full shutdown of this program in the differentiated cells derived from ESC.

Heterogeneous human ESC differentiation propensity

The potential applications of human ESC drove researchers around the world to isolate more human ESC lines. Differentiation assays performed on human ESC with different genetic background yet reveal greater challenges in utilising these cells. An initiative by the International Stem Cell Forum (http://www.stemcellforum.org.uk) performed an extensive characterization of the available human ESC line established worldwide (Adewumi *et al*, 2007). One of the findings in this study is that human ESC lines isolated by different labs express different levels of each lineage markers during EB differentiation, suggesting that

these human ESC lines behave differently under similar differentiation conditions. In a separate study, Osafune *et al* showed that these differences observed during ESC differentiation between the human ESC lines from different labs are not due to the disparity in isolation methods (Osafune *et al*, 2008). They performed human ESC differentiation into pancreatic and cardiac progenies and observed a 100 fold differences in differentiation propensity between 17 human ESC lines isolated by the lab. Such differences were also observed in the differentiation to haematopoietic cells (Melichar *et al*, 2011) and similarly in human iPSC differentiation into the neural lineage (Hu *et al*, 2009). These significant differences in the differentiation propensity between human PSC hamper the effort of establishing universal differentiation protocols; a need to optimize or reinvent protocols for individual human ESC lines as what works for some PSC lines may not work for others. Human PSC lines are largely characterized by the presence of cell surface antigens and few well known pluripotency markers. There is a need for detailed molecular characterization of various human ESC lines to identify molecular cornerstones that reflect their differentiation propensity. This involves dissecting the molecular circuitry governing pluripotency.

The iPSC technology has provided stem cell researchers with limitless source of patient specific PSC. The next major challenge lies in establishing protocols that efficiently direct these cells into high quality and functional somatic cells that are safe for transplantation and reflective of endogenous tissues for drug screening. The above hurdles discussed need to be addressed before these applications become possible. A detailed understanding on how the pluripotency cell state is maintained would very well allow us to devise better strategies in differentiating human PSC.

1.4 Understanding the pluripotent cell state

Core ESC regulator Oct4, Sox2 and Nanog

The higher vertebrates contain a myriad of cell types which exhibit diverse morphologies, unique cellular compositions and perform various functions for the survival of the organism. As almost all cells share the same genome, the cell state would thus be determined by the transcriptional network that establishes the epigenetic status of the cell genome. Since the isolation of the mouse ESC, extensive efforts have been committed by the stem cell community to identify key transcription factors that govern the pluripotent cell fate. These studies serve to identify the regulators of the ESC state to better understand how the cell plasticity is maintained. The importance of these efforts was epitomized by the iPSC reprogramming achievement by Takahashi and Yamanaka. The use of just 4 transcription factors is sufficient to reprogram the committed cell fate to a pluripotent state (Takahashi and Yamanaka, 2006). This study highlighted the importance and critical role of transcription factors in determining the cell state.

Oct4, a POU domain containing transcription factor was the first key regulator of the stem cell fate identified in mouse. Oct4 was found to be expressed specifically in the inner cell mass (ICM) of early blastocyst, the primitive ectoderm in the late epiblast and in germ cells (Rosner *et al,* 1990). Oct4-null embryos failed to form the pluripotent ICM and cells in the blastocyst were mainly trophoblastic (Nichols *et al*, 1998). The importance of Oct4 in maintaining the pluripotent cell state was further exemplified by the sensitivity of ESC towards the level of *Oct4* expression. Ectopic expression of Oct4 in mouse ESC results in differentiation towards the primitive endoderm and mesoderm lineage (Niwa *et al*, 2000). In its role of maintaining the pluripotent cell state, Oct4 was observed to constantly partner

another transcription factor; the HMG-box transcription factor Sox2 was first identified as an interacting partner of Oct4 in regulating the developmental gene *Fgf4* (Yuan *et al*, 1995). This partnership was also observed in the regulation of pluripotency associated genes in mouse ESC (Chew *et al*, 2005; Rodda *et al*, 2005). However, unlike Oct4, Sox2 is dispensable in the formation of the pluripotent cell mass and instead is required for multi-potent cell precursors for post-implantation development (Avilion *et al*, 2003). Sox2 expression is less confined to the embryonic stage and is found in several epithelial tissues in the adult mice (Arnold *et al*, 2011). Subsequent study with inducible Sox2 knockout mouse ESC shows that the molecule is important for regulating key genes that are important for maintaining the *Oct4* expression levels in ESC (Masui *et al*, 2007).

Nanog was identified in two separate studies in an effort to look for regulators of the pluripotent cell state that is independent of external signalling. It was previously established that the mouse ESC is highly dependent on the Lif/gp130/ Stat3 pathway to maintain the pluripotent state (Smith *et al*, 1988; Okada *et al*, 1988). However, this pathway seems to be dispensable for the formation of ICM in the blastocyst (Stewart *et al*, 1992; Ware *et al*, 1995; Takeda *et al*, 1997). In a bid to look for a pluripotent regulator independent of this pathway, two groups identified Nanog as a key regulator of the pluripotent cell state (Mitsui *et al*, 2003; Chambers *et al*, 2003). Ectopic expression of Nanog allowed LIF-independent culture of mouse ESC. Impressively, this signalling-independent phenotype by the overexpression of Nanog was also observed in the human ESC (Xu *et al*, 2008), suggesting the conserved role of Nanog in mouse and human ESC. Interestingly, Nanog-null ESC can be established *in vitro*, and although these cells are prone towards differentiation, they are nevertheless able to contribute towards chimeric mice indicating that they are still pluripotent (Chambers *et al*, 2007). Further in depth studies reveal that Nanog serves as an important gateway to establish

pluripotency *in vitro* (reprogramming) and *in vivo.* Importantly, this delineation of Nanog function in mouse ESC highlights that a transcription factor that is important for establishing pluripotency may not be important for its maintenance.

Mapping the core pluripotency network

While more transcription regulators remain to be discovered, researchers are eager to investigate how the core regulators establish the pluripotent transcriptional network in ESC. In separate studies, Boyer *et al* and Loh *et al* attempt to map the global binding profiles of the core regulators in human and mouse ESC respectively (Boyer *et al*, 2005; Loh *et al*, 2006). Boyer and colleagues established the global promoter binding profiles of human OCT4, SOX2 and NANOG. Their study utilised the chromatin immunoprecipitation assay (ChIP) coupled with whole genome promoter chip arrays (ChIP on chip) to identify potential downstream target genes of the core regulators. Importantly, the mapping studies showed that the three transcription factors have overlapping occupancy on at least 353 genes. Of this, NANOG binds to almost 90% of genes that is co-bound by the OCT4 and SOX2. The genes bound by the core regulators contain both ESC specific genes and also lineage specific genes, which suggest that the core regulators may be playing both activating and repressive roles. Overall, the findings in this study suggest that the three core regulators function cohesively in concert to regulate genes in a context dependent manner (activate or repress) in human ESC.

Loh and colleagues on the other hand, mapped the genome-wide binding profile of Oct4 and Nanog in mouse ESC (Loh *et al*, 2006). They employed a sequencing platform for quantifying the ChIP enriched DNA, using pair end ditags (ChIP-PET). This allows the

mapping of Oct4 and Nanog binding across the entire genome other than the promoters. Similarly, the mouse study showed that Oct4 and Nanog co-bind many genes in mouse ESC. This conserved partnership between Oct4 and Nanog suggests that the core regulators function together to establish a core pluripotency regulatory network to maintain the ESC state. In addition, the mouse study identified potential downstream regulators such as Esrrb and Rif1 whose expressions are important to maintain the ESC identity (Loh *et al*, 2006). Interestingly, other than co-binding downstream regulatory genes, the core regulators bind to their own promoter and to that of one another in both mouse and human ESC. These auto-regulatory, feed forward and backward loops established by the core factors potentially keep their expression in check to achieve a stable equilibrium. This equilibrium forms the apex of the transcriptional network for the activation of downstream regulatory genes in establishing the pluripotency cell state (Illustration 3).

Given the conserved partnership between OCT4 and NANOG in mouse and human ESC, it is surprising that the percentage overlap in the genome-wide binding profile of the human and mouse orthologs is only 9.1% for OCT4 and 13% for NANOG (Loh *et al*, 2006). In a bid to understand this difference in the OCT4 and NANOG regulatory network in mouse and human ESC, Kunarso and colleagues investigated the potential redistribution of regulatory elements between the two species (Kunarso *et al*, 2010). This study used the genome-wide binding datasets generated by ChIP coupled with high throughput sequencing technology (ChIP-Seq). This eliminates potential discrepancy due to different quantitative platforms employed by Loh *et al* and Boyer *et al*. Looking at the top 10% of genes bound by the factors, the overlap between mouse and human OCT4 binding is only 3.8% and 5.3% for NANOG, compared to the 49% for the control CTCF, a conserved transcription factor that

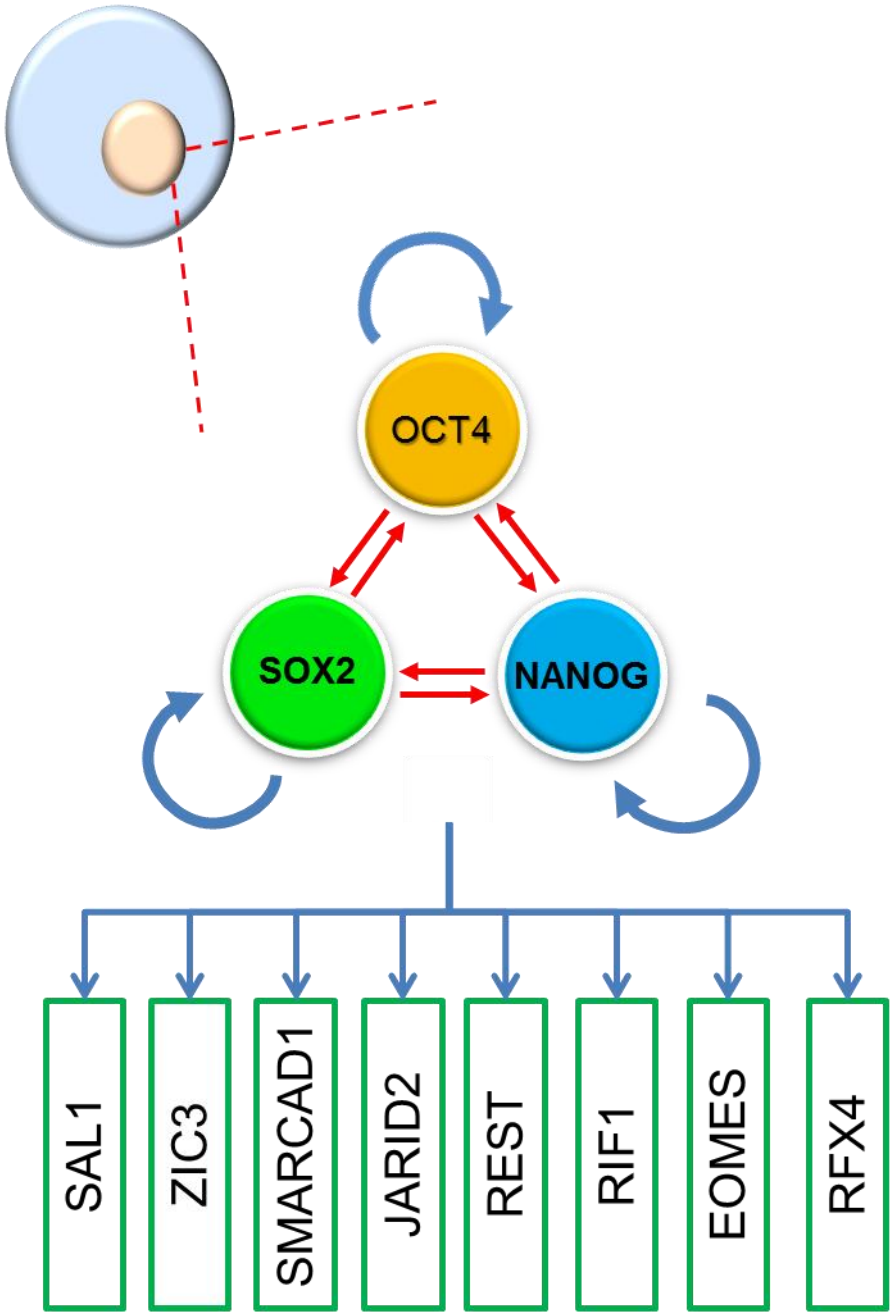


Illustration 3: The transcriptional regulatory network of the core regulators in ESC. The core regulators of the human and mouse ESC cell fate OCT4, SOX2 and NANOG forms a self-regulating circuitry where each factor binds to its on gene and onto the other regulators. This tight regulation keep the gene expression level of the core regulators in check and the core regulators co-bind extensively to downstream pluripotency-associated genes to maintain ESC cell fate. Shown are some of the genes which are co-bound by the core regulators in both human and mouse ESC.

functions as an enhancer insulator (Bell *et al*, 1999). Importantly, their finding indicates that transposon elements account for almost 25% of the OCT4 and NANOG bound sites in human and mouse ESC and these elements have wired new genes into their regulatory network. This study suggests that while the NANOG and OCT4 molecules in human and mouse ESC retains their conserved co-regulatory function, their downstream target genes may have diverged evolutionarily.

The extended pluripotency regulatory network

The growing interest of stem cell researchers on the molecular circuitry governing the pluripotency transcriptional network resulted in a long list of transcription factors identified to play a role in maintaining the ESC state (for a detailed list, see Young, 2011). Such efforts range from small scale screening studies (Ivanova *et al*, 2006; Fazzio, *et al* 2008; Gaspar-Maia *et al*, 2009) to genome-wide RNAi screens in mouse ESC (Ding *et al*, 2009; Hu *et al*, 2009; Kagey *et al*, 2010). In parallel, many potential candidates are also identified through interaction studies with the core regulators Oct4 (Van den Berg *et al*, 2010; Pardo *et al*, 2010) and Nanog (Wang *et al*, 2006). With more pluripotency regulators identified, the pluripotency regulatory network is thus expanded. Kim *et al* included 6 additional transcription factors Klf4, Daz2, Nac1, Zfp281, c-Myc and Rex1 along with the core regulators in their mapping studies in mouse ESC using the ChIP on Chip platform (Kim *et al*, 2008). The extended map revealed that the pluripotency regulators preferentially co-bind to promoter of target genes which are expressed in the ESC whereas genes bound by individual factors alone are usually repressed. The large-scale mapping studies also revealed that the transcription factors co-localized in sub clusters; while Klf4, Daz2, Nac1, Zfp281 showed significant overlap with the core regulators, c-Myc and Rex1 preferentially co-bind

with other genes to form another regulatory node. c-Myc and Rex1 co-bound genes are highly enriched for metabolic and proliferation associated genes compared to the enrichment of pluripotency and developmental associated genes in the core regulatory network.

On the other hand, using the ChIP-Seq platform, Chen *et al* similarly expanded the mouse pluripotency map to include transcription factors Stat3, Smad1, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1 and E2f1 (Chen *et al*, 2008). This study included the downstream effectors of the signalling pathways that maintained the mouse ESC state; the Bmp4/Smad1 and Lif/Stat3 pathway. The mapping results reveal that these signalling effectors integrate with the core signalling network to activate pluripotency genes. Their binding activities are dependent on the core regulators, suggesting that the signalling pathways play a supportive role in maintaining the pluripotency network. Similar to the previous study by Kim *et al*, this study also identified two separate clusters of regulators; an Oct4-centric cluster alongside with Sox2, Nanog, Smad1 and Stat3, and a Myc centric cluster that consists of c-Myc and n-Myc, E2f1 and Zfx. Interestingly, only genomic regions bound by the Oct4-centric cluster displayed enhancer activities in ESC. The mapping of the co-activator p300 accurately predicts enhancers *in vivo* (Visel *et al*, 2009). Correspondingly, the co-activator p300 is shown to co-localize extensively at regions bound by the Oct4-centric clusters but not the Myc-centric clusters (Chen *et al*, 2008). These regions strongly enriched for the Oct4-centric pluripotency factors are proposed to be the enhanceosomes of the mouse ESC genome. These transcriptional hotspots potentially harbour gene regulatory codes that may be essential for maintaining the pluripotent state. Overall, the expanded ESC regulatory network reveals that the pluripotency factors preferentially work in groups to regulate different classes of genes within the pluripotent transcriptional network (Illustration 4).

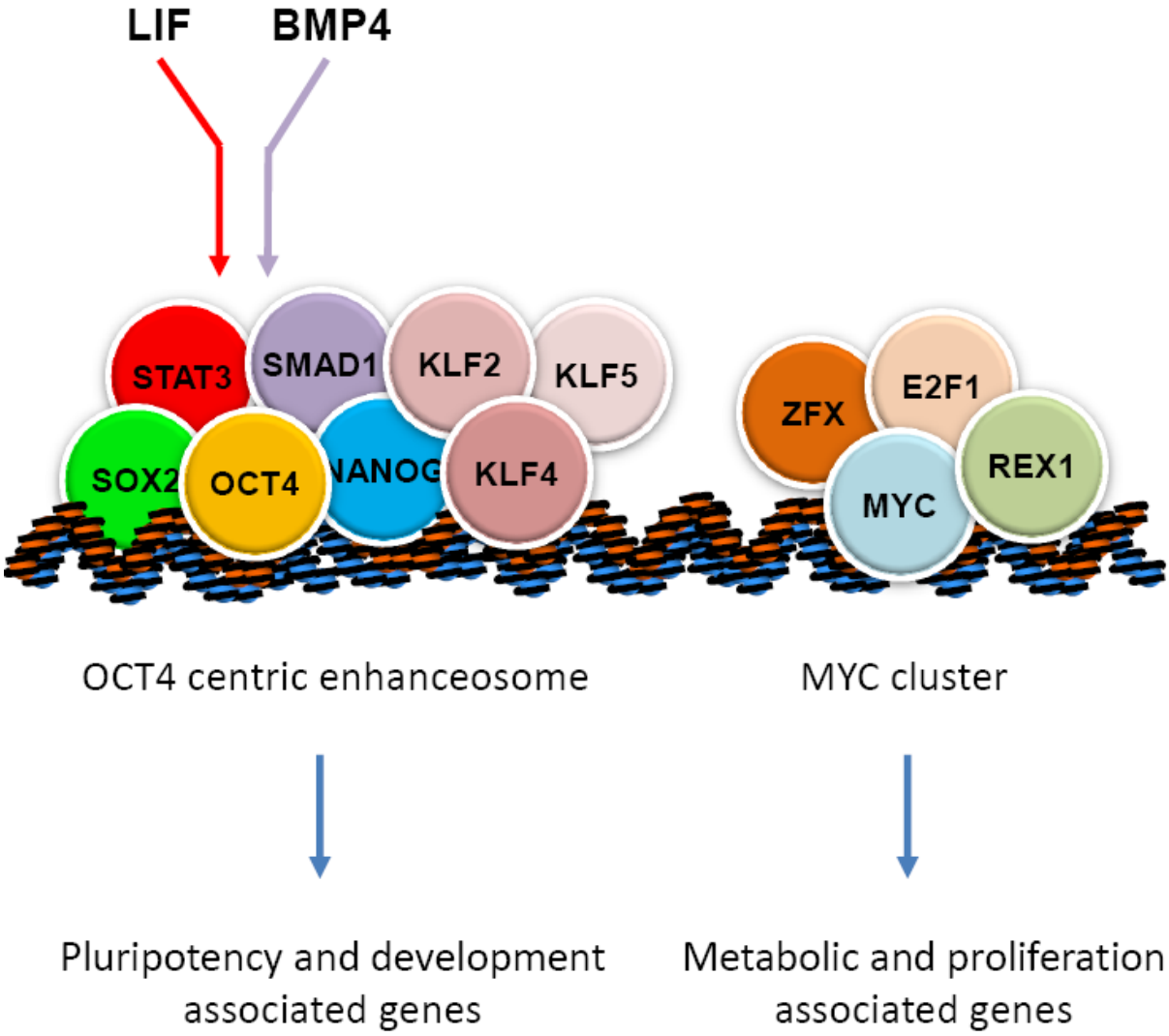


Illustration 4: The expanded pluripotency network in mouse ESC. The identification of more pluripotency regulators allows the further expansion of the transcriptional regulatory network that regulates the ESC fate. The ESC regulators preferentially bind together in mouse ESC to regulate genes. Two clusters of transcription factors, regulating unique set of genes have been identified. An Oct4 centric cluster that include Smad1 and Stat3 which are the downstream effectors of the Lif/Stat3 and Bmp4/Smad1 signaling pathways essential for mouse ESC maintenance. This cluster marks the ESC specific enhanceosome and predominantly regulates pluripotent and development associated genes. On the other hand, the MYC cluster of transcription factors preferentially bind to genes associated with metabolic and proliferative functions.

The epigenome of the pluripotent state

Other than the transcription factors, chromatin modifiers have been identified to play a part in maintaining the pluripotent cell state. Chromatin modifiers such as the polycomb repressive complex 1 (PRC1) and 2 (PRC2) (Boyer *et al*, 2006; Lee *et al*, 2006; Leeb *et al*, 2010), the ESC specific esBAF complex (Ho *et al*, 2009a; Ho *et al*, 2009b), WDR5 of the trithorax group (TrG) (Ang *et al*, 2011), the Tip60-p400 complex (Fazzio *et al*, 2008), and the chromodomain helicase DNA-binding (CHD) family of ATP dependent chromatin remodelers Chd1 (Gaspar-Maia *et al*, 2009) and Chd7 (Schnetz *et al*, 2010) have all been shown to play a role in establishing the epigenetic status of the ESC state (Illustration 5). The genome-wide binding profile of several chromatin modifiers highly overlap with the core pluripotency network, suggesting interactions between the two classes of regulators. While WDR5 and esBAF seem to form a complex with the core regulators, direct interactions between the core regulators and the chromatin modifiers remain to be confirmed. Such interaction studies would potentially reveal how the two classes of regulators work cohesively to control the ESC fate.

The concerted effort of the various chromatin modifiers establish the ESC epigenome and extensive mapping studies for various histone modifications have revealed unique combinatorial histone marks that demarcate functional sites (Azuara *et al*, 2006; Bernstein *et al*, 2006; Mikkelsen *et al*, 2007; Creyghton *et al*, 2008). Azuara *et al* and Berstein *et al* both reported the unique observation that the repressed lineage genes in the mouse ESC are ambiguously marked by the active trimethylated histone H3 lysine 4 modification (H3K4me3) and the repressive H3K27me3, and thus are termed bivalent domains. The repressed genes in the ES cells marked by these bivalent domains are thought to be in a

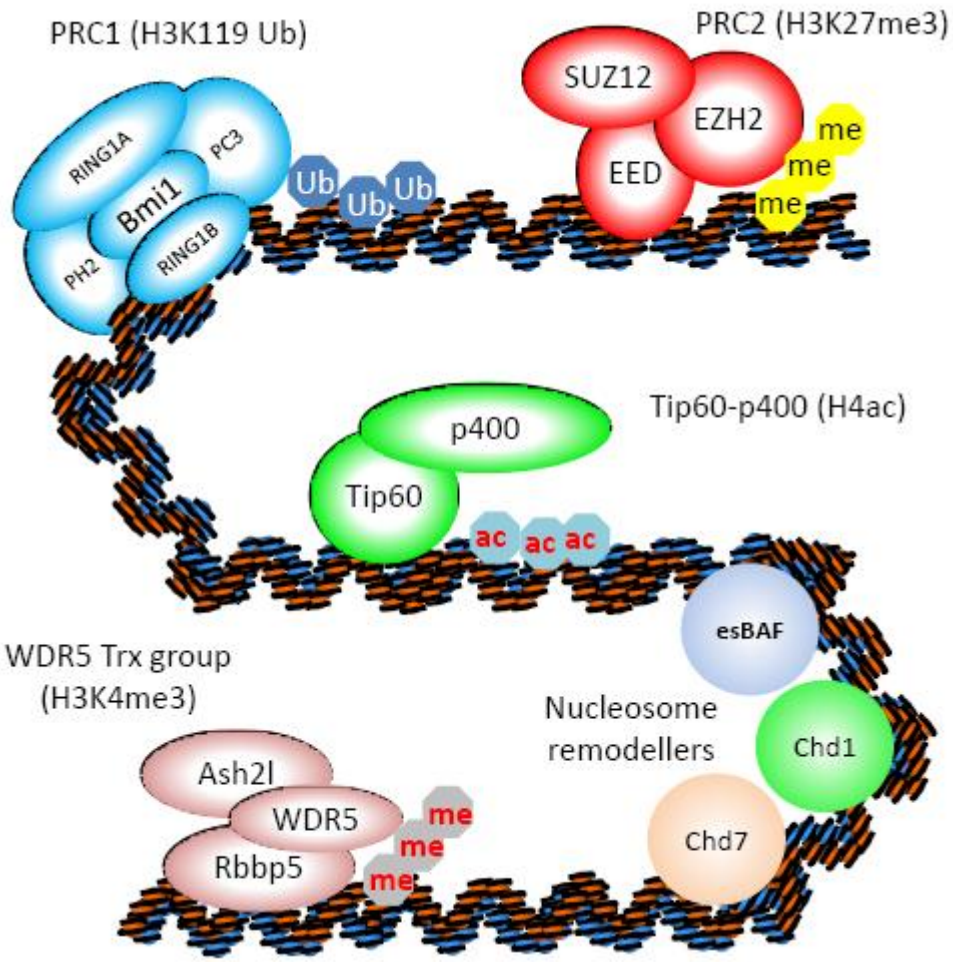


Illustration 5: Chromatin modifiers in ESC regulation. Shown are the various chromatin modifying complexes that have been reported to play a role in maintaining the mouse ESC state. Chromatin modifiers are responsible for specific chromatin modifications which defines the epigenetic status of the ESC genome. PRC1 is responsible for the Ubiquitination of histone 3 lysine 119 (Ub:Ubiquitin), PRC2 tri-methylates the histone 3 lysine 27 (me: methyl group), Tip60/p400 complex acetylates histone 4 (ac: acetylation) and WDR5 Trithorax (Trx) group tri-methylates histone 3 lysine 4. esBAF, Chd1 and Ch7 are chromatin remodelers that regulates the chromatin accessibility.

poised state that allows rapid activation during differentiation. A similar histone modification found to play a repressive role in ESC is the histone variant H2AZ. Creyghton and colleagues found that H2AZ localization across the mouse ESC genome highly overlaps with the repressive H3K27me3 mark (Creyghton *et al*, 2008). H2AZ is essential for the repression of lineage genes in ESC and depletion of H2AZ impedes proper ESC differentiation. This histone variant appears to be found on repressive genes only in mouse ESC and is reported to mark active genes when ESC differentiates to neural progenitors (Creyghton *et al*, 2008).

In summary, for the past three decades, extensive efforts have been devoted in dissecting the ESC pluripotency network since the isolation of the mouse ESC. The advent of technological breakthroughs have greatly assisted the rapid dissection of the pluripotency mechanism; ChIP coupled with high throughput sequencing (Johnson *et al*, 2007) has allowed the extensive genome-wide mapping studies of transcription factors, chromatin modifiers and histone modifications, and RNA mediated interference (RNAi) (Fire *et al*, 1998) has greatly facilitated the rapid screening for pluripotency genes. Since discovery, the genome-wide binding profile of more than 50 transcription factors (a reserved estimation) have been generated (Kim *et al*, 2008; Chen *et al*, 2008; Nishiyama *et al*, 2009) and three separate genome-wide RNAi screens have been carried out (Ding *et al*, 2009; Hu *et al*, 2009; Kagey *et al*, 2010) in the mouse ESC. Most of the above studies discussed have been performed in the mouse ESC and awaits verification in human ESC. A wealth of knowledge has been generated for understanding how pluripotency is maintained in mouse ESC, however much less has been accomplished in dissecting the pluripotency network that governs the human ESC state.

1.5 Dissecting the human ESC transcriptional regulatory network

Intrinsic differences between mouse and human ESC

The need to perform parallel studies in human ESC stems from the intrinsic differences observed between the two species of ESC (Table 2). Human ESC isolated under seemingly identical conditions as of mouse ESC appears much different in morphology (Thomson *et al*, 1998). Human ESC compact into flat colonies rather than the dome-shaped colonies formed by mouse ESC. Human ESC also stains for a different surface marker SSEA4 rather than SSEA1 found in mouse ESC surface. Subsequent studies on the signalling requirements in both ESC revealed that human ESC maintenance is dependent on the FGF and TGFβ/Activin pathway (Levenstein *et al*, 2005; Greber *et al*, 2007; Beattie *et al*, 2005; James *et al*, 2005) whereas mouse ESC requires Lif and BMP4 signalling (Smith *et al*, 1988; Okada *et al*, 1988; Ying *et al*, 2003). Another interesting difference observed between human and mouse ESC is the expanded potency of human ESC to differentiate into the trophoblast lineage (Xu *et al*, 2002). Contrastingly, the differentiation of human ESC to trophoblastic cells is dependent on the BMP4 signalling which is important for mouse ESC maintenance. In culturing the ESC,

Table 2. Differences between human and mouse ESC

Characteristic	Human ESC	Mouse ESC
Colony morphology	Flat colonies	Dome shape colonies
Signaling dependence	bFGF and TGFβ/Activin	Lif and BMP4
Surface marker	SSEA-1 –ve, SSEA4 +ve	SSEA-1 +ve, SSEA4 –ve
Differentiation capacity	3 embryonic germ layers and trophoblast	3 embryonic germ layers
Clonal efficiency	Low; passage in clumps	High; passage as single cell

human ESC are passage in clumps with low clonal efficiency whereas mouse ESC can be passage as single cells with high clonal efficiency. Moreover, as highlighted in the earlier sections, the core regulatory network in both ESC overlaps poorly and there is evidence of evolutionary differences inherited by activities of transposon elements in the respective species. While researchers argued for a species specific difference between the two ESC, the isolation of a pluripotent mouse cell line under human ESC signalling conditions suggest otherwise. These cells first isolated from late post-implanted blastocyst are termed epiblast stem cells (EpiSC) and they display various hallmarks of human ESC (Tesar *et al*, 2007; Brons *et al*, 2007). However, the overlap between the Oct4 regulatory network in human ESC and mouse EpiSC is less than 20%, suggesting intrinsic differences between the 2 cell types. Further studies on FGF signalling in human ESC and mouse EpiSC reveals different requirements of the pathway. While FGF signalling in human ESC modulates *Nanog* and *Oct4* expression, this is not observed in the mouse EpiSC (Greber *et al*, 2010). All these differences observed between the mouse and human pluripotent cells calls for the need to directly dissect the human pluripotency network.

Identifying novel human ESC determinants

Chia *et al* performed the first genome-wide RNAi screen in human ESC to identify potential determinants of the human ESC identity (Chia *et al*, 2010). Using an *Oct4 promoter* driven *eGFP* reporter, the screen aims to identify potential pluripotency regulators with the *OCT4* expression level as a read out. The study unravelled many novel candidate transcription factors, signalling pathways and protein complexes that potentially play a role towards the maintenance of the human ESC state. One such transcription factor identified in the study is the PR domain zinc finger protein 14 (PRDM14), which is in the top 10 candidate genes

identified. Depletion of PRDM14 resulted in a down-regulation of pluripotency marker OCT4 and loss of pluripotency surface marker Tra-1-60, Tra-1-81 and SSEA4 in multiple human ESC lines (Chia *et al*, 2010), suggesting that PRDM14 is an important determinant of the human ESC identity. Moreover, it was also shown that PRDM14 is able to enhance the efficiency of reprogramming fibroblasts back to a pluripotent state (Chia *et al*, 2010).

Known functions of PRDM14 in mouse and human

Prdm14 belongs to the PR domain containing family of transcription factors which are increasingly implicated in cell fate regulation and various diseases (Fog *et al*, 2011); PRDM1 has been shown to be a critical regulator for germ cell development (Ohinata et al, 2005) and PRDM16 is a critical regulator of brown adipocyte specification (Seale *et al*, 2009). Both PRDM3 and PRDM16 have been implicated to be potential oncogenes in leukemias (Wieser *et al*, 2007; Shing *et al*, 2008). On the other hand, PRDM1 is a tumor suppressor in large B cell lymphoma (Mandelbaum *et al* 2010, Calado *et al* 2010). The family of transcription factors is identified by 2 conserved features, a PR domain which is structurally similar to the SET domain (PRDM is a subfamily of the SET domain family of histone methyltransferases) and variable numbers of DNA binding Zinc fingers. However, unlike most SET domain containing histone methyltransferases (HMTases), only PRDM2, PRDM8 and PRDM9 have been shown to exhibit enzymatic activities (Kim *et al* 2003, Eom *et al* 2009, Hayashi *et al* 2005). However, other members of the PRDM family are also involved in epigenetic regulation of gene expression through their interactions with other histone modifying enzymes such as G9a, histone deacetylases and histone acetylase p300 (Fog *et al*, 2011). In addition, the multiple C2H2 type Zinc fingers of each PRDM allows interaction with other

Zinc finger containing proteins (Brayer KJ and Segal DJ, 2008), increasing the versatility and complexity of transcription regulation by the PRDM family of transcription factors.

Prdm14 was first uncovered as a critical germ cell factor, important for establishing germ cells during early development (Yamaji *et al*, 2008). Prdm14 null mice are infertile with the loss of germ cells in the gonads. Molecular characterization of PGC-like cells in the Prdm14 null mice reveals that Prdm14 is important for the reacquisition of pluripotency and epigenetic reprogramming during PGC establishment. This includes the reactivation of *Sox2* expression, and the genome-wide loss of H3K9me2 and gain of H3K27me3 modifications. Although the mouse Prdm14 has not been reported to exhibit enzymatic activity, Yamaji *et al* showed using the GAL4 fusion reporter system that Prdm14 exhibits repressive activity. Similar to Prdm1, another critical regulator of the PGC, PRDM14 is highly expressed in the mouse ESC. However, neither Prdm1 nor Prdm14 have been reported to play a role in maintaining mouse ESC identity.

In human, PRDM14 was first found to be ectopically expressed in cancerous tissue. PRDM14 was found to be a potential proto-oncogene in breast cancer (Nishikawa *et al*, 2007). PRDM14 gene was amplified in two thirds of the tissue examined and the depletion of PRDM14 in these cancer cells induced apoptosis and increased their susceptibility to chemotherapeutic drugs. In a separate study, PRDM14 was also implicated to be a proto-oncogene in lymphomas (Dettman *et al*, 2008). The findings also highlighted that PRDM14 is specifically found in early embryonic tissues up to E6.5 and in ESC cultures. Importantly, tumours found to be ectopically expressing high levels of PRDM14 also showed upregulation of *OCT4* expression, suggesting a role of PRDM14 in pluripotency. Indeed, Tsuneyoshi *et al*

had previously shown that PRDM14 in human ESCs potentially suppresses the expression of differentiation genes, and that overexpression of PRDM14 in human ESC derived embryoid bodies will prevent the upregulation of differentiation genes. While all these phenotypes suggest potential roles of PRDM14 in regulating gene expression in human ESC and cancer cells, the mechanism in which PRDM14 functions remains largely unknown.

1.6 Dissecting the functional roles of PRDM14 in human ESC

This project aims to dissect the mechanism in which PRDM14 maintains the human ESC identity. This is first achieved by mapping the genome-wide binding profile of PRDM14 in human ESC, which will reveal what genes are bound by PRDM14 across the entire human ESC genome. Coupled with expression profiling, potential target genes in which PRDM14 regulates could be determined. In addition, co-analysis with available binding profiles of transcription factors and histone modifications would further reveal how PRDM14 regulates these genes. For instance, the overlap with OCT4 and NANOG binding sites would reveal if the PRDM14 regulatory node is integrated into the core regulatory network. In addition to the mapping studies, biochemical assays are employed to identify the interactive partners of PRDM14 in human ESC. The complementary approaches would reveal how PRDM14 functions to maintain the ESC state. Through the understanding of how PRDM14 functions in regulating pluripotency, this study aims to 1) Increase our knowledge on how human ESC pluripotency is regulated by transcription factors and expand the human ESC transcriptional network, 2) Reveals how different transcription factors and chromatin modifiers can work together to maintain the human ESC state and 3) Identify potential regulatory nodes of the pluripotency network that could be useful in devising better strategy to shut down the network for ESC differentiation or to induce pluripotency in somatic cells.

Chapter 2: MATERIALS AND METHODS

2.1 Cell culture

The human ESC line H1, HES2 (ES-02,), HES3 (ES-03) (WiCell, Madison, Wisconsin, USA) were used for this study. The cells were maintained as feeder-free culture on matri-gel (BD) with conditioned medium. The medium containing 20% KO serum replacement, 1 mM L-glutamine, 1% non-essential amino acids, 0.1 mM 2-mercaptoethanol and 4ng/ml basic fibroblast growth factor (Invitrogen) in DMEM/F12 (GIBCO) was conditioned with mitotically inactivated mouse embryonic fibroblast (MEF) for 24hrs. Additional 8ng/ml of basic fibroblast growth factor (Invitrogen) is supplemented to the conditioned medium before use. The medium was refreshed daily. The human ESC was passaged upon confluency every 5-7 days. Briefly, for a single 6 well of human ESC, the cells were wash with 1XPBS (GIBCO) and dissocated by treatment with 1ml of 1mg/ml collagenase IV (GIBCO) for 6-7mins at 37°C. The collagenase were aspirated and cells were washed twice with 1XPBS before fresh conditioned medium was added. The cells were detached with a cell scraper and colonies were mechanically broken down to large clumps containing 50-100 cells using a 1ml pipette. Cells were passaged at a 1:8 or 1:12 ratio. For transfection, the human ESC were dissociated to single cells with 0.05% Trypsin EDTA (GIBCO). Cells were treated with the trypsin for 1min at 37°C and the trypsin was aspirated. Cells were washed twice with 1XPBS before fresh medium containing 0.5μM of Thiazovivin (STEMGENT) was added. The cells were mechanically dissociated to single cells with a 1ml pipette. E14 mouse ESC were cultured feeder-free on gelatin in DMEM (GIBCO) supplemented with 15% heat-inactivated fetal bovine serum (GIBCO), 1 mM L-glutamine, 1% non-essential amino acids, 0.1 mM 2-mercaptoethanol and 1,000 units/ml of LIF (Chemicon). E14 mouse ESC were passaged every 2-3 days upon confluency. Human embryonic kidney 293T cell and human embryonic

lung fibroblast MRC-5 (ATCC) were cultured in DMEM with 15% FBS. Mouse ESC, 293T and MRC-5 were passage upon confluency with 0.125% trypsin (GIBCO). When the cells were detached from the plates and dissociated to single cells, the trypsin was inactivated with the respective serum containing medium. Cells were passaged at a 1:6 ratio for E14 and 293T and a 1:4 ratio for MRC-5. All cell cultures were maintained at 37°C with 5% CO_2.

2.2 RNA extraction, reverse transcription and quantitative real-time PCR

For expression analysis, total RNA from the cells was extract using the TRIzol reagent (Invitrogen). Total RNA and DNA were extracted with chloroform and precipitated with isopropanol. The precipitated RNA and DNA were centrifuged at 13000rpm at 4°C for 10mins. The RNA and DNA pellet was washed with 70% ethanol and subsequently reconstituted with DEPC-treated water (AMBION). DNA contaminants were digested with DNASE I (Ambion) at 37°C for 30 min. The DNASE I enzyme is heat inactivated at 70 °C for 10 mins. Concentration of the RNA was determine with the NanoDrop 2000 (Thermo Scientific). 500ng of RNA was used for each reverse transcription reaction using the SuperScript II Kit (Invitrogen) to produce the cDNA for subsequent quantitative assays. Quantitative real-time PCR (qPCR) analysis was performed with the SYBR Green Master Mix (KAPA) using the ABI PRISM 7900 sequence detection system.

2.3 Chromatin immunoprecipitation (ChIP)

ChIP was performed using 10e7-10e8 cells per reaction. Crosslinking of protein and DNA in the cells were achieved by treatment with serum free media containing 1% formaldehyde at room temperature for 10 minutes. The crosslinking reaction is quenched with 0.2M final

concentration of glycine for 5 mins. The cells on the plates were scraped and washed in Tris-EDTA buffer before treatment with the cell lysis buffer containing protease inhibitors. The cells were lysed with 0.1% SDS buffers and the nuclei were lysed with 1% SDS buffers. Pelleted chromatin were resuspended in 0.1% SDS buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.1% SDS, 1 mM EDTA) and fragmented via sonication with a BioRuptor (Diagenode) for 16 cycles of 30 seconds with 1 min interval at 40% amplitude on ice. The chromatin sample is precleared at 20,000rpm for 45mins and the chromatin solution was collected for ChIP. ChIP was performed with 5μg of the respective antibodies immobilized to Magnetic Protein A beads (Invitrogen). The Chromatin was incubated with the immobilized antibody overnight. The beads subsequently subjected to multiple wash steps; 3 times with the IP (0.1%SDS) buffer, once with 1M NaCl in the same buffer, once with LiCl buffer and once with Tris-EDTA buffer. The DNA-protein complexes were eluted from the beads with a buffer containing 50 mM Tris-Cl (pH 7.5), 10 mM EDTA, 1% SDS at 69°C for 30mins in an orbital shaker at 14000 rpm. The protein-DNA complex is dissociated with 1.5μg/ml Pronase (Roche) at 42°C for 2 hr followed by 67°C for 6 hr. Human ESC transfected with shRNA targeting *PRDM14, NANOG* or *GFP* was harvested for ChIP 36-40 hours post transfection. Human ESC transfected with HA or HA-PRDM14 overexpression construct was harvested for ChIP 48 hours post transfection. Antibodies used in the ChIP assays include PRDM14 antibody (custom-made), NANOG antibody (R&D), SUZ12 and EZH2 antibody (Active motif), H3K27me3 antibody (Millipore) and histone H3 antibody (Santa Cruz). To generate the ChIP-Seq library, 5-15ng of ChIP enriched DNA was modified with the ChIP-Seq DNA Sample Prep Kit (IP-102-1001, Illumina). Briefly, the DNA end was repaired with exonucleases and an adaptor was ligated to the end, followed by PCR amplification for 15 rounds. The amplified DNA was gel purified and the 200-300bp fragments were selected for subsequent SOLEXA sequencing (Illumina). Enriched DNA from other ChIP assays were

quantitated using qPCR. All values in ChIP-qPCR results presented in this study are means ± s.e.m from 3 independent experiments (n=3). The student t-test was used in the statistical analysis of the results. The results are considered significant if the p value < 0.05.

2.4 Protein expression and purification

cDNA encoding different recombinant proteins used in this study was cloned into the pET42b (Novagen) vector. To generate His-tag only recombinant protein, the cDNA was cloned into the vector using the NdeI and XhoI cloning sites. To generate GST and His tagged proteins, the cloning sites SpeI and XhoI was used. The fusion constructs were transformed into BL21 competent cells (Strategene) according to the supplier's instructions. A 50ml starter cultured was grown overnight at 37°C. 8 to 10ml of the culture was transferred to 600ml of fresh LB/Kanamycin/chloroamphenicol medium and incubated at room temperature until the OD595 read out was 0.5-0.6. Protein expression was then induced with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for another 4-6hrs. Cell pellets were harvested and stored at -80°C or resuspended in 8ml of lysis buffer (50mM Tris pH8, 10% glycerol, 0.3M NaCl, 2mM EDTA and 0.1% TritonX-100) for sonication on ice at 30% amplitude, for 15 times at 2mins interval. For GST-tag purification, soluble lysates were collected and purified with Glutathione (GSH)-sepharose beads (Amersham) in columns. Proteins are retrieved from the columns with the elution buffer (50mM Tris pH8, 10% glycerol, 0.3M NaCl, 2mM EDTA and 0.1% Triton X-100) containing 2mM of GSH. For His-tagged purification, soluble lysates were collected and purified with Nickel-NTA beads (Qiagen) in columns. The protein bound beads were first wash with a buffer containing 20mM HEPES pH7.5, 10% glycerol, 0.3M NaCl and 0.2% Triton-X-100 and subsequently at a higher salt concentration of 1M NaCl. Proteins are retrieved from the columns with the

0.3M NaCl wash buffer containing 0.35M of imidazole. Eluents were dialyzed against a dialysis buffer (10mM Tris–HCl, pH7.4, 100mM NaCl, 10mM ZnCl2 and 10% glycerol) at 4°C for 6hr. Proteins were stored at -80°C in aliquots before use. Protein concentrations were measured with the BIORAD protein measurement assay.

2.5 Electrophoretic mobility shift assay

Recombinant PRDM14 DNA binding domain (His tagged) was used in the gel shift assays. Briefly, a cDNA encoding 179 amino acids of the C terminus of PRDM14 was cloned into the pET42b (Novagen) vector using cloning sites NdeI and XhoI. For GST tagged proteins, cDNA were cloned into the vector with cloning sites SpeI and XhoI. The fusion construct was transformed into BL21 competent cells (Strategene) according to the supplier's instructions. Purified proteins were dialyzed against a dialysis buffer (10mM Tris–HCl, pH7.4, 100mM NaCl, 10mM ZnCl2 and 10% glycerol) at 4°C for 6 hr. DNA probes are generated using oligonucleotides (Proligo) labelled with biotin at the 5' end of the sense strands. The labelled sense strands were incubated with the antisense strands in the annealing buffer (10mM Tris-HCl, pH8.0, 50mM NaCl, 1mM EDTA) for 5 mins at 95°C and 1hr at room temperature. The annealed probes were purified with the DNA extraction kit (Qiagen). DNA concentrations were measure with the NanoDrop ND-1000 spectrophotometer. The gel shift assays were performed using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnologies). 20-100ng of proteins were added to a 5µl reaction mixture (final) containing 1µg of poly(dI-dC) (Amersham), 1ng of biotin-labeled oligonucleotide in the binding buffer (12mM HEPES, pH7.9, 10% glycerol, 60mM KCl, 0.25mM EDTA, 1mM DTT, 10mM $ZnCl_2$). Reaction mixtures were incubated for 20min at room temperature and resolved on pre-run 6% native polyacrylamide gels in 0.5X Tris-buffered EDTA. Gels were

transferred to Biodyne B nylon membranes (Pierce Biotechnologies) using western blot techniques and detected using chemiluminescence.

2.6 Co-immunoprecipitation

The human ESC and transfected 293T cells were lysed with a lysis buffer (50mM Tris-HCl, pH 8.0, 150 mM NaCl, 10µM ZnCl, 0.5% Nonidet P-40, 5% glycerol with protease inhibitor) for 30mins at 4°C. The whole cell lysate was precleared, collected and incubated overnight with protein G beads coated with 5ug of the respective antibodies at 4°C. The beads were washed 4-6 times with the cell lysis buffer and boiled 10mins for elution. The interacting protein bands are resolved with 10% SDS-PAGE gel and transferred to the PVDF membrane, followed by detection with an appropriate primary antibody, a HRP-conjugated second antibody, and an ECL reagent. Anti-NANOG antibody (R&D), anti-GST (Santa Cruz), anti-PRDM14 (custom-made [18]), anti-HA (sc-7392, Santa Cruz), anti-cMyc (sc-40, Santa Cruz) and anti-EZH2 (Active motif) antibodies were used to pull-down the protein complexes.

2.7 Glutathione S-Transferase pull-down assay

The Baculovirus system was used to obtain the PRC2 complex consisting of 4 subunits – Enhancer of Zeste homolog (EZH2), Suppressor of Zeste 12 homolog (SUZ12), PHD finger protein 1(PHF1) and embryonic ectoderm development protein (EED).The Baculovirus constructs of the four members of the PRC2 complex (EZH2, SUZ12, PHF1 and Flag- EED) were transfected into Sf9 insect cells individually and low-titer virus stocks were obtained. Initial low-titer virus stocks (P1) were amplified to subsequently produce P2 and then P3 stocks. The final P3 virus stock was assayed for the virus titer by the plaque assay method.

All P3 stocks obtained were 1-2 x10^8 pfu/ml. The virus stocks were then used to infect Sf9 cells at a multiplicity of infection (MOI) of 10 and at 48h post-infection, cells were harvested. They were then lysed and the PRC2 complex was immuneprecipitated using an anti-Flag antibody. Full length *PRDM14* was cloned into pet42b (Novagen) and expressed in BL21 *Eshcerichia coli*. The GST tagged PRDM14 protein was first purified using glutathione (GSH)-Sepharose beads (Amersham) followed by nickel-nitrilotriacetic acid (Ni-NTA) beads (Qiagen). The purified PRDM14 were bound to the GSH beads and incubated with the purified FLAG tagged PRC2 complex and *vice versa* purified PRC2 complex is bound to the FLAG beads and incubated with the purified PRDM14, for 2 hours at 4°C. The beads were washed 4-6 times and eluents were analysed by western blot.

2.8 Informatics analysis

Peak calling for ChIP-Seq was done with MACS (Zhang *et al*, 2009). All peak calls that could be identified in the control were removed from the results (Quinlan and Hall, 2010). Peaks were associated with their nearest transcription start site using PeakAnalyzer. The enriched sequence motifs were identified using de novo motif discovery programs Weeder, MEME (Bailey *et al*, 2009) and CisFinder (Sharov and Ko, 2009). ChIP-Seq enrichment plots were created by extending the genomic loci of interest to a range of 4000bp. SamTools (Li *et al*, 2009) was used to extract the tag number at every position. The plots show the average number of reads (depth) for all positions (+-2000bp around the center) over all genomic loci of interest, using a sliding window of size 30. ChIP-Seq data was downloaded from the European Nucleotide Archive. Class I and Class II elements were obtained from Rada-Iglesias *et al*, 2011. The reads were mapped to the human genome (hg19) using Bowtie (0.12.5) (Langmead *et al*, 2009). Clustering of ChIP-Seq experiments was done on all

PRDM14 peaks. The sum of tags (+-2000bp around the peak) was first calculated for all peaks and all ChIP-Seq experiments. Pearson correlation was calculated with the log-transformed sums. The heatmap was created on the matrix of all pairwise correlation coefficients using R. Gene expression data was obtained from (GSE22792). Gene Set Enrichment Analysis (GSEA) was performed using the GSEA software from Subramanian *et al*, 2005 and Mootha *et al*, 2003. Gene expression data four days after knockdown of PRDM14 was used. All genes near the 500 most significant PRDM14 binding sites were chosen as the gene set for which the enrichment score was calculated. GO analysis was performed with Panther classification (www.pantherdb.org).

2.9 Short hairpin RNA-mediated knockdown

The pSuper.Puro plasmid (Oligoengine) was used for the RNAi-mediated knockdown. Single strand DNA oligonucleotides containing the forward and reverse target nucleotide sequence was synthesized (Sigma) and annealed in the annealing buffer (10mM Tris-HCl, pH8.0, 50mM NaCl, 1mM EDTA) for 5 mins at 95°C and 1hr at room temperature. The annealed oligos were phosphorylated with the T4 Polynucleotide Kinase (NEB) before ligation with the linearized pSuper.Puro vectors. The nucleotide sequence of the Short hairpin RNA (shRNA) constructs are GAACGGCATCAAGGTGAAC (*GFP*), GATGAAATGGGTAAGTACA (*LUCIFERASE*), GGAGACTGCTATGAGAAAT (*PRDM14*) and GCAAAGCACAGTGCAACAC (*EZH2*). Human ESC were transfected with knockdown constructs using *Trans*IT (Mirus) following the manufacturers protocol. Cells were dissociated to small clumps using 0.05% trypsin-EDTA (GIBCO) and seeded at 1:12 ratio 16-24hr before transfection.

2.10 Luciferase assay

A minimal *POU5F1* proximal promoter region (350bp) was cloned into the PGL3 basic vector (Promega), driving the luciferase gene via the cloning site BglII and NcoI. The CR2 and CR4 fragments (550 and 500bp, respectively) were cloned into the PGL3-POU5F1 proximal promoter vector downstream of the luciferase gene via the cloning sites BamHI and SalI. For the cloning of reporter vector used to test the functional domains of PRDM14, 3 copies of 30bp and 50bp CR2 consensus motif were synthesized and cloned into XhoI and BglII site in front of the minimal promoter of pGL4.23 vector (Promega) in tandem. For the cloning of the vectors used in the repressive activity studies, the 500bp intron of *HES7* and 16bp PRDM14 consensus motif were cloned into the XhoI and BglII site in front of the phosphoglycerate kinase (*PGK*) promoter driven *Luciferase* gene in pGL4.23 vector. Primer used for cloning genomic *HES7* DNA fragment are forward primer GGACCAGGTCAGTCCCTCCGC and reverse primer ATCGCATTTGCGCACTGCCCA, for genomic CR2 fragment are forward primer CCCCACTAAACAAAGCACATC and reverse primer CAGAGTGATAAGACACCCGCT and for genomic CR4 are forward primer TGGGCCTCTTGCCCAGGCCAG and reverse primer TATGTTGCCTCTGTTCGTGTG. H1, HES2 and HES3 human ESCs were transfected with the reporter constructs using Fugene (Roche) or *Trans*IT (Mirus) and E14 mouse ESCs and 293T cells using Lipofectamine 2000 (Invitrogen). Cells were harvested 48-60hrs after transfection and the luciferase activities were quantified using the Dual-luciferase Reporter Assay System (Promega). A *Renilla* luciferase plasmid pRL-SV40 (Promega) was co-transfected as an internal control. Cells were harvested 48hrs post transfection and the luciferase activity of the cell lysate was analysed using the Dual-Luciferase Reporter Assay System (Promega). All values in the reporter assays conducted in this study are means ± s.e.m from 3 independent experiments (n=3).

2.11 Retroviral production and iPSC reprogramming

pMXs retroviral plasmids containing the human *OCT4*, S*OX2*, *KLF4* and *c-MYC* cDNA were obtained from Addgene (plasmids 17217, 17218, 17219 and 17220. S. Yamanaka). The pMXs retroviral plasmids containing cDNA of human PRDM14 gene was obtain as described previously (Chia *et al*, 2010). Retroviruses were packaged using Pantropic Retroviral Expression System (Clontech) in gp293T cells. Briefly, the plasmids with the cDNA of the reprogramming factors were co-transfected with a plasmid expressing the VSV-G protein into the gp293T cells using lipofectamine 2000 (invitrogen). The medium was changed to a 5%FBS containing medium 6 hours post transfection. The virus containing medium was collected 16-24 hrs post transfection and concentrated with centrifugal filter devices (Millipore). The concentrated viruses are stock in aliquots at -80°C. Virus collected from each 15cm dish of transfected gp293T (each reprogramming factors) are sufficient for 8-10 reprogramming experiments. For reprogramming, a single 6well of confluent MRC-5 cells were split into 5 24wells 16-24hrs before and transduced with the retrovirus stock in the presence of 4-6 μg/ml polybrene (Sigma). The virus medium was removed 16-24hr post infection and the cells were passage from one 24-well into two 6-wells pre-seeded with CF-1 feeders upon confluency. The medium was switch to human ESC conditioned medium 1 day after passage on the CF-1 feeders. The medium was changed every 2 days for 2 weeks before human iPSC colonies begin to appear in culture. Human iPSC colonies for each reprogramming experiments were quantitated 3 weeks post infection.

Chapter 3: PRDM14 transcriptional network integrates with the core regulators

3.1 Mapping the PRDM14 binding sites in human ESC

To map the global binding profile of endogenous PRDM14 in human ESC, an antibody was first raised against a PRDM14 antigen. The PRDM14 protein is 571 amino acids (a.a) in size and consists of mainly three domains; the N terminus containing the PR domain, the SET domain and 6 C2H2 type zinc finger (Znf) domains (Figure 1A). A recombinant protein containing the less conserved N terminal domain (220 a.a) was generated and used as the antigen for generating a PRDM14 specific antibody. Since *PRDM14* expression is specific to human ESC (Dettman *et al*, 2008), the specificity of the antibody was first tested with whole cell lysate of differentiated human ESC under various culture conditions and undifferentiated human ESC. The antibody generated from this antigen detected a specific protein band of ~68 kDa in size (approximate size of PRDM14 ~ 65kDa) and this band is absent for the samples of differentiated of human ESC (Figure 1B). The specificity of the PRDM14 antibody was further verified with PRDM14-depleted human ESC (Figure 1C). PRDM14 protein was hardly detected in the 4 PRDM14 depleted samples and along with the depletion of PRDM14, NANOG and OCT4 expression similarly decreased. Thus, the antibody raised is specific for PRDM14 and is used for the subsequent ChIP assay.

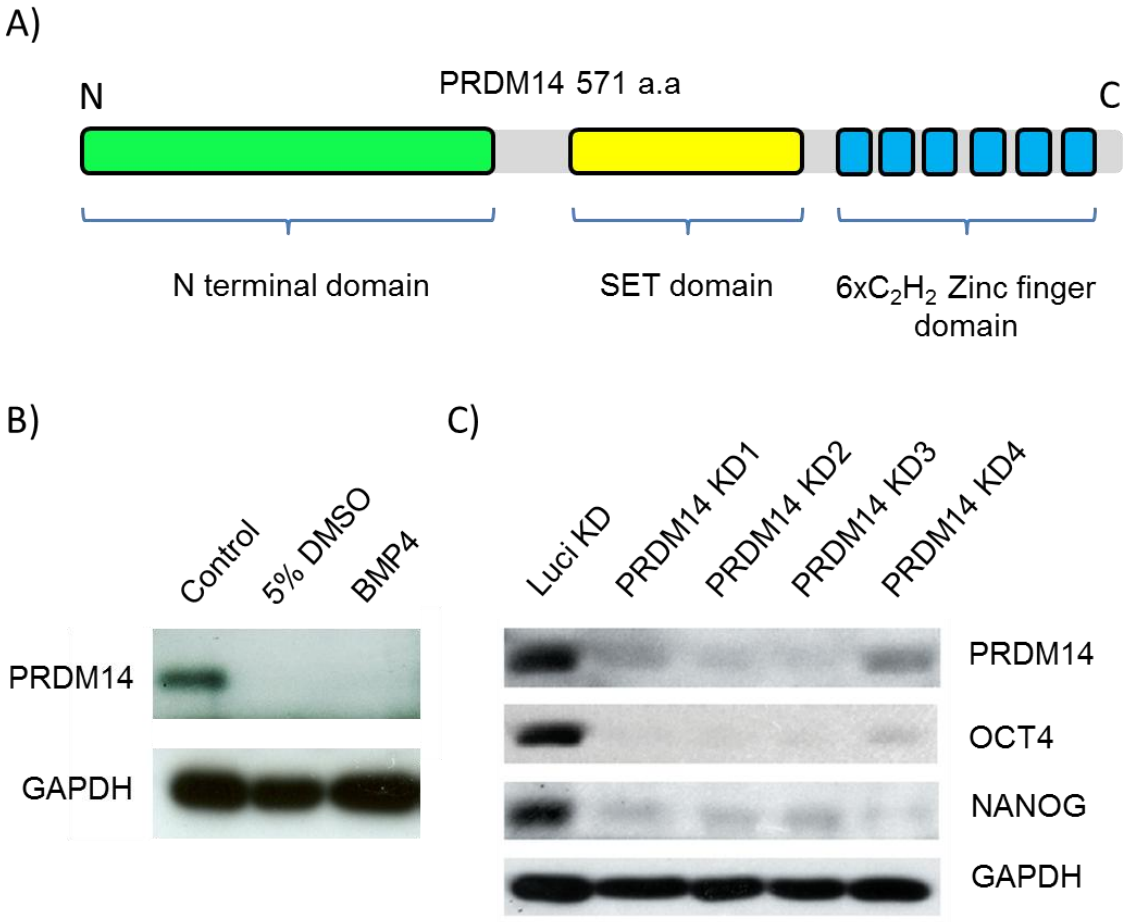


Figure 1. A human PRDM14 specific antibody. A) A schematic representation of the major domains in the PRDM14 protein. An antibody was raised against an antigen containing the first 220 amino acids of the N terminal domain. B) The specificity of the antibody was tested with DMSO and BMP4 differentiated ESC whole cell lysate. A specific PRDM14 band was only detect in the undifferentiated ESC sample (control). C) Whole cell lysates of PRDM14-depleted human ESC (using various PRDM14 targeting shRNA constructs) were used to further validate the antibody specificity. Control samples are transfected with shRNA targeting *GFP*. Corresponding to the loss of pluripotency after PRDM14 depletion, OCT4 and NANOG were also decreased in the PRDM14 depleted samples. Level of GAPDH proteins served as the loading control.

To generate an unbiased map of PRDM14 binding across the human ESC genome, high throughput sequencing was engaged for quantifying the PRDM14 ChIP sample. The ChIP-Seq process is outlined in Illustration 6; briefly, PRDM14 was first cross-linked to the chromatin in the native cells. Cells were subsequently lysed and the harvested chromatin was sheared via sonication. The PRDM14 bound chromatin is enriched using the PRDM14 specific antibody. The PRDM14/chromatin complex is separated and the enriched DNA is modified for the subsequent sequencing process (Illustration 6). The PRDM14 ChIP-Seq generated a large number of potential PRDM14 binding sites in human ESC. Before further analysis of the global binding patterns of PRDM14, there is a need to verify if these enriched sequencing reads are *bona fide* PRDM14 binding sites. For the validation assays, 20 potential PRDM14 sites with various PRDM14 binding strengths (define by the number of binding peaks detected in the sequencing) were selected to test for PRDM14 binding. A HA-tagged PRDM14 was ectopically expressed in the human ESC and cells were harvested for ChIP assay with HA antibody. Specific ChIP enrichment for the 20 loci was observed for the HA-ChIP with cells overexpressing PRDM14 and not the control vector (Figure 2A). This enrichment was also not observed at the 2 control binding sites (no PRDM14 binding detected in ChIP-Seq). To further verify that these enrichments are specific for PRDM14, we depleted PRDM14 in human ESC using shRNA and harvested the cells for ChIP. The enrichment by the PRDM14 specific antibody is strongly reduced in the PRDM14 depleted sample compared to the control sample (Figure 2B). In summary, the PRDM14 binding peaks generated by ChIP-Seq using the PRDM14 antibody is specific and responsive to PRDM14 levels.

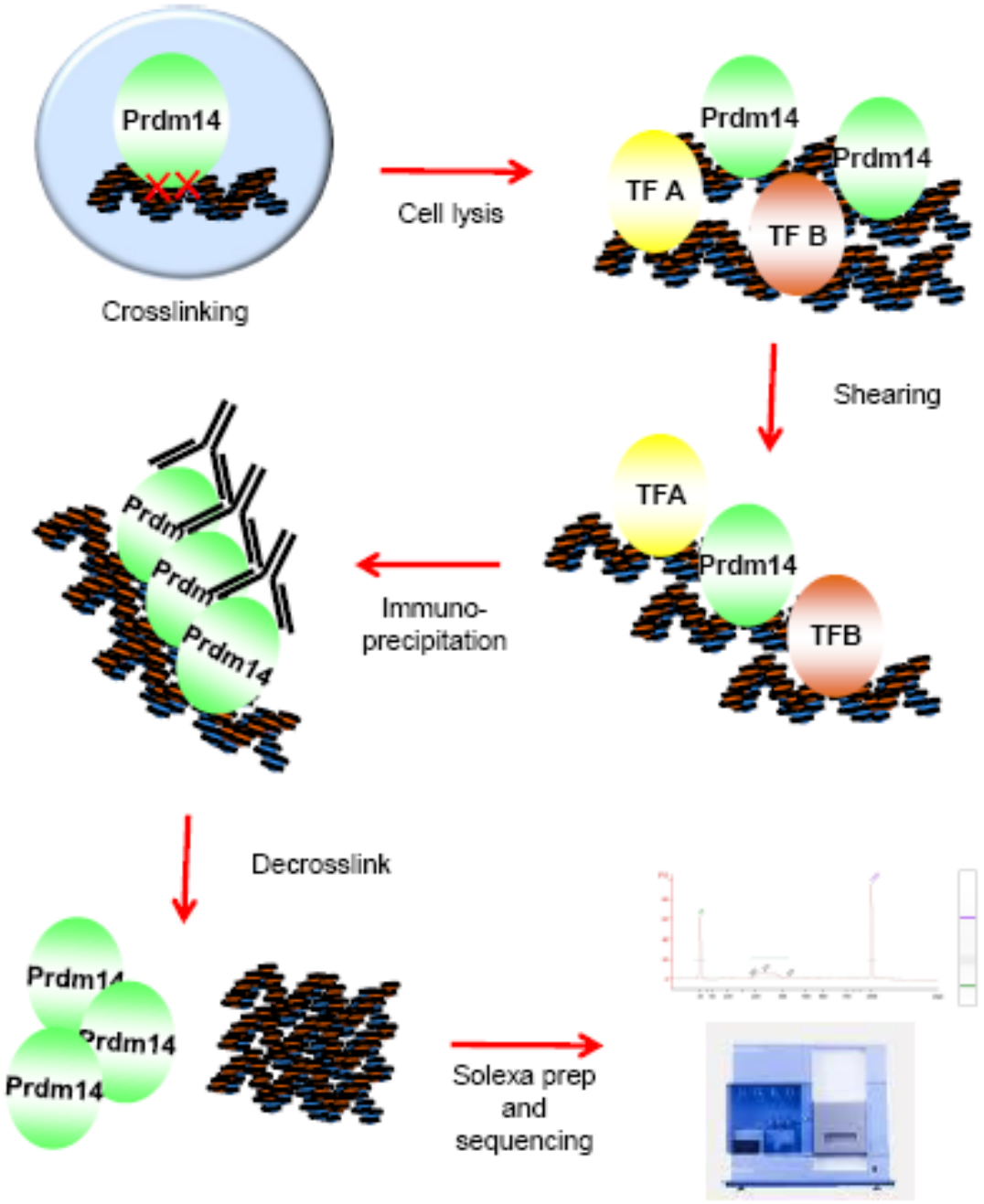


Illustration 6: PRDM14 ChIP-Seq outline. Briefly, Cells were crosslinked with formaldehyde and lysed in SDS buffers and the chromatin harvested was shear via sonication. PRDM14-bound chromatins were enriched with immunoprecipitation using the PRDM14 specific antibody. The enriched DNA was enzymatically decrosslinked from the protein and modified for subsequent sequencing using the SOLEXA high throughput sequencing platform.

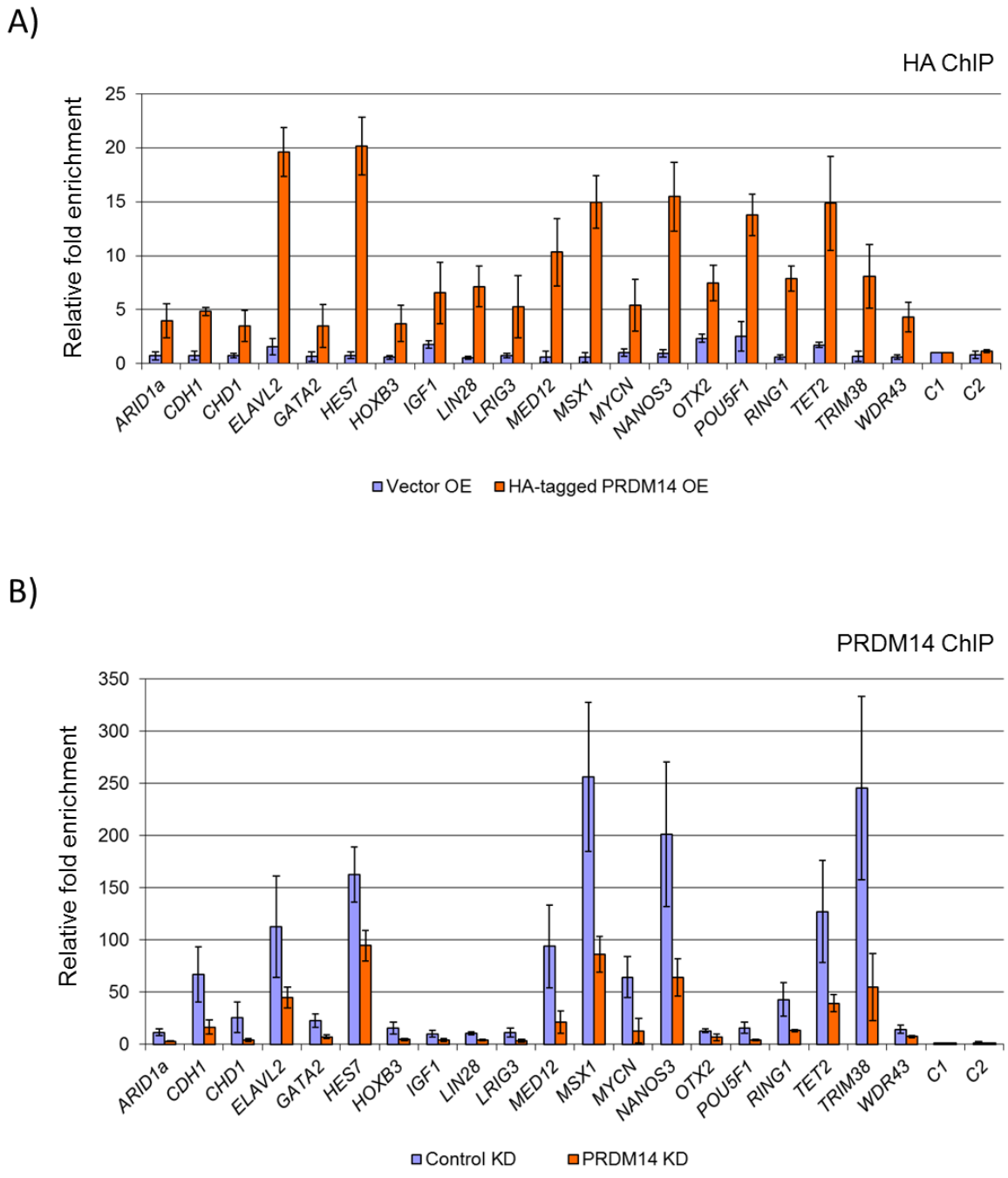


Figure 2. Validation of potential PRDM14 binding loci from ChIP-Seq analysis. A) HA-tagged PRDM14 was overexpressed in the human ESC for 48 hrs and harvested for ChIP with a HA-specific antibody. ChIP was quantitated with qPCR using primers targeting 20 selected PRDM14 ChIP-Seq peaks. Only ChIP with samples overexpressing HA-PRDM14 and not the control vector showed specific enrichment for the 20 loci analyzed. Relative enrichment was calculated via normalization against control regions C1 and C2 which showed only background level of enrichment. B) ChIP was performed using the PRDM14-specific antibody on human ESC transfected with shRNA targeting *PRDM14* or *Luciferase* (control). A marked reduction in the ChIP enrichment at the 20 loci was observed for the PRDM14-depleted sample compared to the control.

3.2 PRDM14 binding profile across the human genome

An analysis of the 12 million sequencing tags generated by ChIP-Seq using the MACS program yielded a total of 24,000 potential PRDM14 binding sites. A stringent cutoff, using only peaks with at least 17 sequencing tags, gave a total of 7,000 high confidence PRDM14 binding sites at p-value < 1.0E-10. These 7,000 binding sites were used for subsequent analysis of the PRDM14 global binding profiles. A look at the distribution of PRDM14 binding sites around human genes showed that 75% of the binding sites can be mapped to annotated genes; 30% are located in the 5' end of the genes, 30% in intragenic regions and 16% in 3' end of the genes (Figure 3A). A comparison of the PRDM14 binding sites to 10,000 random sites in the genome shows strong enrichment of PRDM14 binding sites within 1kb of the transcription start sites (TSS) of genes (Figure 3B). PRDM14 potentially functions in upstream enhancer elements of genes or within the promoter to regulate gene expression. The 7,000 binding sites were mapped to a total of 1700 non redundant gene symbols, forming the potential gene pool regulated by PRDM14. A gene ontology (GO) categorization of these genes revealed a strong enrichment for genes involved in developmental processes (Figure 4). This includes genes that are involved in early embryogenesis, patterning and segmentation of embryos and different germ layer development such as the ectoderm and mesoderm lineages. These results suggest that PRDM14 may play a role in regulating the differentiation potential of human ESC. Interestingly, oncogenes are also among the top class of genes identified. This corresponds to earlier studies identifying PRDM14 as a proto-oncogene and suggests that PRDM14 also plays a role in the proliferative capacity of human ESC.

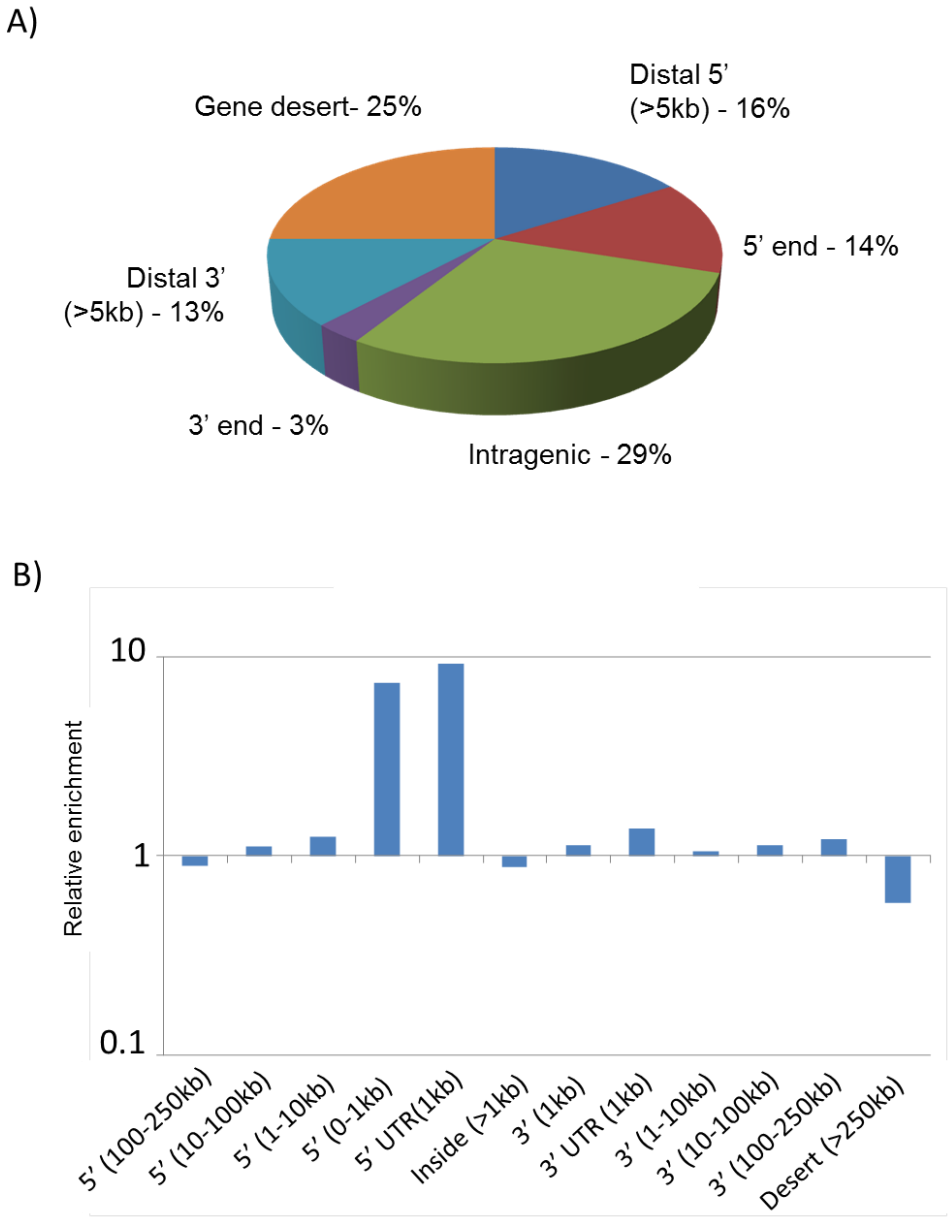


Figure 3. Distribution of PRDM14 binding sites across the genome. A) The distribution of the top 7000 PRDM14 peaks across the genome with reference to annotated genes. 75% of PRDM14 bound loci can be mapped to 5', intragenic and 3' ends of annotated genes while 25% are located in gene deserts. B) Enrichment of PRDM14-bound site in specific loci compared to 10K random sites from the genome. PRDM14 shows preferential binding to ± 1kb of the transcription start sites (TSS).

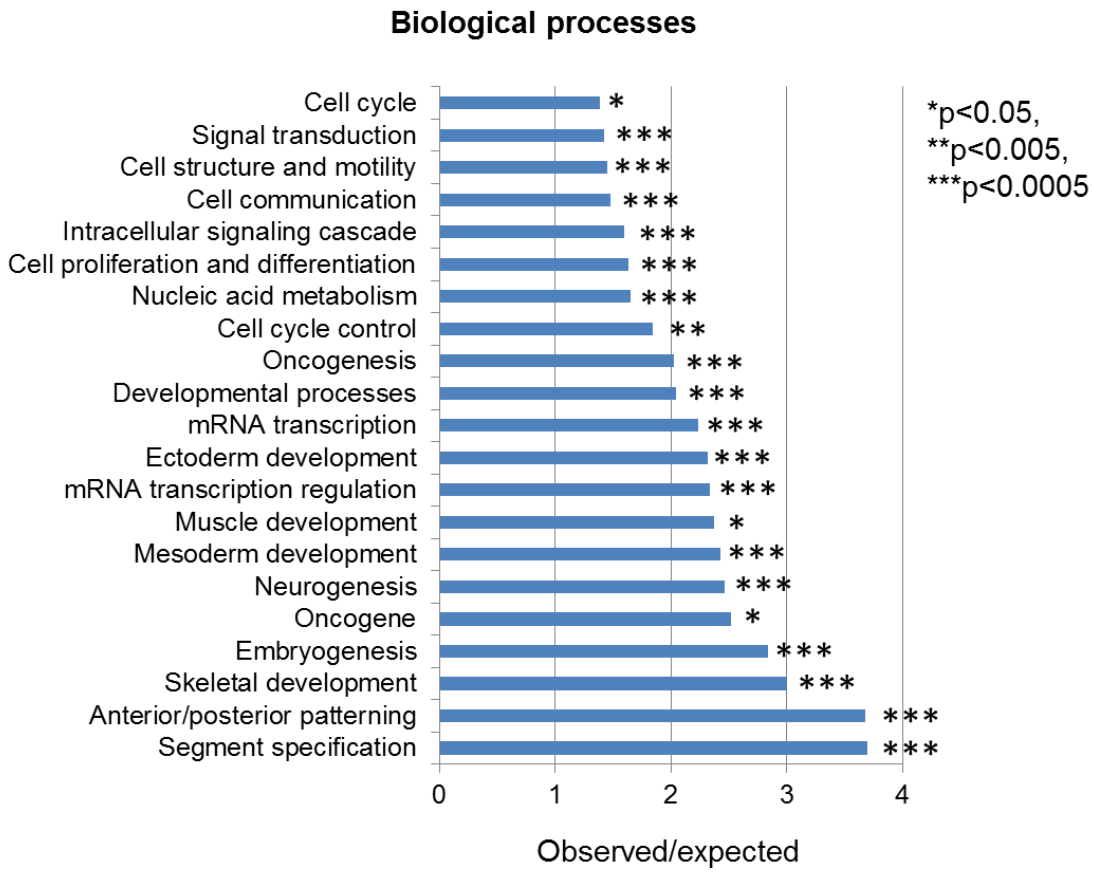


Figure 4. Gene Ontology analysis of putative PRDM14 target genes. Genes which have PRDM14 binding peaks within 20kb to the TSS are taken to be potential PRDM14 targets. A total of 1652 genes were analyzed. Enrichments are defined by the observed number of genes in each category compared to an expected random distribution.

3.3 An enriched DNA motif bound by PRDM14

While PRDM14 contains a potential DNA binding domain (DBD) consisting of 6 C2H2 type Znfs (Figure 1), its DNA binding activity has not been investigated. Using TRANSFAC motif analysis, a list of over represented motifs was generated from the PRDM14 binding peaks (Figure 5). While a few motifs from the TRANSFAC data base including the AP2, SP1 and OCT4 motifs were identified, on top of the list is a *de novo* PRDM14 motif. This motif is highly centralised in PRDM14 binding peaks while the distribution of the other motif identified is much broader. The 18bp extended PRDM14 motif contains a core 9bp central motif GGTCTCTAA (Figure 6A). A His-tagged recombinant protein containing the potential PRDM14 DBD was generated to verify DNA binding activity on the motif using electrophoretic mobility shift assay (EMSA). Two endogenous binding sites near genes *ELAVL2* and *RING1* that contain the conserved binding motif was used as the probe. A protein/DNA complex was formed in the presence of the recombinant protein and probe (Figure 6B). The introduction of point mutations to the conserved motif disrupted the complex formation, suggesting a sequence specific interaction between the PRDM14 DBD and the motif identified.

3.4 PRDM14 binding overlaps with the core regulators

The identification of the OCT4 motif (second most enriched motif) in the TRANSFAC analysis suggests that PRDM14 potentially co-localizes with OCT4 in human ESC. A cluster analysis of the genome-wide binding profile of PRDM14 with other transcription factors OCT4, NANOG, SOX2, p300, KLF4 and c-MYC reveals that PRDM14's global binding profile overlaps extensively with OCT4 and NANOG compared to other factors (Figure 7A).

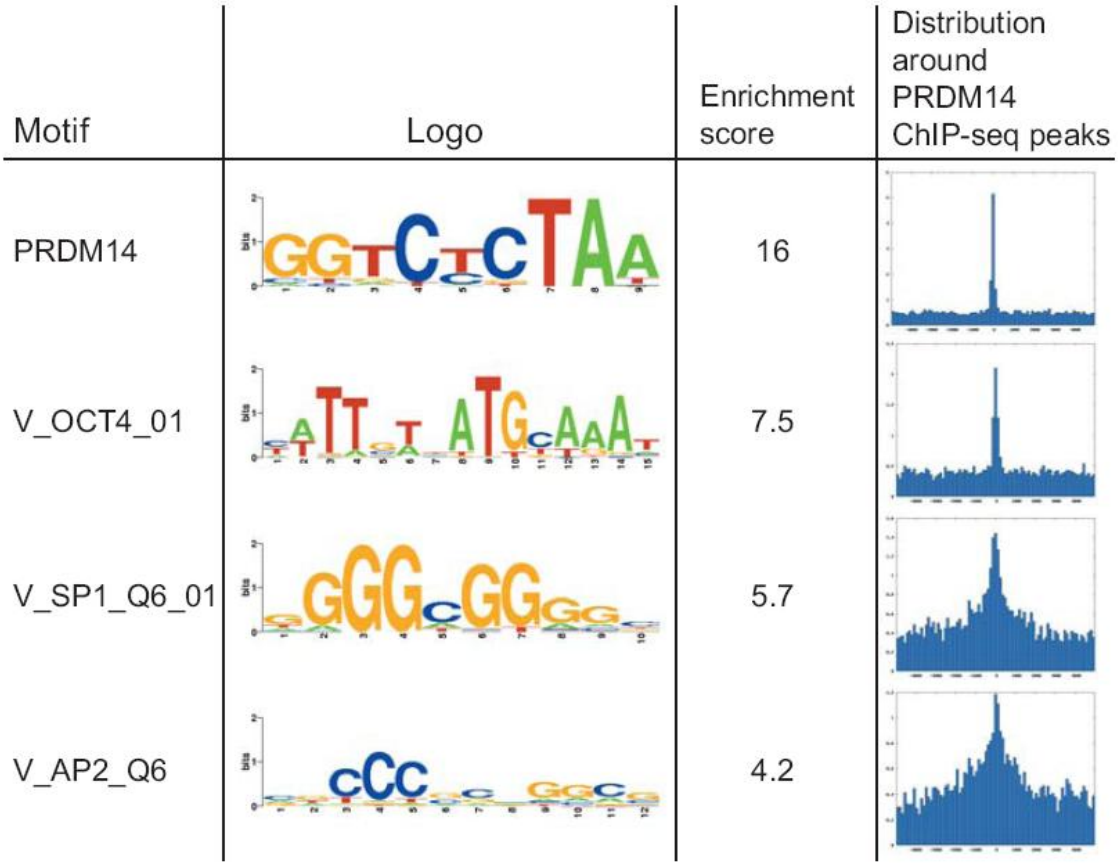
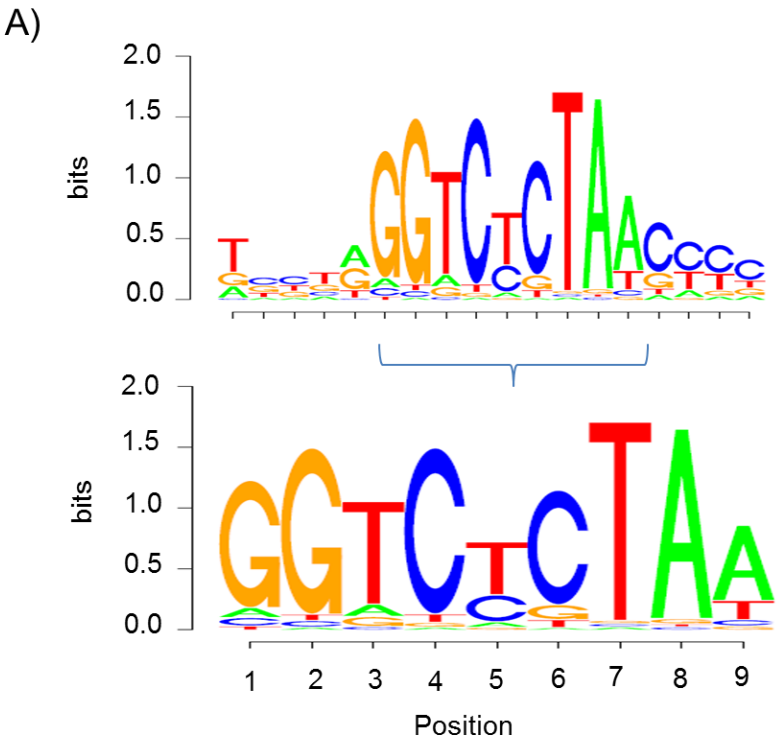
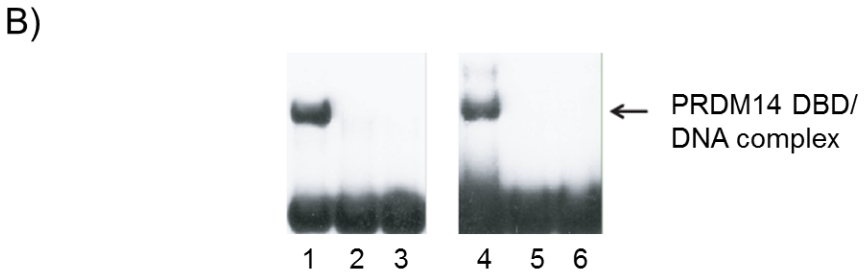


Figure 5. Enriched sequence motif around PRDM14 binding peaks. Co-motif analysis for enrichment of sequence motifs around PRDM14 peaks. Other than the *de novo* PRDM14 motif identified, the OCT4, SP1 and AP2 motifs are enriched around the PRDM14 ChIP-Seq peaks. Enrichment scores are taken to be the number of motif matches compared to the expected number of matches in windows around the PRDM14 peaks. A lower enrichment score is correlated with a broader distribution around the PRDM14 peaks.






```
Probe 1  AGATGCCCATCAGGGTCTCTAACTGCCCCCCACCC  RING1 WT
Probe 2  AGATGCCCATCAGGGTCCCCAACTGCCCCCCACCC  RING1 Mut1
Probe 3  AGATGCCCATCAGGGCCCCCAACTGCCCCCCACCC  RING1 Mut2
Probe 4  ATTCTGTAGTGCAGGTCTCTACCTCAATTCTAGGT  Elavl2 WT
Probe 5  ATTCTGTAGTGCAGGTCTATACCTCAATTCTAGGT  Elavl2 Mut1
Probe 6  ATTCTGTAGTGCAGGTCTCGACCTCAATTCTAGGT  Elavl2 Mut2
```

Figure 6: A PRDM14 specific Motif. A) An extended 16bp motif retrieved from Cisfinder analysis of the PRDM14 peaks. A more conserved central 9bp motif was also retrieved from top 2K PRDM14 peaks. B) EMSA analysis for sequence specific binding by the DNA binding domain (DBD) of PRDM14 on the identified motifs. Two endogenous PRDM14-bound loci near *RING1* and *ELAVL2* contain the conserved 9bp motif and the endogenous gene sequences were used as probes for the gel-shift assay. Point mutations introduced to the PRDM14 conserved motifs (Mut1 and Mut2) for both probes disrupted the protein/DNA complex formed between the recombinant protein and DNA probe.

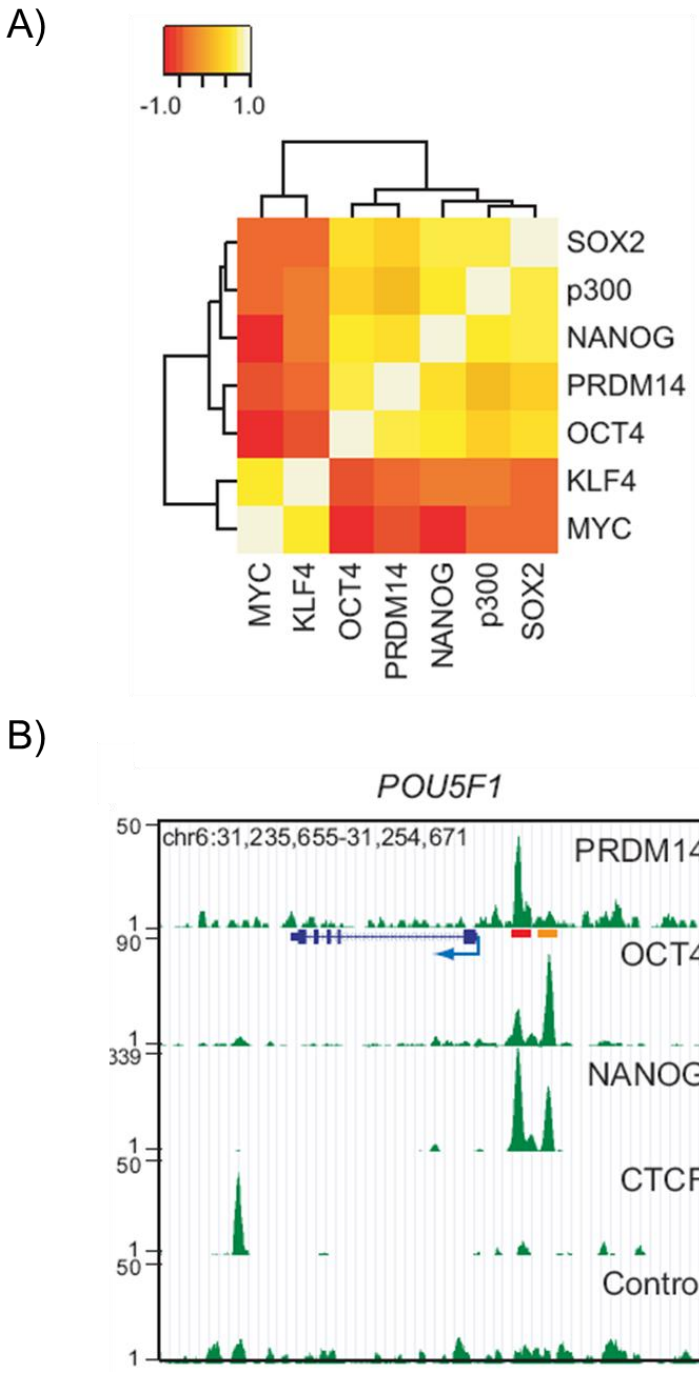


Figure 7: PRDM14 binding co-localizes with the core regulators. A) Pair wise alignment analysis of PRDM14 genome-wide binding profile with known regulators of ESC. PRDM14 show co-binding with the core regulators OCT4 and NANOG. B) A snapshot of ChIP-Seq profiles of PRDM14, OCT4, NANOG and CTCF binding at the *OCT4* loci. PRDM14, NANOG and OCT4 bind on two highly conserved region of the promoter (Orange and red box). Input DNA used for the ChIP assays was sequenced and used as the control.

This overlap suggests that the PRDM14 regulatory node is integrated to the core regulatory network. The presence of p300 in the cluster suggests that PRDM14 potentially regulates enhancer elements with NANOG and OCT4. This co-localization can be observed on the *OCT4* regulatory elements (Figure 7B). As PRDM14 is identified in the screen using the *OCT4* promoter reporter as a read out, it would be intuitive to look at whether this binding translates to a direct regulation of *OCT4* transcription by PRDM14.

3.5 PRDM14 regulates the *OCT4* enhancer elements

A closer look at the binding activities at the *OCT4* promoter highlighted two enhancer regions that are strongly bound by OCT4, NANOG or PRDM14. An element closer to the promoter (red box) is bound by PRDM14 whereas OCT4 preferentially binds to a more distal upstream element (yellow box) (Figure 8A). NANOG binds strongly to both elements on the *OCT4* promoter. The *OCT4* regulatory elements is highly conserved across species in mouse, rat, bovine and human (Figure 8B) (He *et al*, 2009). 4 evolutionary conserved regions identified as conserved region 1 to 4 (CR1-4) are found in the *OCT4* promoter. The two regulatory regions bound by the OCT4, NANOG and PRDM14 are CR2 (red box) and CR4 (yellow box) respectively. In mouse ESC, the binding activity of the mouse core regulators are similarly conserved on these 2 regions (Figure 8C) (Chen *et al*; 2008); Nanog binds strongly to both CR2 and CR4 while Oct4 binds strongly to CR4. In an attempt to identify regulatory elements in the mouse *Oct4* promoter, Yeom *et al* showed that the enhancer activities localized to a proximal enhancer (PE) region and distal enhancer (DE) region (Yeom *et al*, 1996). The CR2 containing PE is highly active in EpiSC and CR4 containing DE is highly active in ESC (Figure 8D) (Tesar *et al*, 2007). I next sought to investigate if the

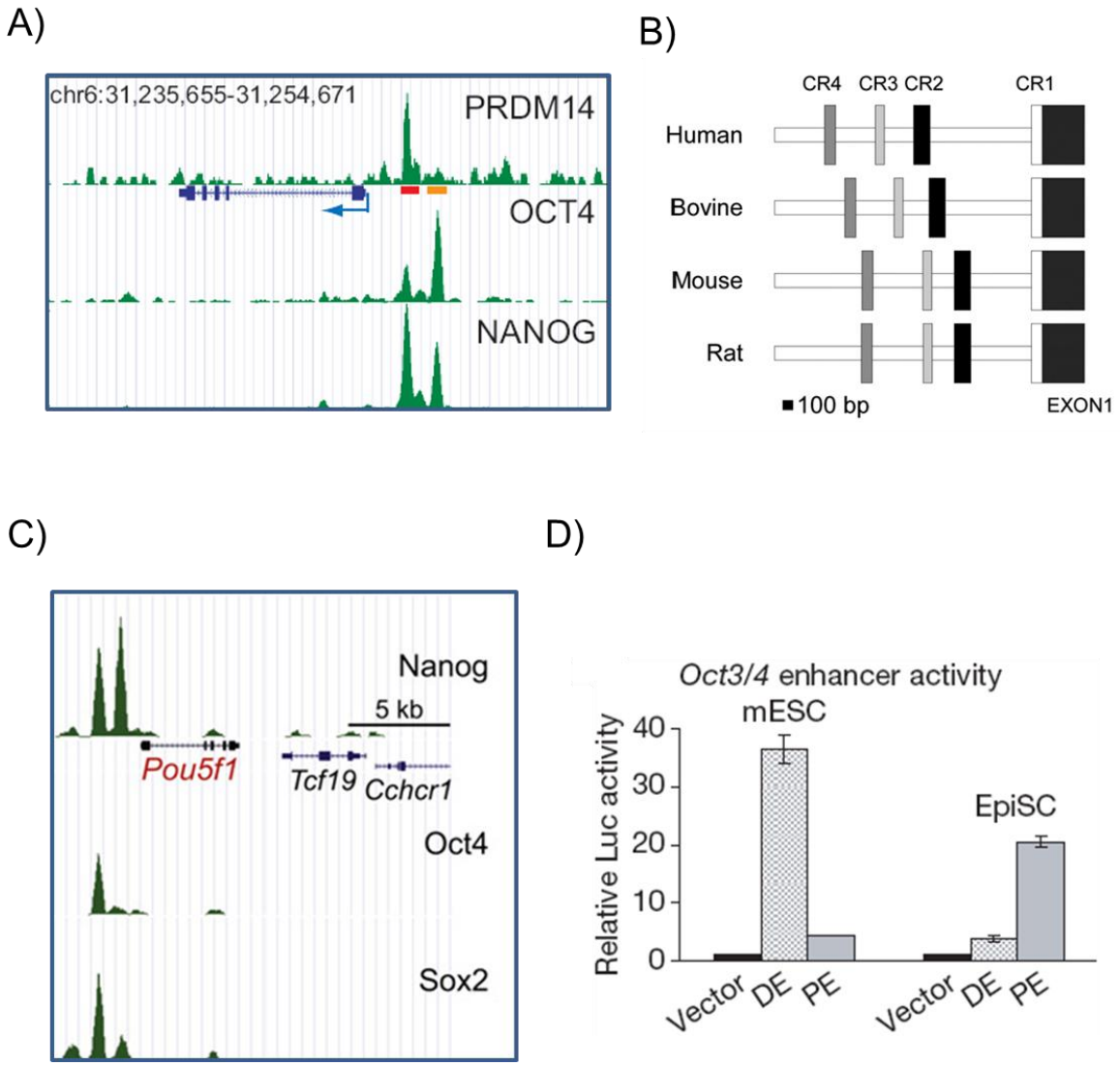


Figure 8: Regulatory elements in *OCT4* promoter. A) PRDM14, NANOG and OCT4 binding on the human *OCT4* promoter. PRDM14, OCT4 and NANOG bind to two conserved regions in the *OCT4* promoter; the conserved region 2 (CR2, red block) is bound strongly by NANOG and PRDM14 and the conserved region 4 (CR4, yellow block) is bound strongly by OCT4 and NANOG. B) Conserved regions 1-4 in the *OCT4* promoter across various species (Adapted from He *et al*, 2009). C) Mouse Nanog and Oct4 similarly binds *Oct4* promoter at the CR2 and CR4 region. Sox2 binds the CR4 enhancer with Oct4 (Adapted from Chen *et al*, 2008). D) Activity of the mouse *Oct4* distal (CR4 containing) and proximal (CR2 containing) enhancers in mouse ESC and EpiSC (Adapted from Tesar *et al*, 2007)**.**

activities of the human CR2 and CR4 regulatory elements are similar in human and mouse ESC and whether PRDM14 is indeed regulating the *OCT4* enhancer elements. The conserved CR2 and CR4 regions in the human *OCT4* promoter were separately cloned downstream of the *Luciferase* reporter gene driven by a minimal *OCT4* promoter (Figure 9A). The enhancer activities of the human CR2 and CR4 are first validated in E14 mouse ESC. Similarly to previous reports, the conserved CR4 element is highly active in mouse ESC whereas CR2 display marginal activities (Figure 9B). This result supports the observation that the two *OCT4* regulatory elements are highly conserved in mouse and human. The CR2 and CR4 enhancer activities are specific to ESC as there is no activity for both constructs in somatic cells MRC-5 (Figure 9C) and HEK-293T (Figure 9D). Next, the activities of the enhancers were tested in 3 separate human ESC lines. Unlike mouse ESC, both enhancers were active in the human ESC (Figure 10). Moreover, the enhancer activity of CR2 seems to be at least twice of CR4 in all three human ESC lines tested. Importantly, only the activity of CR2 and not CR4 is disrupted after PRDM14 depletion (Figure 10). This corresponds with the specific PRDM14 binding on CR2. The binding of PRDM14 on the CR2 fragment cloned into the reporter is also verified by ChIP assay (Figure 11). No binding activity is observed for the CR4 enhancer. The results suggest that PRDM14 regulates *OCT4* expression via its activity on the CR2 enhancer.

Next, using the recombinant PRDM14 DBD generated earlier on, I investigated if PRDM14 binds directly to the CR2 element. Using a set of probes scanning the ChIP-Seq peak of PRDM14 on CR2, a potential PRDM14 binding site was identified (Figure 12A). Point mutations to the putative PRDM14 binding site disrupted the protein/DNA complex (Figure 12B). This specific binding is also observed with endogenous PRDM14 (nuclear extracts) (Figure 12C). With the identification of a PRDM14 binding site on the CR2 element, point

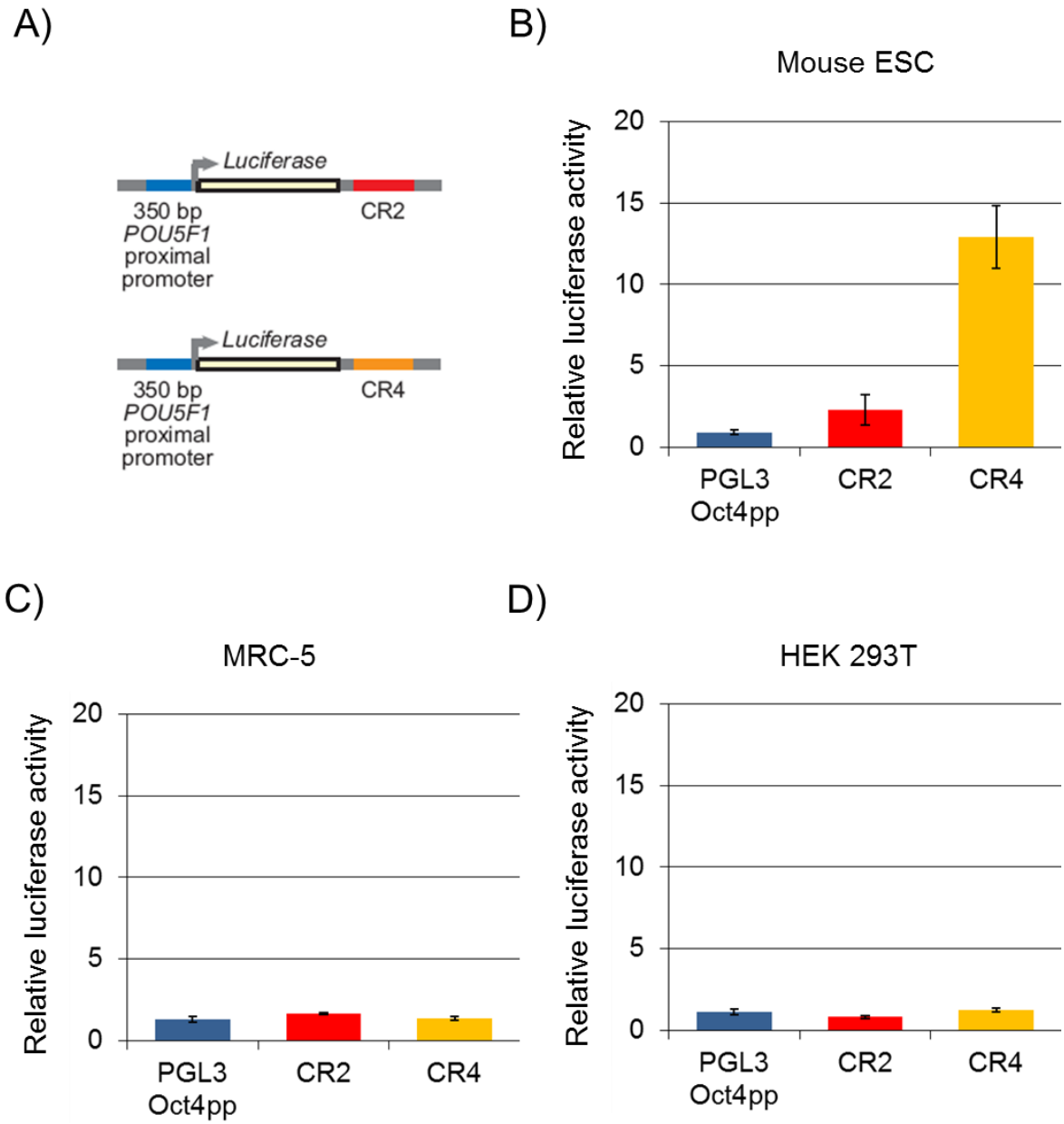


Figure 9: Human CR2 and CR4 enhancer activities in mouse ESC and human somatic cells. A) A schematic representation of the reporter constructs containing the human CR2 and CR4 enhancer. The ~500bp CR2 and CR4 gene fragments were cloned downstream of the *Luciferase* gene driven by a minimal *OCT4* promoter. Activity of the CR2 and CR4 constructs in B) mouse ESC, C) human embryonic lung fibroblast MRC-5 and D) Human Epithelial cell line HEK 293T. The *OCT*4 enhancers are only active in ESC. The relative luciferase activity was obtained by normalizing against the control sample transfected with the basic vector containing only the minimal *OCT4* promoter (PGL3 Oct4pp).

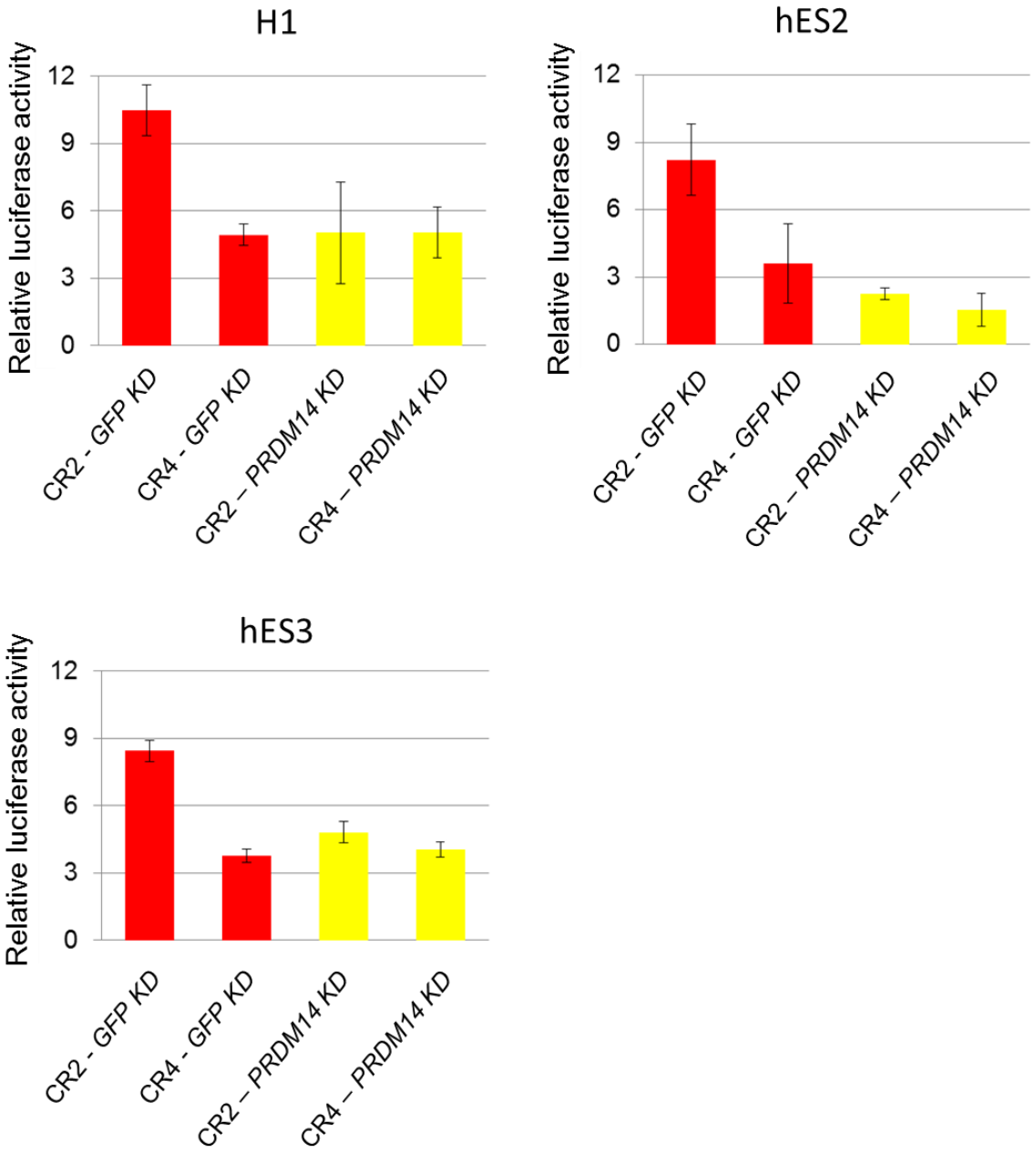


Figure 10: PRDM14 regulates CR2 but not CR4 enhancer activity in human ESC. The CR2 and CR4 constructs are transfected into 3 human ESC lines H1, hES2 and hES3 to test for enhancer activities. shRNA construct targeting *PRDM14 or GFP* was co-transfected with the reporter construct. The relative luciferase activity was obtained by normalizing against the control sample co-transfected with the respective shRNA and the basic vector containing only the minimal *OCT4* promoter.

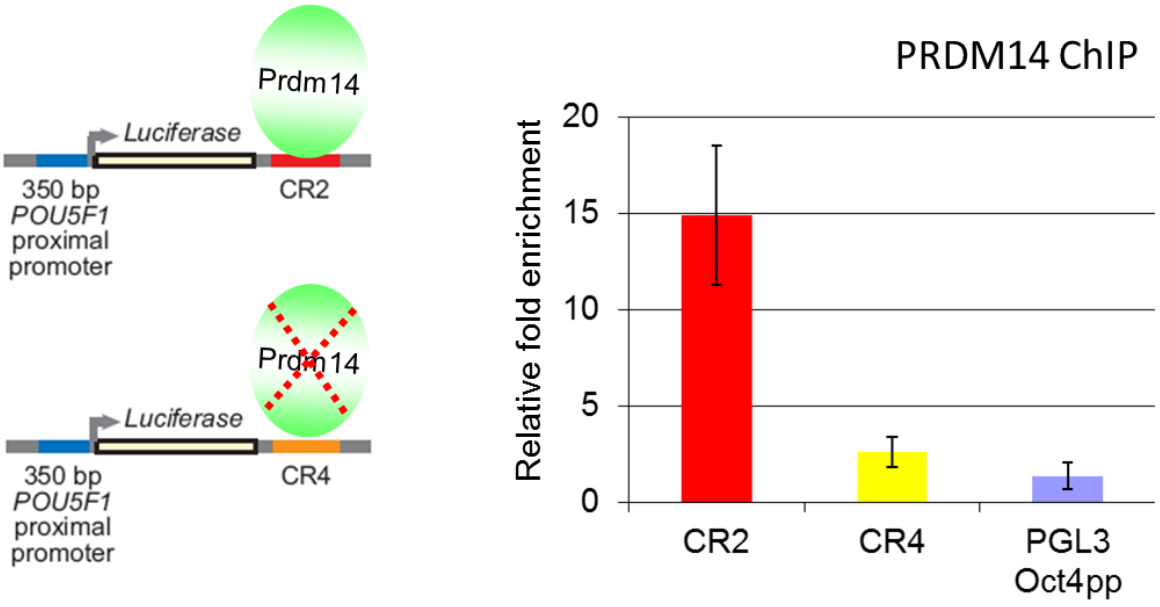


Figure 11: PRDM14 binds to the CR2 enhancer on the reporter construct. Human ESC transfected with the CR2 or CR4 reporter constructs were harvested 48hrs post transfection for ChIP assay with the PRDM14 specific antibody. The relative binding enrichment at CR2 or CR4 was obtained by normalizing against the background enrichment on the minimal *OCT4* promoter.

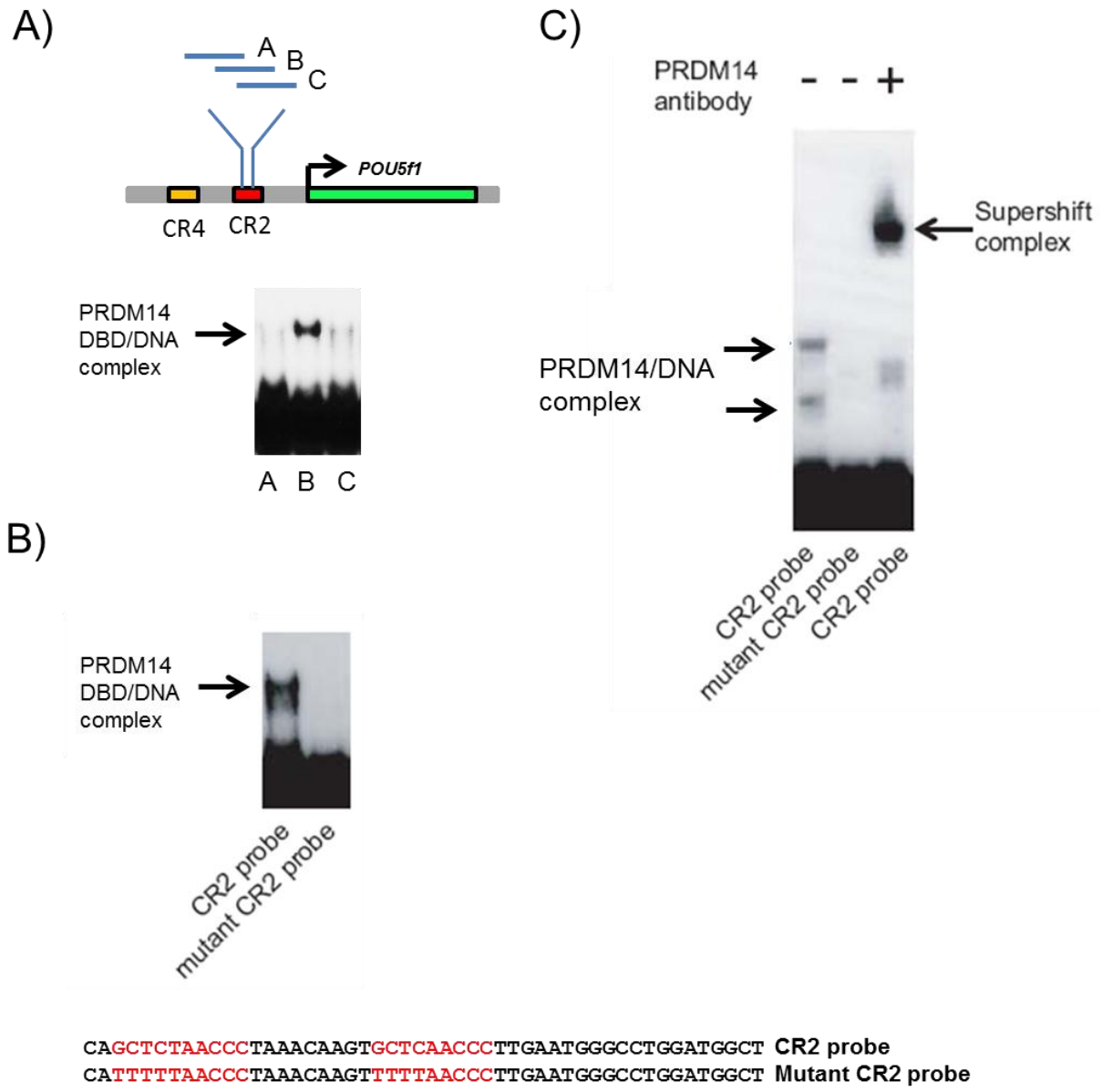


Figure 12: PRDM14 directly binds to CR2 enhancer. A) A set of probes A-C was design to scan for potential PRDM14 binding sites on the CR2 enhancer. PRDM14 DBD binding was detected in probe B. B) Two potential PRDM14 binding motifs were identified in probe B. Point mutations to the 2 binding motifs disrupted the PRDM14 DBD/DNA complex. C) Similar gel shift assay with human ESC nuclear extract. Two specific Protein/DNA complexes were detected. Similar point mutations to the PRDM14 binding sites disrupted both Protein/DNA complexes. Super shift assay was performed with a PRDM14-specific antibody. A complex of low mobility is detected.

mutations that disrupted the protein DNA complex was introduced into the CR2 enhancer in the reporter construct. The point mutations in PRDM14 bound motif drastically reduced CR2 activity in human ESC (Figure 13). As the CR2 enhancer studied is ~500bp in size, there may be other regulatory elements bound by other transcription factors such as NANOG (Figure 8), that may be co-regulating the enhancer activity with PRDM14. To investigate PRDM14 mediated enhancer activity, three copies of PRDM14 motif found on CR2 is cloned upstream of the *OCT4* minimal promoter to test for enhancer activity (Figure 14A). The reporter construct is co-transfected with vectors overexpressing NANOG, PRDM14 or the empty vector as a control, into HEK 293T cells. Only PRDM14 can enhance the reporter activity compared to NANOG, supporting PRDM14-specific activities on its bound sequence in the CR2 enhancer (Figure 14B).

Overall, the data presented in this section suggest that PRDM14 directly binds to the CR2 enhancer element in the *OCT4* promoter to regulate *OCT4* transcriptional activity. Moreover, it is revealed for the first time that the CR2 and CR4 elements are both active in human ESC cells compared to selected activities in mouse ESC and EpiSC. This further highlights intrinsic differences between human and mouse pluripotent stem cells. As human ESC has been proposed to be a human equivalent of the mouse EpiSC, it would be intuitive to look at whether PRDM14 is also highly expressed in EpiSC. Contrastingly, PRDM14 is highly upregulated in mouse ESC compared to EpiSC (Figure 15). Other genes upregulated in mouse ESC or EpiSC are included as controls (Tesar *et al*, 2007). While this study is performed, Ma and colleagues dissected the role of PRDM14 in mouse ESC (Ma *et al*, 2010). Contrastingly, mouse PRDM14 binds strongly to CR4 in the *OCT4* promoter, and PRDM14 depletion did not result in mouse ESC differentiation or *OCT4* reduction. Thus, the regulation of *OCT4*

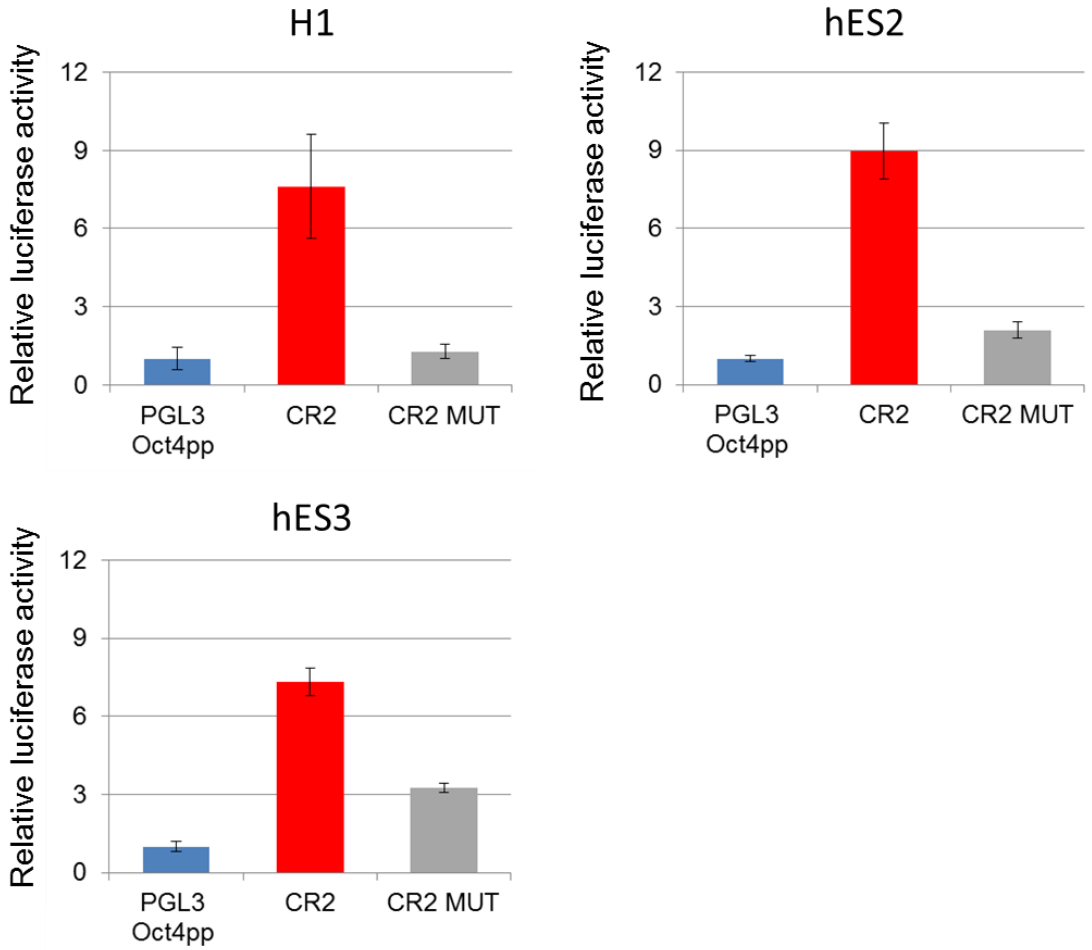


Figure 13: The CR2 enhancer activity is dependent on the PRDM14-bound sequences. Point mutations which disrupted the PRDM14 DBD/DNA complex in the gel shift assay was introduced to the CR2 construct (CR2 MUT). The CR2 and CR2 MUT construct are transfected into 3 human ESC lines H1, hES2 and hES3 to test for enhancer activity. The relative luciferase activity was obtained by normalizing against the control sample transfected with the basic vector containing only the minimal *OCT4* promoter.

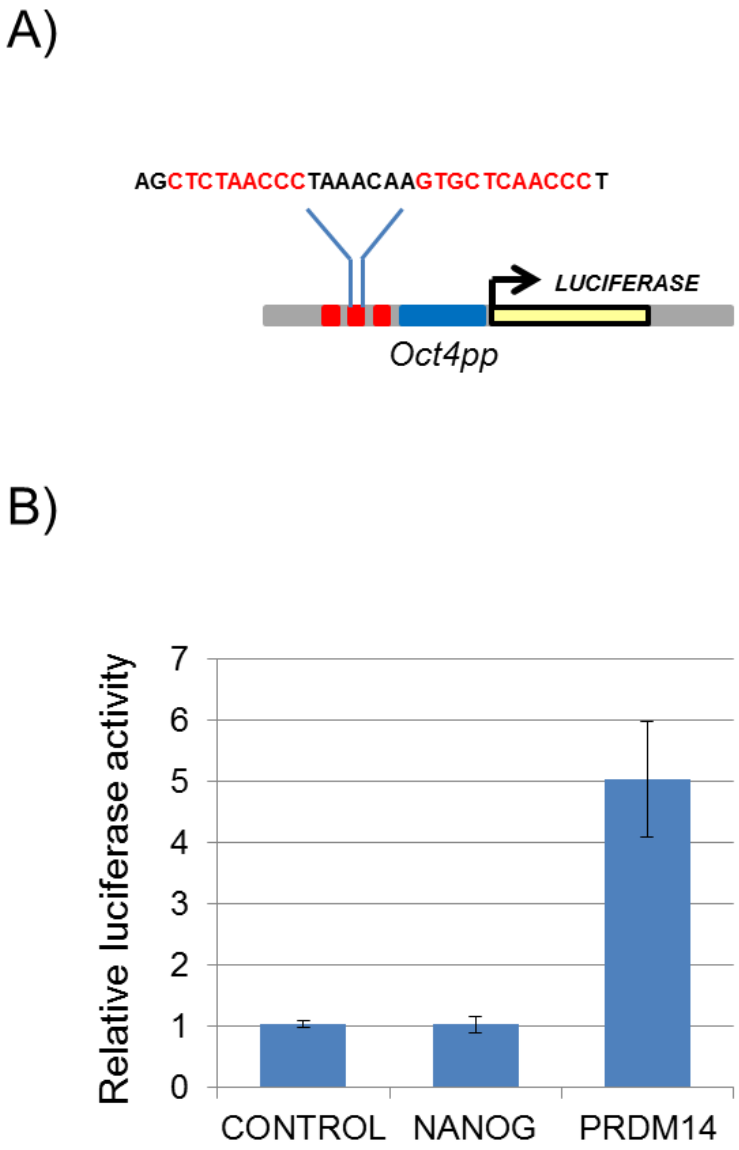


Figure 14: PRDM14 specific enhancer activity on CR2 motif repeats. A) Three copies of 30bp PRDM14 binding motif in the CR2 enhancer was cloned in tandem, upstream of the minimal *OCT4* promoter. B) The reporter construct was co-transfected with vectors ectopically expressing NANOG or PRDM14 into HEK 293T cells. Relative luciferase activity was obtained via normalization against the sample co-transfect with the empty vector.

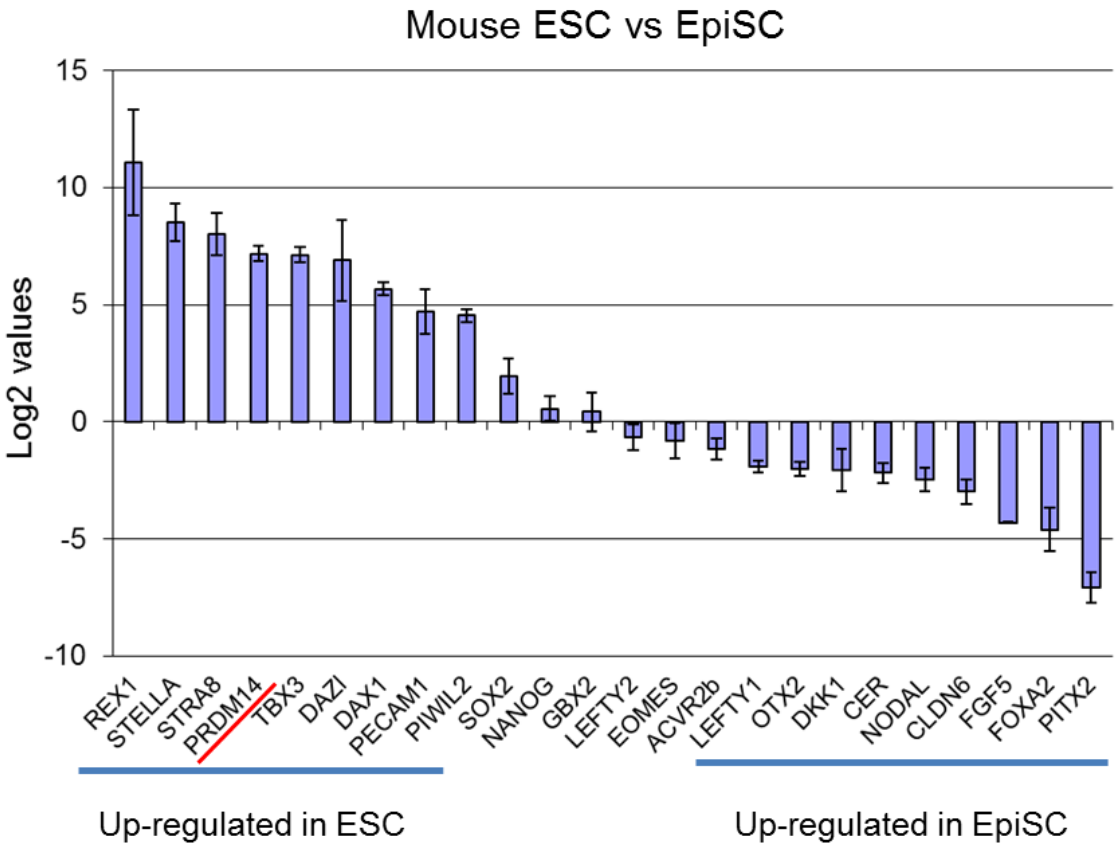


Figure 15: *Prdm14* expression in mouse ESC and EpiSC. The expression level of *Prdm14* and other markers highly expressed in either mouse ESC (up-regulated in ESC) or EpiSC (up-regulated in EpiSC), were quantitated in mouse ESC and EpiSC. Shown is the relative enrichment of each transcript in the mouse ESC normalized against the expression levels of the each transcript in mouse EpiSC. Relative enrichment is presented in Log2 values.

promoter by PRDM14 seems to be human specific and highlights PRDM14 as a potential ESC specific regulator in human.

3.6 PRDM14 partners NANOG in human ESC

To investigate what other genes PRDM14 regulates in the human ESC, the global expression of PRDM14 depleted human ESC was profiled. While not many ESC regulators have been identified in human ESC, Assou *et al* identified a list of pluripotency associated genes in human ESC through a meta-analysis of expression profiles of human ESC differentiated under various conditions (Assou *et al* 2007). The downregulation of PRDM14 resulted in the decrease in some of these genes as early as 48hrs after PRDM14 depletion (Figure 16A), suggesting direct regulation by PRDM14. The co-localization of PRDM14 with NANOG and OCT4 in the earlier analysis (Figure 8A) highlights potential interaction between PRDM14 and the core regulators. In addition, PRDM14 and NANOG showed an overlapping binding profile at the CR2 promoter (Figure 8B). This overlapping binding profile between the two regulators is also observed in some of these pluripotency associated genes regulated by PRDM14 (Figure 16B), indicating a potential interaction between the two proteins. Using a NANOG specific antibody for immunoprecipitation, PRDM14 was found to co-immunoprecipitate with NANOG from human ESC whole cell extract (Figure 17A). This interaction between PRDM14 and NANOG is confirmed in HEK-293T cells ectopically expressing the HA-tagged PRDM14 and NANOG; NANOG co-immunoprecipitates with HA-tagged PRDM14 and *vice versa* (Figure 17B). Interestingly, the strong correlation between PRDM14 and NANOG binding is also observed in the mouse Prdm14 studies (Ma *et al*, 2010). However, the interaction between the mouse Prdm14 and Nanog remains to be validated. The interaction between PRDM14 and NANOG suggests potential co-binding

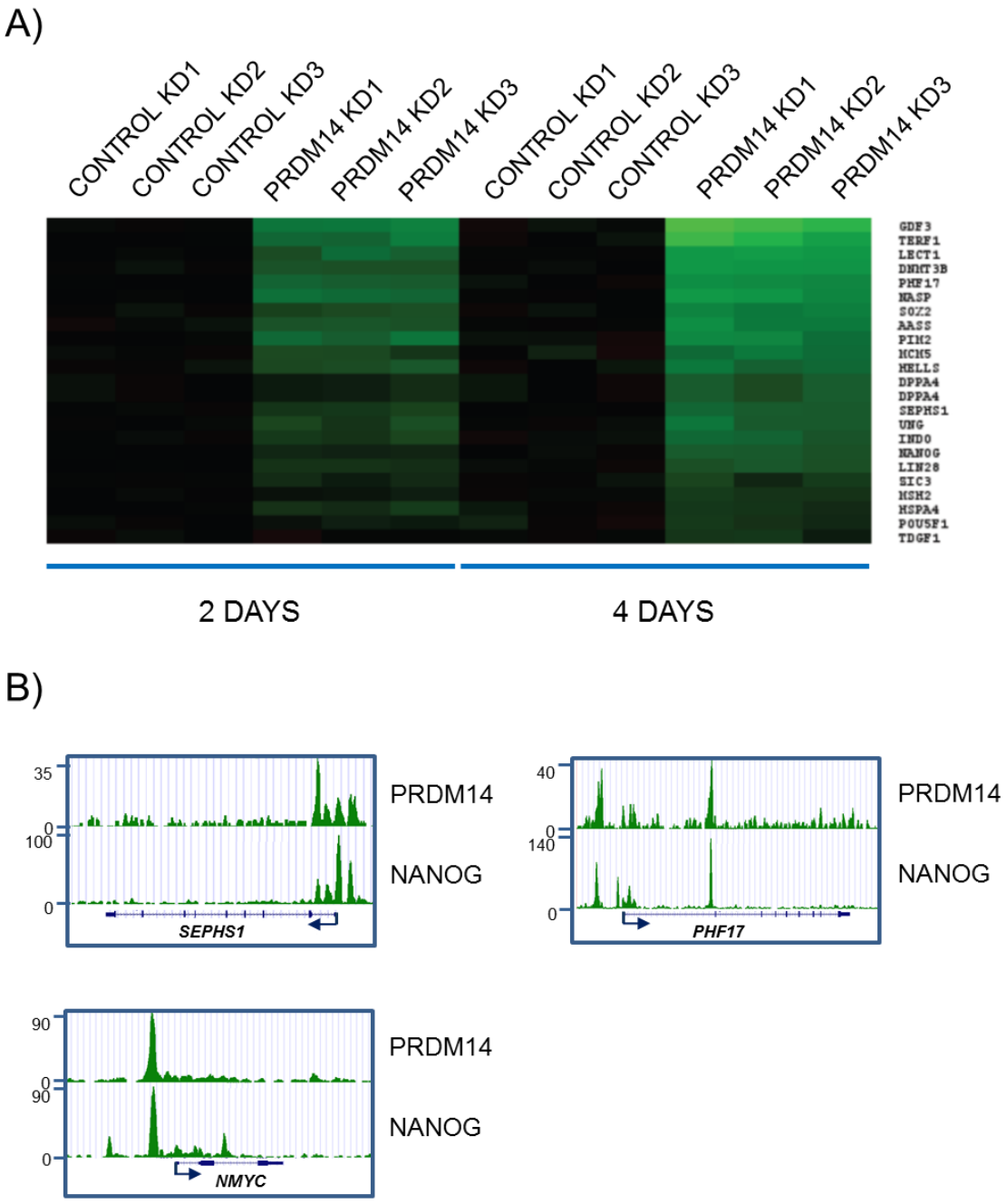


Figure 16: PRDM14 and NANOG ChIP-Seq profiles highly overlap on pluripotency associated genes. A) A heatmap representation of pluripotency associated genes down-regulated after PRDM14 depletion for 2 and 4 days. Decreasing levels of transcripts range from black to green. B) Snapshots of PRDM14 and NANOG overlapping ChIP-Seq profiles in down-regulated pluripotency associated genes *SEPHS1*, *PHF17* and *NMYC*.

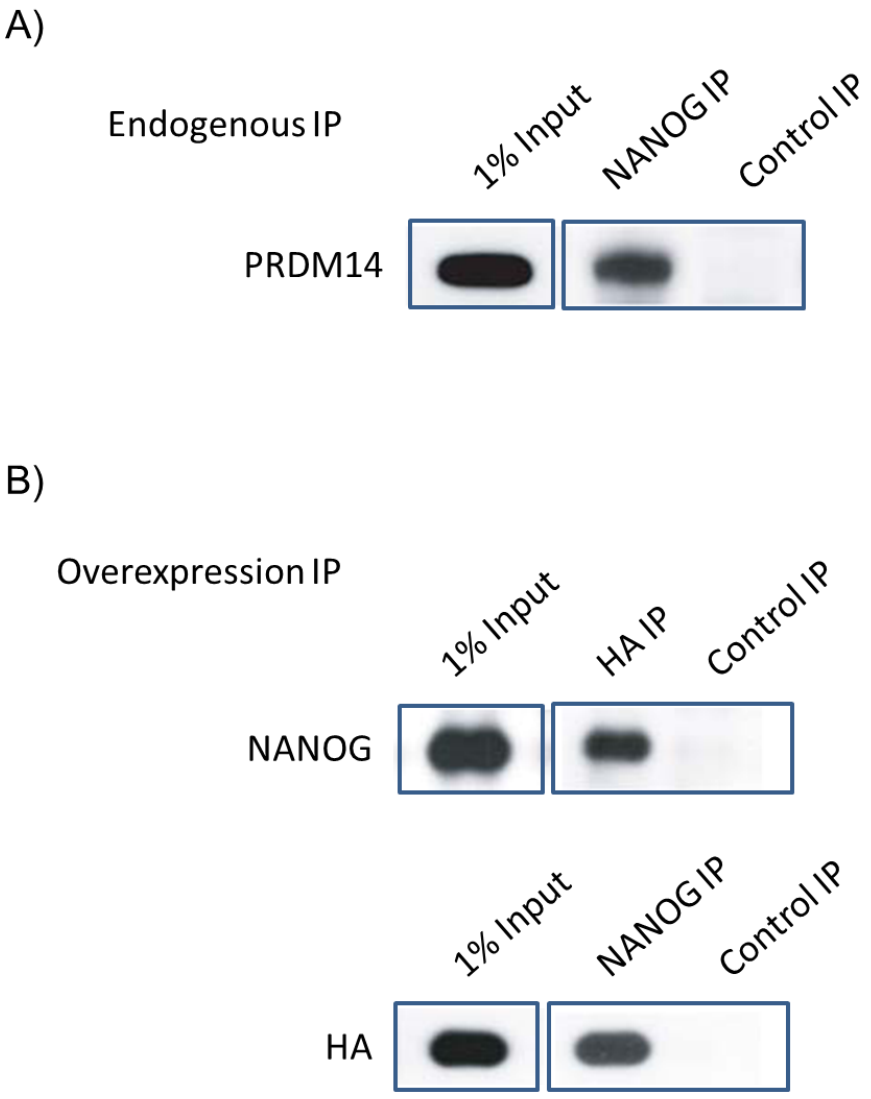


Figure 17: PRDM14 interacts with NANOG. A) Co-IP assay using human ESC whole cell lysate with anti-NANOG antibody. Western blot analysis was carried out with PRDM14-specific antibody. Control IgG antibody was used in the control IP. PRDM14 is detected in the IP with NANOG antibody and not the control IgG IP. B) NANOG associates with PRDM14 in 293T cells. 293T cells were co-transfected with vectors over-expressing HA-tagged PRDM14 and NANOG protein. Whole cell lysate was used for co-IP with anti-HA and anti-NANOG antibody. Anti-GST antibody was used as a control for the anti-HA IP.

activity between the two proteins on these pluripotency associated genes. ChIP assay was performed for NANOG under PRDM14 depleted conditions and *vice versa* to investigate if the binding activities of the two proteins are dependent on each other. While NANOG binding on a number of co-bound genes are unaffected under PRDM14 depleted conditions (Figure 18), PRDM14 binding on these sites are affected by the loss of NANOG (Figure 19A). The loss of PRDM14 binding is specific to NANOG co-bound sites as PRDM14 binding on other sites not bound by NANOG remain unperturbed (Figure 19B). Thus, the results suggest that the NANOG and PRDM14 interaction may be important for stable PRDM14 binding to these pluripotency associated genes. While NANOG is important for PRDM14 stable binding to these genes, depletion of PRDM14 within 48hrs resulted in a strong downregulation of these genes including *OCT4* (Figure 20).

3.7 Synergistic activity of PRDM14 and NANOG

To establish the functional relevance of this interaction, I investigated the activity of PRDM14 and NANOG at the *OCT4* CR2 element. The direct binding of NANOG to the CR2 enhancer was first investigated using EMSA with recombinant GST-NANOG protein. Using the same set of probes design for PRDM14 on the CR2 region, NANOG was found to bind strongly to the same probe as PRDM14 (Figure 21A). Previous genome-wide mapping studies for mouse Nanog identified a Nanog binding motif CATT (Loh *et al*, 2006). On this 50bp probe bound by PRDM14 and NANOG, a CATT motif was found 3bp adjacent of the PRDM14 motif (Figure 21B). The mutation of this motif abolishes the strong NANOG binding to the probe (Figure 21B). This result shows that NANOG binds closely to PRDM14 on the CR2 enhancer and suggests potential synergistic activity between the two proteins in regulating this enhancer. Two copies of this 50bp sequence containing both the NANOG and

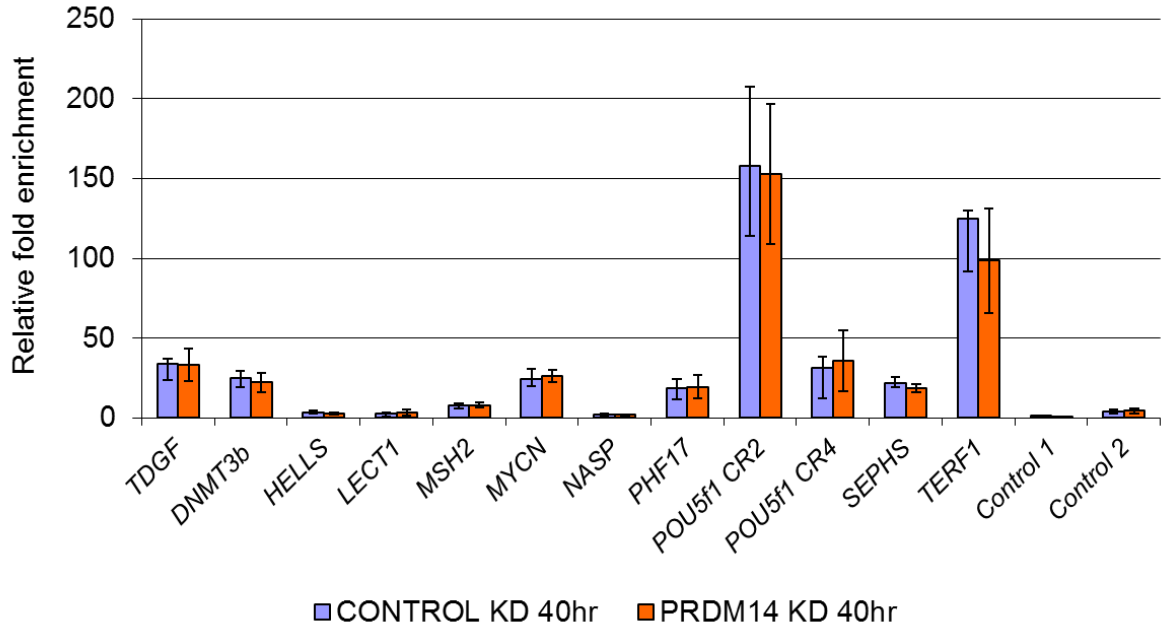
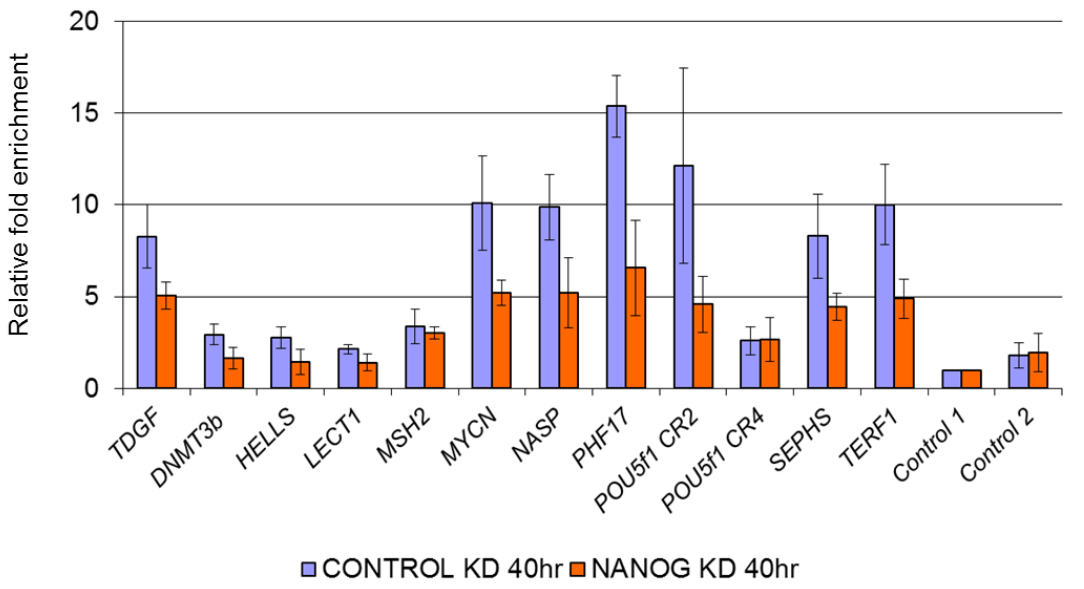


Figure 18: PRDM14 does not affect NANOG binding *in vivo.* Cells were transfected with shRNA targeting *PRDM14* or *Luciferase* and harvested 40hrs post-transfection for ChIP with NANOG antibody. NANOG binding enrichment was tested for 12 selected PRDM14 and NANOG co-bound loci. Relative enrichment was calculated by normalizing against control regions 1 and 2 which display minimal background enrichment.

A)



B)

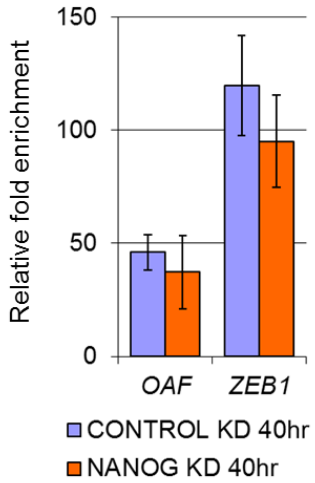


Figure 19: NANOG affects PRDM14 binding on co-bound sites *in vivo*. A) Cells were transfected with shRNA targeting *NANOG* or *Luciferase* and harvested for ChIP using PRDM14 specific antibody 40hrs post-transfection. PRDM14 binding at the 12 selected NANOG-bound loci was decreased after NANOG depletion. B) PRDM14 binding at *OAF* and *ZEB1* (not bound by NANOG) was not affected by NANOG depletion.

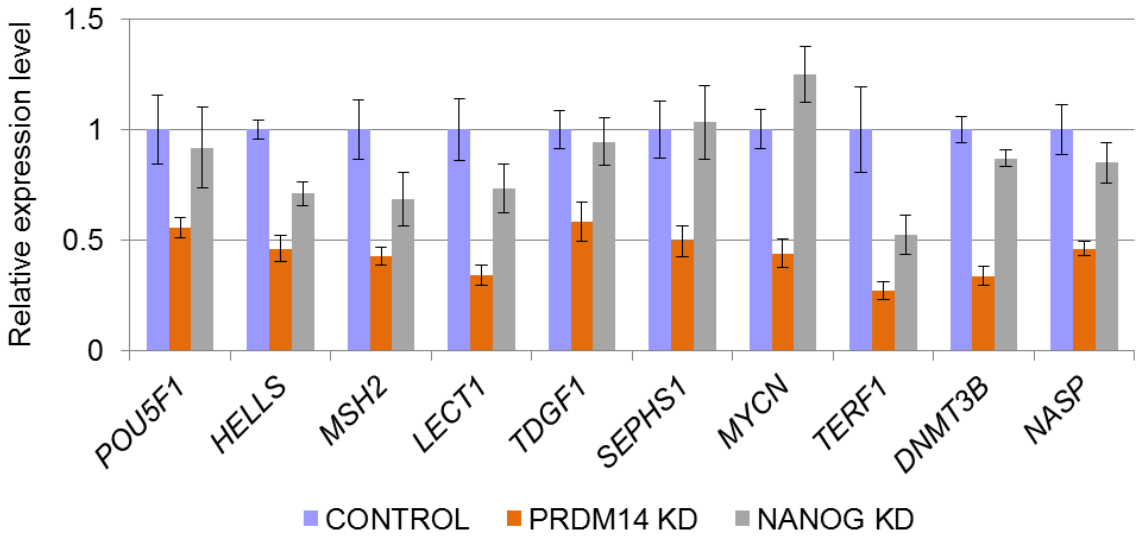


Figure 20: PRDM14 strongly regulates expression of NANOG co-bound genes. The expression of the pluripotency associated genes were quantitated 48hrs after NANOG or PRDM14 depletion. Relative expression was obtained by normalizing against sample transfected with shRNA targeting *Luciferase* gene (control).

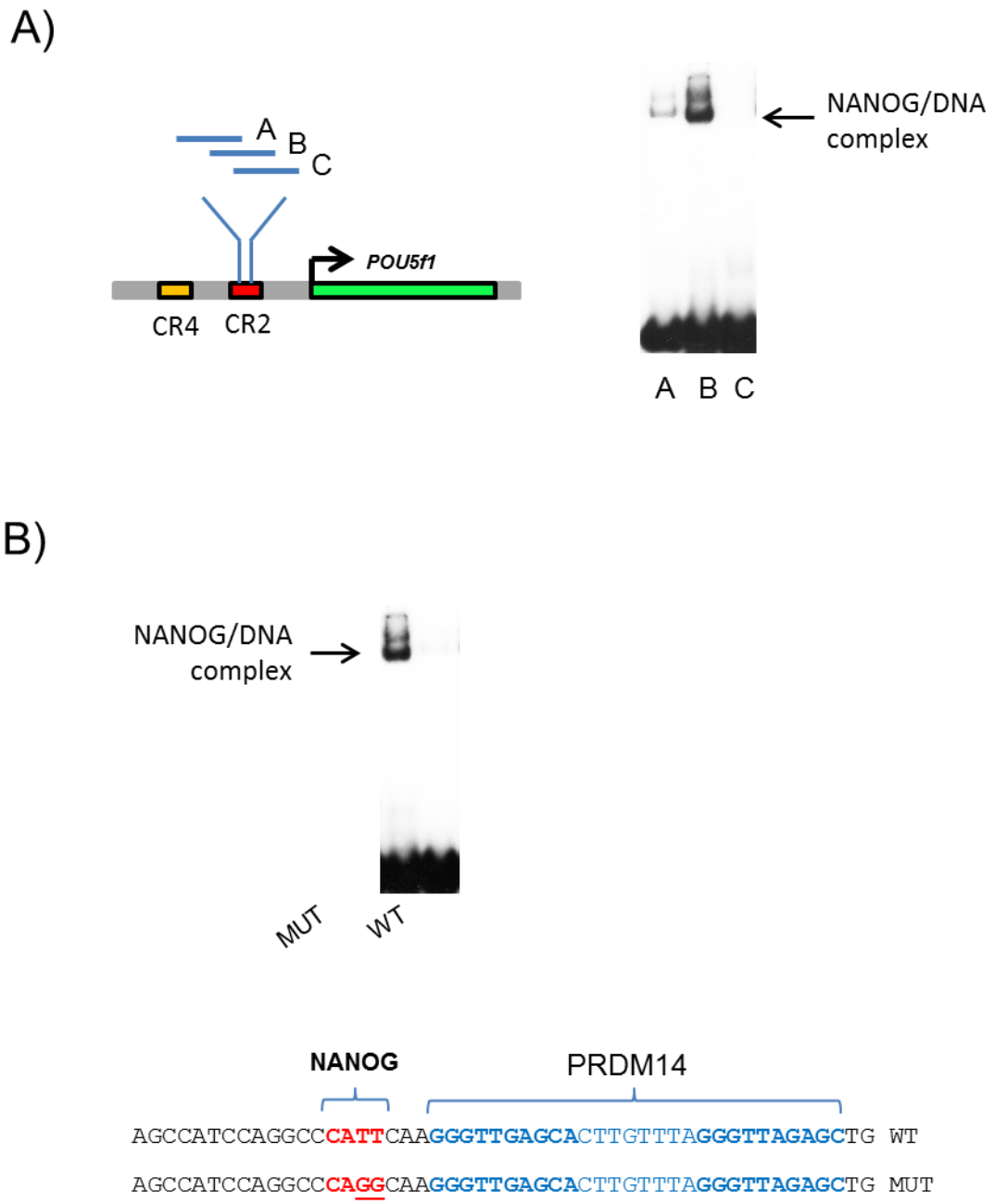


Figure 21: NANOG binds adjacent to PRDM14 on CR2 enhancer. A) The NANOG binding site at the CR2 enhancer was screened using the similar probe set for PRDM14 with recombinant NANOG protein. NANOG binds similarly to Probe B. B) A 4bp conserved NANOG motif CATT (red) was identified. PRDM14 binding motif identified earlier on (blue) is 3bp adjacent to the NANOG binding site. Point mutation to the NANOG binding site abolishes the NANOG/DNA complex.

PRDM14 binding motif were cloned into the reporter construct in place of the 3x30bp motif used earlier on which only contain the PRDM14 motif (Figure 22A). This construct was used to evaluate if NANOG and PRDM14 display synergistic activity on their binding site in the CR2 enhancer in HEK 293T cells. While ectopic expression of NANOG or PRDM14 alone increases the luciferase activity by only two fold, the co-expression of the two proteins increases the luciferase activity by more than six fold (Figure 22A). This data suggest that PRDM14 and NANOG synergistically maintain the *OCT4* expression in human ESC.

In addition to maintaining the human ESC state, PRDM14 was shown to enhance reprogramming efficiency with OCT4, SOX2 and KLF4 (OSK) (Chia *et al*, 2010). It would thus be interesting to see if NANOG and PRDM14 would show similar synergistic activity in enhancing reprogramming. The co-transduction of PRDM14 or NANOG alongside with OSK increases the reprogramming efficiency (measured by number of TRA-1-60 positive colonies) by 5-fold (Figure 22B). Impressively, the reprogramming efficiency was increased by more than 25-fold when PRDM14 and NANOG was induced together with OSK (Figure 22B). This result shows that the PRDM14 and NANOG duo can synergistically improve reprogramming efficiency. The results thus far showed that PRDM14 potentially interacts with NANOG and this partnership is important for their synergistic activity in human ESC and iPSC reprogramming. The global binding profile of NANOG revealed strong binding of NANOG on the *PRDM14* promoter (Figure 23A). NANOG binds strongly to both the distal enhancer and proximal region of the *PRDM14* promoter, verified by ChIP-qPCR. The expression of *PRDM14* greatly decreases after 72hr depletion of NANOG (Figure 23B). Hence, in the maintenance of the human ESC identity, NANOG maintains *PRDM14* expression and the pair of transcription factors synergistically regulates pluripotency genes such as *OCT4* in human ESC.

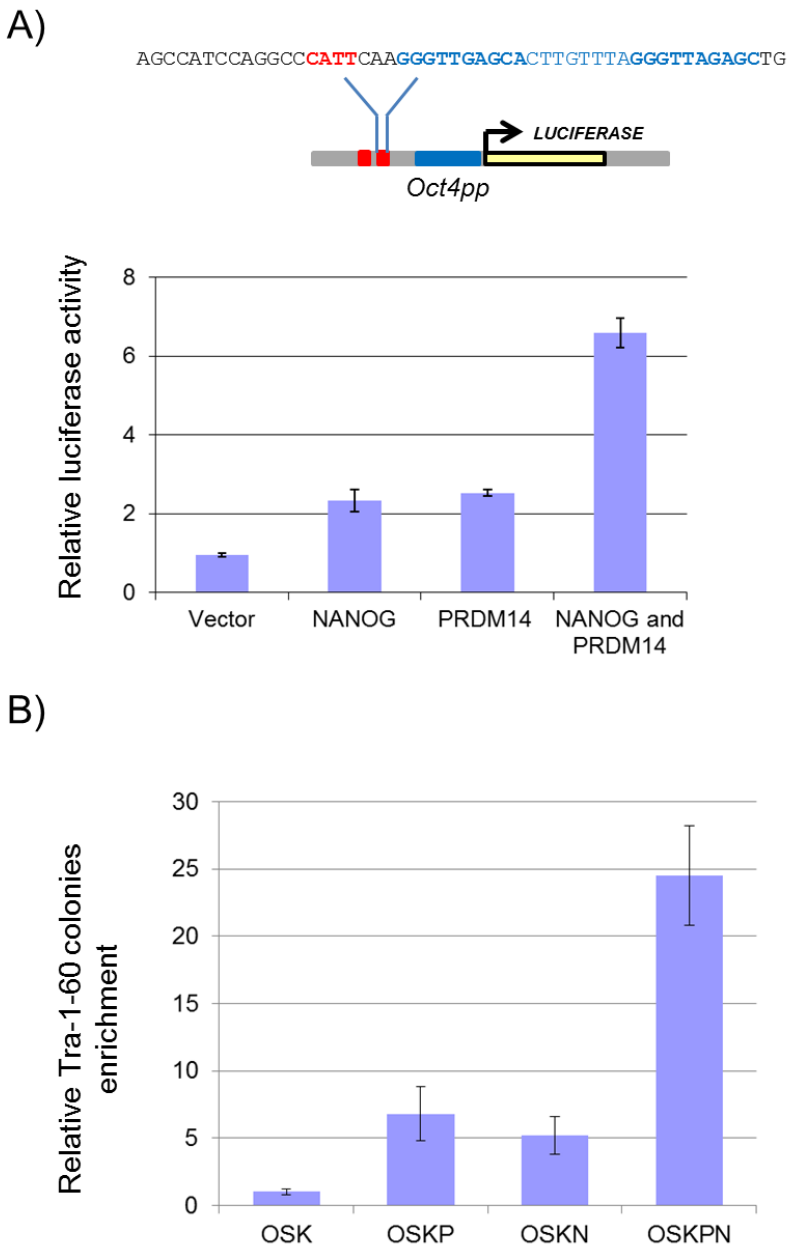


Figure 22: Synergistic activity of NANOG and PRDM14. A) Two copies of the 50bp DNA sequence containing the NANOG and PRDM14 motifs were cloned in tandem, upstream of the minimal *OCT4* promoter driven *Luciferase* reporter construct. The reporter construct was co-transfected with the vectors expressing PRDM14 or NANOG, into HEK293T cells to test for enhancer activity. Enhancer activity was also tested with the co-expression of both proteins. Relative luciferase activity was obtained via normalization against the sample co-transfect with the empty vector. B) Reprogramming of human fibroblast MRC-5 with OCT4 (O), SOX2 (S) and KLF (K). PRDM14 (P) and NANOG (N) was co-transduce with OSK to test for reprogramming activity. Reprogramming efficiency was measured by the number of Tra-1-60 positive colonies formed after 3 weeks post infection. Relative enrichment was obtained via normalization against number of colonies obtained under OSK mediated reprogramming.

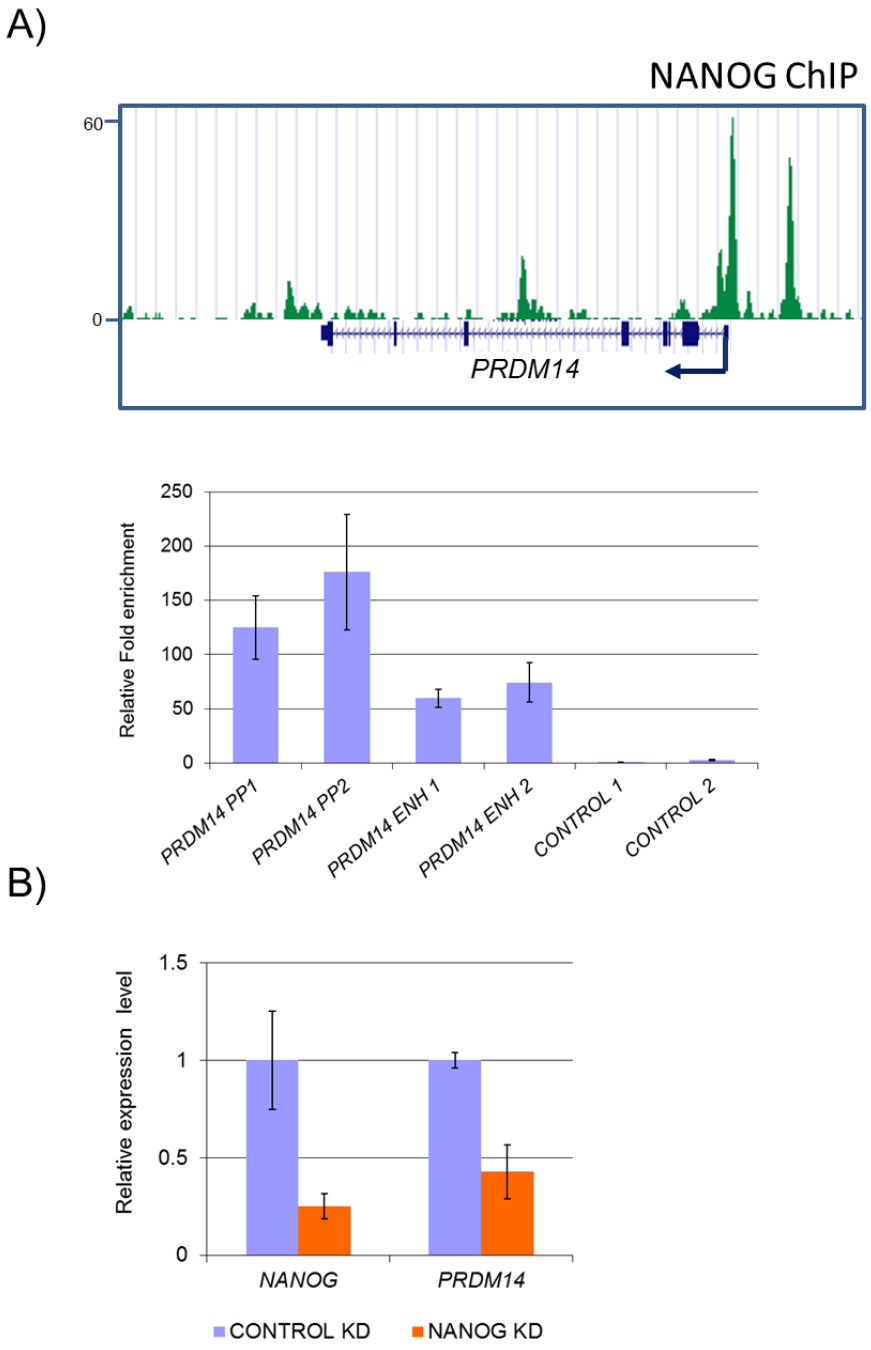


Figure 23: NANOG regulates *PRDM14* Expression. A) A Snapshot of NANOG ChIP-Seq profile on *PRDM14* promoter. Two distinct peaks were identified in the proximal promoter and upstream enhancer. B) The ChIP-Seq peaks are validated with ChIP-qPCR using two primer pairs for each of the NANOG peaks in the promoter and enhancer. C) The PRDM14 transcript is down-regulated after NANOG depletion in human ESC for 72hrs. Control samples were transfected with shRNA targeting *Luciferase.*

3.8 Summary and discussion 1

The results presented thus far show that PRDM14 is a potential core regulator of the human ESC fate. PRDM14 binds predominantly to genes associated with embryogenesis in human ESC. Its genome-wide binding profile overlaps with OCT4, NANOG and the co-activator p300 which suggests that PRDM14 activates pluripotency associated genes with the core regulators. Importantly, PRDM14 is shown to be a direct regulator of *OCT4* transcriptional activity in human ESC. Corresponding to the overlapping global binding profiles, PRDM14 partners NANOG in regulating pluripotency associated genes. This partnership is important for their synergistic activity in activating the OCT4 enhancer and also in the conversion of fibroblasts back to a pluripotent cell state. The study thus far expanded the pluripotency network in human ESC that includes PRDM14 as a close partner of the core regulators in maintaining the human ESC fate (Illustration 7). The results also potentially explain the *OCT4* expression in lymphomas that overexpress PRDM14 (Dettman *et al*, 2008). PRDM14 expression may have reactivated the pluripotency network in these cancerous cells. Thus, understanding how PRDM14 functions in the human ESC may potentially reveal new strategies in combating these PRDM14 expressing lymphomas.

The results in this first half of the study also highlight PRDM14 as a potential human ESC specific regulator. The differences in CR2 and CR4 activity in human ESC compared to mouse ESC and EpiSC further supports the notion that there exist intrinsic differences in the regulatory elements in human ESC that could have rewired the pluripotency regulatory network. The direct regulation of the CR2 enhancer by PRDM14 is not seen in mouse ESC; mouse Prdm14 binds CR4 in mouse ESC (Ma *et al,* 2010) and CR2 is not active in mouse ESC. While Prdm14 is also expressed in mouse PGC where Oct4 is also expressed, only the

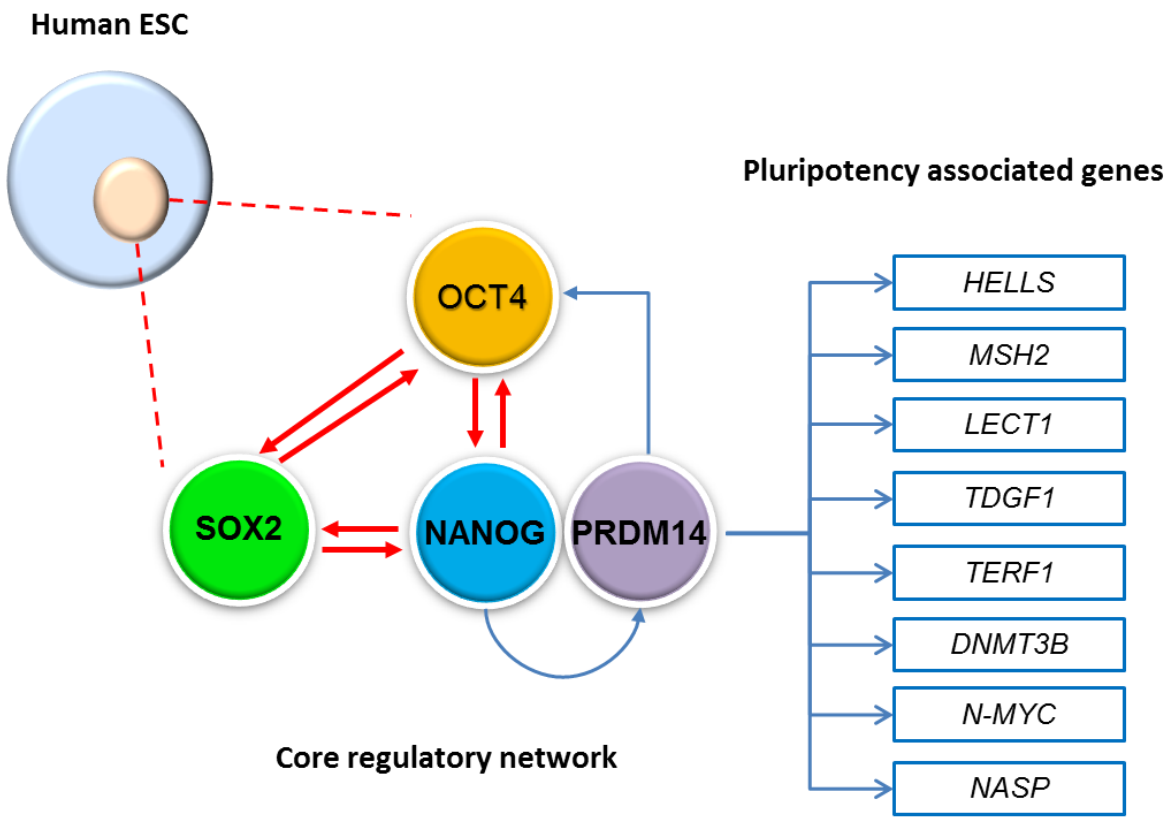


Illustration 7: PRDM14 integrates with the core regulatory network. A graphical representation of the core regulatory network that includes PRDM14 as a new member of the ESC transcriptional network regulators. PRDM14 partners NANOG in this network in regulating pluripotency associated genes including *OCT4*. NANOG also regulates *PRDM14* expression to maintain its level in human ESC.

CR4 of the *Oct4* promoter is active in the germ cells (Yeom *et al*, 1996). Although CR2 is active in mouse EpiSC, Prdm14 is not expressed in these cells (Figure 15). However, the presence of other members of the PRDM family of regulators may have replaced the activity of Prdm14 in mouse EpiSC. As an example, the dissection on the role of Klf4 in mouse ESC highlighted the redundant function between the Klf family of transcription factors Klf2, Klf4 and Klf5 in maintaining the mouse ESC identity (Jiang *et al*, 2008). While Prdm14 is not expressed in mouse EpiSC, Prdm2, Prdm4, Prdm5 and Prdm10 are expressed in mouse EpiSC (Figure 24). Whether these close homologs of PRDM14 have replaced its activity in mouse EpiSC remains to be investigated. In addition, the human and mouse PRDM14 amino acid sequence is only 70% conserved (Yamaji *et al*, 2008). This potentially accounts for divergences in functions displayed by the 2 orthologs in mouse and human ESC.

Lastly, the interaction between PRDM14 and NANOG is potentially important for stabilizing PRDM14 binding to its loci *in vivo*. Depletion of NANOG results in the decrease in PRDM14 binding near pluripotency-associated genes (Figure 19A). The need for NANOG stabilization at these loci is further supported by the lower ChIP enrichment at these sites compared to other loci near developmental genes (Figure 19B): an average of 10 fold enrichments observed in these pluripotency genes compared to the strong 50-100 folds in the developmental genes not co-bound and influenced by NANOG. As such, NANOG may function to direct and stabilize PRDM14 binding to pluripotency-associated genes in human ESC. On the other hand, the enhancer activity of PRDM14 in the reporter assays suggests that PRDM14 is able to bind to the DNA in the absence of NANOG. Thus, how the interaction between the NANOG and PRDM14 protein results in their synergistic activity in the OCT4 enhancer and reprogramming remains to be investigated.

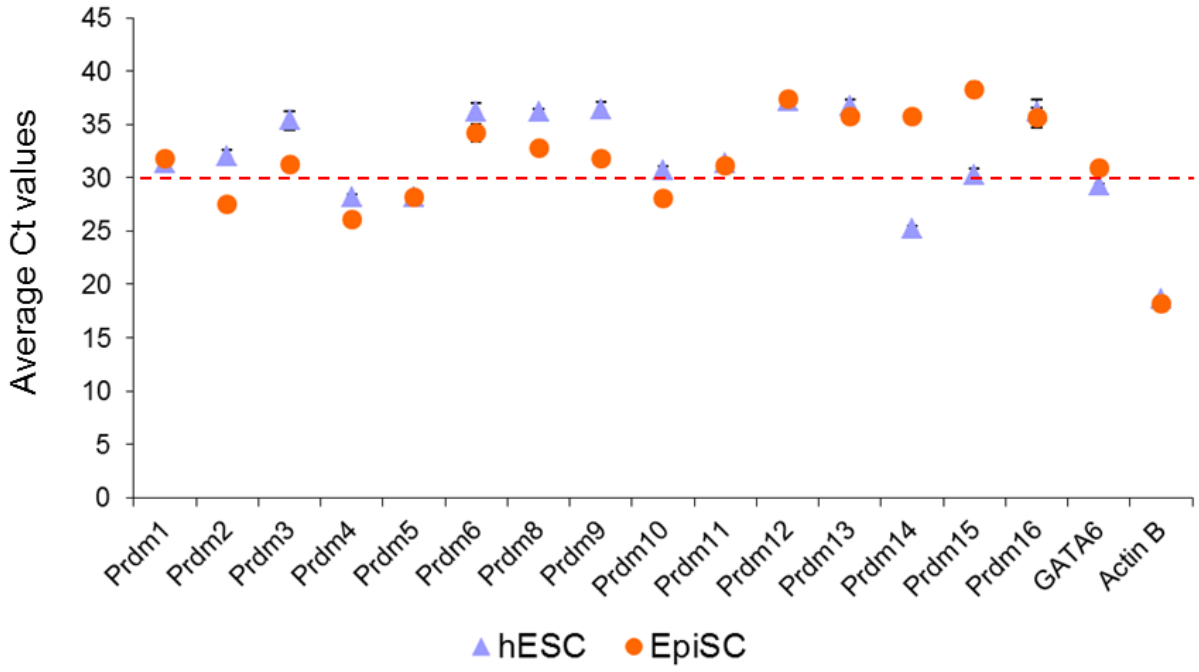


Figure 24: Expression of PRDM family of regulators in human ESC and mouse EpiSC. The expression of the orthologs of human and mouse PRDM family of regulators in human ESC and mouse EpiSC were examined. Presented is the average Ct values detected in the qPCR. A house keeping gene such as β-actin has an average Ct value of 20 and a lowly expressed gene such as GATA6 has an average Ct value of 30. The GATA6 Ct value is used as a cutoff for determining whether the PRDM family of regulators is expressed in either cell type.

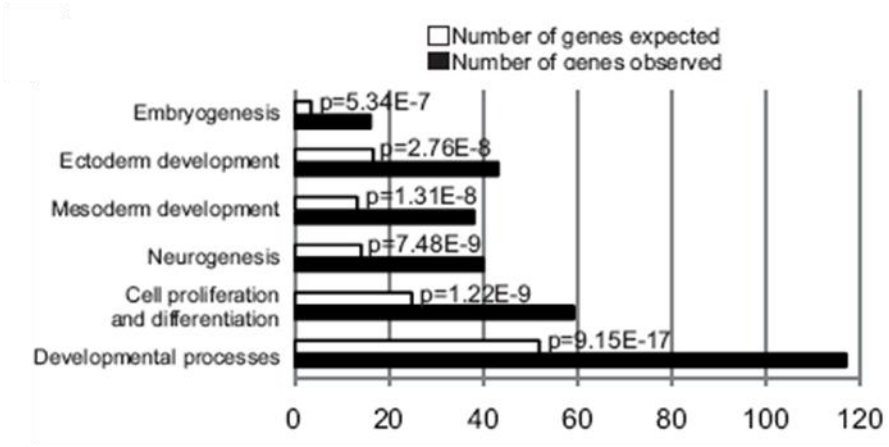
Chapter 4: A repressive role of PRDM14 in human ESC and iPSC reprogramming

4.1 PRDM14 binds repressed genes and poised elements in human ESC

The gene ontology analysis of PRDM14 bound genes revealed an enrichment of developmental genes involve in various lineage specifications (Figure 4). Concurrently, the depletion of PRDM14 in human ESC resulted in significant upregulation of developmental genes (Figure 25A). These genes are also enriched for transcriptional regulators that potentially mediate cell fate specification of human ESC (Figure 25B). A global view of the expression profiles of genes close to all PRDM14 binding sites reveals that more genes are upregulated after PRDM14 depletion in human ESC (Figure 26A). To investigate whether the PRDM14 bound genes are highly upregulated after PRDM14 depletion, a Gene Set Enrichment Analysis (GSEA) was performed (Subramanian *et al*, 2005; Mootha *et al*, 2003). The analysis revealed that PRDM14 bound genes are highly enriched for upregulated genes (Figure 26B), supporting a repressive role of PRDM14 in human ESC.

In comparison to the few transcription regulators studied in human ESC, the epigenome has been extensively profiled (Heintzman *et al*, 2009; Creyghton *et al*, 2010; Hawkins *et al*, 2010; Hawkins *et al*, 2011; Rada-Iglesias *et al*, 2011). Regulatory regions within the human ESC genome are marked with unique signatures consisting of different chromatin modifications. Promoters of lineage specific genes in both human and mouse ESC are marked by the active tri-methylated histone H3 lysine 4 (H3K4me3) and the repressive tri-methylated histone H3 lysine 27 (H3K27me3), termed bivalent domains (Bernstein *et al*, 2006; Zhao *et al*, 2007). Genes marked by bivalent domains have low or no expression in the ESC and are

A)



B)

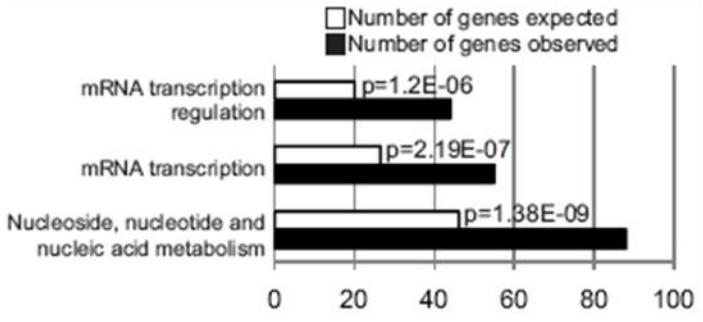
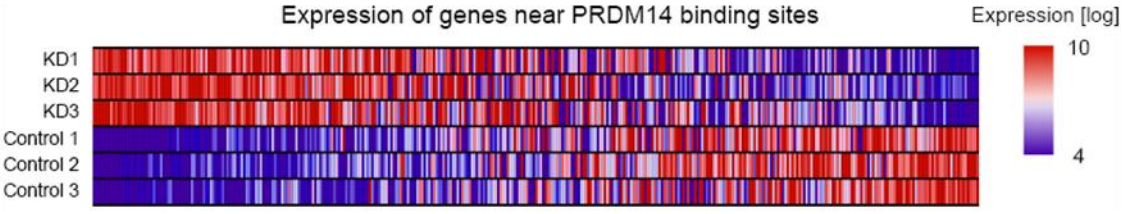


Figure 25: Strong upregulation of developmental genes in PRDM14 depleted human ESC. Gene ontology (GO) analysis of up-regulated genes in PRDM14 depleted human ESC. A) PRDM14 depletion results in up-regulation of genes associated with developmental processes. B) Transcription regulators are highly enriched among these genes.

A)



B)

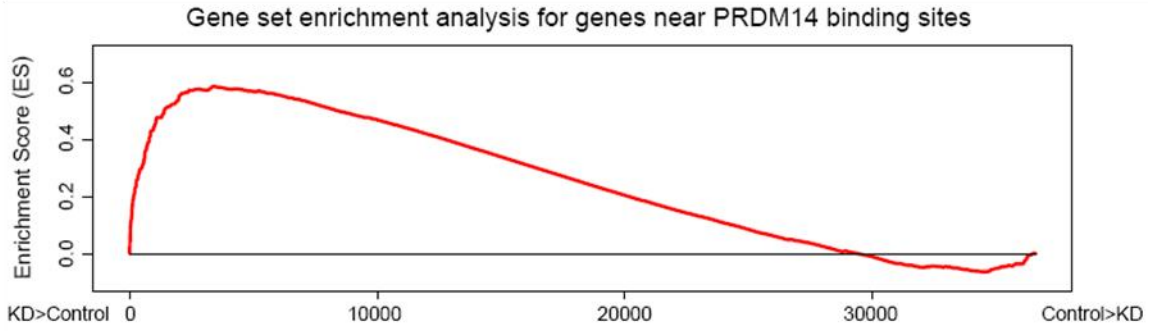


Figure 26: De-repression of genes near PRDM14 binding site. A) Shown is the expression of genes near 500 of the highest scoring PRDM14 peaks in the ChIP-Seq profile. Genes are arranged according to their relative enrichment after PRDM14 depletion from lowest (blue) to highest (red). B) Gene Set Enrichment Analysis (GSEA) of PRDM14 target genes. The running-sum statistic for all the genes on the array is shown; PRDM14 target genes increase the statistic while non-target genes decrease it. The maximum deviation from zero indicates the enrichment score (ES). PRDM14 target genes are significantly enriched at the top of the list (KD>Control, ES 0.59, p-value=0.0, FDR=0.0).

proposed to be in a poised state for activation. A similar observation was also reported for enhancers in human ESC (Rada-Iglesias *et al*, 2011). The active and poised enhancers are both marked by the presence of p300, mono-methylation of histone H3 lysine 4 (H3K4me1) and with low nucleosome levels. Active enhancers are however marked by acetylated H3 lysine 27 (H3K27ac) while poised enhancers are marked by H3K27me3. These poised enhancers are inactive in human ESC and are shown to be only active during late developmental stages (Rada-Iglesias *et al*, 2011). To further explore the potential repressive role of PRDM14 in human ESC, we expanded the analysis of PRDM14 genome-wide binding overlap with known histone modification marks (Figure 27A). In addition to OCT4 and NANOG, PRDM14 binding seems to cluster with H3K27me3, H3K4me1 and p300 (black box), which together marks poised enhancers. An overlap of PRDM14 binding sites with the active and poised enhancers identified in Rada-Iglesias *et al* showed that PRDM14 indeed preferentially binds poised enhancers (Figure 27B). Thus, in addition to the pluripotency genes, PRDM14 is found to bind strongly to poised elements of repressed genes in human ESC.

4.2 PRDM14 binding correlates with the repressive H3K27me3 mark

The repressive H3K27me3 mark is a hallmark for poised enhancers and bivalent domains and is responsible for silencing these elements. The high overlap of PRDM14 with poised enhancers highlights PRDM14's potential regulation of this chromatin modification for its repressive activities. Thus, it would be intuitive to investigate if PRDM14 binding is correlated with this repressive mark. For this analysis, PRDM14 bound sites are first divided into 4 quartiles with different binding intensities (Figure 28A). Interestingly, it was observed

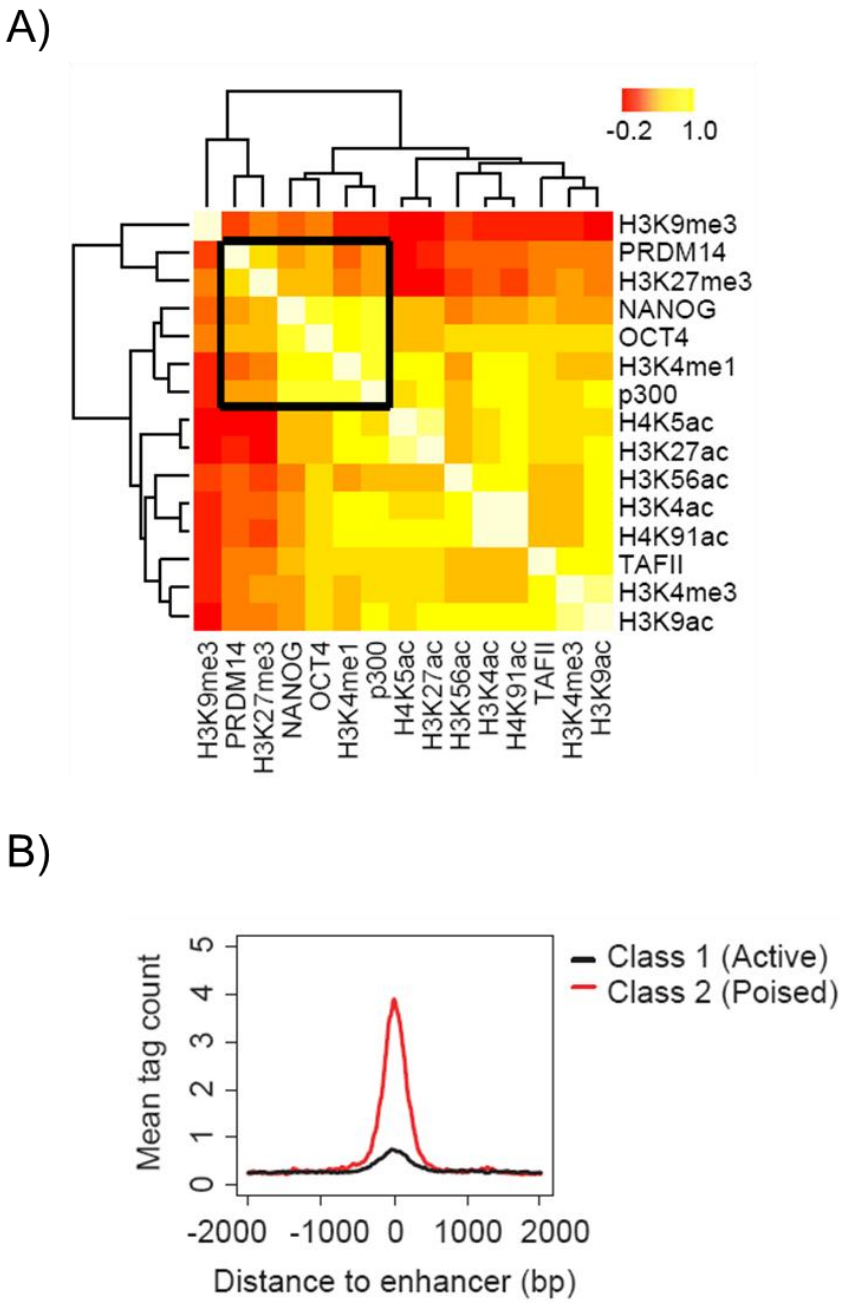


Figure 27: PRDM14 binding overlaps with poised enhancers. A) Heatmap showing correlations of PRDM14 with histone modifications and transcription factors. The heatmap shows two main clusters; a repressive cluster that includes H3K9me3, PRDM14 and H3K27me3 and an active cluster that include OCT4, NANOG, p300 and histone marks associated with active enhancers and promoter. The two cluster show overlaps that include H3K27me3, H3K4me1 and p300 which are hallmarks of poised enhancers (black box). B) Relative enrichment of PRDM14 binding peaks in Class 1 active enhancer and Class 2 poised enhancers.

that the level of the H3K27me3 mark increases with the binding intensities of PRDM14 (Figure 28A). This result suggests a strong correlation between PRDM14 binding and the H3K27me3 mark. While p300 is associated with the H3K27me3 mark at poised enhancers, its correlation with H3K27me3 is weak compared to PRDM14 (Figure 28B). Conversely, p300 binding intensity is highly correlated with the active H3K27ac (Figure 28B), which corroborates with its use for identifying enhancers in vivo (Visel *et al*, 2009). Both p300 and PRDM14 binding intensity correlated well with the H3K4me1 mark that is located in both classes of enhancers (Figure 28B). The difference in H3K27me3 correlation between p300 and PRDM14 suggest that PRDM14 association with the repressive marks extend beyond just the poised enhancers. Indeed, a Pearson correlation analysis showed that the H3K27me3 is the most highly enriched chromatin modification mark compared to 10 other histone modifications mapped in human ESC (Figure 29). In summary, a strong correlation between PRDM14 binding and H3K27me3 modification was observed and H3K27me3 modification is highly enriched at PRDM14-bound sites. These observations suggest that PRDM14 potentially mediates silencing through H3K27me3. While PRDM14 contains the SET domain, no enzymatic activity has been reported thus far for this domain. Whether any PRDM14 enzymatic activity is responsible for H3K27me3 remains to be validated.

4.3 PRDM14 interacts with PRC2

The H3K27me3 modification plays a central role in PRC2 mediated repression (Cao *et al*, 2002; Müller *et al*, 2002; Czermin *et al*, 2002). The drosophila PRC2 component Enhancer of Zest (E(Z)) exhibits methyltransferase activity that is responsible for H3K27me3 in PRC2 bound genes (Müller *et al*, 2002; Czermin *et al*, 2002). This activity is also conserved in the human homolog EZH2 (Cao *et al*, 2002). The strong correlation between the H3K27me3

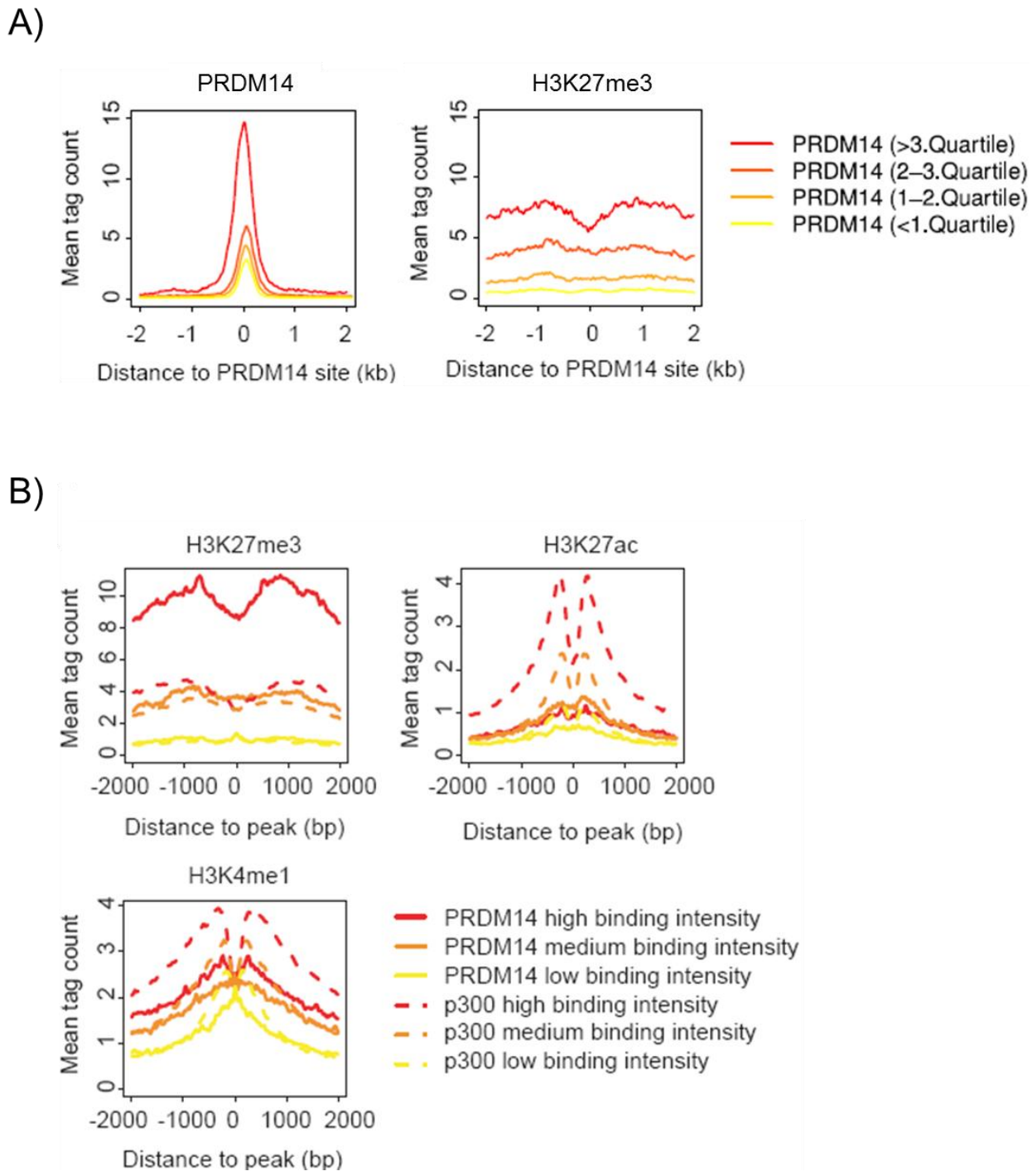


Figure 28: PRDM14 binding correlates with H3K27me3 modification. A) PRDM14-bound sites are divided into 4 quartiles according to their binding strength. The relative enrichment of H3K27me3 modification is tabulated for each quartile. B) Similar analysis of the p300-bound loci with histone modification H3K27me3, H3K27ac and H3K4me3. p300 bound sites were divided into 3 subgroups according to their binding strength (low, medium and high). PRDM14-bound sites were similarly analyzed for parallel comparison.

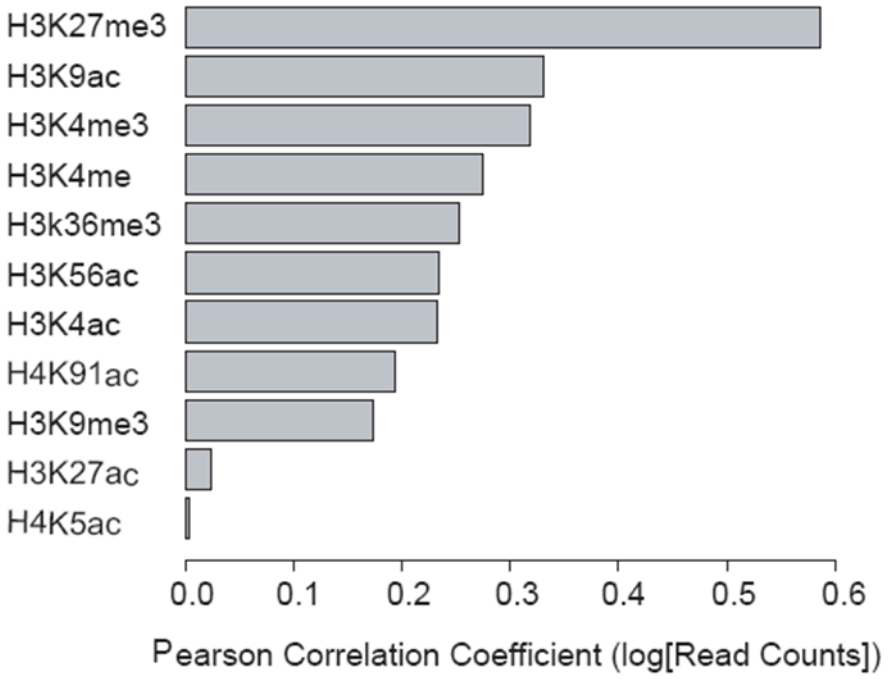


Figure 29: Correlation of PRDM14-bound sites with known histone modifications. The Pearson correlation between ChIP-Seq signal strength of PRDM14 and 11 other known histone modifications in human ESC within 4kb of PRDM14-bound sites. Significant enrichment of H3K27me3 modification in PRDM14 binding sites compared to other histone modifications.

modification and PRC2 has also been established in both human and mouse ESC (Boyer *et al*, 2006; Lee *et al*, 2006); PRC2 binds extensively to lineage specific genes marked with H3K27me3 in ESC. Thus, it would be apparent to investigate whether PRDM14 recruits PRC2 for the H3K27me3 modifications at its binding sites. Potential PRDM14 and PRC2 interaction was first determined using co-immunoprecipitation (Co-IP) assays with whole cell lysate of human ESC. Both SUZ12 and EZH2 were identified in the immunoprecipitation with PRDM14 antibody (Figure 30A) and *vice versa*, PRDM14 was co-immunoprecipitated with EZH2 (Figure 30B). This interaction was further supported with Co-IP using HEK-293T cells ectopically expressing MYC-tagged EZH2 and HA-tagged PRDM14 (Figure 30C). These results suggest that PRDM14 potentially interacts with the PRC2 complex. To probe whether PRDM14 directly interacts with the PRC2 complex, a GST-tagged recombinant PRDM14 protein was generated for pull-down assays. A FLAG-tagged PRC2 complex was purified from Sf9 cells ectopically expressing the PRC2 components EZH2, SUZ12, PHF1 and FLAG-tagged EED. PRDM14 was identified in the immunoprecipitation with a FLAG-specific antibody, along with other components of PRC2 (Figure 31A). Similarly, EZH2, SUZ12 and FLAG-tagged EED were also identified in the immunoprecipitation with a GST-specific antibody (Figure 31B). Overall, the biochemical studies supported the direct interaction between PRDM14 and the PRC2 complex.

4.4 PRDM14 regulates PRC2 binding at its bound loci

The direct interaction between PRDM14 and PRC2 prompts further investigation for PRC2 activity at PRDM14-bound loci in human ESC. To look at the overlap between PRC2 binding and PRDM14, I mapped the genome-wide binding profile of EZH2 in human ESC. EZH2 global binding generated was subsequently overlapped with the NANOG and PRDM14

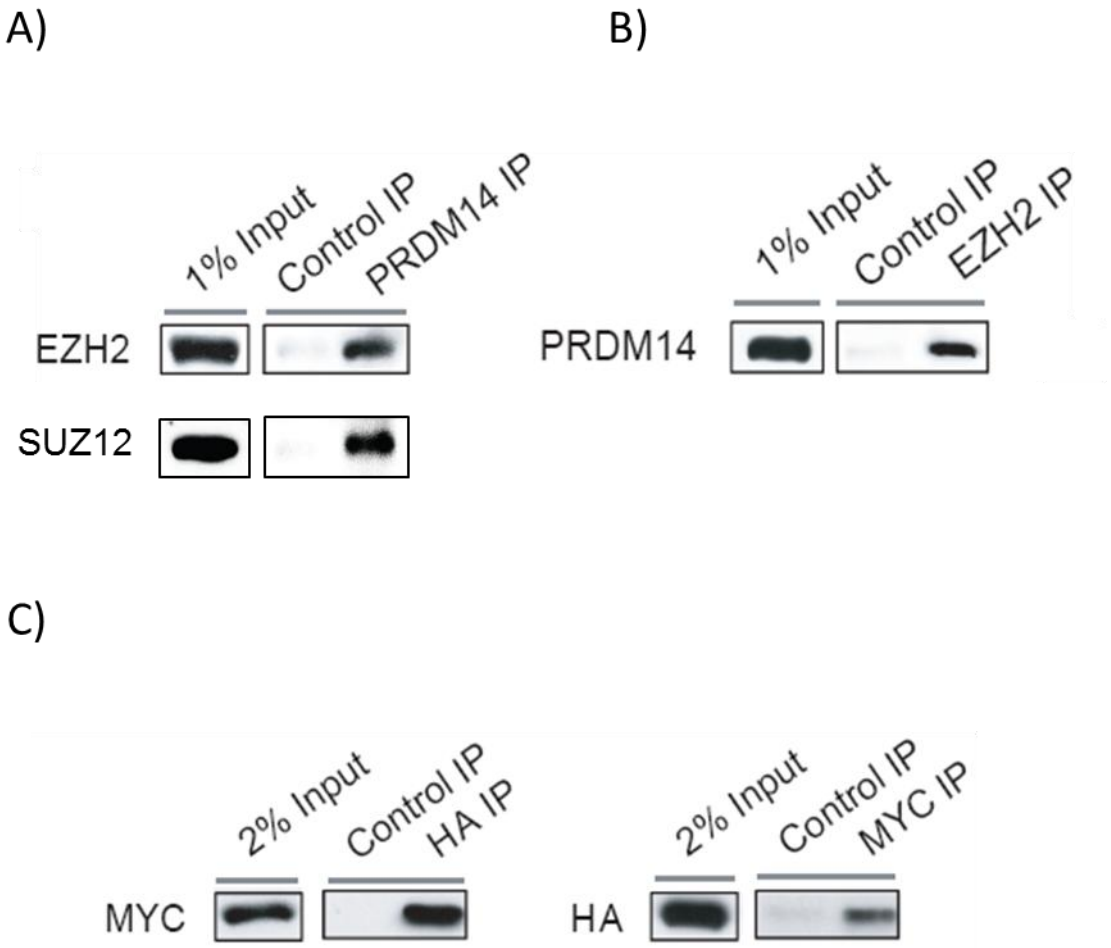


Figure 30: PRDM14 interacts with PRC2 complex. A) Endogenous co-immunoprecipitation (Co-IP) assay using human ESC whole cell lysate. PRDM14 specific antibody was used for the IP and anti-GST as the control. EZH2 and SUZ12 were co-immunoprecipitated with PRDM14. B) Endogenous Co-IP with EZH2 antibody. PRDM14 is detected in the EZH2 IP and not the control IP. C) Co-IP assay with Myc-tagged EZH2 and HA-tagged PRDM14 ectopically expressed in 293T cell. IP was performed with antibodies against the MYC and HA tag on the respective proteins. MYC-tagged EZH2 is detected in the HA-IP and *vice versa.*

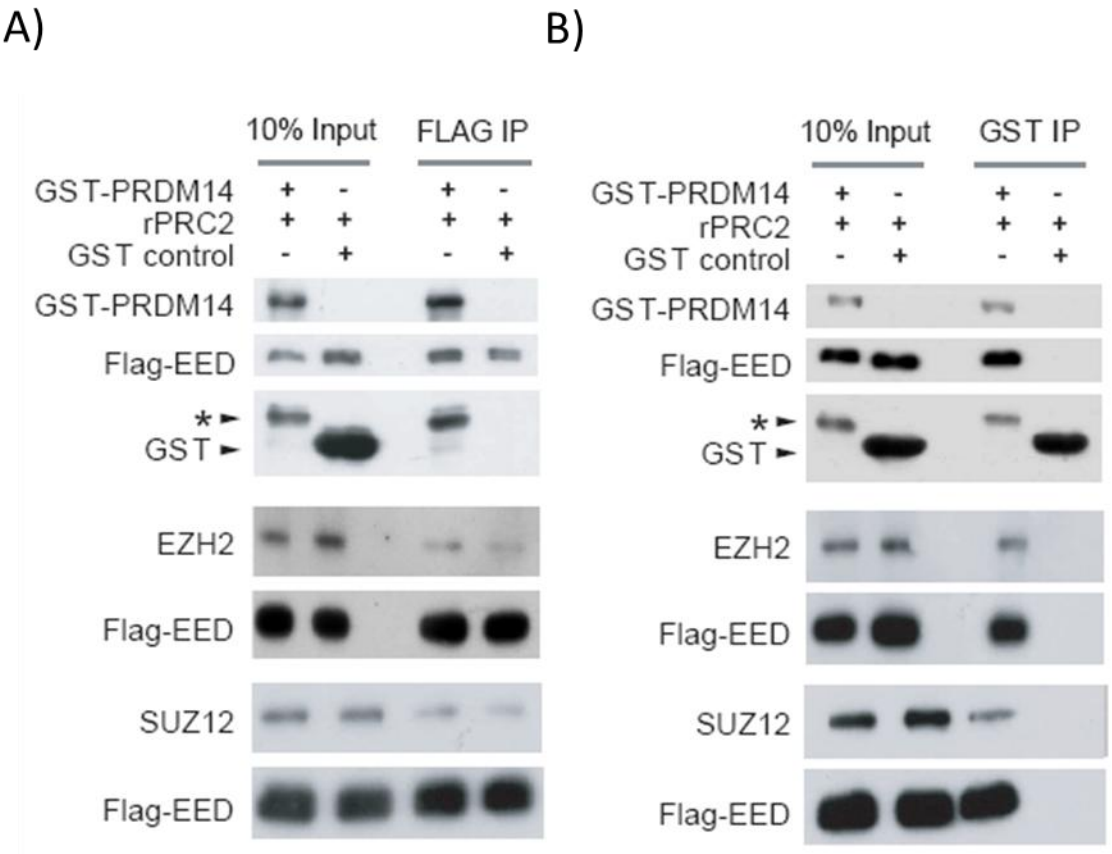


Figure 31: Direct interaction between PRDM14 and PRC2 complex. A) Recombinant GST-PRDM14 or GST was incubated with purified PRC2 complex containing Flag-tagged EED and subjected to IP with A FLAG tag specific antibody. GST-PRDM14 and not GST is detected in the FLAG-IP. Other components of the PRC2 complex SUZ12 and EZH2 are also detected. * A specific background band (likely to be degraded GST-PRDM14) in the GST-PRDM14 sample that is also detected in the FLAG-IP. B) Similar pull-down assay with GST-specific antibody. Components of PRC2 complex is only detected in IP with samples where PRC2 is incubated with GST-PRDM14.

binding profiles. Compared to the non-enriched DNA used for ChIP (INPUT), there is clear enrichment of EZH2 binding at PRDM14 binding sites (Figure 32). This enrichment is however not observed for NANOG, suggesting a PRDM14 specific interaction with PRC2 across the genome. This overlap is supported by the enrichment of active H3K27ac for NANOG binding sites and an enrichment of H3K27me3 at PRDM14 binding sites. Overall, the analysis of global binding profiles suggests that PRC2 may be present at PRDM14 binding sites. Indeed, EZH2 and SUZ12 binding was detected at 16 PRDM14-bound loci near development genes with H3K27me3 modifications, using ChIP-qPCR (Figure 33). To validate if PRC2 binding at these loci are indeed dependent on PRDM14, the ChIP assays were repeated with PRDM14 depleted human ESC. The depletion of PRDM14 resulted in a decrease in both EZH2 and SUZ12 binding at these PRDM14 loci and this is accompanied by the decrease in H3K27me3 levels (Figure 33). Importantly, this reduction in EZH2 and SUZ12 binding and H3K27me3 levels is not due to the decrease in total protein levels (Figure 34A); neither is it due to the differentiation of ESC as NANOG levels are not reduced when the cells are harvested for ChIP (Figure 34A). The result of the histone H3 ChIP (Figure 34B) also showed that the decrease in binding is not a result of nucleosome depletion at these sites after PRDM14 depletion. Overall, the ChIP assays show that PRC2 binds to PRDM14 bound sites in the human ESC and its binding at these sites are dependent on PRDM14.

4.5 PRDM14-bound loci exhibit repressive activity

Thus far, it is observed that many PRDM14 bound sites on developmental genes are also bound by PRC2 and is highly enriched with H3K27me3. However, there has been no report so far on the repressive activity of PRDM14 bound elements. I successfully isolated a ~500bp

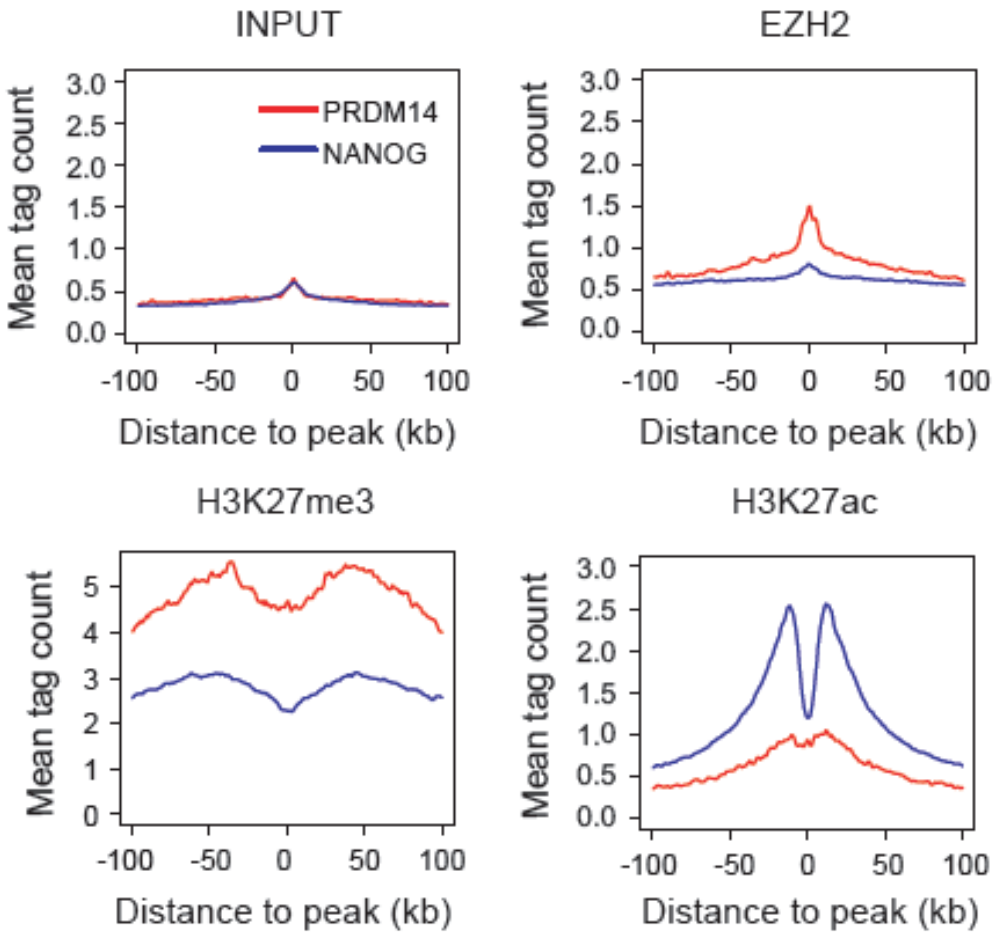


Figure 32: Enrichment of EZH2 binding at PRDM14 binding sites. Enrichment of ChIP-Seq tag counts of EZH2, H3K27me3 and H3K27ac within 100bp of PRDM14 and NANOG binding sites. Significant enrichment of EZH2 at PRDM14 and not at NANOG binding sites compared to input. This corresponds to a stronger enrichment of H3K27me3 at PRDM14-bound loci and H3K27ac at NANOG-bound loci.

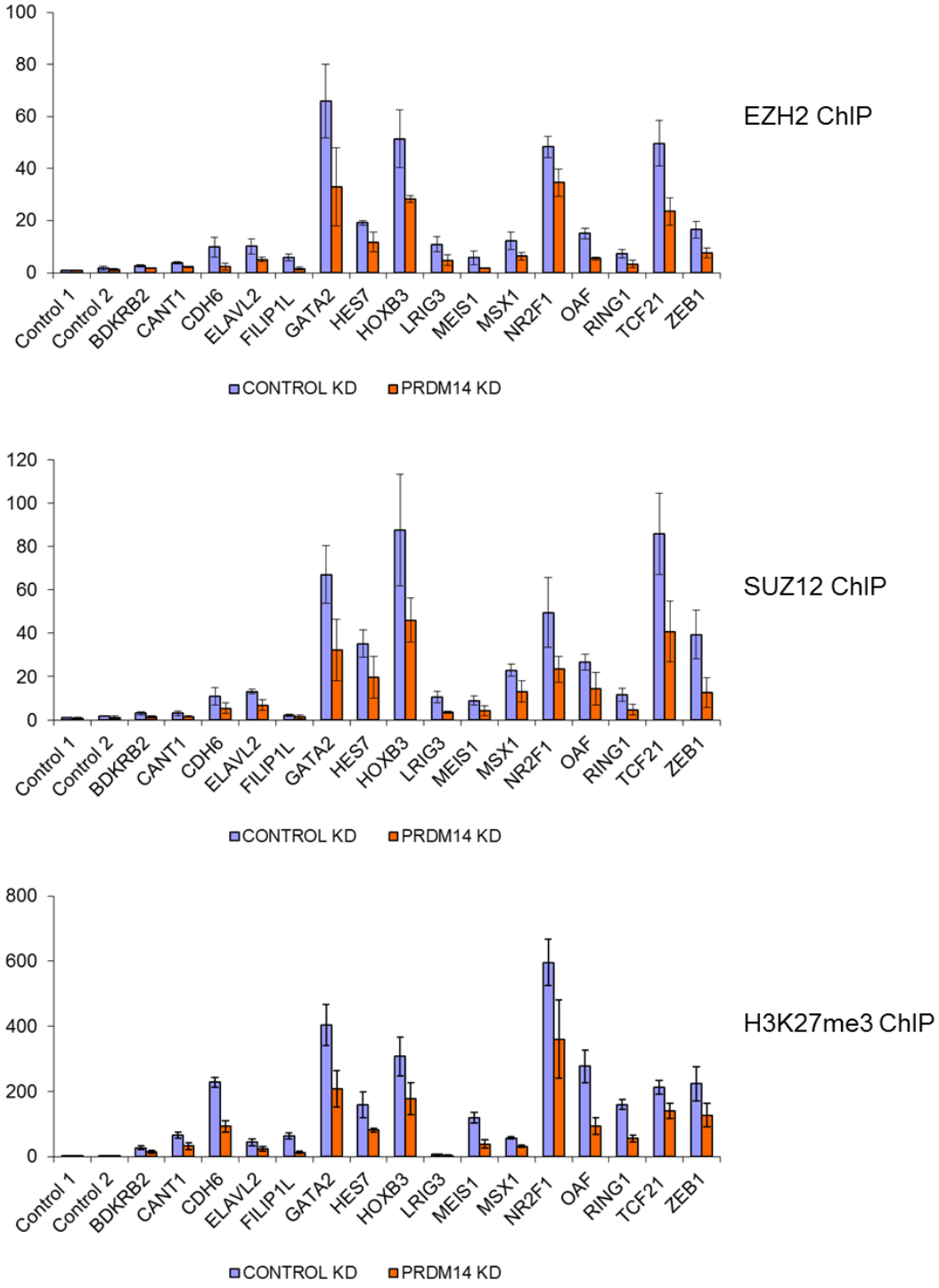


Figure 33: PRC2 binding and H3K27me3 modification at PRDM14-bound loci. ChIP assays using antibodies specific for EZH2, SUZ12 and H3K27me3 with human ESC transfected with shRNA targeting *Luciferase* (control) or *PRDM14*. EZH2 and SUZ12 binding were detected at 16 selected PRDM14-bound loci near developmental genes. The binding of EZH2 and SUZ12 were decreased after PRDM14 depletion. The levels of H3K27me3 in these loci were also decreased after PRDM14 depletion.

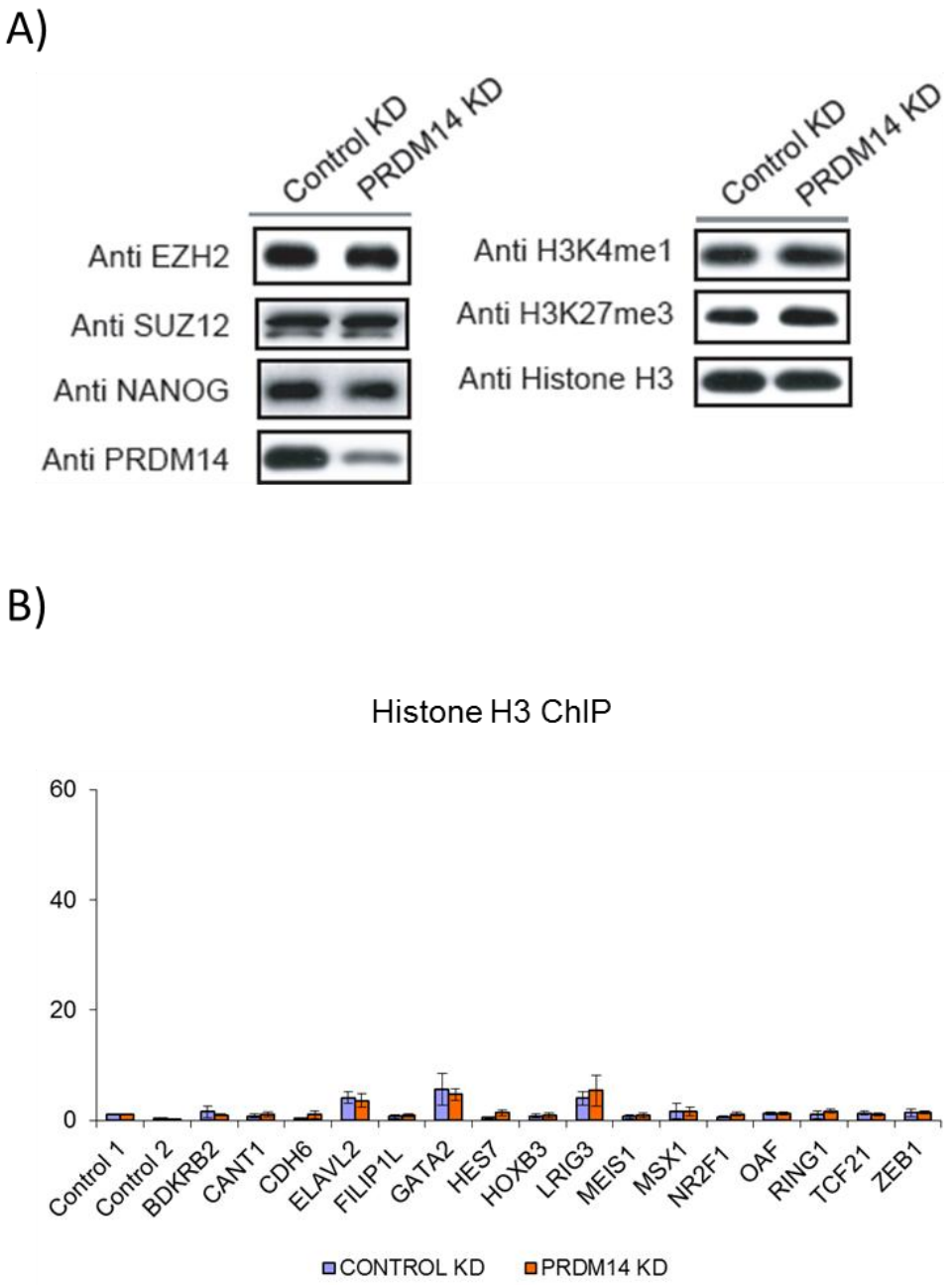


Figure 34: PRC2, NANOG and histones protein level during early PRDM14 depletion. A) The total protein level of EZH2, SUZ12, NANOG, PRDM14, H3K4me1, H3K27me3 and histone H3 in chromatin samples from human ESC transfected with *PRDM14* or *Luciferase* shRNA for 40hrs. The total level for each protein remains largely unchanged compared to PRDM14 in cells transfected with the *PRDM14* targeting shRNA. B) Histone H3 ChIP in PRDM14 depleted cells. Levels of histone H3 remains unchanged at the 16 loci analyzed for PRC2 binding.

PRDM14-bound loci on the *HES7* gene and cloned this short DNA fragment upstream of a highly active phosphoglycerate kinase (*PGK*) gene promoter driving a reporter *Luciferase* gene (Figure 35A). This DNA fragment contains a conserved PRDM14 binding motif. The reporter was used to test for PRDM14 specific repressive activity in human ESC. The presence of this PRDM14-bound locus was able to reduce 50% of the *PGK* promoter activity. This repressive activity is dependent on the PRDM14 motif as two point mutations on the motif abolish the repressive activity of the entire 500bp loci. More importantly, this activity is dependent on both PRDM14 and PRC2 as the depletion of either resulted in de-repression of the *PGK* promoter activity (Figure 35B). Thus, the results showed that PRDM14-bound loci in the human ESC have repressive activities and this repression is dependent on both PRDM14 and PRC2.

The use of a ~500bp DNA fragment may contain regulatory elements that could be unknowingly mediating the PRC2 activity. To further show that this repressive activity is specific to PRDM14, I cloned two tandem repeats of the PRDM14 binding sites in place of the 500bp DNA fragment in the reporter construct (Figure 36A). The tandem repeats of the 16bp PRDM14 motif also repressed the *PGK* promoter activity (Figure 36B) and point mutations introduced into both motifs abolished the repression (Figure 36B). Similarly, the repressive activity of this tandem repeats of PRDM14 motif is dependent on both PRDM14 and PRC2. Importantly, this result suggests that the PRDM14 motif alone is sufficient for PRDM14 binding and recruitment of PRC2 to mediate gene repression.

In summary, there is strong evidence to suggest that PRDM14 plays a repressive role in human ESC. PRDM14 represses developmental genes in the human ESC, thus accounting for the observation that PRDM14 over expression in human ESC derived embryoid bodies supressed the expression of lineage-specific genes (Tsuneyoshi *et al*, 2008). In addition, PRDM14 interacts with PRC2 and mediates its binding at its loci in human ESC,

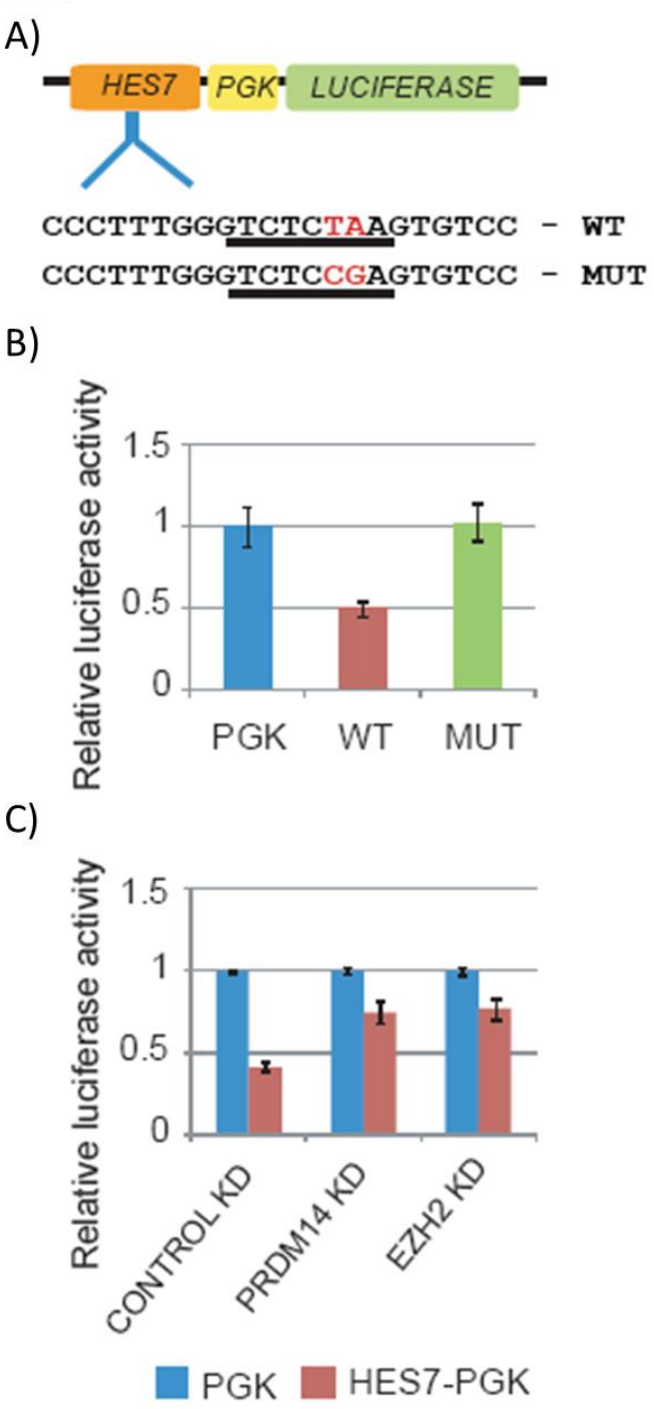


Figure 35: Repressive function of PRDM14-bound site in *HES7* gene. A) A schematic representation of the reporter construct. A ~500 bp genomic loci in the intron of *HES7* gene containing PRDM14 binding site was cloned before the *PGK* gene promoter driving the *Luciferase* gene in the reporter cassette. B) The *HES7* genomic fragment containing the PRDM14 binding motif results in a repression of the *PGK* promoter activity (WT). Point mutations in the PRDM14 binding motif led to a loss of repression activity (MUT). Relative luciferase activity is obtained via normalization against the sample transfected with the control PGK reporter construct. C) Repressive activity of the *HES7* loci is dependent on PRDM14 and EZH2. The reporter construct was co-transfected with shRNA construct targeting *PRDM14* or *EZH2*.

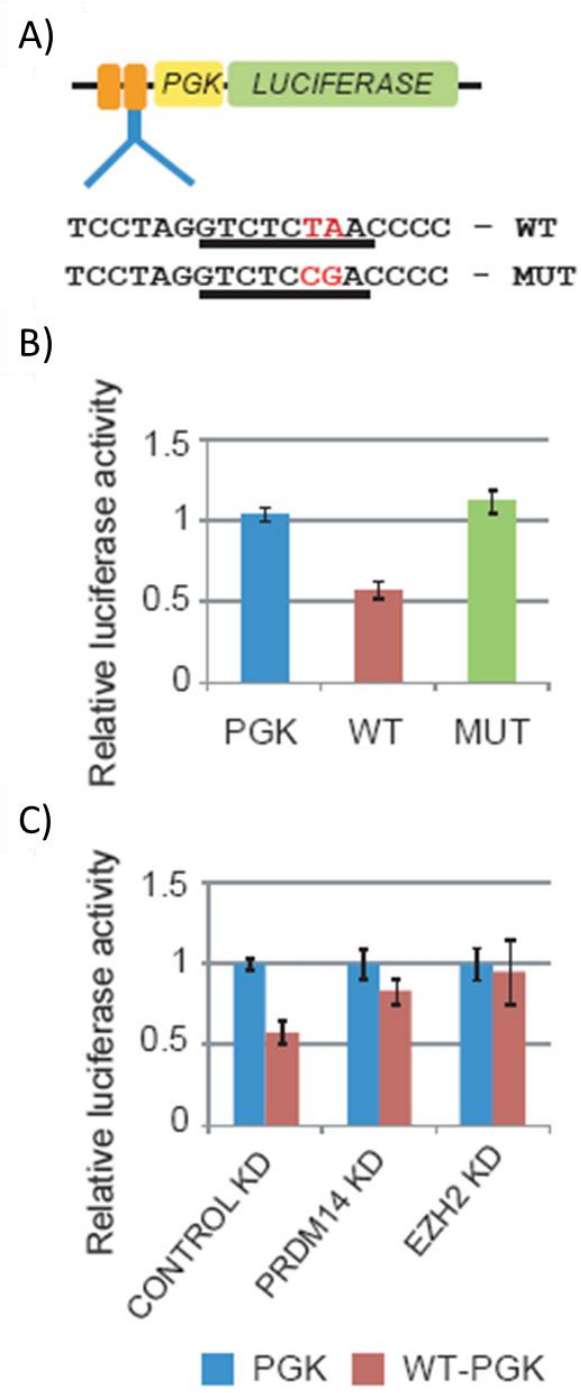


Figure 36: Repressive function of PRDM14 binding motif elements. A) A schematic representation of the reporter construct with two copies of the PRDM14 binding motif in place of the *HES7* DNA fragment. B) Addition of 2 tandem sequences of PRDM14 binding motif results in the repression of the *PGK* promoter activity. Point mutations in the two motifs results in a loss repression. Relative luciferase activity is obtained via normalization against the sample transfected with the control PGK reporter construct. C) Repressive activity of the PRDM14 motif is dependent on PRDM14 and EZH2. The reporter construct was co-transfected with shRNA construct targeting *PRDM14* or *EZH2*.

accounting for the H3K27me3 enrichment at PRDM14 binding sites. Functionally, PRDM14 interaction with PRC2 is important for its repressive activities in human ESC.

4.6 A repressive role for PRDM14 during iPSC reprogramming

As PRDM14 enhances iPSC reprogramming, it would be of interest to investigate if the repressive role of the PRDM14 could help explain this enhancement. Understanding how PRDM14 increases reprogramming efficiency potentially identifies rate limiting steps in the process. In a bid to understand the molecular changes occurring during iPSC reprogramming, Samavarchi-Tehrani *et al* analysed the expression profiles throughout the reprogramming process and divided the process into 3 major stages (Samavarchi-Tehrani *et al*, 2010); the initiation, maturation and stabilization stages. Our studies on PRDM14 reprogramming revealed an enhanced upregulation of markers for each stages of reprogramming (Chia *et al*, 2010); however, the exact mechanism remains unclear.

To determine of PRDM14's repressive function plays a role in enhancing the reprogramming process, it would be intuitive to look at whether PRDM14 could be repressing its target developmental genes that are expressed in the fibroblast. To look at the specific ectopic effect of PRDM14 under the 4F reprogramming process, I first determined the induction of *PRDM14* by the 4F during reprogramming. Endogenous *PRDM14* was induced 7 days post infection (d.p.i) in the human fibroblast MRC-5 (Figure 37). Thus, the ectopic effect of PRDM14 under 4F mediated reprogramming was measured at earlier time points. PRDM14 target genes *OAF*, *LMNA* and *USP3* were not repressed or even increased under 4F reprogramming at 4 d.p.i (Figure 38A). The co-induction of PRDM14 rapidly repressed these

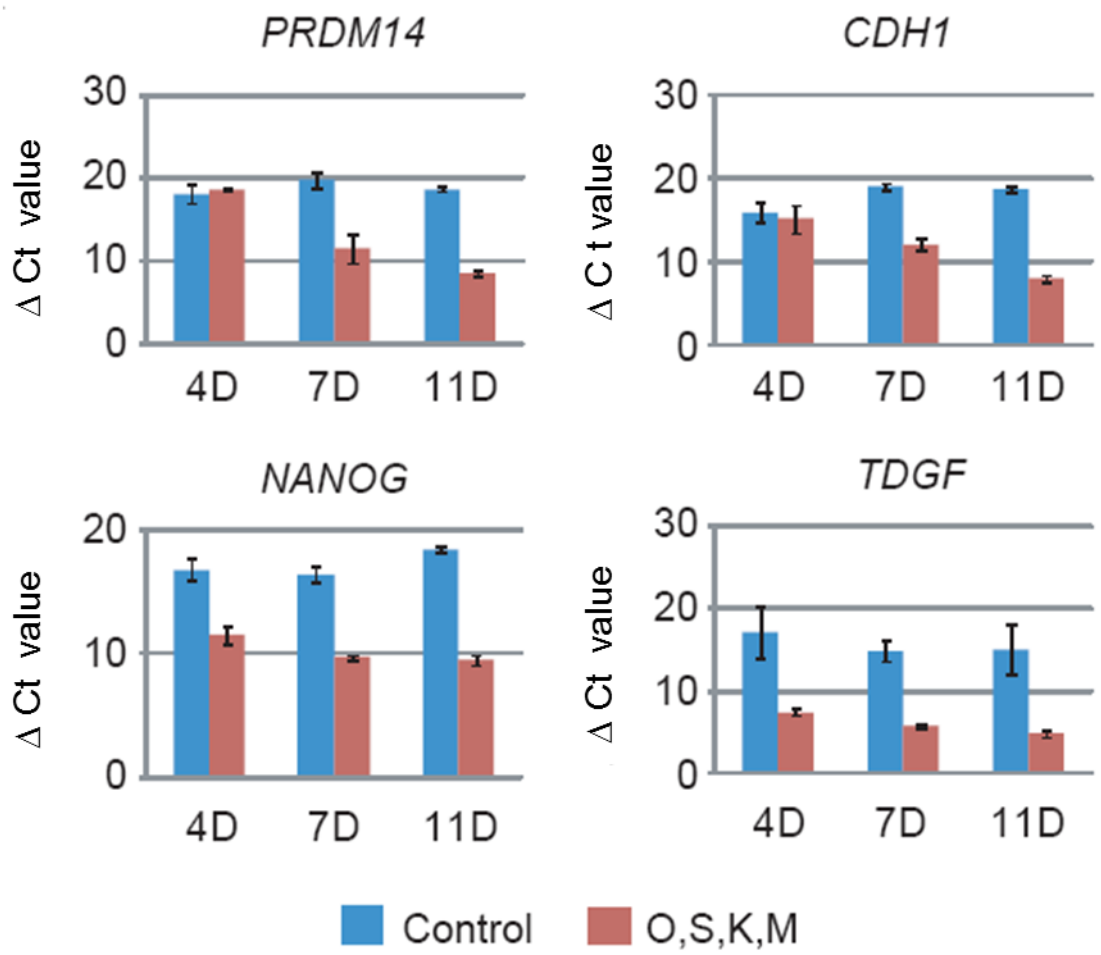


Figure 37: Induction of *PRDM14* expression during reprogramming. Time course analysis of the gene expression of *PRDM14* and pluripotency associated genes in MRC-5 fibroblast infected with retrovirus ectopically expressing the 4 reprogramming factors OCT4 (O), SOX2 (S), KLF4 (K) and c-MYC (M) or GFP (control). The ΔCt value of each gene is obtained after subtraction with the Ct value of GAPDH. An increase expression is represented by a decrease in ΔCt value.

three genes (Figure 38A). While PRDM14 target genes *IGFBP5*, *CDC92*, *MYH9* were repressed under the 4F reprogramming, the co-induction of PRDM14 resulted in a stronger repression of these three genes (Figure 38B). Thus, PRDM14 can indeed suppress its target genes during reprogramming with the 4F.

The early stage of reprogramming or the initiation of reprogramming involves a mesenchymal-to-epithelial transition (MET) process (Samavarchi-Tehrani *et al* 2010; Li *et al* 2010). This MET process during iPSC reprogramming involves the repression of the key mesenchymal regulators *SNAIL*, *SLUG* and *ZEB1*; Li *et al* reported the importance of BMP4 signalling during early stages of reprogramming to suppress *Zeb1* and *Zeb2*. On the other hand, Samavarchi-Tehrani *et al* highlighted the importance of suppressing the TGFβ signalling that activates the expression of these mesenchymal genes. Although *ZEB1*, *SNAIL* and *SLUG* are rapidly repressed by the 4F at 4 d.p.i, the co-induction of PRDM14 further enhanced the repression of *ZEB1*, *SNAIL* and *SLUG* (Figure 39). The genome-wide binding profile of PRDM14 in human ESC suggests that PRDM14 binds *ZEB1* promoter (Figure 40), along with enrichment in H3K27me3. To verify if PRDM14 directly binds *ZEB1* promoter in fibroblasts, I infected MRC-5 cells with a retrovirus ectopically expressing PRDM14 for ChIP assays. It is observed that the ectopic PRDM14 could bind directly to the *ZEB1* promoter in fibroblast (Figure 41A). PRDM14 binding was also detected at another target gene *OAF* compared to the two control regions, supporting the specificity of this PRDM14 binding activity in the fibroblast. Importantly, the binding of PRDM14 resulted in a significant enrichment of EZH2 at these two PRDM14 binding sites (Figure 41B). Consequently, the H3K27me3 levels also increased at these two loci (Figure 41C). This potential repression mediated by PRC2 recruitment to the loci and subsequent H3K27me3 enrichment is supported by a decrease in *ZEB1* expression in the fibroblasts ectopically expressing PRDM14 (Figure 41D).

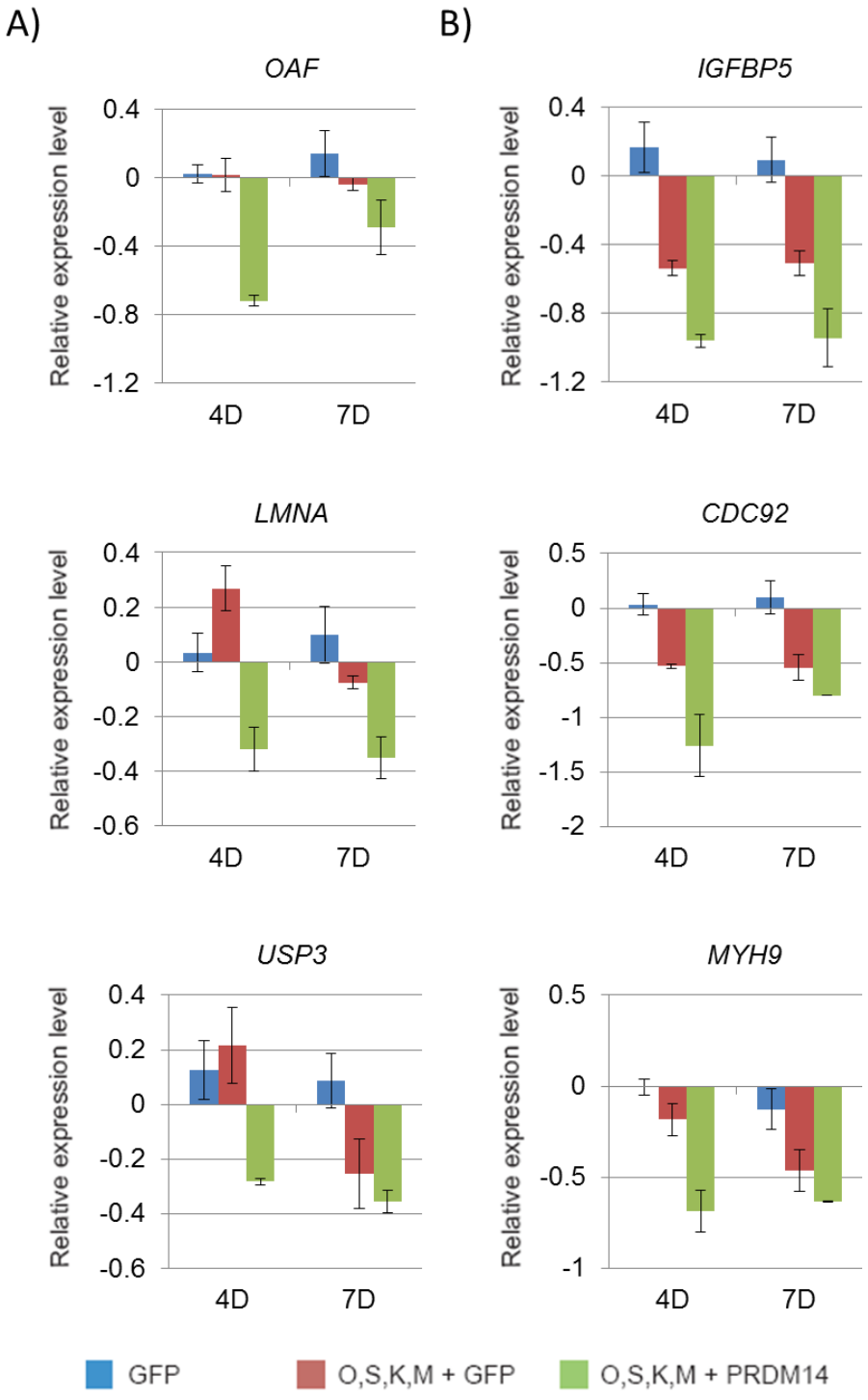


Figure 38: PRDM14 represses target genes in fibroblast during reprogramming. Time course expression analysis of PRDM14 target genes in fibroblast reprogrammed by O,S,K,M with or without PRDM14. A) *OAF*, *LMNA* and *USP3* which are not repressed at 4 d,p.i under the OSKM reprogramming are highly repressed in the presence of PRDM14. B) *IGFBP5*, *CDC92* and *MYH9* are repressed at 4 d,p.i under the OSKM reprogramming and are further repressed in the presence of PRDM14.

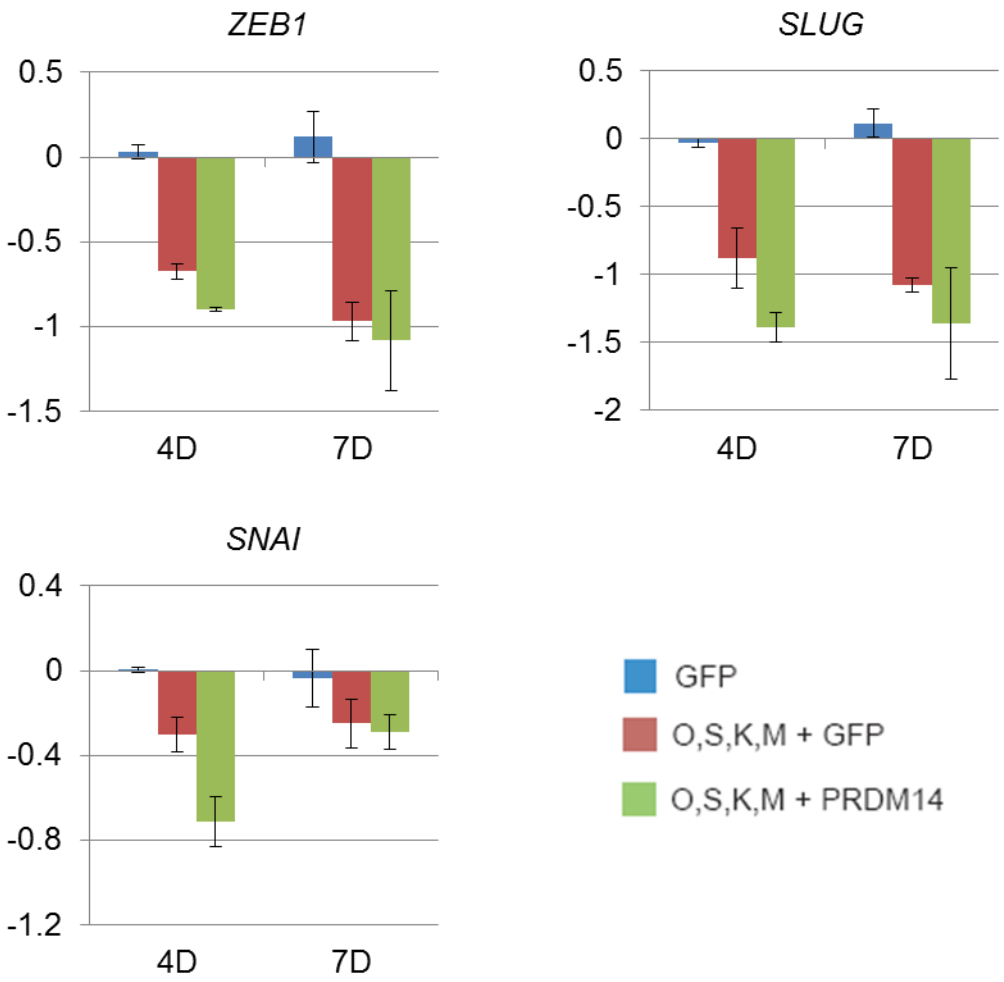


Figure 39: Enhanced repression of MET genes by PRDM14. Expression levels of key mesenchymal regulators *ZEB1*, *SNAI* and *SLUG* during the process of reprogramming by O,S,K,M with and without PRDM14. Co-induction of PRDM14 further suppresses the expression of *ZEB1*, *SNAI* and *SLUG* at 4 d.p.i.

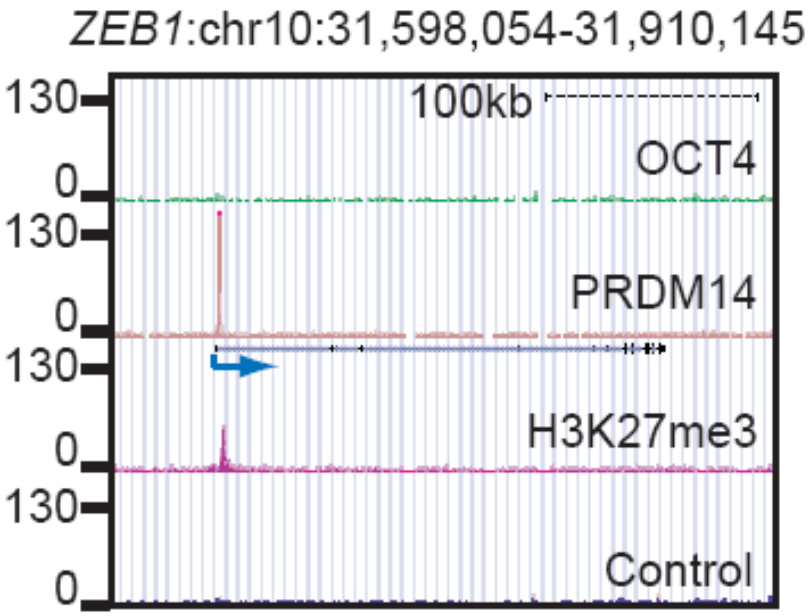


Figure 40: PRDM14 binds to *ZEB1* gene in human ESC. A Snapshot of PRDM14, OCT4 and H3K27me3 ChIP-Seq profiles at *ZEB1* loci in human ESC. Strong PRDM14 binding at *ZEB1* promoter is accompanied by enriched H3K27me3 modification. This enrichment is not seen in the ChIP-Seq profile of OCT4 and control input DNA.

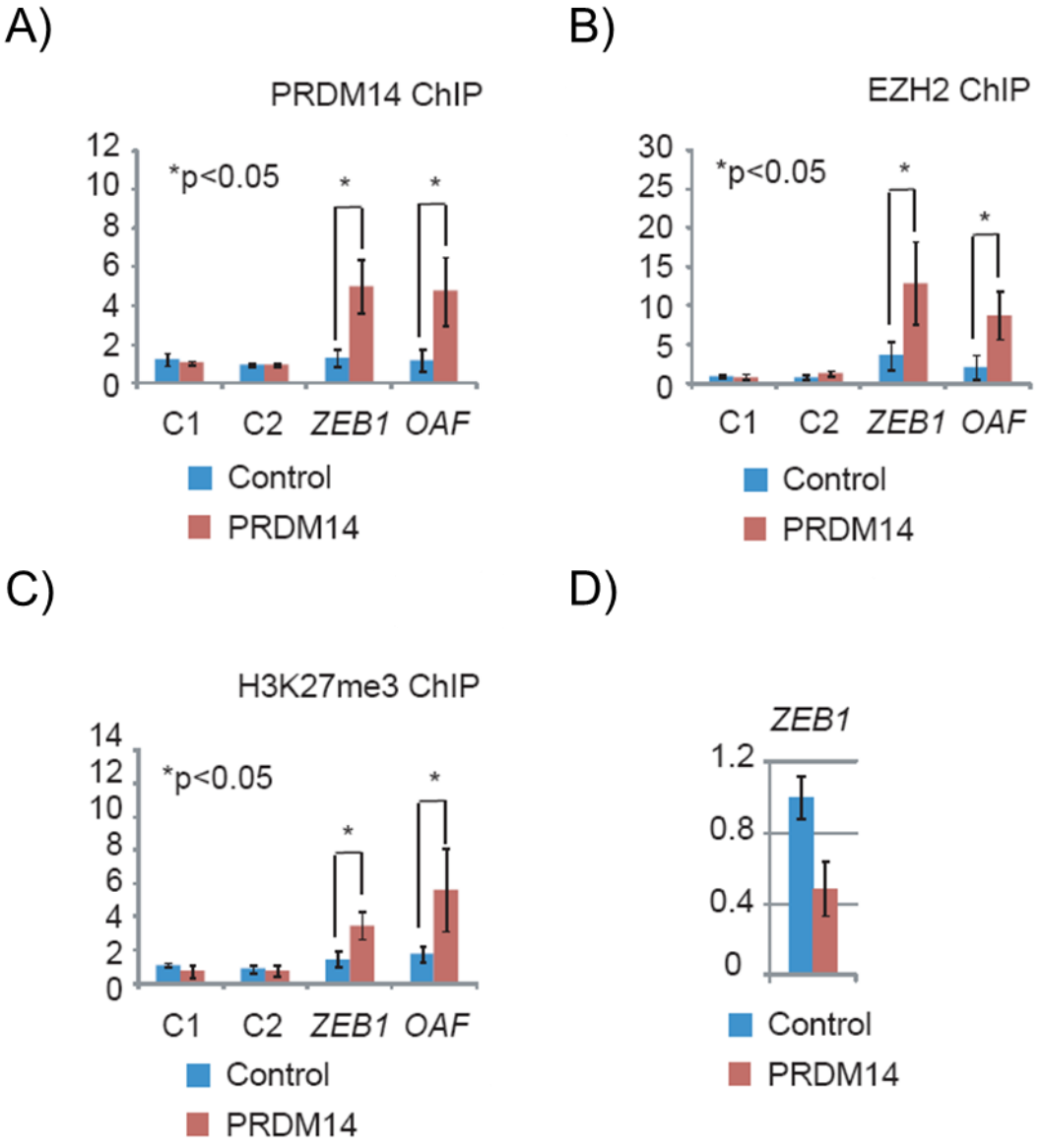


Figure 41: PRDM14 binds and recruits PRC2 to repress *ZEB1* expression in the fibroblast. MRC-5 fibroblasts are infected with retrovirus ectopically expressing PRDM14 or GFP and harvested for ChIP assays 5 d.p.i. A) PRDM14 ChIP. Enrichment of PRDM14 binding is detected at the *ZEB1* and *OAF* loci compared to control region 1 and 2 which shows background level of enrichment. B) EZH2 ChIP shows enrichment of EZH2 at the PRDM14 binding sites. C) H3K27me3 ChIP shows corresponding increase in H3K27me3 levels at these regions. D) *ZEB1* expression level is decreased in the PRDM14 overexpressing cells.

Overall, the results in this section showed that the repressive function of PRDM14 also plays a role in PRDM14 mediated reprogramming. PRDM14 is able to repress its target genes in the fibroblast to efficiently suppress the expression of these developmental genes under 4F mediated reprogramming. PRDM14 can bind directly to its target genes and recruit PRC2 for the repression of the gene promoter. This includes a key mesenchymal gene *ZEB1* whose suppression is important for the initiation of reprogramming. Onder *et al* highlighted the effect of ZEB1 in inhibiting reprogramming efficiency in a screen for chromatin modifying enzymes that regulates reprogramming efficiency (Onder *et al*, 2012). The findings in the paper showed that the inhibition of DOT1L resulted in increased MET efficiency and overall reprogramming efficiency, but the overexpression of ZEB1 abolishes the effect. In summary, PRDM14's repressive role potentially enhances reprogramming by the repression of developmental genes in fibroblasts, including a key mesenchymal regulator to facilitate MET.

4.7 Summary and discussion 2

Interestingly, while PRDM14 partners NANOG in human ESC to regulate pluripotency genes such as *OCT4*, it also partners PRC2 in human ESC to repress developmental genes. The enrichment of EZH2 at PRDM14-bound loci and not at NANOG-bound loci (Figure 32) suggests that PRDM14 selectively partners NANOG and EZH2 in regulating different classes of genes in human ESC. While it is not known how PRDM14 selectively partners NANOG and EZH2 in regulating different classes of genes, the data presented here strongly supports a repressive function for PRDM14 in human ESC. PRDM14 binding sites are highly correlated with H3K27me3 and this could be accounted for by the recruitment of PRC2 to its binding sites. This interaction between PRDM14 and PRC2 is functionally significant as the repressive activity of PRDM14 at its bound sites is dependent on PRC2. In human ESC,

PRDM14 recruits PRC2 to well-known developmental genes such as HES7 and Homeobox genes MSX1/Hox7 and HoxB3 (Figure 33). The mouse HES7 is specifically expressed in presomitic mesoderm and plays and essential role in its formation (Bessho et al, 2001). PRDM14 binds directly to the HES7 gene and recruits PRC2 to the gene locus. This locus displays PRDM14 and PRC2 dependent repressive activity (Figure 35). These results collectively show that PRDM14 plays an important role in the repression of key developmental genes in maintaining ESC pluripotency. Correspondingly, PRDM14 enhances the repression of these developmental genes that are expressed in the fibroblast during reprogramming (Figure 38 and 39). This repressive role of PRDM14 is important for the repression of a key mesenchymal gene ZEB1 to facilitate the initial MET event during the process of reprogramming. In summary, PRDM14 interacts with PRC2 to modulate repressive function in maintaining and inducing the pluripotent cell state (Illustration 8).

While PRDM14 adds to the list of regulators of human ESC pluripotency, it is also an important regulator of one the most conserved chromatin modifier in eukaryotes. Since the discovery of the PcG proteins in *Drosophila* (Struhl G, 1981), the PRC2 complex has been the most conserved complex among the polycomb group proteins from *Drosophila* to mammals (Margueron and Reinberg, 2011). Homologs of the core subunits of PRC2 can even be found in unicellular fungi (Shaver *et al*, 2010) and plants (Birve *et al*, 2001; Hennig and Derkacheva, 2009). The highly conserved complex has been reported to play multiple roles; maintaining the cell state, differentiation of stem cells and cell proliferation in simple multiple cellular organisms to higher vertebrates (Margueron and Reinberg, 2011). Other than developmental roles, increasing number of studies have shown that the deregulation of PRC2 activity aids tumour progression; EZH2, the H3K27 methylase of the complex, is important for tumour suppression in acute T cell leukaemia (Ntziachristos *et al*, 2012; Simon

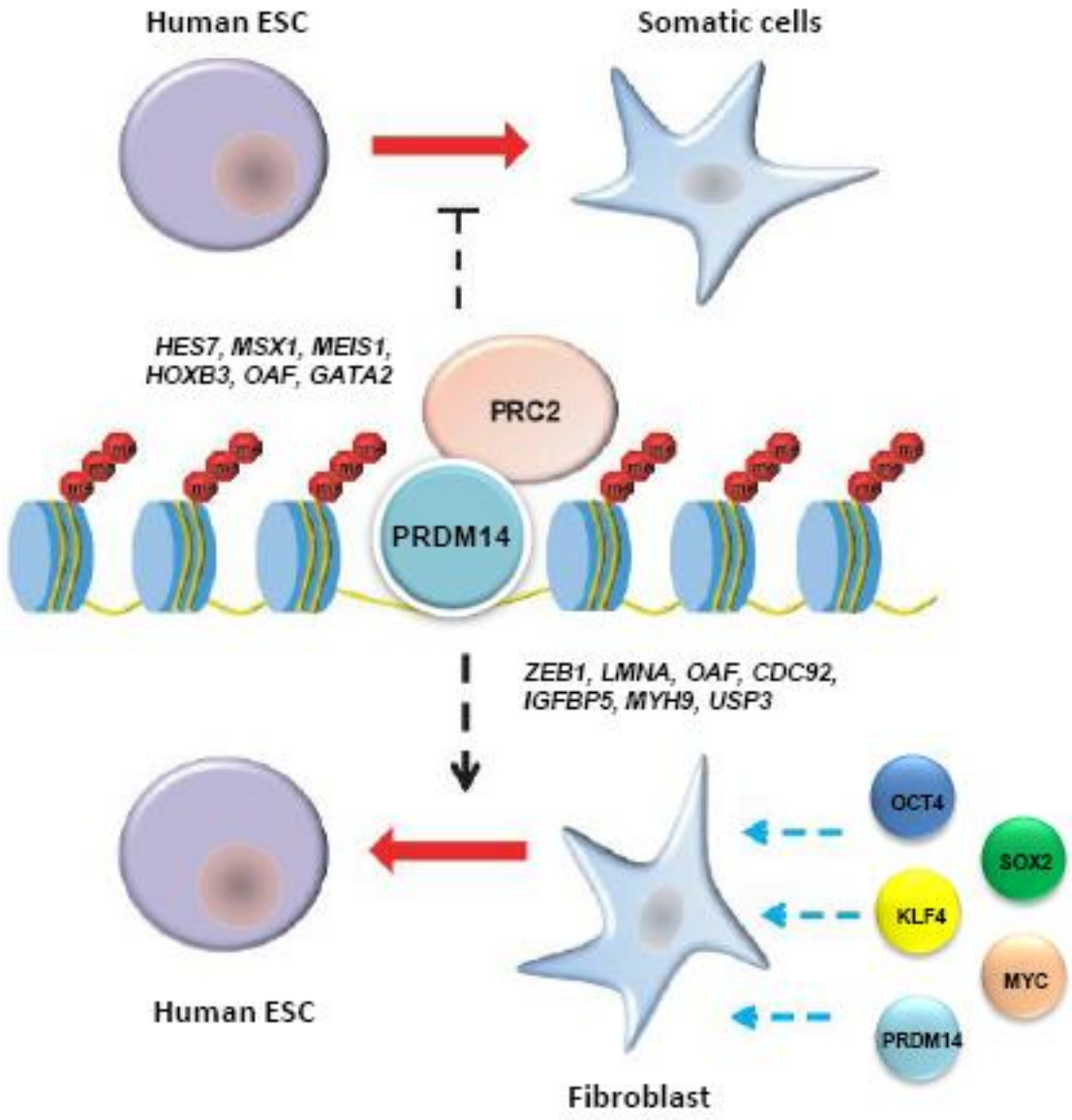


Illustration 8: A repressive function of PRDM14 in human ESC and reprogramming. PRDM14 recruits PRC2 to its bound loci for H3K27me3 mediated gene repression. This occurs for both its role in repressing developmental genes to block human ESC differentiation and in the repression of these genes expressed in fibroblast to facilitate iPSC reprogramming.

et al 2012) and conversely, upregulation of EZH2 promotes tumour progression via suppression of tumour suppressors in liver, breast and prostate cancer (Au *et al*, 2012; Ren *et al*, 2012). This context-dependency of PRC2 function further highlights the need to understand how this evolutionary conserved chromatin regulator regulates repression.

The recruitment of PRC2 to specific genomic loci has been a topic of interest. Specifically, little is known about how the mammalian PRC2 is recruited to the DNA (Simon and Kingston, 2009; Margueron and Reinberg, 2011). Jarid2 was identified as a new component

of PRC2 in mouse ESC (Peng *et al*, 2009; Shen *et al*, 2009; Pasini *et al* 2010). The genome-wide binding profile of Jarid2 extensively overlaps with PRC2 in mouse ESC and the molecule is shown to interact with the complex. Importantly, Jarid2 display a binding preference for (G+C)-rich sequences (Li *et al*, 2010), which is enriched in PRC2 binding in mouse ESC (Lee *et al*, 2006; Boyer *et al*, 2006). As Jarid2 is regarded as the potential DNA binding component of the PRC2 complex and generally binds to almost all PRC2 sites, how PRC2 activity is specifically targeted to selective genes in the cells to mediate repression remains largely unknown. The transcription factors which play specific roles in each cell type serve as ideal candidates in guiding PRC2 to specific target genes. In *Drosophila*, PRC2 binding to the polycomb responsive elements (PRE) is mediated by the PHO domain containing proteins (Mihaly *et al*, 1998; Brown *et al*, 1998). The mammalian ortholog Ying and Yang 1 (YY1) was thought to play similar roles at the mammalian PRE (Sing *et al*, 2009; Woo *et al*, 2009; Wilkinson *et al*, 2006). However, the genome-wide binding profile of YY1 did not significantly overlap with PRC2 (Xi *et al*; 2007) and the YY1 motif was not enriched in EZH2 binding sites in ESC (Ku *et al*, 2008). Till date, a role for transcription factors in regulating PRC2 recruitment in mammal remains poorly supported. Thus, this study served to

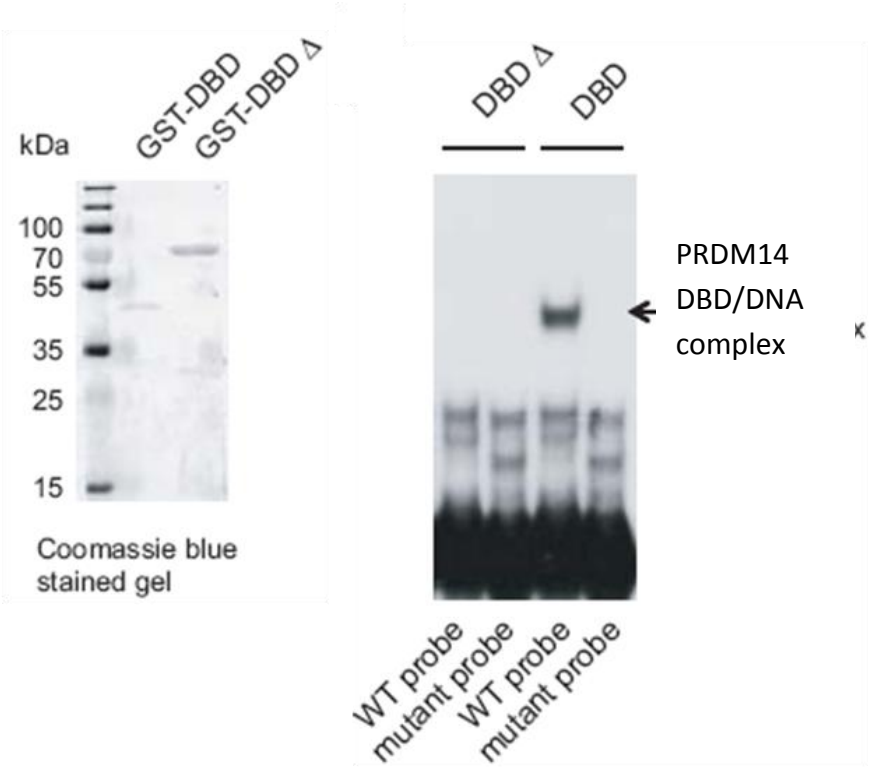
reinforce the role of transcription factors in mediating PRC2 binding activities. The specific DNA binding activity of PRDM14 helps to recruit PRC2 to specific developmental genes in human ESC and during the course of iPSC reprogramming.

On the other hand, the global binding overlap between PRDM14 and EZH2 is not as extensive as Jarid2, suggesting that PRDM14 is not a core component of PRC2 and PRDM14 specifically regulates PRC2 activity at a subset of PRC2 regulated genes. This observation fits the notion that PRDM14 plays a dual function in human ESC; PRDM14 activation of pluripotency associated genes with the core regulators does not involve PRC2. Thus, similar to YY1, PRDM14 may be regulating PRC2 activity only in a specific subset of PRC2 bound genes in human ESC. Importantly, this observation suggests that transcription factor mediated PRC2 recruitment is not likely to be global; multiple transcription factors could be responsible for directing PRC2 to specific subset of genes. Each transcription factor may be interacting with similar or different core components of the PRC2 complex to mediate the specific recruitment to its target genes. This proposed mode of regulation provides certain advantages to the regulation of PRC2 activity; it minimizes the disruption of PRC2 activity after the loss of a single transcription factor while allowing subsets of PRC2 activity to be regulated without perturbing the levels of core PRC2 components. The identification of more PRC2 recruiters would shed light on how the activity of this important chromatin modifier is regulated in mammals.

Chapter 5: Dual function of PRDM14 in human ESC

Previous mapping studies on the OCT4, NANOG and SOX2 suggest that the core regulators bind and mediate the repression of lineage genes in human ESC (Boyer *et al*, 2005). The global binding profile of SUZ12 shows that the core regulators and PRC2 co-binds extensively at the promoter of these developmental genes in human ESC (Lee *et al*, 2006). Moreover, it is shown the OCT4 is needed for the binding of PRC1 components RING1A/B at the OCT4 target genes in mouse ESC (Endoh *et al*, 2008). However, there has been little biochemical evidence for direct interaction between the core regulators and the PcG in mediating gene repression. Moreover, there is no clear evidence that OCT4, NANOG and SOX2 molecules have repressive functions. Thus, whether the core regulators OCT4 and NANOG play a dual function in maintaining the ESC state remains unclear. In this study, PRDM14 is shown to play a dual function in human ESC. The molecule is shown to display both activating and repressive activities in human ESC at different loci; activating the *OCT4* enhancer (Figure 11 and 14) and repressing *HES7* promoter (Figure 35). The data presented thus far suggest that PRDM14 engage different partners in mediating this dual function at different classes of genes in human ESC.

In a bid to understand how the molecule could alternate between the two functional roles, I attempted to map the functional domains of the PRDM14 protein. The DNA binding activity of the molecule is specifically regulated by the DBD containing the 6 Znf (Figure 42). Specifically, the first 5 zinc fingers (from N to C terminal of the protein) are essential for DNA binding activity while the 6^{th} seems to be redundant (Figure 43A). Correspondingly, mutation of the first 5 Znf domains in the full protein abolished PRDM14 enhancer activity



WT probe	AGATGCCCATCAGGGTCTCTAACTGCCCCCCACCC
Mutant probe	AGATGCCCATCAGGGTCCCCAACTGCCCCCCACCC

Figure 42: DNA binding activity of PRDM14 resides in the 6 Zinc finger domains. A recombinant GST-tagged PRDM14 DBD (GST-DBD) and a recombinant GST-tagged PRDM14 with DBD deleted (GST-DBDΔ), were generated to validate if the DNA binding activity of PRDM14 only resides in the 6 Zinc fingers. EMSA shows that only the DBD containing the 6 zinc finger forms a complex with the probe containing the PRDM14 motif (Wild type (WT) probe). This binding is disrupted when point mutations are introduced into the PRDM14 motif (mutant probe)

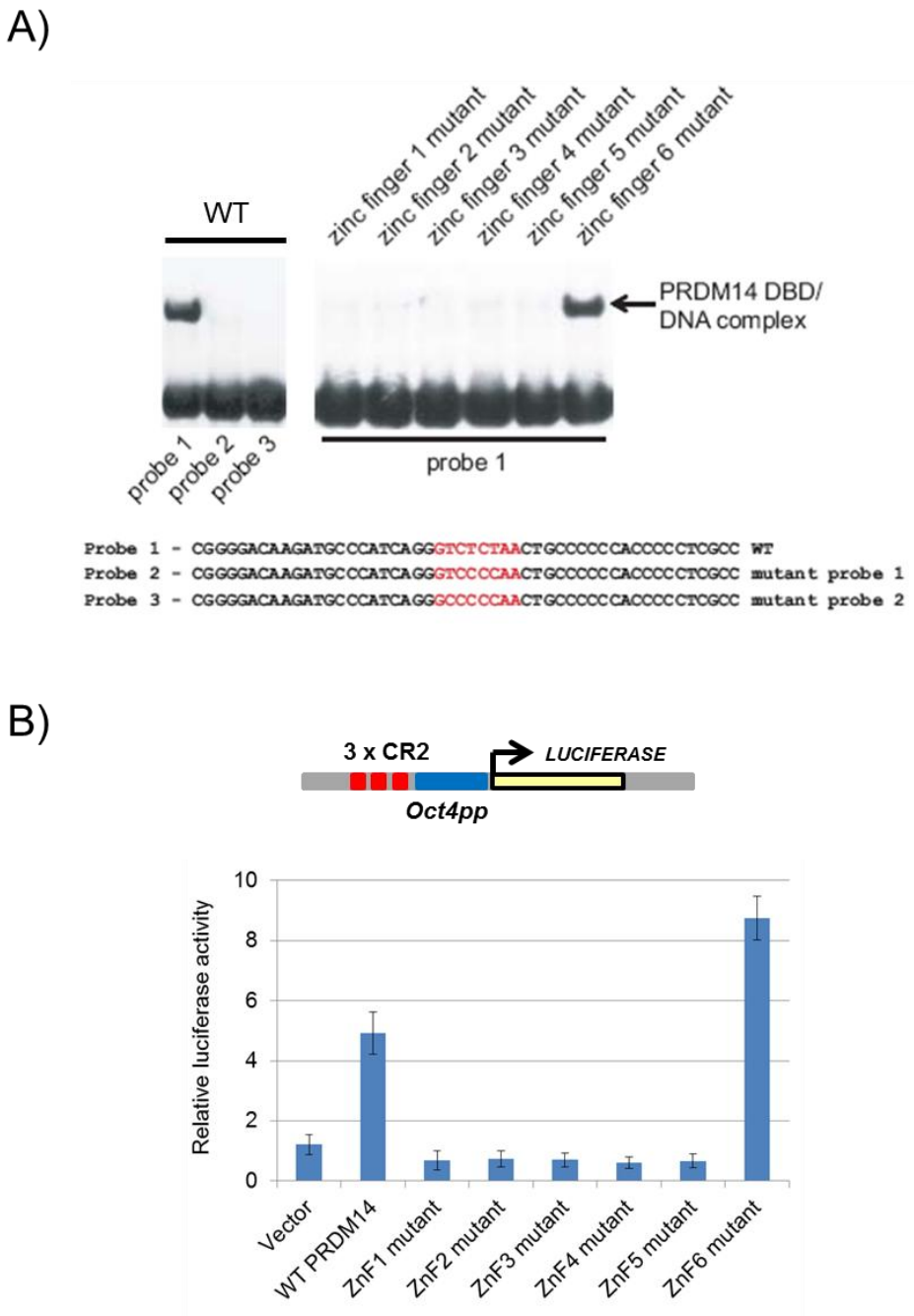
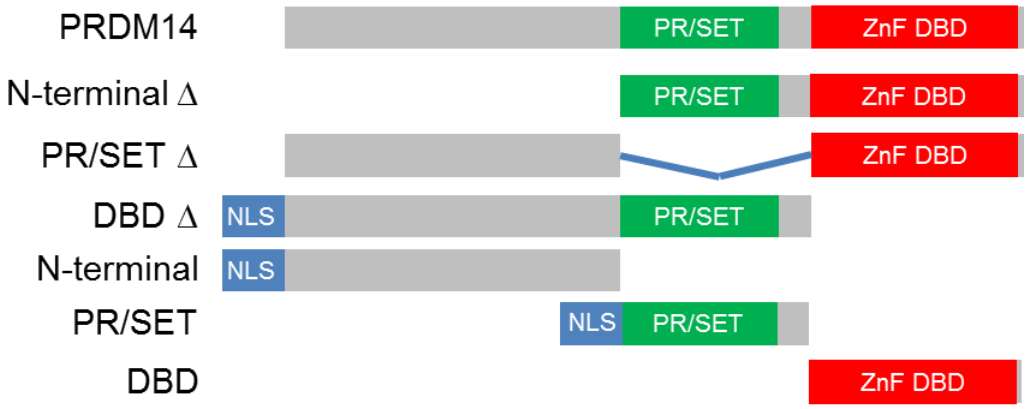


Figure 43: The first 5 zinc fingers are important for PRDM14 binding and transcriptional activity. A) Six recombinant proteins of the PRDM14 DBD with each Znf (1-6) mutated were generated for EMSA. Wild type (WT) DBD created earlier on was used as control (Figure 6). Only the 6th Znf (N to C terminus) is not essential for PRDM14 DBD binding activity. B) The individual Zinc finger mutations were introduced into the full PRDM14 protein and used for luciferase assay with the 3x30bp CR2 reporter (Figure 14A). Enhancer activity of the PRDM14 protein was loss after mutation of any of the first 5 Znfs.

and not the 6th Znf (Figure 43B). Going further, I examined the activity of the protein with each domain deleted and the enhancer activity of each individual domain (Figure 44A). As the nuclei localization signal of the molecule resides in the DNA binding domain, a nuclei localization signal is tagged to the protein fragments without the DNA binding domain (Figure 44A). None of the domains exhibited the enhancer activity by itself (Figure 44B). Importantly, deletion of the N terminal domain and DNA binding domain abolished the PRDM14 enhancer activity. This suggests that the PRDM14 enhancer activity may reside in the N terminal domain. In contrast, the deletion of the SET domain results in a 5 fold increase in enhancer activity. This intriguing result suggests that the SET domain may be suppressing the PRDM14 enhancer activity in the N terminal domain. While it is tempting to suggest that the SET domain potentially have repressive activity or interacts with the repressive PRC2, this remains to be determined.

The dual function of PRDM14 molecule could potentially be mediated by different partners in human ESC; activation activity with NANOG and repressive activity with PRC2. PRDM14 may be utilising different function domains to interact with its partners at different loci. I attempted to map the interaction sites between PRDM14 with NANOG and PRC2 with the different functional domains. Interestingly, both NANOG and EZH2 bind to the SET domain of PRDM14 (Figure 45). Whether the 2 protein binds to similar residues within the domain or not remains to be investigated. Nevertheless, this result suggests that NANOG and EZH2 may be competing for the same binding site on PRDM14. A tripartite complex of EZH2, NANOG and PRDM14 is unlikely as EZH2 is not enriched at NANOG binding sites (Figure 32). In addition, the repressive activity of sites bound by PRDM14 alone supports the notion by previous large scale mapping studies that genes bound by single factors are usually repressed in ESC (Kim *et al*, 2008; Chen *et al*, 2008).

A)



B)

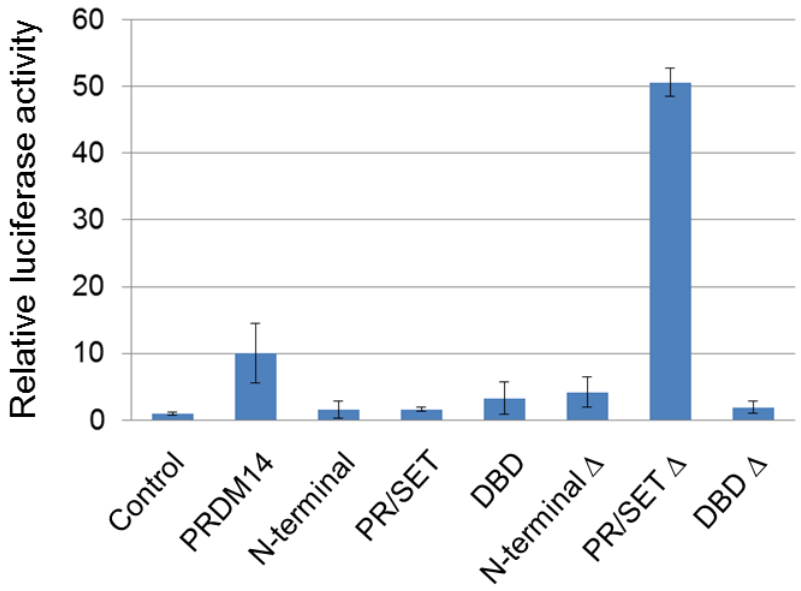


Figure 44: Transcriptional activities of various PRDM14 domains. A) Constructs expressing various domains of the PRDM14 protein and constructs expressing full PRDM14 protein with various domains deleted. A nuclear localizing signal (NLS) was added to the proteins without the Znf DBD domain. B) The constructs were co-transfected with the 3x30bp CR2 reporter in HEK 293T cells to test for enhancer activity.

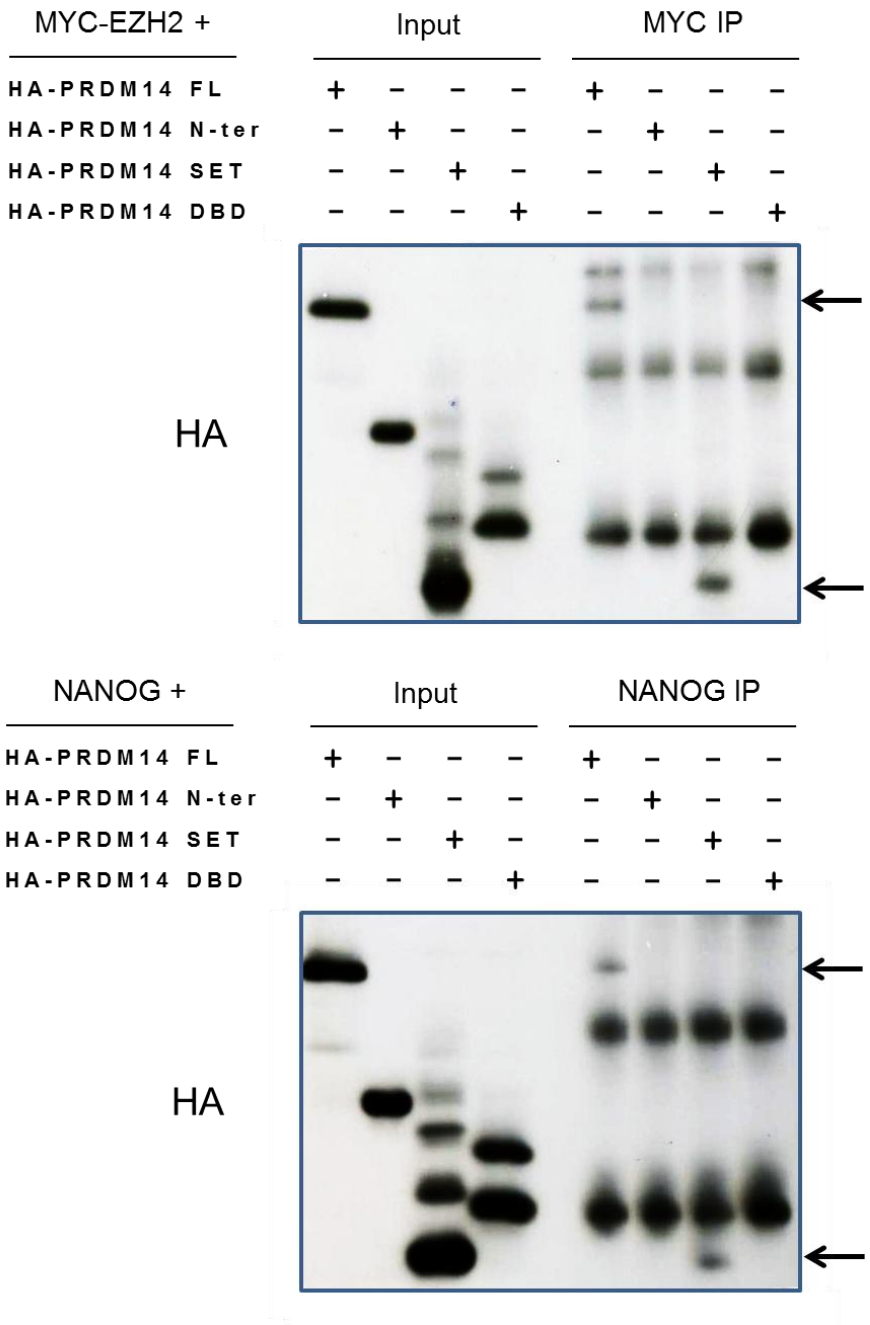


Figure 45: EZH2 and NANOG binds PRDM14 SET domain. Co-IP assays with HA-tagged full length (FL) PRDM14 and various PRDM14 domains: N-terminal domain (N-ter), SET domain (SET) and DNA binding domain (DBD) (Figure 1). The HA tagged PRDM14 proteins were co-transfected with MYC-tagged EZH2 or NANOG in HEK 293T cells and the cell lysate was used for IP with MYC and NANOG specific IP. Both the FL PRDM4 protein and SET domain is immunoprecipitated with NANOG and MYC-tagged EZH2.

In summary, the results in this last section suggest that the DNA binding activity of the PRDM14 resides in the first 5 Znfs in the C terminal domain. The N terminal domain of the protein is important for PRDM14 enhancer activity and the SET domain is potentially crucial for its interaction with NANOG for gene activation or with PRC2 for gene repression. Taking into consideration the results presented in the earlier sections, NANOG may help direct PRDM14 to the pluripotency associated genes and via competing for the same binding site on PRDM14, reduces PRC2 binding at these loci. However, the observed overlapping ChIP-Seq profiles of all 3 factors at some sites (data not shown) suggest that the regulation of these interactions may be much more complex than thought and may involve the interaction with other transcription factors. Further studies in future would be essential to understand how PRDM14 could alternate between different partners at different loci to activate or repress gene expression.

CONCLUSIONS

This study has in large achieved its aim in understanding how PRDM14 functions to regulate human ESC pluripotency. The identification of PRDM14 as a regulator of *OCT4* expression corroborated with the fact that PRDM14 was identified as a key human ESC determinant in a RNAi screen using the activity of *OCT4* promoter as the read out. In addition to regulating *OCT4* expression, PRDM14 regulates pluripotency associated genes closely with the core regulators OCT4 and NANOG across the genome. While multiple transcription factors such as Sall4 (Wu *et al*, 2006), Daz2, Nac1 and Zfp281 (Wang *et al*, 2006) have been identified to partner mouse Nanog in ESC, there is little known on the partners of human NANOG. Although NANOG co-binds closely with the heterodimer OCT4/SOX2, there has been no direct interaction reported between NANOG and OCT4 or SOX2. Thus, this study uncovers a novel partner of NANOG in regulating human ESC pluripotency. Moreover, it is shown that this interaction is functionally important for their synergistic activity in regulating the *OCT4* elements and iPSC reprogramming. On the other hand, PRDM14 binds many developmental genes and is important for their repression. These PRDM14-bound loci are enriched with the H3K27me3 repressive modification. While the enzymatic activity of PRDM14's SET domain remains to be investigated, PRDM14 recruits the PRC2 to its binding to mediate the H3K27me3 repression. Repressive activities in these loci are dependent on PRDM14 and PRC2. Importantly, PRDM14 exhibits similar repressive functions during the process of iPSC reprogramming. Overall, a dual function of PRDM14 in maintaining human ESC pluripotency has been uncovered in this study.

Dissecting PRDM14 functions have expanded our knowledge on human ESC pluripotency. Mapping of PRDM14 binding sites have expanded the core human pluripotency network

which would be useful for the identification of human ESC specific enhanceosomes. PRDM14 interaction studies have displayed how key regulators in human ESC functions together to maintain gene expression and more importantly, how the transcription factors interact with the chromatin regulators in maintaining the pluripotent cell state. While many transcription factors and chromatin regulators were identified to play a role in maintaining the ESC state, little is known whether the two classes of regulator function cohesively. PRDM14 interaction with PRC2 shows how the transcription factor could direct the repressive complex to mediate silencing in specific developmental genes in human ESC. Importantly, this interaction has provided the much needed evidence to support the involvement of transcription factor in PRC2 recruitment. The results of the activities of *OCT4* promoter elements CR2 and CR4 in human ESC and the regulation of CR2 by PRDM14 further highlighted the intrinsic differences that exist between the human and mouse pluripotent stem cells. The active CR4 in human ESC and lack of PRDM14 in EpiSC adds to the mounting evidence that shows intrinsic differences between human ESC and mouse EpiSC. The presence of PRDM14 in mouse ESC supports human ESC to be a *bona fide* developmental equivalent of mouse ESC. The conserved repressive role of Prdm14 in mouse ESC and its redundancy in regulating *OCT4* expression further suggest evolutionary differences between the human and mouse ESC transcriptional regulatory network. Taking the above discussions in consideration, PRDM14 potentially is a human ESC specific regulator and this further support the need for direct dissection of human ESC pluripotency.

In addition to understanding how human ESC pluripotency is maintained, the regulation of *OCT4* by PRDM14 is important for understanding how PRDM14 potentially functions in the tumours (NishiKawa et al, 2007; Dettman et al, 2008). PRDM14 could potentially induce the pluripotency networks in the cancer cells to promote cell proliferation. Together with its

repressive function to suppress cell fate determinants, PRDM14 potentially promotes the formation of stem cells within the tumours. Thus, this study reveals insights to how PRDM14 potentially functions as a proto-oncogene. PRDM14 could be a good drug target for therapies as it is not expressed in mature adult tissues (Dettman et al, 2008). Lastly, the repressive function of PRDM14 in iPSC reprogramming suggests that the repression of developmental programs in the fibroblast could be a rate limiting step in reprogramming. This calls for the addition of factors into the reprogramming cocktail or the use of small chemical inhibitors to help silence developmental genes in fibroblast to improve reprogramming efficiency. In summary, the findings in this study have expanded our knowledge on human ESC pluripotency which could allow us to better manipulate cell plasticity in human ESC and iPSC for therapeutic purposes and also provides new avenues to combat cancer cells that overexpress PRDM14.

BIBLIOGRAPHY

Adewumi O, Aflatoonian B, Ahrlund-Richter L, Amit M, Andrews PW, Beighton G, Bello PA, Benvenisty N, Berry LS, Bevan S, Blum B, Brooking J, Chen KG, Choo AB, Churchill GA, Corbel M, Damjanov I, Draper JS, Dvorak P, Emanuelsson K, Fleck RA, Ford A, Gertow K, Gertsenstein M, Gokhale PJ, Hamilton RS, Hampl A, Healy LE, Hovatta O, Hyllner J, Imreh MP, Itskovitz-Eldor J, Jackson J, Johnson JL, Jones M, Kee K, King BL, Knowles BB, Lako M, Lebrin F, Mallon BS, Manning D, Mayshar Y, McKay RD, Michalska AE, Mikkola M, Mileikovsky M, Minger SL, Moore HD, Mummery CL, Nagy A, Nakatsuji N, O'Brien CM, Oh SK, Olsson C, Otonkoski T, Park KY, Passier R, Patel H, Patel M, Pedersen R, Pera MF, Piekarczyk MS, Pera RA, Reubinoff BE, Robins AJ, Rossant J, Rugg-Gunn P, Schulz TC, Semb H, Sherrer ES, Siemen H, Stacey GN, Stojkovic M, Suemori H, Szatkiewicz J, Turetsky T, Tuuri T, van den Brink S, Vintersten K, Vuoristo S, Ward D, Weaver TA, Young LA, Zhang W. (2007) Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol. 25, 803-16.

Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, Wang J, Rendl M, Bernstein E, Schaniel C, Lemischka IR. (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell 145, 183-97.

Arnold K, Sarkar A, Yram MA, Polo JM, Bronson R, Sengupta S, Seandel M, Geijsen N, Hochedlinger K. (2011) Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. Cell Stem Cell. 9, 317-29.

Asahara, T, Murohara, T, Sullivan, A, Silver, M, van der Zee, R, Li, T, Witzenbichler, B, Schatteman, G, and Isner, JM (1997). Isolation of putative progenitor endothelial cells for angiogenesis. Science 275, 964–967.

Assou S, Le Carrour T, Tondeur S, Ström S, Gabelle A, Marty S, Nadal L, Pantesco V, Réme T, Hugnot JP, Gasca S, Hovatta O, Hamamah S, Klein B, De Vos J. (2007) A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells 25, 961-73.

Au SL, Wong CC, Lee JM, Fan DN, Tsang FH, Ng IO, Wong CM. (2012) Enhancer of zeste homolog 2 epigenetically silences multiple tumor suppressor microRNAs to promote liver cancer metastasis. Hepatology 25679.

Avilion AA, Nicolis, SK, Pevny, LH, Perez, L, Vivian, N and Lovell-Badge, R (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17,126–140

Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. (2006) Chromatin signatures of pluripotent cell lines. Nat Cell Biol. 8, 532-8.

Bagutti C, Wobus AM, Fässler R, Watt FM. (1996). Differentiation of embryonal stem cells into keratinocytes: comparison of wild-type and beta 1 integrin-deficient cells. Dev Biol. 179, 184-96

Beattie, GM, Lopez, AD, Bucay, N, Hinton, A, Firpo, MT, King, CC, and Hayek, A (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem Cells 23, 489–495

Bell, AC, West, AG & Felsenfeld, G (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98, 387–396.

Beltrami, AP, Barlucchi, L, Torella, D, Baker, M, Limana, F, Chimenti, S, Kasahara, H, Rota, M, Musso, E, Urbanek, K., Leri, A, Kajstura, J, Nadal-Ginard, B, Anversa, P (2003). Adult cardiac stem cells are multipotent and support myocardial regeneration. Cell 114, 763-76.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326.

Bessho Y, Miyoshi G, Sakata R, Kageyama R. (2001) Hes7: a bHLH-type repressor gene regulated by Notch and expressed in the presomitic mesoderm. Genes Cells 6, 175-85.

Birve A, Sengupta AK, Beuchle D, Larsson J, Kennison JA, Rasmuson-Lestander A, Müller J. (2001) Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development 128, 3371-9.

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349-53

Boyer, LA, Lee, TL, Cole, MF, Johnstone, SE, Levine, SS, Zucker, JP, Guenther, MG, Kumar, RM, Murray, HL, Jenner, RG, Gifford, DK, Melton, DA, Jaenisch, R, and Young RA (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956.

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

Brayer KJ, Segal DJ. 2008. Keep your fingers off my DNA: protein–protein interactions mediated by C2H2 zinc finger domains. Cell Biochem Biophys 50: 111–31.

Brinster, RL (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. J. Exp. Med. 140, 1049–1056

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature. 448, 191-5.

Brown, JL, Mucci, D, Whiteley, M, Dirksen, ML & Kassis, JA (1998) The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol. Cell 1, 1057–1064.

Cai J, Zhao Y, Liu Y, Ye F, Song Z, Qin H, Meng S, Chen Y, Zhou R, Song X, Guo Y, Ding M, Deng H. (2007) Directed differentiation of human embryonic stem cells into functional hepatic cells. Hepatology. 45, 1229-39.

Calado DP, Zhang B, Srinivasan L, Sasaki Y, Seagal J, Unitt C, Rodig S, Kutok J, Tarakhovsky A, Schmidt-Supprian M, Rajewsky K. (2010). Constitutive canonical NF-kappaB activation cooperates with disruption of BLIMP1 in the pathogenesis of activated B cell-like diffuse large cell lymphoma. Cancer Cell 18. 580–9.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. (2002) Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science 298, 1039-43

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

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell 133, 1106–1117

Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P, Ng HH. (2005) Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol. 25, 6031-46.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 107, 21931-6

Creyghton MP, Markoulaki S, Levine SS, Hanna J, Lodato MA, Sha K, Young RA, Jaenisch R, Boyer LA. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell 135, 649-61.

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. (2002) Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. Cell 111, 185-96.

D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, Baetge EE. (2005) Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol. 23, 1534-41.

D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat Biotechnol. 24, 1392-401.

Dani C, Smith AG, Dessolin S, Leroy P, Staccini L, Villageois P, Darimont C, Ailhaud G. (1997). Differentiation of embryonic stem cells into adipocytes in vitro. J Cell Sci. 110, 1279-85.

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Rust AG, Pan Zj, Schilstra MJ, Clarke PJ, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H (2002). A genomic regulatory network for development. Science. 295,1669-1678

Dettman EJ, Justice MJ. (2008). The zinc finger SET domain gene Prdm14 is overexpressed in lymphoblastic lymphomas with retroviral insertions at Evi32. PLoS One. 3, e3823.

Ding L, Paszkowski-Rogacz M, Nitzsche A, Slabicki MM, Heninger AK, de Vries I, Kittler R, Junqueira M, Shevchenko A, Schulz H, Hubner N, Doss MX, Sachinidis A, Hescheler J, Iacone R, Anastassiadis K, Stewart AF, Pisabarro MT, Caldarelli A, Poser I, Theis M, Buchholz F (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. Cell Stem Cell 4, 403-415

Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J Embryol Exp Morphol. 87, 27-45.

Ebert AD, Yu J, Rose FF, Mattis VB, Lorson CL, Thomson JA, Svendsen CN. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature 457, 277–280.

Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T., and Sasai, Y. (2011). Self-organizing optic-cup morphogenesis in three-dimensional culture. Nature 472, 51-6.

Endoh M, Endo TA, Endoh T, Fujimura Y, Ohara O, Toyoda T, Otte AP, Okano M, Brockdorff N, Vidal M, Koseki H. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development. 135, 1513-24.

Endoh M, Endo TA, Endoh T, Fujimura Y, Ohara O, Toyoda T, Otte AP, Okano M, Brockdorff N, Vidal M, Koseki H. (2008) Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development. 135, 1513-24.

Eom GH, Kim K, Kim SM, Kee HJ, Kim JY, Jin HM, Kim JR, Kim JH, Choe N, Kim KB, Lee J, Kook H, Kim N, Seo SB. (2009). Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis. Biochem Biophys Res Commun 388, 131–6.

Evans, MJ & Kaufman, MH (1981). Establishment in culture of pluripotential cells from

Fazzio, TG, Huff, JT, and Panning, B (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell 134, 162–174

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806-11

Fog C K, Galli G G, Lund A H. (2011) PRDM proteins: important players in differentiation and disease. Bioessays 34, 50-60

Fraichard A, Chassande O, Bilbaut G, Dehay C, Savatier P, Samarut J. (1995). In vitro differentiation of embryonic stem cells into glial cells and functional neurons. J Cell Sci. 108, 3181-8.

Fuchs, E. and Segre, J.A. (2000) Stem Cells: A New Lease on Life. Cell 100, 143–155

Fujikawa T, Oh SH, Pi L, Hatch HM, Shupe T, Petersen BE. (2005). Teratoma formation leads to failure of treatment for type I diabetes using embryonic stem cell-derived insulin-producing cells. Am J Pathol. 166, 1781-91.

Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E, Ramalho-Santos M. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460, 863–868.

Greber B, Wu G, Bernemann C, Joo JY, Han DW, Ko K, Tapia N, Sabour D, Sterneckert J, Tesar P, Schöler HR. (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. Cell Stem Cell. 6, 215-26.

Greber, B, Lehrach, ., and Adjaye, J (2007). Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (human ESCs) to support human ESC self-renewal. Stem Cells 25, 455–464.

Green MD, Chen A, Nostro MC, d'Souza SL, Schaniel C, Lemischka IR, Gouon-Evans V, Keller G, Snoeck HW. (2011). Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. Nat Biotechnol. 29, 267-72.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B. (2010). Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. Cell Stem Cell 6, 479-91.

Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK, Ngo QM, Klugman S, Ching KA, Edsall LE, Ye Z, Kuan S, Yu P, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Ren B. (2011). Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. Cell Res. 2011;21:1393-409

Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell 146, 519-32.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature 438, 374–8

He H, McHaney M, Hong J, Weiss ML (2009). Cloning and Characterization of 3.1kb Promoter Region of the Oct4 Gene from the Fischer 344 Rat. Open Stem Cell J. 1, 30-39.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459, 108-12.

Hennig L, Derkacheva M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet 25, 414-23.

Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A. 106, 5187-91.

Ho L, Ronan JL, Wu J, Staahl BT, Chen L, Kuo A, Lessard J, Nesvizhskii AI, Ranish J, Crabtree GR. (2009a). 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. 106, 5181-6.

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B, Meng X, Miller JC, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. 27, 851-7.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol. 29, 731-4.

Hu BY, Weick JP, Yu J, Ma LX, Zhang XQ, Thomson JA, Zhang SC. (2010) Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proc Natl Acad Sci U S A. 107, 4335-40.

Hu, G, Kim, J, Xu, QK, Leng, Y, Orkin, SH and Elledge SJ (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. 23, 837-848

Ivanova, N, Dobrin, R, Lu, R, Kotenko, I, Levorse, J, DeCoste, C, Schafer, X, Lun, Y, and Lemischka, IR (2006). Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533–538.

James, D, Levine, AJ, Besser, D, and Hemmati-Brivanlou, A (2005). GFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273–1282.

Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, Lim CA, Robson P, Zhong S, Ng HH. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol. 10, 353-60.

Johnson DS, Mortazavi A, Myers RM, Wold B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science. 316, 1497-502.

Jones, EA, Tosh, D, Wilson, DI, Lindsay, S. and Forrester, LM (2002). Hepatic differentiation of murine embryonic stem cells. Exp Cell Res. 272, 15-22.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435.

Kamiya D, Banno S, Sasai N, Ohgushi M, Inomata H, Watanabe K, Kawada M, Yakura R, Kiyonari H, Nakao K, Jakt LM, Nishikawa S, Sasai Y. (2011). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. Nature. 470, 503-9.

Khan, IF, Hirata, RK, Wang, PR, Li, Y, Kho, J, Nelson, A, Huo, Y, Zavaljevski, M, Ware, C and Russell, DW (2010). Engineering of human pluripotent stem cells by AAV-mediated gene targeting. Mol Ther. 18, 1192-9.

Kim KC, Geng L, Huang S. (2003). Inactivation of a histone methyltransferase by mutations in human cancers. Cancer Res 63, 7619–23.

Kim, J, Chu, JL, Shen, XH, Wang, JL, and Orkin, SH (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell 132, 1049–1061

Körbling, M and Estrov, Z (2003). Adult Stem Cells for Tissue Repair - A New Therapeutic Concept. N Engl J Med 349, 570-82.

Kriegstein A, Alvarez-Buylla A. (2009). The glial nature of embryonic and adult neural stem cells. Annu Rev Neurosci. 32, 149-84.

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 4, e1000242.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 42, 631-4.

Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, Ganat YM, Menon J, Shimizu F, Viale A, Tabar V, Sadelain M, Studer L. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. Nature 461, 402-6.

Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP,

Melton DA, Gifford DK, Jaenisch R, Young RA. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 125, 301-13.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298, 799-804

Leeb, M, Pasini, D, Novatchkova, M, Jaritz, M, Helin, K, and Wutz, A (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev. 24, 265–276

Levenstein, ME, Ludwig, TE, Xu, RH, Llanas, RA, Vandenheuvel-Kramer, K, Manning, D, and Thomson, JA (2005). Basic fibroblast growth factor support of human embryonic stem cell self-renewal. Stem Cells 24, 568–574.

Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. 24, 368–380.

Li JY, Christophersen NS, Hall V, Soulet D, Brundin P. (2008). Critical issues of clinical human embryonic stem cell therapy for brain repair. Trends Neurosci. 31, 146-53.

Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q, Qin B, Xu J, Li W, Yang J, Gan Y, Qin D, Feng S, Song H, Yang D, Zhang B, Zeng L, Lai L, Esteban MA, Pei D. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell 7, 51-63.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. 38, 431–440.

Lumelsky, N, Blondel, O, Laeng, P, Velasco, I, Ravin, R and McKay, R (2001). Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. Science. 292, 1389-94.

Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J (2011). Sequence-specific regulator Prdm14 safeguards mouse ESC from entering extraembryonic endoderm fates. Nat Struct Mol Biol 18, 120-7.

Mandelbaum J, Bhagat G, Tang H, Mo T, Brahmachary M, Shen Q, Chadburn A, Rajewsky K, Tarakhovsky A, Pasqualucci L, Dalla-Favera R. (2010). BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell 18, 568–79.

Margueron Rand Reinberg D (2011). The Polycomb complex PRC2 and its mark in life. Nature 469, 343-9.

Martin, GR & Evans, MJ (1974). The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. Cell 2, 163–172.

Martin, GR (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78, 7634-8.

Masui, S, Nakatake, Y, Toyooka, Y, Shimosato, D, Yagi, R, Takahashi, K, Okochi, H, Okuda, A, Matoba, R, Sharov, A, Ko, M, and Niwa, H (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature Cell Bio. 9, 625–635.

McNeish J (2004). Embryonic stem cells in drug discovery. Nat Rev Drug Discov. 3,70-80.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766-70.

Melichar H, Li O, Ross J, Haber H, Cado D, Nolla H, Robey EA, Winoto A. (2011). Comparative study of hematopoietic differentiation between human embryonic stem cell lines. PLoS One 6, e19854

Mihaly, J, Mishra, RK & Karch, F (1998). A conserved sequence motif in Polycomb-response elements. Mol. Cell 1, 1065–1066.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-60.

Mintz, B & Illmensee, K (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proc. Natl Acad. Sci. USA 72, 3585–3589.

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

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 34, 267-73.

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 34, 267-73.

Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. (2002). Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. Cell. 111, 197-208

Nakano, T., Kodama, H. and Honjo, T. (1996). In vitro development of primitive and definitive erythryocytes from different precursors. Science 272, 722–724.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, K.D., Chambers, I., Schöler, H.R. and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379–391

Nishikawa N, Toyota M, Suzuki H, Honma T, Fujikane T, Ohmura T, Nishidate T, Ohe-Toyota M, Maruyama R, Sonoda T, Sasaki Y, Urano T, Imai K, Hirata K, Tokino T. (2007) Gene amplification and overexpression of PRDM14 in breast cancers. Cancer Res. 67, 9649-57.

Nishiyama A, Xin L, Sharov AA, Thomas M, Mowrer G, Meyers E, Piao Y, Mehta S, Yee S, Nakatake Y, Stagg C, Sharova L, Correa-Cerro LS, Bassey U, Hoang H, Kim E, Tapnio R, Qian Y, Dudekula D, Zalzman M, Li M, Falco G, Yang HT, Lee SL, Monti M, Stanghellini I, Islam MN, Nagaraja R, Goldberg I, Wang W, Longo DL, Schlessinger D, Ko MS (2009). Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell 5, 420-433.

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

Ntziachristos P, Tsirigos A, Van Vlierberghe P, Nedjic J, Trimarchi T, Flaherty MS, Ferres-Marco D, da Ros V, Tang Z, Siegle J, Asp P, Hadler M, Rigo I, De Keersmaecker K, Patel J, Huynh T, Utro F, Poglio S, Samon JB, Paietta E, Racevskis J, Rowe JM, Rabadan R, Levine RL, Brown S, Pflumio F, Dominguez M, Ferrando A, Aifantis I. (2012). Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. Nat Med. 18, 298-301

Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA. (2005) Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436, 207-13.

Oldershaw RA, Baxter MA, Lowe ET, Bates N, Grady LM, Soncin F, Brison DR, Hardingham TE and Kimber SJ. (2010). Directed differentiation of human embryonic stem cells toward chondrocytes. Nat Biotechnol 28, 1187-94.

Onder TT, Kara N, Cherry A, Sinha AU, Zhu N, Bernt KM, Cahan P, Marcarci BO, Unternaehrer J, Gupta PB, Lander ES, Armstrong SA, Daley GQ. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature. 483, 598-602

Osafune K, Caron L, Borowiak M, Martinez RJ, Fitz-Gerald CS, Sato Y, Cowan CA, Chien KR, Melton DA. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat Biotechnol 26, 313-5.

Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell 6, 382-95.

Park, IH, Arora, N, Huo, H, Maherali, N, Ahfeldt, T, Shimamura, A, Lensch, MW, Cowan, C, Hochedlinger, K and Daley, GQ (2008). Disease-specific induced pluripotent stem cells. Cell 134, 877-86.

Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 464, 306–310.

Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 139, 1290–1302.

Pessach IM, Ordovas-Montanes J, Zhang SY, Casanova JL, Giliani S, Gennery AR, Al-Herz W, Manos PD, Schlaeger TM, Park IH, Rucci F, Agarwal S, Mostoslavsky G, Daley GQ, Notarangelo LD. (2011). Induced pluripotent stem cells: a novel frontier in the study of human primary immunodeficiencies. J Allergy Clin Immunol. 27, 1400-7.

Pittenger, MF, Mackay, AM, Beck, SC, Jaiswal, RK, Douglas, R, Mosca, JD, Moorman, MA, Simonetti, DW, Craig, S, and Marshak, DR (1999). Multilineage potential of adult human mesenchymal stem cells. Science 284, 143–147.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279-83

Raya A, Rodríguez-Pizà I, Guenechea G, Vassena R, Navarro S, Barrero MJ, Consiglio A, Castellà M, Río P, Sleep E, González F, Tiscornia G, Garreta E, Aasen T, Veiga A, Verma IM, Surrallés J, Bueren J, Izpisúa Belmonte JC. (2009). Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. Nature 460, 53-9.

Ren G, Baritaki S, Marathe H, Feng J, Park S, Beach S, Bazeley PS, Beshir AB, Fenteany G, Mehra R, Daignault S, Al-Mulla F, Keller E, Bonavida B, de la Serna I, Yeung KC. (2012). Polycomb Protein EZH2 Regulates Tumor Invasion via the Transcriptional Repression of the Metastasis Suppressor RKIP in Breast and Prostate Cancer. Cancer Res. 72, 3091-3104.

Reynolds, B.A., and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. Science 255, 1707–1710.

Rheinwald, J.G., and Green, H. (1975). Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. Cell 6, 331–343.

Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323, 445-8.

Rodaway A, Patient R. (2001) Mesendoderm. an ancient germ layer? Cell 105, 169-72.

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem. 280, 24731-7.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature 345, 686–692.

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell. 7, 64-77.

Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T, Tesar P, Wei CL, Scacheri PC. (2010). CHD7 targets

active gene enhancer elements to modulate ES cell specific gene expression. PLoS Genet. 6, e1001023

Seale P, Kajimura S, Spiegelman BM. (2009) Transcriptional control of brown adipocyte development and physiological function--of mice and men. Genes Dev 23, 788-97.

Shaver S, Casas-Mollano JA, Cerny RL, Cerutti H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. Epigenetics 5, 301-12.

Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 139, 1303–1314

Shing DC, Trubia M, Marchesi F, Radaelli E, Belloni E, Tapinassi C, Scanziani E, Mecucci C, Crescenzi B, Lahortiga I, Odero MD, Zardo G, Gruszka A, Minucci S, Di Fiore PP, Pelicci PG. (2007). Overexpression of sPRDM16 coupled with loss of p53 induces myeloid leukemias in mice. J Clin Invest 117: 3696–707

Simon C, Chagraoui J, Krosl J, Gendron P, Wilhelm B, Lemieux S, Boucher G, Chagnon P, Drouin S, Lambert R, Rondeau C, Bilodeau A, Lavallée S, Sauvageau M, Hébert J, Sauvageau G. (2012). A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. Genes Dev. 26, 651-6.

Simon JA, Kingston RE. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol. 10, 697-708.

Sing A, Pannell D, Karaiskakis A, Sturgeon K, Djabali M, Ellis J, Lipshitz HD, Cordes SP. (2009) A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell 138, 885–897.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336, 688–690

Smithies O, Gregg, R. G., Boggs, S. S., Doralewski, M. A., Kucherlapati, R. S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature. 317, 230-4.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., Zhang, L., Guschin, D., Fong, L.K., Vu, B.J., Meng, X., Urnov, F.D., Rebar, E.J., Gregory, P.D., Zhang, H.S. and Jaenisch, R. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318-31.

Solter D. (2006). From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. Nat Rev Genet. 7, 319-27.

Song, H., Chung, S.K. and Xu, Y. (2010) Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. Cell Stem Cell. 6, 80-9.

Spence. J.R., Mayhew, C.N., Rankin, S.A., Kuhar, M.F., Vallance, J.E., Tolle, K., Hoskins, E.E., Kalinichenko, V.V., Wells, S.I., Zorn, A.M., Shroyer, N.F. and Wells, J.M. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nature 470, 105-9.

Stewart, C.L., Kaspar, P., Brunet, L.J., Bhatt, H., Gadi, I., Kontgen, F., and Abbondanzo, S.J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature 359, 76–79.

Strübing C, Ahnert-Hilger G, Shan J, Wiedenmann B, Human ESCheler J, Wobus AM. (1995). Differentiation of pluripotent embryonic stem cells into the neuronal lineage in vitro gives rise to mature inhibitory and excitatory neurons. Mech Dev. 53, 275-87.

Struhl G. (1981). A gene product required for correct initiation of segmental determination in Drosophila. Nature. 293, 36-41.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 102, 15545-50

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-76

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861-72.

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T., and Akira, S. (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proc. Natl. Acad. Sci. USA 94, 3801–3804.

Temple, S. (1989) Division and differentiation of isolated CNS blast cells in microculture. Nature 340, 471–473

Temple, S. (2001) The development of neural stem cells. Nature 414, 112–117

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature. 448, 196-9.

Thomas KR, Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 51, 503-12.

Thomson, JA, Itskovitz-Eldor, J, Shapiro, SS, Waknitz, MA, Swiergiel, JJ, Marshall, Jones JM. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282, 1145-7.

Tsuneyoshi N, Sumi T, Onda H, Nojima H, Nakatsuji N, Suemori H. (2008). PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. Biochem Biophys Res Commun. 367, 899-905

V. S. & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts.

Van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 6, 369-81.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854-8.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature 444, 364–368.

Ware, C.B., Horowitz, M.C., Renshaw, B.R., Hunt, J.S., Liggitt, D., Koblar, S.A., Gliniak, B.C., McKenna, H.J., Papayannopoulou, T., Thoma, B., Cheng L.Z., Donovan P.J., Peschon

J.J., Bartlett P.F., Willis C.R., Wright B.D., Carpenter M.K., Davison B.L. and Gearing P.L.(1995). Targeted disruption of the low-affinity leukemia inhibitory factor receptor gene causes placental, skeletal, neural and metabolic defects and results in perinatal death. Development 121, 1283–1299.

Wieser R (2007) The oncogene and developmental regulator EVI1: expression, biochemical properties, and biological functions. Gene 396, 346–57.

Wilkinson, F. H., Park, K. & Atchison, M. L. (2006) Polycomb recruitment to DNA in vivo by the YY1 REPO domain. Proc Natl Acad Sci U S A. 103, 19296–19301.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336, 684–687.

Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J. & Kingston, R. E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell 140, 99–110.

Wu Q, Chen X, Zhang J, Loh YH, Low TY, Zhang W, Zhang W, Sze SK, Lim B, Ng HH. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem. 281, 24090-4.

Xi H, Yu Y, Fu Y, Foley J, Halees A, Weng Z. (2007). Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. Genome Res. 17, 798–806.

Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C, Zwaka TP, Thomson JA. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nat Biotechnol. 20, 1261-4.

Xu, RH, Sampsell-Barron, TL, Gu, F, Root, S, Peck, RM, Pan, G, Yu, J, Antosiewicz-Bourget, J, Tian, S, Stewart, R and Thomson, JA (2008). NANOG Is a Direct Target of TGFb/Activin-Mediated SMAD Signaling in Human ESCs. Cell stem cell 3, 196–206

Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y and Saitou M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet. 40, 1016-22

Yang, JS, Lavker, RM, and Sun, TT (1993). Upper human hair follicle contains a subpopulation of keratinocytes with superior in vitro proliferative potential. J. Invest. Dermatol. 101, 652–659.

Ye L, Chang JC, Lin C, Sun X, Yu J, Kan YW. (2009). Induced pluripotent stem cells offer new approach to therapy in thalassemia and sickle cell anemia and option in prenatal diagnosis in genetic diseases. Proc Natl Acad Sci U S A. 106, 9826-30

Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hübner K, Schöler HR. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122, 881-94.

Ying, QL, Nichols, J, Chambers, I & Smith, A (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115, 281–292.

Young, RA (2011) Control of the embryonic stem cell state. Cell 144, 940-54

Yuan, H, Corbi, N, Bascilico, C and Dailey, L (1995). Developmental-specificity activity of the FGF-4 enhancer requires the synergistic action of Sox-2 and Oct-3. Genes Dev. 9, 2635–2645.

Zambrowicz, BP and Sands, AT (2003) Knockouts of the 100 best-selling drugs — will they model the next 100? Nature Rev. Drug Discov. 2, 38–51.

Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung WK, Shahab A, Kuznetsov VA, Bourque G, Oh S, Ruan Y, Ng HH, Wei CL. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell 1, 286-98.

Zhao, XY, Li, W, Lv, Z, Liu, L, Tong, M, Hai, T, Hao, J, Guo, CL, Ma, QW, Wang, L, Zeng, F and Zhou, Q (2009). iPS cells produce viable mice through tetraploid complementation. Nature 461, 86-90.

Zwaka, TP and Thomson, JA (2003) Homologous recombination in human embryonic stem cells. Nat Biotechnol. 21,319-21