IDENTIFICATION OF OCT4 AND SOX2 TARGETS IN MOUSE EMBRYONIC STEM CELLS

CHEW JOON LIN

(M.Sc., NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2007

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Dr Ng Huck Hui, who has taught me a great deal about working in a very competitive field. Thank you for your leadership and guidance throughout the four and a half years working with you. Thank you for seeing my thesis through.

I am indebted to my committee members:
Professor Hew Choy Leong for your untiring counsel, objectivity, support and feedback
Associate Professor Larry Stanton for untiring counsel, objectivity, support and feedback
Associate Professor Gong Zhiyuan for your kindness, support and feedback

Special thanks to:
Dr Paul Robson for invaluable discussions and feedback particularly on the Sox2 project
Associate Professor Lim Bing for encouragements and feedback
Associate Professor Thomas Lufkin for some discussions on ESC culture
Dr Neil Clarke for your student counsel and support
Associate Professor Nallasivam Palanisamy for laughter shared during late nights and weekends
Dr Edwin Cheung for presentation feedback
Prof Alex Ip for your invaluable knowledge on teaching and presentation methods
Prof Larry Stanton, Prof Hew Choy Leong, Dr Patrick Ng Wei Pern, Wong Meng Kang, Lo Ting Ling and Wong Kee Yew for reviewing and commenting on this thesis

Many thanks to collaborators in this work and side projects, particularly:
Dr Ruan Yijun, Dr Wei Chia-Lin, Dr Paul Robson, Associate Professor Larry Stanton, Associate Professor Lim Bing, Vinsensius Berlian Vega, Dr Bernard Leong, Charlie Lee, Dr Leonard Lipovich, Dr Vladamir Kuznetsov, Wong Kee Yew, Dr Zhao Xiaodong, Lim Leng Hiong, Loh Yuin Han, Li Pin. Thank you to the GIS sequencing facility for massive sequencing of the ChIP-PETs and the Bioinformatics group for high throughput computational work.

Special thanks *to* the Singapore Millennium Foundation, Temasek Holdings and Mr. John deRoza for financial support and counsel.

I am very blessed to have great labmates, past and present, especially:
Chen Xi, Dr Wu Qiang, Lim Ching Aeng, Dr Zhang Wensheng, Winston Chan, Dr Yan Junli, Dr Fengbo, Dr Yuan Ping, Kenny Chew, Chia Nayu, Tay Hwee Goon, Fan Yi, Dr Julia Zhu, Katty Kuay and Tan Qiu Li: *thanks for the many discussions and outings!*

To all GIS inhabitants, especially:
Clara Cheong, Sumantra, Evan, Serene, Alicia, Sandy, Say Li, Dr Patrick Ng, Pauline, Govind, Dr Sanjay Gupta, Dr Mani, Dr Srini, Meng, Dr Majid, Dr Andrew Thomson, and all the administrative personnel at level 2: *thank you for all the good memories.*

Thanks to the Genome Institute of Singapore and National University of Singapore for facility support, administrative assistance and good services rendered.

Thank you SMF-ers Chia Jer Ming, Lynn Chiam, Chang Kai Chen, Azhar Ali, Chang Ti Ling and many others, for a great fellowship!

Special thanks to my family members who sacrificed the most but ironically, may not understand much beyond this page: I love you.

TABLE OF CONTENTS

CHAPTER 4

Genome-wide Mapping of Oct4-DNA Interactions in Mouse Embryonic Stem Cells

CHAPTER 5

Genome-wide Identification of Sox2-DNA Interactions in Mouse Embryonic Stem Cells

CHAPTER 6

Analyses of the Combined Oct4 and Sox2 DNA Binding Sites

CHAPTER 7

Discovery of Oct4 and Sox2 Collaborating Factors and Demonstrating a Link between Different Pathways in Mouse Embryonic Stem Cells

SUMMARY

Embryonic stem cells (ESCs) are derived from the inner cell mass (ICM) of the mammalian blastocyst. They are capable of indefinite self-renewing cell division under specific cell culture conditions for extended periods and they have the ability to differentiate into all cell types of the adult organism. This property of ESCs holds great promise for regenerative therapeutic medicine. The fundamental understanding of the molecular biology of ESCs will be essential for the eventual rational design of methods to control the self-renewal and differentiation of these cells.

Oct4 and Sox2 are key transcription factors that are important in maintaining the ESC state. In order to understand the roles of these factors and how they collaborate with each other, it is essential to know which genes are directly associated with these proteins *in vivo*. In this study, chromatin immunoprecipitation (ChIP) was used in a small scale study to map the binding circuitry of Oct4 and Sox2 on *Oct4*, *Sox2* and *Nanog* genes. Oct4 and Sox2 were shown to bind to the genes encoding Oct4, Sox2 and Nanog.

Subsequently, the ChIP-PET (paired end ditag) technology was used to map the whole genome binding sites of Oct4 and Sox2. Thousands of novel Oct4 and Sox2 binding sites were identified, mapping to genes implicated in important cellular functions including maintenance of pluripotency. Oct4 and Sox2 were found to co-occupy a substantial number of genes. A fifteen nucleotide Sox2-Oct4 motif was identified and characterized.

In order to identify other DNA binding transcription factors that may work together with Oct4 and Sox2, 500 bp DNA sequences centered on the sites bound by Oct4 and Sox2 were matched against matrices from the TRANSFAC database to reveal putative transcription factor co-motifs. Two of the extrinsic signaling substrates, Stat3 and Smad1, were shown to bind in the

vicinity of Oct4 and Sox2 DNA binding sites associated with genes important in cell cycle regulation, pluripotency and self-renewal. Stat3 and Smad1 were further demonstrated to be Oct4 partners, revealing the connection between the intrinsic transcription factors and extrinsic signaling pathways in mouse embryonic stem cells (mESCs). Stat3 and Oct4 were shown to bind to the promoter of *Nanog*, and subsequent results suggest that Stat3 may regulate the transcription of *Nanog*. This may provide a link between the three transcriptional pathways in mESCs. Together, these results suggest that Oct4 and Sox2 (1) have multiple DNA binding sites, (2) may exert transcriptional effects from distal locations, (3) work in collaboration with other transcription factors and (4) their binding may be a pre-requisite for the binding of other transcription factors.

In conclusion, this study identified the Oct4 and Sox2 DNA interactions, characterized the joint DNA motif, expanded the Oct4 and Sox2 transcription binding network to the Smad1 and Stat3 signalling networks and utilized these binding data in an attempt to answer a long standing hypothesis of a crosstalk between the Oct4, Stat3 and Nanog transcriptional pathways. The findings obtained here may contribute to the eventual mapping of the whole transcriptional regulatory network and elucidation of transcriptional mechanisms in mESCs.

LIST OF TABLES

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

Ab	antibody
BMP	bone morphogenic protein
bp	base pairs
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
ChIP-on-chip	chromatin immunoprecipitation coupled to microarray chip
ChIP-PET	chromatin immunoprecipitation coupled to paired end ditag sequencing
Co-IP	co-immunoprecipitation
DNA	Deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
EB	embryoid body
ECC	embryonic carcinoma cells
EDTA	ethylenediaminetetraacetic acid
EGC	embryonic germ cells
EMSA	electrophoretic mobility shift assay
FBS	fetal bovine serum
GFP	green fluorescent protein
gp-130	glycoprotein-130
GST	glutathione S-transferase
hESC	human embryonic stem cell
IB	immunoblot
ICM	inner cell mass
kDa	kilo Dalton
LB	Luria-Bertani
LIF	leukemia inhibitory factor
LIFR	leukemia inhibitory factor receptor
mESC	mouse embryonic stem cell
Min	minute
mM	milli Molar
mRNA	messenger RNA
MW	molecular weight
ng	nanogram
Oct	Octamer binding protein
PCR	polymerase chain reaction
PET	paired-end ditag
Q-PCR	quantitative polymerase chain reaction
RA	Retinoic acid
RNA Pol II	RNA polymerase II
RNA	Ribonucleic acid
RNAi	RNA interference
Rpm	rotation per minute
RT-PCR	Reverse transcription polymerase chain reaction
seqChIP	Sequential chromatin immunoprecipitation
shRNA	short hairpin RNA
Sox	Sry-related HMG box
Stat	signal transducer and activator of transcription
TFBS	Transcription factor binding site
TNF	Tumor necrosis factor

LIST OF PUBLICATIONS (FROM THIS THESIS)

1. 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. ***Molecular and Cellular Biology*** 25(14):6031-46.

2. Rodda DJ*****, Chew JL*****, Lim LH*****, Loh YH, Wang B, Ng HH, Robson P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. ***Journal of Biological Chemistry*** 280(26):24731-7.

3. 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. ***Nature Genetics*** 38(4):431-40.

*these authors contributed equally to this work

CHAPTER 1

GENERAL INTRODUCTION

1.1 Stem cells

Stem cells are classically defined as cells that can generate daughter cells identical to their parent cells (self-renewal) as well as can produce progeny with restricted potential (differentiation) (Smith, 2001; Weissman *et al.*, 2001). Stem cells can be isolated from the embryo, umbilical cord and adult tissues, with different levels of potency or differentiation ability.

In early mouse embryogenesis, the zygote and blastomeres from two- to four- cell embryo are totipotent. Totipotent cells can differentiate into any cell types of the body including the placenta, and can give rise to a whole organism. As the embryo continues to cleave, the blastomeres lose the potential to differentiate into all cell lineages. At 3.5 days *postcoitus* (dpc) during the formation of the blastocyst, the outer cells of the embryo develop into the trophectoderm (TE) from which the placenta is derived. The cells on the inside are called the inner cell mass (ICM) that can generate all cell lineages of the embryo proper, but they cannot give rise to totipotent stem cells nor a whole organism, and thus are considered pluripotent. Then at 4.5dpc, a primitive endoderm layer is observed at the surface of the ICM, and the remaining pluripotent cell population that is covered by the primitive endoderm is termed as the epiblast, which would produce all adult tissues (Figure 1.1). After implantation, the epiblast cells start to proliferate rapidly and increase in size. At 6.0 dpc, apoptotic cell death occurs in the central part of the epiblast, resulting in the formation of an epithelialized monolayer of pluripotent cells designated as the primitive ectoderm. At 7.0dpc, only primordial germ cells (PGCs) retain pluripotency, from which embryonic germ cells (EGCs) can be established *in vitro* (Coucouvanis and Martin, 1995; Matsui *et al.*, 1992) (Figure 1.1). Multipotent stem cells, with restricted differentiation potential, can be found in the umbilical cord and adult tissues such as haematopoietic stem cells that give

rise to blood cells (Moore and Lemischka, 2006; Petersen and Terada, 2001) (Figure 1.2). Stem cells, particularly embryonic stem cells (ESCs), offer enormous potential for a diverse range of cell-replacement therapies, in addition to their use as research tools for understanding self-renewal, lineage commitment and cellular differentiation.

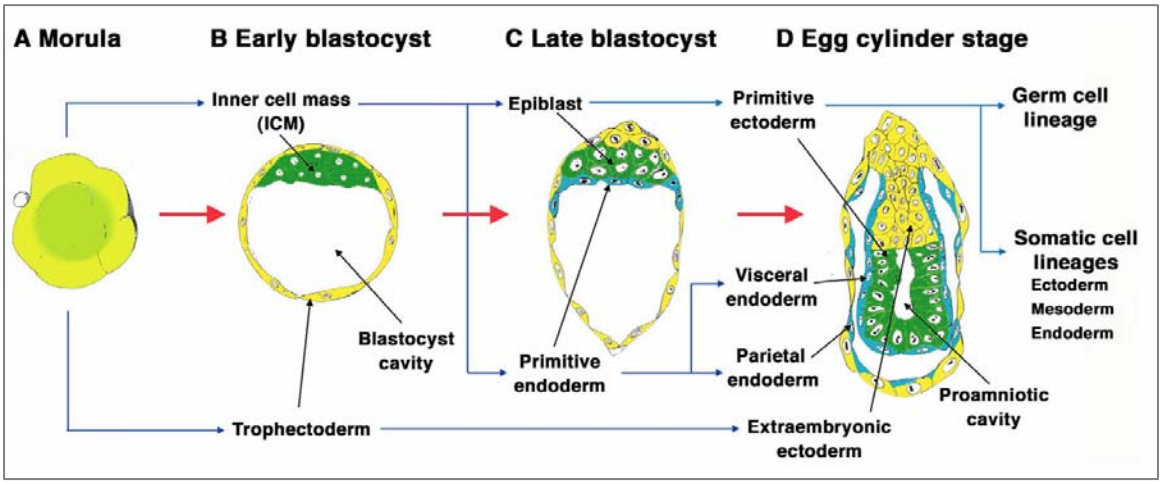


Figure 1.1: A schematic diagram of the mouse pre-implantation development. Pluripotent stem cells (green) appear in the morula as inner cells (A), which then form the inner cell mass (ICM) of the blastocyst (B). After giving rise to the primitive endoderm, pluripotent stem cells form the epiblast and start to proliferate rapidly after implantation (C). The pluripotent cells then form the primitive ectoderm, a monolayer epithelium that has restricted pluripotency that goes on to give rise to the germ cell and somatic lineages of the embryo (D). Figure was adopted from Niwa (2007).

1.2 Embryonic stem cells (ESCs)

Cells from the ICM can be isolated and cultured *in vitro* under permissive conditions as embryonic stem cell (ESC) lines (Bongso *et al.*, 1994; Evans and Kaufman, 1981; Martin, 1981; Thomson *et al.*, 1998) (Figure 1.2). These ESCs are able to give rise to all derivatives of the three primary germ layers (ectoderm, endoderm and mesoderm) as well as possessing the ability to differentiate into the trophectoderm lineage (Niwa *et al.*, 2005; Niwa, 2007; Pierce *et al.*, 1988). *In vivo*, ESCs can produce germ line chimeras. ESCs have a doubling time of about 12 hr, similar to that of the epiblast, and have similar expression pattern of specific marker genes (Pelton *et al.*, 2002).

ESCs exhibit unique properties, including unlimited self-renewal, long term proliferation *in vitro*, stable karyotype, highly efficient and reproducible differentiation potential *in vivo* and *in vitro* and germ line colonization. They also exhibit clonogenicity where the cells are capable of growing as separate colonies. They demonstrate high versatility in the types of cells they can differentiate into upon specific culture conditions. These properties of ESCs provide an enormous potential for clinical treatments of degenerative illnesses such as Parkinson's disease, whereby diseased or dysfunctional cells can be replaced with healthy, functional ones (Pera *et al.*, 2000). However, fundamental to the realization of this goal is a better understanding of the nature and basic biology of ESCs, such as genetic switches and molecular mechanisms that govern pluripotency. To this end, both human ESCs (hESCs) and mouse ESCs (mESCs) can be used as experimental models to further our understanding of the molecular basis of differentiation.

Human ESCs are relatively more difficult to culture and manipulate compared to mESCs (Pera *et al.*, 2000; Reubinoff *et al.*, 2000; Thomson *et al.*, 1998). Notwithstanding ethical issues in deriving hESCs, *in vivo* assessment of pluripotency is not possible for hESCs as these cells must be reintroduced into the human body to determine their ability to produce all somatic lineages and germ line chimerism. These cells must also generate embryoid bodies and teratomas containing differentiated cells of all three germ layers. Comparatively, mESCs are easier to culture, transfect and differentiate; and are thus used as a convenient cellular model to study the various aspects of ESC biology such as the mechanisms of pluripotency.

1.3 Properties of mouse ESCs (mESCs)

mESCs were originally isolated and maintained by co-culture on a feeder layer of mitotically inactivated mouse embryonic fibroblasts (MEFs) (Evans and Kaufman, 1981; Martin, 1981). They are characterized by their expression of distinctive cellular markers and possession of functions that relate to their uncommitted state. When cultured mESCs are reintroduced into a developing

blastocyst, a chimeric mouse is produced in which ESC-derived progeny can be found in all adult tissues. In addition, transplantation of undifferentiated mESCs into the adult results in the formation of teratomas, which are tumours that contain an array of different cell types, representative of each of the three embryonic germ layers. Upon cell fusion, mESCs can also dominantly reprogram somatic cells to re-express markers of earlier embryonic stages (Tada *et al.*, 2001). In addition to their developmental potential *in vivo*, mESCs display a remarkable capacity to form many differentiated cell types in culture (Keller, 2005; Smith, 2001). Differentiation of mESCs *in vitro* provides a powerful model system to compare between the undifferentiated and differentiated states, as well as to address questions related to cell fate determination.

1.3.1 Differentiation of mESCs

Mouse ESCs can be induced to differentiate into derivatives of the three embryonic germ layers (Figure 1.2) by altering the growth matrix, changing the chemical composition of the culture medium, co-culture with cell lines, or modifying the cells by genetic manipulation. Differentiation of the mESCs can be observed morphologically and also measured by reduced pluripotency markers such as Oct4 or increased lineage-specific markers such as Cdx2.

When mESCs are removed from contact with the feeder cells or gelatin and grown in suspension culture, they will clump together and spontaneously differentiate to form three-dimensional colonies called embryoid bodies (EBs) (Doetschman *et al.*, 1985; Keller, 2005; Smith, 2001). Cells within the EBs can differentiate into various committed cell types including cardiomyocytes (Maltsev *et al.*, 1993), skeletal muscle (Miller-Hance *et al.*, 1993), endothelial cells (Vittet *et al.*, 1996), neuronal cells (Bain *et al.*, 1995), adipocytes (Dani *et al.*, 1997) and hematopoetic precursors (Schmitt *et al.*, 1991; Smith *et al.*, 1988). In addition, removal of factors crucial for the maintenance of mESCs, such as leukemia inhibitory factor (LIF), also causes the mESCs to differentiate (Smith *et al.*, 1988). Besides that, addition of commonly used inducers of

mESCs differentiation such as retinoic acid (RA), also induces different types of differentiation when applied at various concentrations (Soprano *et al.*, 2007). Lower levels of retinoic acid have been found to promote the formation of epithelial-like cells, whereas higher levels favored the differentiation of mESCs into fibroblastic-like cells (Faherty *et al.*, 2005). Another way to induce differentiation of mESCs is through genetic manipulation. Some examples include the ectopic expression of Gata6 (into primitive endoderm) and Cdx2 (into trophectoderm), or RNAi-mediated depletion of factors which are important in maintaining pluripotency and self-renewal (Ding and Buchholz, 2006; Fujikura *et al.*, 2002; Hay *et al.*, 2004; Hough *et al.*, 2006b; Lim *et al.*, 2007; Niwa *et al.*, 2005; Tolkunova *et al.*, 2006).

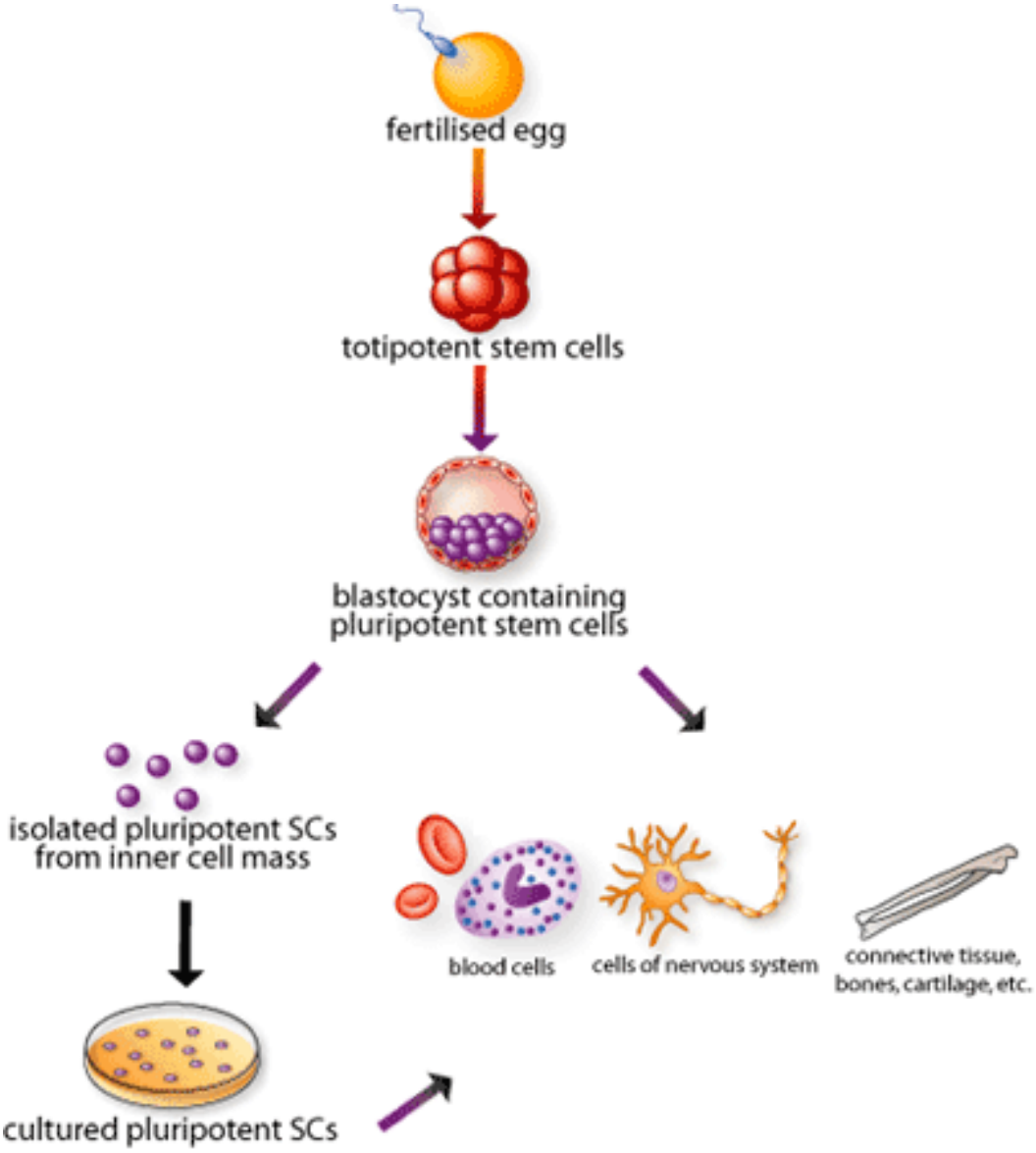


Figure 1.2: Mouse embryonic stem cell (mESC) culture, self-renewal and differentiation. The fertilized oocyte and blastomeres stage embryos are totipotent. The inner cell mass from the blastocyst gives rise to pluripotent mESCs. Undifferentiated mESCs can be cultured *in vitro* in the presence of LIF and FBS. Changing the culture conditions of mESC will allow differentiation into different cell types such as blood cells, neural cells and muscle cells. (Figure was adopted from the University of British Columbia at http://www.scq.ubc.ca/?p=317)

1.4 Maintaining mESCs in their undifferentiated state

The propagation of undifferentiated mESCs is dependent on extrinsic factors that induce signalling pathway cascades into the cell nucleus as well as intrinsic transcription factors that control the expression repertoire of the cell (Figure 1.3). In this model, cell-surface receptors initiate signals that are conveyed to the nucleus and affect key transcription factors such as Oct4 and Nanog, as well as self-renewal transcription factors such as signal transducer and activator of transcription-3 (STAT3) which in turn promotes the ESC state and inhibits differentiation (Figure 1.3).

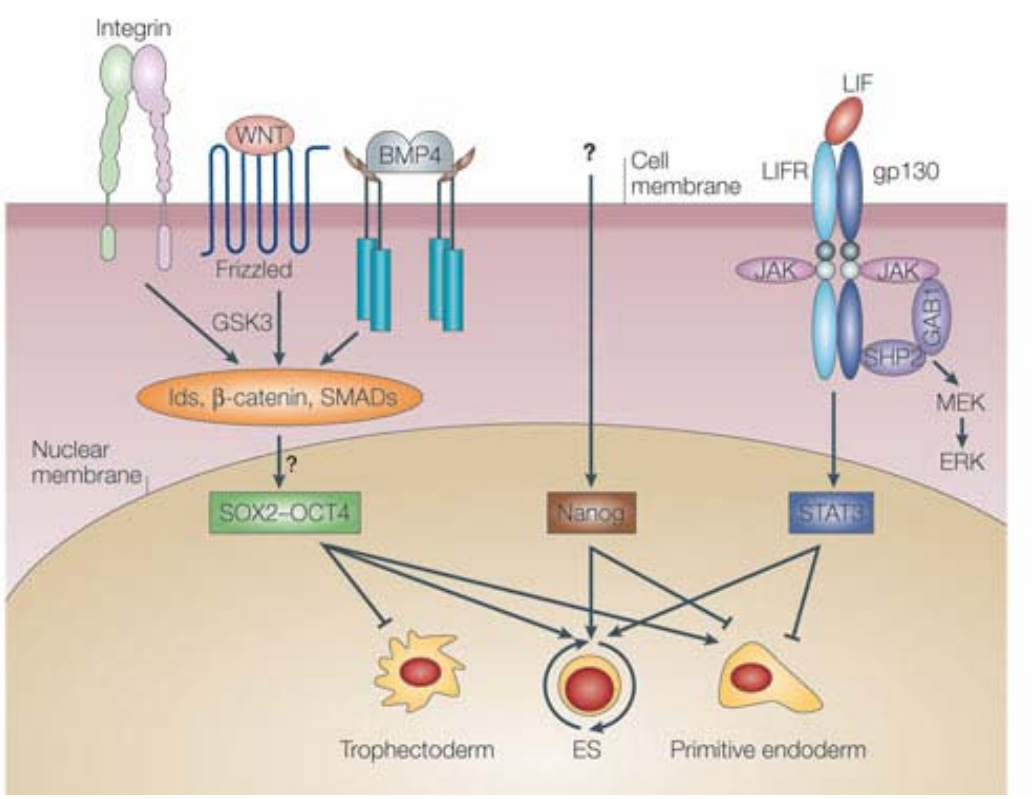


Figure 1.3: A model for extrinsic signalling and intrinsic transcription factor pathways in the maintenance of mESC pluripotency and self-renewal. Leukemia inhibitory factor (LIF) and bone morphogenic protein (BMP) pathways send signals into the nucleus. In the nucleus, transcription factors such as Oct4/Sox2, Nanog and Stat3 promote the ESC state and inhibit differentiation. Question marks denote lack of information. GAB1: GRB2-associated binding protein-1; GSK3: glycogen synthase kinase-3; Id: inhibitor of differentiation; JAK: janus kinase; MEK: mitogen-activated protein kinase (MAPK) and extracellular signal regulated kinase (ERK) protein kinase; SMAD: similar to mothers against decapentaplegic homologue; SHP2: SH2-domain-containing protein tyrosine phosphate-2; WNT: wingless type protein. Figure was adopted from Boiani and Scholer (2005).

1.4.1 Signalling pathways

In addition to general pathways of signal transduction that operate in most cell types, such as extracellular matrix (ECM) signalling that originates from the cell-membrane receptors of the integrin family, several exogenous factors that are involved in nucleus-directed signalling pathways are known to modulate ESC pluripotency both *in vivo* and *in vitro* (Boiani and Scholer, 2005).

Mouse ESCs are typically cultured on fibroblast feeder cells or gelatin with serum and leukemia inhibitory factor (LIF), a cytokine that activates the essential gp130/Stat3 signalling pathway (Smith *et al.*, 1992; Williams *et al.*, 1988). Combined with LIF, bone morphogenetic proteins (BMP) support unlimited ESC self-renewal in the absence of serum or feeders via the BMP pathway (Ying *et al.*, 2003). So far, the intracellular signalling cascades initiated by both LIF and BMP have been worked out extensively (Chambers and Smith, 2004). Other less studied pathways such the WNT pathway (Figure 1.3) has also been reported to be activated in ESCs (Hao *et al.*, 2006; Sato *et al.*, 2004), but it is known to have dichotomous effects on stem cells (both proliferative and roles in differentiation). It is likely that LIF, BMPs and other pathways exert their effects on mESCs by controlling the expression and activity of transcription factors such as Oct4 and Nanog, but evidence to back these facts is still lacking.

1.4.1.1 LIF-STAT3 signalling

Leukaemia inhibitory factor (LIF; also known as differentiation inhibitory activity, Dia) is a member of the interleukin-6 cytokine family which helps to maintain the pluripotency of mESCs (Smith *et al.*, 1988). LIF can be provided by MEFs (Rathjen *et al.*, 1990) or as a recombinant protein (Williams *et al.*, 1988). MEFs deficient for the LIF gene cannot maintain ESC self-renewal (Stewart *et al.*, 1992; Stewart and Cullinan, 1997) and LIF can replace the requirement for feeders in mESC culture (Smith *et al.*, 1988) and *de novo* derivation of mESCs (Nichols *et al.*, 1990). The

removal of LIF results in differentiation of mESCs, mainly toward the primitive endoderm lineage (Niwa *et al.*, 1998).

In the LIF-Stat3 pathway (Figure 1.3), LIF functions by binding to LIF receptor (LIFR) at the cell surface, which causes the receptor to heterodimerize with another transmembrane protein, glycoprotein-130 (gp130). On binding LIF, the intracellular domains of the LIFR–gp130 heterodimer can recruit the non-receptor Janus tyrosine kinase (JAK) and the anti-phospho tyrosine immunoreactive kinase (TIK) and become phosphorylated. The phosphorylated intracellular domains of the LIFR–gp130 heterodimer function as docking sites for proteins that contain the Src-homology-2 (SH2) domains, which include the latent transcription factor Stat3. Stat3 binds to the phosphotyrosine residues on the activated LIFR-gp130 and undergoes phosphorylation and dimerization. Subsequently, phosphorylated Stat3 dimers translocate to the nucleus where they function as transcription factors (Matsuda *et al.*, 1999; Niwa *et al.*, 1998). Signalling of gp130 is not limited to activation of Stat3 but includes stimulation of the Ras/ mitogen-activated protein kinase pathway (MAPK) which activates extracellular regulated kinase (ERK) when cell surface receptors are stimulated by a complex containing Grb2 adaptor and Sos guanine-nucleotide-exchange factor (Kolch, 2000). Activation of ERK has a pro-differentiation effect and is antagonistic to ESC self-renewal (Burdon *et al.*, 1999).

Activated Stat3 formed downstream of LIF induction are crucial in preventing the differentiation of ESCs. Previous co-immunoprecipitation studies have indicated that Stat3 is a component of the tyrosine-phosphorylated complex that forms upon LIF induction (Boeuf *et al.*, 1997; Hocke *et al.*, 1995). Moreover, tyrosine kinase inhibitors were shown to impair the formation of the activated Stat3 complex and cause differentiation of ESCs (Boeuf *et al.*, 1997). Blockage of activation of Stat3 by overexpression of its dominant-negative mutant in the presence of LIF induces mesoderm-endoderm differentiation similar to that induced by the withdrawal of

LIF, indicating that Stat3 is essential for LIF action (Niwa *et al.*, 1998). Matsuda *et al.* (1999) reported that activation of Stat3 is sufficient to maintain ESC self-renewal in the presence of serum without LIF.

The knocking-out of LIF, Stat3 or gp130 in mice results in severe developmental disorders indicating that these factors play important roles in embryonic development of mice. LIF mutant female mice are infertile as the adherence of the embryo to the uterus wall is dependent on the production of estrogen and LIF by the uterus (Stewart *et al.*, 1992). Consequently, $LIF^{-/-}$ females fail to support embryonic implantation. However, $LIF^{-/-}$ embryos can implant and develop to full term foetuses in wild type mice possessing a normal uterus. LIFR-null embryos died shortly after birth and exhibit reduced bone mass and profound loss of motor neurons (Li *et al.*, 1995; Ware *et al.*, 1995). Embryos homozygous for the gp130 mutation died between 12-18 dpc due to placental, myocardial, hematological and neurological disorders (Nakashima *et al.*, 1999; Yoshida *et al.*, 1996). Epiblast cells of the ICM of $gp130^{-/-}$ embryos differentiate into parietal endoderm and undergo apoptosis. Moreover, $LIFR^{-/-}$ and $gp130^{-/-}$ delayed embryos are unable to resume embryogenesis after 12 and 6 days of diapause (arrest of embryonic development in lactating mothers), respectively. Targeted disruption of the Stat3 gene *in vivo* also leads to embryonic lethality due to the failure to establish metabolic exchanges between the embryo and maternal blood (Takeda *et al.*, 1997). Stat3-null embryos develop into elongated egg cylinders and degenerate around embryonic day (E) 7.0. Thus, Stat3 plays a unique and crucial role in embryonic development that cannot be compensated for by other members of the Stat family (Stat1, 2, 5a, 5b).

Collectively, these studies show that the LIF-Stat3 pathway is important in maintaining pluripotency and self-renewal. However, LIF alone is not sufficient for mESC self-renewal as the cells still undergo differentiation when cultured without serum in the presence of LIF (Ying *et al.*,

2003). Studies by Ying *et al.* (2003) have identified the bone morphogenic proteins (BMPs) to be one of the components in serum which contributes to mESC self-renewal. Together with LIF, BMP can maintain mESCs in their undifferentiated state without the use of serum or feeder cells (Ying *et al.*, 2003).

1.4.1.2 BMP signalling

Bone morphogenic proteins (BMPs) belong to the transformation growth factor beta (TGFβ) superfamily. BMP4, BMP2 or growth and differentiation factor 6 (GDF6) have been demonstrated to suppress differentiation and results in concomitant self-renewal in mESCs by blocking neuronal differentiation (Finley *et al.*, 1999; Tropepe *et al.*, 2001; Ying *et al.*, 2003) and promote non-neuronal (mesoderm, endoderm, trophoblast) differentiation (Ying *et al.*, 2003). BMP4 can also replace serum during *de novo* derivation of ESCs. Loss of BMP4 function results in embryonic development disorders in mice such as gastrulation defects, disorganized posterior structure and devoid of primordial germ cells (PGCs) (Dunn *et al.*, 1997; Lawson *et al.*, 1999; Winnier *et al.*, 1995).

In the canonical BMP signalling pathway, BMPs bind to the BMP receptors (BMPRs) which are serine/threonine kinases composed of type I and II subtypes. Subsequently, BMPR type I phosphorylates the C-terminus of Smad1, Smad5 or Smad8. Upon phosphorylation, Smad1/5/8 forms dimers and complexes with the common Smad protein, Smad4. The heterotrimeric complex then translocates into the nucleus and cooperates with other transcription factors to modulate the transcription of target genes such as *Inhibitor of differentiation* (*Id*) genes (Derynck and Zhang, 2003; Miyazono *et al.*, 2005; Moustakas *et al.*, 2001; Shi and Massague, 2003; Wotton *et al.*, 1999; Zhang and Li, 2005). Smad1 and BMP4 are found to be crucial for mediating the embryonic development of mice. $Smad1^{-/-}$ embryos die by 10.5 dpc due to their inability to connect to the

placenta (Tremblay *et al.*, 2001b). BMP4 may also be involved in the maintenance of ESCs by inhibition of the ERK/MAPK (Qi *et al.*, 2004).

BMP and LIF pathways are important in maintaining mESCs by cascading signals into the cell nucleus. However, little is known about how LIF and BMPs control intrinsic molecular determinants of pluripotency and self-renewal.

1.4.2 Key transcription factors controlling pluripotency

The pluripotency of ESCs is intrinsically regulated by ESC-specific transcription factors which includes Oct4, Sox2 and Nanog (Chambers, 2004; Chambers and Smith, 2004; Niwa, 2001; Wang *et al.*, 2006). Oct4, Sox2 and Nanog-deficient ESCs are not capable of extensive self-renewal and spontaneously differentiate, even in the presence of LIF and serum (Chambers *et al.*, 2003; Mitsui *et al.*, 2003; Masui *et al.*, 2007; Nichols *et al.*, 1998). All three transcription factors including the substrate of LIF, Stat3, help to maintain mESCs in its undifferentiated condition (Figure 1.3). Nanog and LIF/Stat3 were reported to suppress primitive endoderm differentiation while Oct4 over-expression promotes it. Oct4 and Sox2 have been shown to inhibit differentiation into the trophectodermal lineage (Figure 1.3; Masui *et al.*, 2003; Mitsui *et al.*, 2007; Niwa *et al.*, 1998).

1.4.2.1 Oct4

1.4.2.1.1 Oct4 structure

Oct4 (also referred to as Oct-3/4), encoded by *Pou5f1*, is a homeodomain transcription factor belonging to class V of POU (Pit, Oct, Unc) factors. Two domains spanning the N- and C-terminal portion of Oct4 protein define the transactivation capacity of the POU transcription factor (Imagawa *et al.*, 1991; Okamoto *et al.*, 1990; Vigano and Staudt, 1996). The N-terminal region is a proline- and acidic residue- rich region, whereas the C-terminal region is rich in proline, serine

and threonine residues. The N-terminal domain can function as an activation domain in heterologous cell systems, while the C-terminal domain consists of a POU-domain which mediates cell-type specific functions (Brehm *et al.*, 1997). In a complementation assay, Niwa *et al.* (2002) used a conditional Oct4-null ESC line to establish the essential domains of the Oct4 protein that are crucial in maintaining ESCs in an undifferentiated state (Niwa *et al.*, 2002). The complementation assay was based on the ability of a protein to rescue the self-renewal capability of cells that would otherwise differentiate because of Oct4 downregulation. Oct4 was found to be the only POU domain containing protein to have the ability to rescue the self-renewing phenotype, as Oct2 and Oct6 were found to have no effect on cell fate in this system. Intriguingly, truncated Oct4 protein containing the Oct4 POU domain and either the C- or N-terminal domain can support ESC self-renewal. Gene expression analysis revealed that Oct4 transactivation domains have the ability to elicit the activation of different target genes such as *Sox2*, *Fgf4*, *Utf1*, *Zfp42*, *Lefty1* and *Otx2*.

Oct proteins have a common conserved DNA binding domain called the POU domain, which is a bipartite module comprising of two structurally independent subdomains: a 75 amino acid N-terminal POU-specific domain (POU_s) and a 60 amino acid C-terminal POU homeodomain (POU_H) connected by a flexible linker of 15-56 amino acids (Herr and Cleary, 1995). Both of these domains are required for DNA binding through a helix-turn-helix structure. The POU domain binds DNA via the POU_s subdomain to an octamer DNA consensus sequence ATGC(A/T)AAT (Scholer, 1991). The POU domain also binds to DNA via the interaction of the third recognition helix of the POU_H subdomain with bases in the DNA major groove at the TAAT core site. This DNA region interacts with the Oct proteins at a lower affinity compared to the earlier mentioned octamer DNA sequence. The POU_s domain exhibits a site-specific, high affinity DNA binding and bending capability (Verrijzer *et al.*, 1991). Besides mediating the binding of

POU factors to DNA, both the subdomains can also participate in protein-protein interactions (Brehm *et al.*, 1999; Vigano and Staudt, 1996).

1.4.2.1.2 Oct4 expression and function

Oct4 is expressed in totipotent and pluripotent cell populations during development (Nichols *et al.*, 1998). It is initially expressed as a maternal transcript and is essential for the formation of the pluripotent ICM (Nichols *et al.*, 1998). Oct4 is expressed at low levels in all blastomeres until the four-cell stage (Palmieri *et al.*, 1994). At this particular stage, the *Oct4* gene undergoes zygotic activation resulting in high Oct4 protein levels in the nuclei of all blastomeres until compaction (Yeom *et al.*, 1991). After cavitation, Oct4 expression is maintained only in the ICM of the blastocyst and is downregulated in the differentiated trophectoderm (Okamoto *et al.*, 1990; Rosner *et al.*, 1990; Scholer *et al.*, 1990). After implantation, Oct4 expression is restricted to the epiblast. During gastrulation at 6.0-6.5dpc, it is downregulated and from 8.5dpc, Oct4 is restricted to precursors of the gametes or PGCs. Oct4 is also expressed in undifferentiated mouse ESC, EGC and ECC (embryonic carcinoma cell) lines (Okamoto *et al.*, 1990; Rosner *et al.*, 1990; Smith *et al.*, 1992; Yeom *et al.*, 1996).

Oct4 plays a central function in embryonic development and in maintaining undifferentiated ESCs. Oct4-deficient embryos die at the peri-implantation stage and form empty deciduas or implantation sites that contain trophoblastic cells that are devoid of yolk sac or embryonic structures (Nichols *et al.*, 1998). *In vitro* cultures of the cells from the inner region of the Oct4-deficient blastocyst contain trophoblastic giant cells and not pluripotent cells nor extraembryonic endoderm. Moreover, a critical amount of Oct4 has been found to be crucial for the maintenance of ESC self-renewal (Niwa *et al.*, 2000). In the study, Niwa *et al.* used ESCs with inactivated endogenous Oct4 alleles that were maintained by tetracycline regulated transactivator constructs activating a transgene expressing Oct4. They showed that downregulation of

endogenous Oct4 to below 50% of its original levels in undifferentiated mESCs resulted in the cells to be committed towards trophectoderm lineages due to the upregulation of transcription factors Cdx2 and Eomesodermin (Eomes) (Niwa *et al.*, 2005). This result is consistent with $Oct4^{-/-}$ embryos (Nichols *et al.*, 1998). On the other hand, an increase beyond 50% of the endogenous levels of Oct4 leads to the concomitant differentiation of ESCs into extraembryonic endoderm and mesoderm, similar to that observed upon LIF withdrawal (Niwa *et al.*, 2000). Therefore, Oct4 governs commitment of embryonic cells along three distinct cell fates: self-renewal, trophectoderm, extraembryonic endoderm and mesoderm.

Retinoic acid triggers the rapid downregulation of Oct4 in ESC and ECC (Okamoto *et al.*, 1990; Rosner *et al.*, 1990; Scholer *et al.*, 1989). In ESCs, Oct4 activates gene transcription irrespective of the distance of the octamer motif from the transcriptional initiation site (Okamoto *et al.*, 1990; Scholer *et al.*, 1989). In differentiated cells, Oct4 can transactivate gene transcription only from a proximal location. However, interaction between adenovirus protein EA1 or human papillomavirus E7 oncoprotein and the Oct4 POU domain is sufficient for Oct4 to elicit transcriptional activation from remote binding sites (Scholer *et al.*, 1991). E1A and E7 proteins would therefore mimic unidentified ESC-specific coactivators that serve a similar function in pluripotent cells.

1.4.2.1.3 Regulation of Oct4 expression

The *Oct4* gene comprises a TATA-less proximal promoter as well as two stem-cell-specific regulatory elements, a distal enhancer and a proximal enhancer (Pikarsky *et al.*, 1994; Schoorlemmer *et al.*, 1994; Sylvester and Scholer, 1994; Yeom *et al.*, 1996). Many positive and negative regulators are recruited to *Oct4* at these sites. Among these are the Liver receptor homolog 1 (Lrh1 or Nr5a2) which is a putative positive regulator of *Oct4*. Oct4 expression is lost in the epiblast of Lrh1-null embryos and is quickly downregulated after the induction of

differentiation in Lrh1-null ESCs (Gu *et al.*, 2005a). Conversely, the germ cell nuclear factor (Gcnf or Nr6a1) is a potential *Oct4* negative regulator. Low Gcnf expression is detected in the whole mouse embryo at 6.5dpc where Oct4 expression is high. At 7.5dpc, increasing Gcnf and correspondingly decreasing Oct4 mRNA levels are observed in neural folds and at the posterior of the embryo. In Gcnf-deficient embryos, Oct4 mRNA is detected in the putative hindbrain region and posterior of the embryo (Chung and Cooney, 2001; Fuhrmann *et al.*, 2001). This indicates that loss of Gcnf leads to loss of Oct4 repression in somatic cells and loss of Gcnf-induced restriction of Oct4 in the germ line. The same phenotype is observed in Gcnf-deficient mice containing a targeted deletion of the DNA-binding domain of Gcnf (Lan *et al.*, 2002). Oct4 repression following the induction of differentiation is also delayed in Gcnf-null ESCs (Gu *et al.*, 2005b). In addition to Gcnf, Chicken ovalbumin upstream promoter-transcription factors (Coup-tf) I and II, encoded by *Nr2f1* and *Nr2f2*, respectively, also function as negative regulators of Oct4 expression (Ben Shushan *et al.*, 1995; Gu *et al.*, 2005b). In addition to Oct4 itself, Sox2 and Nanog have also been implicated to regulate Oct4 expression, which will be discussed in detail in the following chapters. FoxD3 has also been implicated as an activator for Oct4 expression (Pan *et al.*, 2006). The balance between these positive and negative regulators might determine the precise level of Oct4 expression.

1.4.2.2 Sox2

Sox2 is a member of the mammalian testis-determining factor Sry-related high mobility group (HMG box-containing) transcription factor family that binds to DNA through its 79 amino acid high mobility group (HMG) domain (Kamachi *et al.*, 2000). In contrast to most DNA-binding proteins, which access DNA through the major groove, the HMG box interacts with the minor groove of the DNA helix, and introduces a bend in the DNA molecule.

Sox2 is indispensable for maintaining pluripotency. Sox2 is co-expressed with Oct4 in the ICM of preimplantation embryos, ESCs, ECCs, and EGCs (Avilion *et al.*, 2003). Sox2 is also expressed in neuronal multipotent cells (Zappone *et al.*, 2000). Sox2 homozygous null embryos stop development in the peri-implantation stage (as found in Oct4 mutants) and die immediately after implantation (Avilion *et al.*, 2003). Using an inducible system, Masui *et al.* (2007) recently showed that Sox2 null mESCs differentiate primarily into trophectoderm-like cells. In concurrence, knockdown of Sox2 in mESCs also drives the cells towards the trophectodermal lineage (Ivanova *et al.*, 2006). In contrast, Sox2 overexpression did not impair propagation of undifferentiated mESCs. However, when mESCs were released from self-renewal, they differentiated into the neurectoderm. More recently, Sox2 was also implicated in controlling transcriptional regulators of Oct4, thereby affecting Oct4 expression levels in mESCs (Masui *et al.*, 2007).

The two regulatory regions termed Sox2 regulatory region 1 (SRR1) and SRR2, were identified based on their activities in pluripotent ESCs where they exert their function specifically when cells are in an undifferentiated state (Tomioka *et al.*, 2002). It was shown that SRR2 has a regulatory core sequence comprising octamer and Sox-2 binding sequences and that SRR2 exhibits its activity by recruiting the Oct4-Sox2 complex to it in ESCs. The Sox2 HMG domain binds to the DNA recognition sequence CTTTGTT through numerous specific hydrogen bonds. Sox2 is able to bind to DNA on its own, but with a significantly lower affinity compared with binding to DNA as part of a ternary complex with POU proteins (Remenyi *et al.*, 2003). Studies on Sox2 as a transcriptional regulator have mostly been in combination with Oct4, where Oct4 and Sox2 have been shown to interact and bind to multiple binding sites as a complex (Ambrosetti *et al.*, 2000; Remenyi *et al.*, 2003).

1.4.2.2.1 Oct4 and Sox2 partnership

Oct and Sox proteins selectively interact with each other via their conserved domains POU and HMG, respectively, which also bind to DNA. The C-terminal domains of both Sox2 and Oct4 contribute to the functional activity of Sox2/Oct4 complex (Ambrosetti *et al.*, 2000). The C-terminal domain of Oct4 is active when Oct4 exists as an Oct4/Sox2 complex. The synergistic action of Sox2 and Oct4 results from two distinct yet concerted events. Cooperative binding of Sox2 and Oct4 to the DNA via their respective DNA binding domains, and upon tethering of each factor to the enhancer region forming a ternary complex, new DNA-protein and protein-protein interactions induce conformational changes that may lead to activation of latent domains and constitute a new, distinct platform for the recruitment of other coactivators (Dailey and Basilico, 2001). Their functional partnership has been characterised on regulatory elements in various species including that of human and mice (Dailey and Basilico, 2001). Recently, the importance of Oct4 and Sox2 in the maintenance of pluripotency was further reinforced in an artificial system when Takahashi and Yamanaka (2006) reported that the expression of Oct4 and Sox2 with two other transcription factors, c-Myc and Klf4, could induce murine fibroblasts to become pluripotent ES-like cells (Takahashi and Yamanaka, 2006; Wernig *et al.*, 2007).

An example of a target gene with Sox2-Oct4 composite binding element is the *Fgf4* (fibroblast growth factor 4) enhancer, which was the first DNA element that was described to contain a composite DNA element binding Sox2 and Oct4 (Yuan *et al.*, 1995). The *Fgf4* enhancer in the untranslated region of exon 3 consists of a closely juxtaposed Oct4 binding site and Sox2 binding site. Subsequent biochemical work showed the cooperative nature of Sox2 and Oct4 interaction and the requirement of a specific arrangement of binding sites within the *Fgf4* enhancer DNA (Ambrosetti *et al.*, 1997). The different degrees of Sox2/Oct4 cooperativity on regulatory elements *in vitro* are in congruence with the regulation of *Fgf4* during development, indicating that

varying the amount of cooperativity in complex formation could result in distinct functional properties *in vivo*, such as varying amount of transcription level production.

Remenyi *et al.* (2003) analysed the crystal structure of the ternary Oct1/Sox2/*Fgf4* (fibroblast growth factor 4) enhancer element complex and then used homology modelling tools to construct an Oct4/Sox2/*Fgf4* as well as an Oct4/Sox2/*Utf1* structural model. These models revealed that the *Fgf4* and the *Utf1* enhancers mediate the assembly of distinct POU/HMG complexes, leading to different quaternary arrangements by swapping protein–protein interaction surfaces of Sox2. Sox2 is also able to bind with other Oct proteins *in vitro* and this may confer the specificity of Sox-Oct complexes on different target genes. For example, the POU domains of several family members, including the prototype member Oct1, bind cooperatively with the HMG domain of Sox2 onto the *Fgf4* enhancer (Ambrosetti *et al.*, 1997).

The specific expression of Utf1 in ESCs is also regulated by the synergistic action of Sox2 and Oct4, where the binding sites of these factors have no intervening spacing (Nishimoto *et al.*, 1999; Okuda *et al.*, 1998) . In addition, a non-canonical binding site for Oct4 and Sox2 has been found in the 3' regulatory region of the *Sox2* gene, SRR2 and is involved in Sox2 expression in ESCs (Tomioka *et al.*, 2002). Oct4 or Oct6 (but not Oct1) has been shown to increase Sox2 dependent transcription of the *Sox2* gene. Cooperative binding of Oct4 and Sox2 results in embryonic expression of the F-box containing protein 15 (Fbx15) and mutation of either binding site is known to abolish the activity of the enhancer (Tokuzawa *et al.*, 2003). In addition, other pluripotency-associated genes such as *Lefty1*, *Oct4*, *Sox2* and *Nanog* have been shown to be regulated by an enhancer element that contains the Oct4 and Sox2 binding motif and these genes are highly expressed in undifferentiated ESCs but not in differentiated cells (this study; Chew *et al.*, 2005; Kuroda *et al.*, 2005; Nakatake *et al.*, 2006; Okumura-Nakanishi *et al.*, 2005; Rodda *et al.*, 2005; Tomioka *et al.*, 2002). Chromatin immunoprecipitation (ChIP) studies have also

demonstrated that Oct4 and Sox2 co-occupy a few thousand regulatory sites in the genome of ESCs (this study; Boyer *et al.*, 2005; Loh *et al.*, 2006).

1.4.2.3 Nanog

Nanog was identified as another transcription factor whose functions are essential in maintaining the ESC state (Chambers *et al.*, 2003; Mitsui *et al.*, 2003). It is an NK2-family homeobox transcription factor containing an N-terminal domain, homeobox domain and C-terminal domain and is expressed *in vivo* in the interior cells of compacted morulae, ICM or epiblast, and germ cells. *In vitro*, Nanog is utilised as a marker for all pluripotent cell lines such as ESC, EGC and ECC. Nanog expression is downregulated upon differentiation of these cells.

Nanog is essential for the maintenance of a pluripotent phenotype and endoderm specification is caused by Nanog downregulation. Nanog-deficient embryos die after implantation due to a failure in specification of the pluripotent epiblast, which is diverted to the endodermal fate. ESCs derived from $Nanog^{-/-}$ blastocysts differentiate into parietal-endoderm lineages (Mitsui *et al.*, 2003). Knockdown of Nanog in mESCs also showed similar differentiation phenotype (Hough *et al.*, 2006a; Hough *et al.*, 2006b). However, recently, Chambers *et al.* (2007) reported that Nanog null ESCs are more susceptible to differentiation but could still proliferate as pluripotent stem cells, whereas Nanog expression appears to suppress differentiation. Overexpression of Nanog in mESC circumvents the necessity of either LIF or BMP stimulation (Chambers *et al.*, 2003; Ying *et al.*, 2003), suggesting that Nanog may be a downstream effector for extrinsic signalling molecules. Nanog has also been shown to reinstate pluripotency in somatic cells after fusion (Silva *et al.*, 2006). Nanog expression has been reported to be positively regulated by Oct4 and Sox2 (Kuroda *et al.*, 2005; Rodda *et al.*, 2005) and FoxD3 (Pan *et al.*, 2006), while negatively regulated by p53 (Lin *et al.*, 2005) and Tcf3 (Pereira *et al.*, 2006). Recently, Stat3 and T (Brachyury) binding sites were also identified in the *Nanog* enhancer region

and Nanog was found to interact with Smad1 to inhibit the activity of BMP signalling (Suzuki *et al.*, 2006). Wang *et al.* (2006) has also identified Nanog-associated proteins such as zinc finger proteins that form protein interaction networks in mESCs. In addition, Sall4 is another transcription factor that was shown to interact with Nanog and associate together at genomic sites *in vivo* (Wu *et al.*, 2006).

1.4.2.4 Other transcription factors in the maintenance of mESC

The identification of the Oct4, Sox2, Nanog 'triad' as master regulators has been an important advancement in stem-cell biology, although the expression of the triad does not, in itself, guarantee pluripotency. For example, ECCs express these three factors at appreciable levels, but are able to develop along only a limited range of specific developmental pathways. This indicates that additional regulators are required to establish or efficiently retain the pluripotent state.

Other regulators of pluripotency have been identified in screens for genes that give ESCs a selective advantage in self-renewal. RNAi was used to screen genes that were required to maintain ESCs in an undifferentiated state (Ivanova *et al.*, 2006). This study identified four genes, *estrogen-related receptor-β* (*Esrrb*), *T-box 3* (*Tbx3*), *T-cell lymphoma breakpoint 1* (*Tcl1*) and *developmental pluripotency-associated 4* (*Dppa4*), in addition to *Oct4*, *Sox2* and *Nanog*. By analysing changes in ESC transcription after the knockdown of each of these six genes, the authors identified three sets of target genes: 800 genes that were either up- or downregulated in response to most knockdowns; 474 that were affected only by the knockdowns of *Oct4*, *Sox2* and *Nanog*; and 272 that responded to the *Esrrb*, *Tbx3*, *Tcl1* and *Dppa4* knockdowns. These findings indicated that at least two separate pathways control ESC self-renewal. *Zfx* have also been shown to be involved in regulating ESC self-renewal (Galan-Caridad *et al.*, 2007). Other studies have also characterized downstream targets that contribute to the maintenance of ESCs, such as *Zfp206* (Wang *et al.*, 2007) and *Zic3* (Lim *et al.*, 2007). In a recent report, the ectopic expression of four genes, *Oct4*,

Sox2, *Klf4* and *c-Myc*, have been shown to convert mouse embryonic fibroblasts to ES-like pluripotent stem cells (Takahashi and Yamanaka, 2006). However, the low frequency of this conversion suggests that other factors might be required for 'resetting' developmental potential. Alternatively, these converting factors might be effective in only a minority of fibroblasts that might have already acquired stem-like properties.

1.5 Cell cycle and proliferation of mESCs

Under optimized culture conditions, mESCs divide symmetrically every 12 hr. During self-renewal, most ESCs are in the S phase of the cell cycle, with only a few in G1 (Burdon *et al.*, 2002; Prost *et al.*, 1998). When ESCs begin to differentiate, the G1 phase of the cell cycle becomes longer and the rate of cell division slows. A number of pathways and factors have been reported to play a role in promoting the proliferation, survival and/or differentiation of mESCs, and this includes the phosphoinositide-3-kinase (PI3K)/Akt/Eras/Tcl1 signalling (Ivanova *et al.*, 2006; Matoba *et al.*, 2006; Sun *et al.*, 1999; Takahashi *et al.*, 2003; Watanabe *et al.*, 2006), transcription factor b-Myb (Iwai *et al.*, 2001; Tanaka *et al.*, 1999) and Myc (also known as c-Myc) which acts via the activation of cyclin E expression to promote G1-S transition (Cartwright *et al.*, 2005; Hooker and Hurlin, 2006). It was shown that c-Myc is a direct target of Stat3, and that overexpression of a dominant-active form of c-Myc attenuates self-renewal of mESCs independent of LIF, which suggests that the regulation of the G1-S transition may be linked to the maintenance of pluripotency (Burdon *et al.*, 2002). In addition, mESCs with reduced expression of Undifferentiated embryonic cell transcription factor 1 (Utf1) and mESCs that lack Sall4 showed reduced proliferation ability (Nishimoto *et al.*, 2005; Sakaki-Yumoto *et al.*, 2006).

1.6 Epigenetic modifications in mESCs

Epigenetic processes such as DNA methylation, histone modifications and RNAi may also be required for proper ESC differentiation (Niwa, 2007) or to facilitate access for transcription factor

action (Azuara *et al.*, 2006; Smith, 2005). Recently, studies examining specific epigenetic features of ESCs such as the abundance of modified histones, Polycomb group (PcG) protein-binding patterns, replication timing and chromatin accessibility have revealed that ESCs manage their pluripotent status by priming important regulator genes for future expression (Bernstein *et al.*, 2006; Boyer *et al.*, 2006a; Boyer *et al.*, 2006b; Lee *et al.*, 2006).

1.7 Building the transcriptional network in mESCs

1.7.1 Transcriptional regulators

Transcriptional regulation is controlled by the sequence-specific DNA binding activity of transcription factors (Tjian and Maniatis, 1994; McKnight and Tjian, 1986; Blackwood and Kadonaga, 1998). They determine the expression repertoire and thus the cellular phenotype of a cell. Regulated eukaryotic gene transcription involves the assembly of an initiation complex at the core promoter region and the coordinated binding of multiple transcription factors and regulatory complexes to the promoter and enhancer regions. Transcription factors affect the basal transcriptional machinery and regulate the transcriptional rate. Hence, they influence the expression level of the corresponding gene. Their ultimate action may be positive (activators) or negative (repressors). The activity of many of these transcription factors is regulated in a number of different ways by distinct signal transduction pathways: by changing the rate of synthesis or degradation of the protein, by post-translational modification of the transcription factor or by altering its subcellular localisation.

Proximal promoter bound transcription factors are thought to play a direct role in transcriptional initiation, since many transcription factors contain distinct domains that directly contact the basal transcriptional machinery. Enhancer elements are *cis*-acting sequences that stimulate transcription in an orientation-independent manner and can operate from long distances

(Dynan and Tjian, 1985). Studies have highlighted the importance of higher order structure in regulating gene expression, particularly in the context of chromatin architecture (Cosma, 2002; Wolffe *et al.*, 1997). Besides proximal promoter and enhancer elements, other elements such as locus control region, insulator and matrix attachment regions may also associate with transcription factors in order to facilitate gene expression (Carey and Smale, 2000).

With the completion of the mouse (Waterston *et al.*, 2002) and human (Lander *et al.*, 2001; Venter *et al.*, 2001) genome, emerging studies are now defining the transcriptome to decipher the orchestration of gene expression, which is fundamental to the phenotype of a cell and ultimately the identity of the organism. Mapping and building regulatory networks in ESCs began with Oct4, Sox2 and Nanog as the building blocks and continued to gain impetus with the advent of high throughput technologies.

1.7.2 Technologies for studying the transcriptome

The transcriptome is a complex collection of transcripts, due to alternative splicing, temporal and spatial expression. Powerful high-throughput technologies have been developed to dissect the transcriptome and to identify genes potentially involved with ESC self-renewal such as (1) transcriptional profiling for global gene expression, (2) high-throughput RNA interference (RNAi) that allows systematic perturbation of biological systems, and (3) genome-wide analysis of protein-DNA interactions which facilitates the study of transcriptional networks (Liu, 2005; Ruan *et al.*, 2004). The advent of these high-throughput tools together with computational bioinformatics allows the study of biological processes in a global scale.

1.7.2.1 Transcriptional profiling

Transcriptional profiling refers to the simultaneous analysis of gene expression of all the transcripts encoded by the genome. Techniques that are currently being used for transcriptional

profiling include serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS), and expression microarrays (Ruan *et al.*, 2004). SAGE is a technology specifically designed to digitally quantify expression of genes annotated by short cDNA tags. MPSS couples a SAGE-like approach with a novel restriction–ligation, bar code identification, and bead-based detection system to identify 17–20 bp tags of every transcript in an RNA sample (Brenner and Livak, 1989; Brenner *et al.*, 2000). Both SAGE and MPSS involve the sequencing of transcripts. The primary feature of expression microarrays is the arrangement of DNA probes precisely placed on a two dimensional surface, upon which the RNA targets of analysis are overlaid (Stears *et al.*, 2003).

In the field of stem cells, attempts to uncover a common set of molecular properties that define the uncommitted state were pioneered several years ago in a series of microarray expression studies (Ivanova *et al.*, 2002; Ramalho-Santos *et al.*, 2002; Rao and Stice, 2004). Systematic, genome-wide interrogations have identified hundreds of genes, including several transcription factors, which have expression patterns tightly correlated with ESC differentiation. Examples of methods used include gene expression profiling (Furusawa *et al.*, 2006; Tanaka *et al.*, 2002), EST (Brandenberger *et al.*, 2004; Palmqvist *et al.*, 2005; Sharov *et al.*, 2003), MPSS (Wei *et al.*, 2005) and SAGE (Richards *et al.*, 2004). In addition, global expression profiling was used to delineate the downstream target genes of Oct4 in Oct4-manipulated ESCs (Matoba *et al.*, 2006).

Although transcriptional profiling provides information about the genes that are expressed by a particular cell type and their relative abundance, it is unable to define the functional contribution of a particular gene to a certain biological process. In this aspect, one approach used to determine the role of a particular gene in a biological process is to remove, reduce or inactivate the gene and subsequently observing the effects on the system.

1.7.2.2 RNAi screen

Classical loss-of-function genetic approaches involve the random mutagenesis of genomic DNA followed by identification of resulting mutant phenotypes, or targeted disruption of a desired genetic locus. Such forward and reverse genetic screens are time-consuming, technically challenging and costly (Rajewsky *et al.*, 1996). This can be circumvented by using a technique called RNA interference (RNAi). RNAi is an evolutionary conserved phenomenon discovered in *Caenorhabditis elegans* (Fire *et al.*, 1998) and has since been shown to operate in other organisms ranging from yeast to mammals (Sen and Blau, 2006). RNAi can be mimicked experimentally whereby the introduction of double-stranded RNA (dsRNA) corresponding to a particular mRNA causes the specific and rapid degradation of that mRNA in cells. Large-scale RNAi screens involving the systematic knockdown can be done in mammalian cells (Paddison *et al.*, 2004). Recently, Zhang *et al.* (2006) used an RNAi library constructed from subtracted mESC cDNAs in an attempt to discover new players in maintaining the ESC state. More interestingly, (Ivanova *et al.*, 2006) used short hairpin RNA (shRNA)-mediated depletion to identify *Esrrb*, *Tbx3*, *Tcl1* and other transcripts as additional important factors for ESC self-renewal.

Although loss-of-function RNAi genetic screens are useful for adding new players to the stem-cell regulatory network, it cannot distinguish between direct and indirect regulators of gene activity. To fully understand the network architecture of ESCs, we need to know much more about how these players interact with DNA. In this aspect, chromatin immunoprecipitation (ChIP) assays can be used to map out the targets of transcription regulators to better define the ESC interactome.

1.7.2.3 *In vivo* analysis of transcription factor-DNA interactions

An important strategy in understanding the transcriptome is to uncover its regulatory control that is directed by transcription factors (TFs). *In silico*, the DNA-binding feature of transcription factors were used in computational methods to predict the location of promoters and find targets of

transcription factors based on consensus sequences (Davuluri *et al.*, 2001). However, information from this approach will remain inadequate without experimentation. Nevertheless, multiple bioinformatics methods have aided the experimental analysis for reconstruction of gene networks.

1.7.2.3.1 Chromatin immunoprecipitation (ChIP)

A method currently employed to identify the endogenous or *in vivo* direct targets of a transcriptional regulator is chromatin immunoprecipitation (ChIP), which was modified from the method described in Solomon *et al.* (1988). ChIP is a powerful tool for analysing target sites of protein binding to DNA as well as patterns of histone modifications (Orlando, 2000; Roh *et al.*, 2004). ChIP involves the treatment of living cells with formaldehyde, a procedure that crosslinks DNA to proteins that are in physical contact with it, thereby preserving the endogenous DNA-protein interactions. Chromatin is then fragmented either by physical or enzymatic means to an average size of 500bp. DNA fragments associated with a particular protein are then isolated by immunoaffinity purification using a specific antibody against the protein. After crosslinking is reversed, the ChIP-enriched DNA can be analysed by quantitative polymerase chain reaction (qPCR) using specific primers targeting the loci of interest. ChIP-qPCR is not high-throughput and requires test targets or binding sites to be pre-determined or assumed. Therefore, more advanced technologies have been used to identify ChIP DNA (target sequences) on a genome-wide scale such as paired end ditag (PET) sequencing and DNA microarray (Figure 1.4).

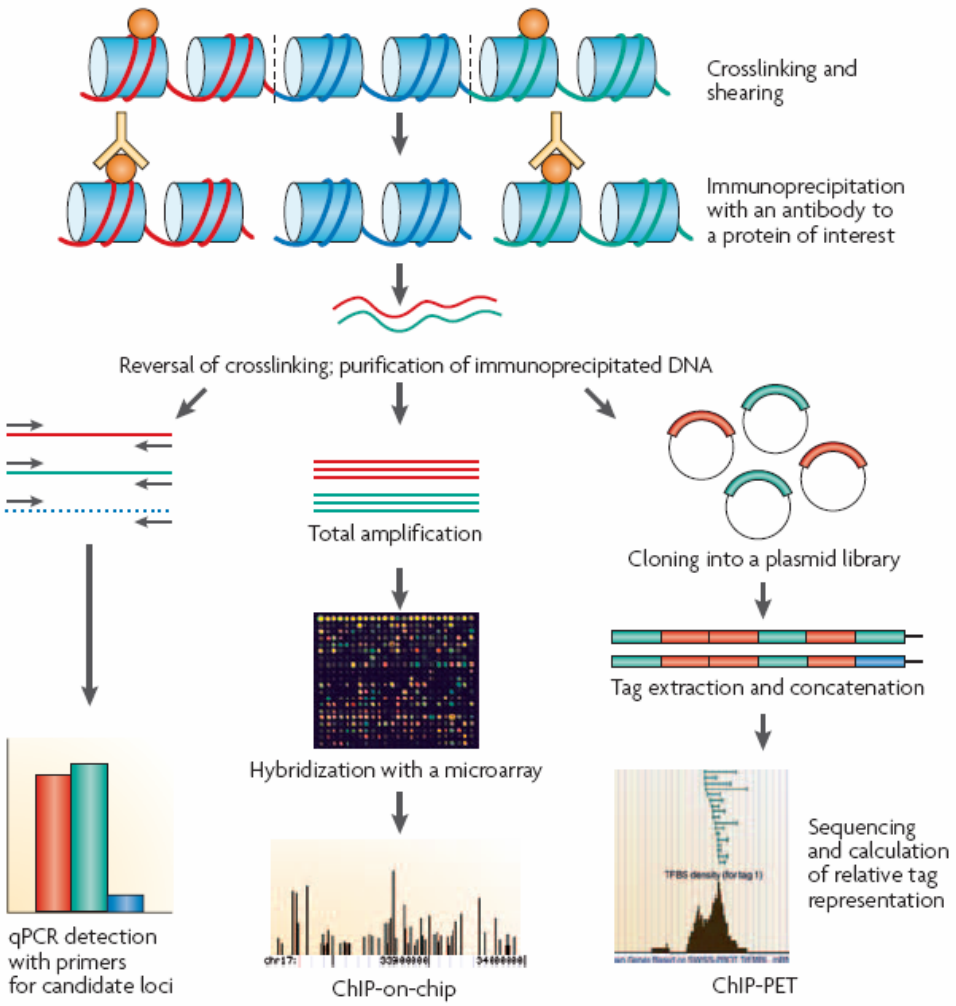


Figure 1.4: Diagram depicting the chromatin immunoprecipitation (ChIP) assay for the study of transcription factor DNA binding sites (TFBSs) in living cells. Formaldehyde is used to crosslink proteins to proteins and DNA. The crosslinked chromatin is then sheared by sonication or nuclease treatment and immunoprecipitated with specific antibodies against the factor of interest. DNA sequences that are bound by this factor are pulled out. After the crosslinking is reversed, the IP DNA can be analysed by quantitative polymerase chain reaction (qPCR). Two approaches which provide a bigger coverage are the ChIP-on-chip which takes advantage of genomic microarrays and ChIP-PET (paired-end tag) where short sequence tags of immunoprecipitated fragments are cloned into a plasmid library and then directly analysed by sequencing. Figure was obtained from Spivakov and Fisher (2007).

1.7.2.3.2 ChIP-Paired-end ditag (PET) technology

Chromatin immunoprecipitation - paired-end ditag (ChIP-PET) technology is a method that can be used in the genome-wide location analysis of protein-DNA interactions (Figure 1.5). It was developed by Dr Ruan Yijun's group in GIS. The ChIP-enriched chromatin fragments are blunt-ended before being cloned into a primary vector library that appends the cloned fragments with *Mme*I restriction sites. Using *Mme*I, 18 bp tags from each end of the cloned DNA fragments are retained. These tags are then re-ligated to generate paired-end ditags (PETs). These PETs are released from the primary vector library by restriction digestion before being concatenated and re-cloned into a PET sequencing library. The PETs are subsequently sequenced and the sequences obtained are computationally mapped onto the genome. ChIP-PET can unambiguously assign enriched genomic fragments to a specific region in the genome and also define the exact length of each enriched genomic fragment. Due to the random sampling of DNA fragments during PET cloning and sequencing, genomic fragments that are enriched by ChIP are more likely to be sequenced and mapped. Therefore, protein-DNA interaction sites are represented by genomic loci that contain multiple overlapping PETs, with overlap regions demarcating the protein binding site. A vast amount of sequencing is required to achieve sufficient representation of immunoprecipitated fragments (each one must be sequenced many times in order to calculate relative enrichment). Nevertheless, the ChIP-PET approach allows genome-wide, high-throughput and unbiased identification of transcription factor binding sites. Although currently the ChIP-PET is expensive and labour-intensive relative to ChIP-on-chip approaches, future advances in DNA sequencing could redress this balance. The ChIP-PET approach was first employed to map the binding sites of p53 in a colorectal cancer cell line (Wei *et al.*, 2006). It was also used in this study to map Oct4 and Nanog global binding sites in mESCs.

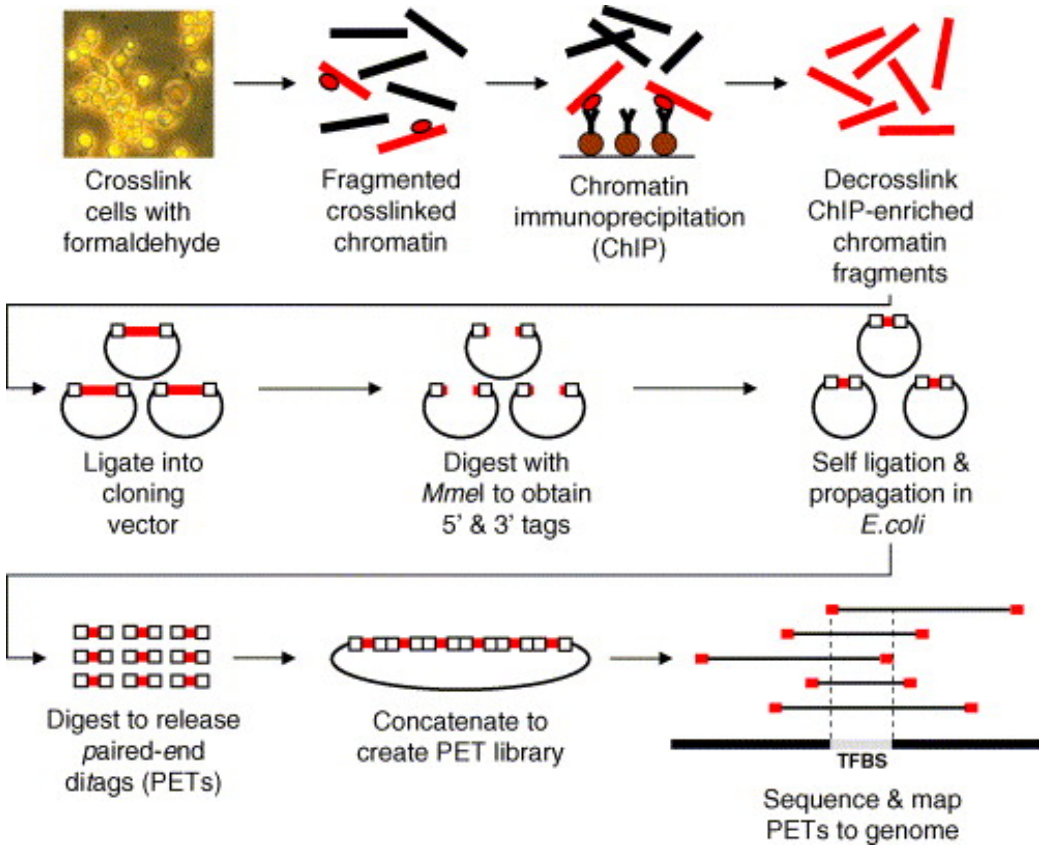


Figure 1.5: Schematic diagram of the chromatin immunoprecipitation-paired end ditag (ChIP-PET) approach. Cells are treated with formaldehyde to mediate covalent crosslinks that preserves endogenous interactions between DNA and proteins. The chromatin is fragmented followed by immunoprecipitation using beads coupled to a specific antibody to capture the transcription factor bound to target sites (shown in red). The ChIP-enriched chromatin fragments are then decrosslinked and subjected to proteinase digestion to liberate ChIP-enriched DNA. The DNA is then blut-ended and cloned into a primary plasmid-based library that affixes the fragments with *Mme*I restriction sites. Using *Mme*I, tags of 18 bp in length are generated from the 5'-most and 3'-most ends of the ChIP-enriched DNA fragments. These tags are then religated to generate paired end ditags (PETs). These PETs are released from the primary vector library by restriction digestion before being concatenated and recloned into a second library called the PET sequencing library. This second library increases the throughput of analysis, as each sequencing read identifies 10 to 15 PETs representative of 10 to 15 ChIP-enriched genomic fragments. Subsequently, the concatenated PETs are sequenced and computationally mapped to the mouse genome to demarcate genomic regions bound by the protein of interest (indicated by overlap regions within PET clusters). Figure was obtained from Lim and Ng (2007).

1.7.2.3.3 ChIP-on-chip

Another approach that allows for whole genome assessment of the DNA binding sites for transcription factors is the coupling of chromatin immunoprecipitation with oligonucleotide microarrays (ChIP-on-chip). These microarrays are similar to those used for transcriptional profiling, except that the oligonucleotides on the microarray are designed to probe genomic sequences instead of transcript sequences. ChIP-on-chip uses total amplification before hybridization; therefore a main concern for using this approach is the introduction of bias (Buck and Lieb, 2004; Negre *et al.*, 2006).

ChIP-on-chip identification of *bona fide* targets of transcriptional regulators in a genome-wide manner was reported in yeast (Iyer *et al.*, 2001; Ren *et al.*, 2000). In the mammalian context, early microarrays designed to analyse protein-DNA interactions only probed selected regions of the mammalian genome such as core promoter microarrays that probe genomic sequences flanking the transcription start sites of genes (Balciunaite *et al.*, 2005; Scacheri *et al.*, 2006), CpG microarrays that analyse genomic CpG islands (Weinmann *et al.*, 2002; Wells *et al.*, 2003), ENCODE microarrays (Takayama *et al.*, 2007; Wormald *et al.*, 2006), and tiling microarrays that probed the entire non-repeat regions of chromosomes 21 and 22 (Cawley *et al.*, 2004; Euskirchen *et al.*, 2004; Martone *et al.*, 2003). Rick Young's group used oligonucleotide-based human promoter arrays in ChIP-on-chip studies on transcription factors and Polycomb complexes in hESCs (Boyer *et al.*, 2005; Boyer *et al.*, 2006b; Lee *et al.*, 2006). Kim *et al.* (2005) used microarrays containing approximately 15 million 50-mer probes covering all non-repeat regions of human DNA at a 100bp resolution to generate a genome-wide map of active promoters in human fibroblasts by determining the binding sites of transcriptional pre-initiation complexes. Very recently, whole mouse and human genome microarrays for ChIP-on-chip were made available from Nimblegen and Affymetrix. Nimblegen chips come in survey sets of 10 (human) and 10 (mouse) covering the repeat masked regions of the human or mouse genome at an average probe

spacing of 100 bp, while the Affymetrix array sets cover the whole genome on 7 arrays, at 35 bp resolution.

1.8 Aim and experimental approach

This study focuses on Oct4 and Sox2, which are two key transcription factors that are essential in maintaining pluripotency of ESC. In order to gain a deeper understanding of the functions of Oct4 and Sox2 in governing the ESC fate, it is essential to know which genes are directly targeted by these factors *in vivo*. Prior to this study, only a few Oct4 and Sox2 targets have been identified.

Therefore, the main aim of this study is to map the *in vivo* genomic binding sites of Oct4 and Sox2 using mouse ESC as a model. The main methodology employed is chromatin immunoprecipitation (ChIP). Subsequently, information is extracted from these binding datasets to elucidate how these factors may work in ESCs. The outline of the study is as follows:

(i) Gene-by-gene ChIP-qPCR to optimize the ChIP technique and to establish the transcriptional network consisting of Oct4, Sox2 and Nanog

(ii) Genome-wide mapping of the Oct4 and Sox2 binding sites using the ChIP-PET sequencing technology

(iii) Identification of the Oct4 and Sox2 co-bound targets and characterization of the joint consensus motif by *in vitro* assays

(iv) Discovery of factors collaborating with Oct4 and Sox2 at *cis*-regulatory elements by computational search, ChIP-qPCR and ChIP-on-chip

(v) Investigation of the molecular mechanisms that provide insights into the juncture between different transcription factor pathways that confer mouse ESC characteristics

CHAPTER 2

MATERIALS AND METHODS

2.1 Chemicals and reagents

All chemicals and reagents used were of analytical grade unless otherwise stated. The phosphate-buffered saline (PBS) used in all assays was composed of 137 mM NaCl, 2.7 mM KCl, 4.3 mM KH_2PO_4 and 1.4 mM $K_2HPO_4.7H_2O$ at pH 7.3. Luria-Bertani (LB) medium and ampicillin were purchased from Sigma Aldrich (St. Louis, MO). All other chemicals were purchased from either the A*STAR Biopolis Shared Facility (BSF), Sigma (St. Louis, MO) or BDH Laboratory Supplies (Poole, England).

2.2 Antibodies

Sox2 (Y-17) sc-17320 goat polyclonal, Oct4 (N-19) sc-8628 goat polyclonal, Stat3 (C-20) sc-482 rabbit polyclonal, Stat3 (K-15) sc-483 rabbit polyclonal, Smad1 (A-4) sc-7965 mouse monoclonal, Ap-2α (3B5) sc-12726 mouse monoclonal, AP-2α (C-17) sc-6312 goat polyclonal, E2F1 (C-20) sc-193 rabbit polyclonal, p53 (Pab-240) sc-100 mouse monoclonal, c-Fos (6-2H-2F) sc-447 mouse monoclonal, JunB (C-11) sc-8051 mouse monoclonal, NF_kB p65 (c-20) sc-372 rabbit polyclonal, GST (Z-5) sc-459 rabbit polyclonal, GFP (FL) sc-8334 rabbit polyclonal, Ena1 (yC-20) sc-15542 goat polyclonal and MLL (c-20) sc-18214 goat polyclonal antibodies were purchased from Santa Cruz Biotech, CA. Sox2 AB5603 rabbit polyclonal antibody was purchased from Chemicon (now Millipore Corporation). Oct4 in-house antibody was obtained from colleague, Tay Hwee Goon. Smad1 rabbit monoclonal [EP435E] (ab33902), histone H3 (tri methyl K4) rabbit polyclonal (ab8580), histone H4 (tri methyl K20) rabbit polyclonal (ab9053) and RNA polymerase II CTD repeat YSPTSPS (phospho S5) (ab5131) antibodies were purchased from Abcam. Anti-E2F1, clones KH20 & KH95 (mixed mouse monoclonal antibody) (#05-379), anti-Stat3 rabbit polyclonal (#06-596) and anti-Smad4/DPC4 rabbit polyclonal (#06-693) antibodies were

purchased from Upstate. Phospho-Smad1/5 (Ser463/465)(41D10) rabbit monoclonal (#9516) and phospho-Stat3 (Tyr705) rabbit polyclonal (#9131) antibodies were purchased from Cell Signaling Technology Inc.

2.3 Recombinant DNA manipulations

General recombination DNA manipulations were performed as described by Sambrook and Russell (2001). Restriction enzyme digests and other DNA modifications such as ligation of DNA and de-phosphorylation of cloning vectors were performed using the appropriate buffers specified and supplied by the manufacturers (New England Biolabs, Inc., MA). Polymerase chain reactions (PCR) were carried out in a thermal cycler (Thermal cycler PTC-200 from MJ Research Inc. MA) using high fidelity *Pfu* DNA polymerase (Promega, Madison, WI) or Expand High Fidelity^{PLUS} System (Roche Diagnostics, Germany). DNA fragments from agarose gels were purified using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, FRG), according to the manufacturer's protocol. Rapid ligation kit from Roche or T4 ligase from New England Biolabs were used for DNA ligation. Plasmid DNA was purified using the QIAprep procedure (QIAGEN, FRG) that is based on alkaline lysis of bacterial cells and clearing the lysates under high salt conditions. The high salt concentration causes denatured proteins, cellular debris and chromosomal DNA to precipitate, while the plasmid DNA stays in solution. The latter is isolated by adsorption of DNA onto silica-based columns and subsequent elution in a low salt buffer. Competent cells used were SoloPack Gold Supercompetent cells from Stratagene unless otherwise stated. DNA sequencing was carried out by the GIS Sequencing Facility with the capillary Applied Biosystems sequencers ABI 3730xl and Applied Biosystems Big Dye Terminator version 3.1 Sequencing kit. Analysis of DNA was carried out by agarose gel electrophoreses using 1-1.5% (w/v) agarose (Seakem #50004) in 1x TAE buffer (40 mM Tris-acetate, 10 mM EDTA) containing 1 μg/ml ethidium bromide (EtBr) at 100V, 30 min unless otherwise stated. The DNA samples were mixed with DNA loading buffer [0.1% Orange G, 30% Ficoll (type 400)] and loaded onto agarose gel alongside 100bp and 1kb

DNA ladders from New England Biolabs, Inc., MA. The detection of the DNA bands was carried out using the UV trans-illuminator. Analyses of DNA sequences and design of cloning and sequencing primers were carried out using the Vector NTI® version 2.0 software (VNTI Suite 8.0, Invitrogen Corp., CA).

2.4 SDS-PAGE, Western blots and immunodetection

Protein concentrations were measured using the Protein Assay Dye Reagent Concentrate (Coomassie® Brilliant Blue G-250 dye) from Bio-Rad which is based on the Bradford method. Bovine serum albumin standards (2 mg/ml) were purchased from Pierce, IL. Proteins were typically analysed by means of a one-dimensional 7.5% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) using the Bio-Rad Mini-Protean II system, according to standard protocols (Sambrook and Russell, 2001). To prepare the samples for SDS-PAGE, equal volumes of protein samples and 2x Laemmli buffer [125 mM Tris-HCl (pH 6.8), 4% (v/v) SDS, 10% (v/v) β-mercaptoethanol, 20% glycerol and 0.4% (w/v) bromophenol blue] were mixed thoroughly and boiled at 95°C for 5 min before loading into the wells. The components of the electrophoresis buffer were 25 mM Tris-HCl (pH 8.3), 192 mM glycine and 0.1% SDS.

Chromatin extracts or whole cell lysates were resolved by SDS-PAGE and electrophoretically transferred onto Immuno-Blot™ PDVF membranes (pre-activated with methanol) (Bio-Rad) using Western transfer apparatus (Bio-Rad Trans-Blot systems) in a transfer buffer containing 25 mM Tris-base (pH 8.3), 192 mM glycine, 10% SDS and 20% methanol for 1 hr at 4°C. Subsequently, the membranes were incubated in blocking buffer [PBS containing 1% (w/v) BSA and 0.1% (v/v) polyoxyethylene-sorbitan monolaurate (Tween 20)] for at least 1 hr at room temperature. The membranes were then incubated for 1 hr at room temperature with primary antibody (0.1-1 g/ml). Unbound antibodies were removed by extensive washing in wash buffer (PBS containing 0.1% Tween-20). Membranes were then incubated with 0.5 μg/ml horseradish peroxidase-conjugated (HRP) secondary antibody for at least 1 hr at room temperature and then

washed in wash buffer 3 times. Membranes were then developed using an enhanced chemiluminescence (ECL) kit (GE Healthcare) for 1 min. HRP catalysed-oxidative degradation of luminal occurs, resulting in light emission at a wavelength of 428 nm, which was detected on an autoradiography film (GE Healthcare). Bands were also visualised using 0.1% Ponceau S (Sigma) in 5% acetic acid. Stripping of blots was carried out in a stripping buffer (2% SDS, 100 mM 2-mercaptoethanol, 62.5 mM Tris pH6.7) at 50°C, 2 hr and blocked again before re-probing with another antibody.

2.5 Cell Culture

2.5.1 Feeder-free mESC culture

All mESCs used in this project were E14 mESCs grown in feeder-free conditions. The HPRT-deficient E14 mESC line were cultured without feeders on 0.1% gelatin coated plates and maintained in Dulbecco's modified Eagle's medium (DMEM; GIBCO), supplemented with 15% heat-inactivated ES fetal bovine serum (FBS; Invitrogen), 55 mM 2-mercaptoethanol (GIBCO), 200 mM L-glutamine (GIBCO), 10 mM minimal essential medium with nonessential amino acids (GIBCO) and 1,000 U/ml of ESGRO® leukemia inhibitory factor (LIF) (Chemicon International CA, #ESG1107). Cells were routinely passaged at a 1:5 ratio after subjecting the cells to 0.25% trypsin (GIBCO) treatment at 37°C for 2 min. Cells were cultured in 500 cm^2 dishes (Falcon) for ChIP assays, 6-well plates for protein lysate and RNA extraction, and 96-well plates for reporter assays.

2.5.2 Differentiation of mESCs

Differentiation of mESCs was induced by addition of retinoic acid, withdrawal of LIF or by allowing the cells to spontaneously differentiate into embryoid bodies. In retinoic acid differentiation assays, E14 mESCs were treated with 0.5 μM of retinoic acid (RA) (Sigma #R2625) in mESC media without LIF for 3, 6 and 9 days. In formation of embryoid bodies, 15 x 10^6 E14 mESCs were seeded onto a 15 cm bacterial petri dish using ES media without LIF and

incubated in a shaker at 37°C/ 5% CO_2 for up to 10 days. In LIF withdrawal assays, mESCs were cultured in mESC media without LIF for up to 2 days.

2.5.3 Defined serum-free mESC culture

ESGRO Complete™ Clonal grade medium (#SF001-500) and ESGRO Complete™ Basal medium (#SF002-500), Accutase™ (#SCR005) were purchased from Chemicon International. ESGRO Complete™ Clonal Grade Medium is a defined serum-free medium containing BMP4 and LIF which has been optimized to grow and maintain undifferentiated mESCs in the absence of serum. Adaptation of feeder-free mESCs to serum-free cell culture conditions was carried out over at least 3 passages. E14 mESCs were grown to 60% confluence in serum-supplemented ESC medium in a 10 cm dish in the absence of feeder cells. Cells were washed once with PBS. To dissociate cells, 1 ml Accutase was added to mESCs and incubated at 37°C to allow cells to detach (5-10 min). 5 ml of Basal Medium was added, mixed and cells were spun at 1000 rpm. Pellet was resuspended in 5 ml Clonal Grade Medium and $1x10^6$ cells were seeded onto a pre-gelatin coated 10 cm dish containing 10 ml pre-warmed Clonal Grade Medium. When mESCs were about 60% confluent, a 1 in 5 split of the 10 cm dish culture was done into another coated 10 cm dish containing Clonal Grade Medium.

2.5.3.1 Low density plating assay

Early passage E14 mESCs that had adapted to serum-free culture conditions were seeded into a 0.1% gelatin-coated 10 cm plastic tissue culture dish at $1x10^3$ cells/ dish in 10-20 ml of Clonal Grade Medium. This medium was changed every 3 days. Colony formation was observed after culturing the cells for 5 days. The majority of the mESC colonies showed no signs of differentiation.

2.5.4 LIF and BMP treatment of serum-free, feeder-free mESC

Human recombinant BMP4 was purchased from Sigma, MI (#B2680) and R&D Systems (#314-BP) and ESGRO® LIF (#ESG1107) was purchased from Chemicon International. Feeder-free mESCs that have adapted to serum-free conditions and grown in ESGRO Complete™ Clonal

grade medium were washed with PBS twice and added with the following different media: (1) ESGRO Complete™ Clonal grade medium, (2) ESGRO Complete™ Basal medium, (3) ESGRO Complete™ Basal medium with 1000 U/ml ESGRO® LIF (#ESG1107), (4) ESGRO Complete™ Basal medium with 30 or 50 ng/ml BMP4 and (5) ESGRO Complete™ Basal medium with 1000 U/ml LIF and 30 ng/ml BMP4. Cells were grown for specific time-points and treated in 500 cm^2 culture dishes (Falcon) for ChIP assays, and 6-well plates (Falcon) for protein lysate and RNA extractions.

2.5.5 Human ESC culture

Human ESC line HUES-6 was obtained from Doug Melton (Harvard University) and cultures were obtained from Yeap Leng Siew from Assoc. Prof. Lim Bing's group, GIS. HES-3 (46X,X) (ES Cell International) cell cultures were obtained from Andre Choo (Biotechnology Institute, BTI). The cells were cultured feeder-free on matrigel with conditioned-medium from primary mouse embryonic fibroblasts (MEFs) in medium supplemented with basic fibroblast growth factor, recombinant human LIF, serum replacement, and a human plasma protein fraction (plasmanate) at 37°C/ 5% CO_2 according to methods described in Cowan *et al.* (2004) and Xu *et al.* (2001).

2.5.6 HEK293T cell culture

Human kidney HEK-293T cells (HEK 293T/17, CRL-11268) were purchased from American Type Culture Collection (ATCC, MD) and cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO) supplemented with 10% heat-inactivated fetal bovine serum (FBS; GIBCO) and maintained at 37°C in the presence of 5% CO_2. The cells were routinely passaged in a 1:5 ratio using 0.25% trypsin.

2.6 Cell Images

Cultured cells were routinely examined on the Nikon Eclipse TS100 inverted microscope (objective lens magnifications: 4X, 10X, 20X, 40X; eyepiece lens magnification: 10X) and images

were acquired with the Nikon Coolpix 4500 digital camera attached to the microscope. Images were also taken using the Leica DM-IRB fluorescent microscope (objective lens magnifications: 5X, 10X, 20X, 40X, 63X, 100X, eyepiece lens magnification: 10X).

2.7 Transfection of mammalian cells

HEK293T or mESCs were seeded at ~1 x 10^6 cells into 10 cm culture plates for 14 hr before they were subjected to transfection. Adherent cells at ~70% confluency were rinsed with PBS and transfected with plasmid DNA by liposome mediated transfection with Lipofectamine™ 2000 reagent (2.5 µl per 1 µg DNA) (Gibco-BRL) according to the manufacturer's instructions. DNA and lipofectamine were each incubated with 1.5 ml OptiMEM (GIBCO, Invitrogen Corp) in separate tubes for 5 min at room temperature. Subsequently, the solutions were mixed together and incubated for another 20 min before adding to the cells. After 4-6 hr, the transfection reagent was aspirated and the cells were incubated for another 24-36 hr in complete growth medium before they were harvested.

2.8 Preparation of nuclei extracts from mESCs

Feeder-free E14 mESC nuclear extracts were prepared according to methods described in Dignam *et al.* (1983) with modifications as follows: mESCs were harvested in cold PBS. After centrifugation, cells were resuspended in high salt buffer (20 mM HEPES-KOH pH 7.9, 26% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF) containing protease inhibitor and incubated on ice for 10 min before being subjected to lysis with a Wheaton dounce homogenizer. Pelleted nuclei were resuspended in low salt buffer (10 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT and 0.2 mM PMSF) with protease inhibitor. They were incubated on ice for 30 min and vortexed every 10 min. After centrifugation, cell supernatant were dialyzed against dialysis buffer (20 mM HEPES-KOH pH 7.9, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF) at 4°C for 2 hr.

2.9 Preparation of whole cell lysates

Cultured HEK293T or mESCs were rinsed twice with cold PBS before being lysed in cold HEPES lysis buffer [20 mM HEPES (pH 7.5); 140 mM NaCl; 10% Glycerol; 1% Triton X-100; 1.5 mM $MgCl_2$; 1 mM EGTA (pH 8.0) with EDTA-free Complete protease inhibitors (Roche, Mannheim, FRG)] and the cells were collected by manual scraping with a cell scraper (Costar, Corning, NY). Whole cell lysates were clarified by centrifugation at 14000 rpm for 15 min at 4°C and the cell pellet was discarded.

2.10 RNA extraction and Reverse Transcription (RT)-PCR

RNA extraction was carried out using Trizol (Invitrogen, CA) according to the manufacturer's protocol. In brief, cells were washed with PBS, and were lysed with Trizol for 5 min, while shaking at room temperature. Chloroform extraction was carried out, followed by isopropanol precipitation of the RNA. The pellet was resuspended in DEPC water and treated with DNase for 30 min at 37°C. The reaction was then purified using the RNeasy Mini Kit (Qiagen, Hilden, FRG). 20µl RT reaction containing of 1 µg RNA, 0.5 µg $Oligo(dt)_{12\text{-}18}$ primer, 0.5 mM dNTP mix, 5x first strand buffer, 10 mM DTT, 200 U SuperScript™ II Reverse transcriptase was set up. Incubation was done at 65°C 10 min, 42°C 1 hr, and heat inactivation at 70°C 15 min. All reagents were from Invitrogen, CA.

Real-time quantitative PCR was utilised to analyse mRNA expression in the reverse-transcribed cDNA samples as described in 2.14. RT-PCR primers were designed based on the 3'end of unigene mRNA sequences from NCBI (http://www.ncbi.nlm.nih.gov/) using OLIGO 6 with the same parameters described in 2.13.

2.11 Chromatin Immunoprecipitation (ChIP)

2.11.1 Crosslinking of cells and chromatin extract preparation

In ChIP assays, cells were grown to 70% confluency in 500 cm^2 dishes and were crosslinked with 1% formaldehyde in culture media for 10 min at room temperature. Formaldehyde was then inactivated by adding glycine into the medium to a final concentration of 0.2 M. Cells were washed with TBSE (20 mM Tris pH7, 0.15 M NaCl, 1 mM EDTA pH8) for 3 times and were scraped into a 50 ml tube. The cells were then washed twice with cell lysis buffer (10 mM Tris pH8, 0.25% Triton X-100, 10 mM EDTA pH8, 0.1 M NaCl) for 15 min, nutation at 4°C and centrifugation was done at 3000 rpm, 4°C, 5 min. The cell pellet was then transferred into a 50 ml Oakridge tube and washed once with high SDS FA lysis buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% NaDOC, 1% SDS) and twice with low SDS FA lysis buffer (50 mM HEPES-KOH pH7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% NaDOC, 0.1% SDS) and subjected to 30 min of nutation at 4°C and centrifugation at 20000 rpm, 4°C for 1hr. Complete, EDTA-free protease inhibitor tablets (Roche #1873580) were added into all wash buffers. The chromatin pellet was then sonicated in low SDS FA lysis buffer (1ml buffer for every 100 µl pellet) together with 1 ml of 0.5 mm glass beads (Biospec Products Inc. #11079105) using the Branson sonifier with tapered tip probe to obtain DNA fragments of an average size of 500 bp. Optimization was carried out for different types of cells. The typical sonication parameters used were: 40% amplitude, 15 pulses, 1 min rest in between pulses, 4°C. After sonication, the chromatin extracts were centrifuged at 20000 rpm, 4°C, 1hr to remove the unsonicated chromatin. In order to obtain input DNA, chromatin extracts were decrosslinked at 42°C for 2 hr, and 68°C for 6 hr, in a reaction containing 100 µl extracts, 200 µl ChIP elution buffer (50 mM Tris pH 7.5, 10 mM EDTA, 1% SDS), 90 µl TE buffer, 30 µl pronase (Roche, #11459643001). Phenol:chloroform extraction, ethanol precipitation with glycogen (Roche #901393) and RNAse digestion (DNAse-free Rnase, Roche #1119915) were carried out to purify

the DNA further. The input DNA was ran on 1.5% agarose gel, 50 V, 1hr, 1x TAE buffer to analyse the size distribution of the DNA fragments.

2.11.2 Immunoprecipitation (IP)

30 µl of protein G sepharose 4 fast flow (Amersham Biosciences, #17-0618-02) or 50 µl of Dynabeads® protein G (Dynal Biotech, Norway #100.03) beads were washed 3 times with low SDS FA lysis buffer and incubated with 6 µg antibodies of interest for 3 hr, nutating at room temperature. Specificity of all the antibodies was checked by Western blot before they were used for ChIP. After antibody binding, the beads were washed twice using low SDS FA lysis buffer. Subsequently, 500 ml chromatin extract which was pre-cleared (3 hr, nutating at 4°C in 100 µl protein G sepharose or 50 µl Dynal beads slurry) was added to the antibody-bound beads and nutated at 4°C, overnight. The beads were then washed 3 times with low SDS FA lysis buffer, once with low SDS FA lysis buffer containing 0.35 M NaCl, once with NP-40/LiCl wash buffer (10 mM Tris pH 8, 0.25 M LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% NaDOC), once with TE buffer (10mM Tris, 1mM EDTA, pH8) and eluted with 270µl ChIP elution buffer at 68°C, 30 min, 1400 rpm. 260 µl supernatant was added to a reaction containing 220 µl TE buffer and 20 µl pronase and incubated at 42°C for 2hr, and 68°C for 6hr. Subsequently, phenol:chloroform extraction and ethanol precipitation were carried out and the pellet was resuspended in 100 µl TE buffer to obtain the IP DNA.

2.12 Picogreen DNA quantitation

The ChIP DNA was routinely quantified using the Quant-iT™ PicoGreen® dsDNA quantitation kit (Molecular Probes, OR) following the manufacturer's instructions. Fluorescence measurement was carried out in the TECAN GENios multidetection microplate reader (Tecan Trading AG, Switzerland) using the fluorescein wavelengths: excitation= 485 nm, emission= 535 nm, gain: 60, number of flashes: 10, measurement mode: Fluorescent Top.

2.13 Q-PCR primer designs

Genomic DNA sequences with masked repeats were obtained from the UCSC genome browser (http://genome.ucsc.edu) for ChIP real-time quantitative PCR primer designs. The primers were designed using OLIGO 6 (version 6.6.4.0; Molecular Biology Insights, Inc http://www.oligo.net). The designs were based on 'very high' search stringency parameters with the following modifications: oligonucleotide length of 29-30 nucleotides, product length of 150-250bp, oligo T_m> 72°C. All primers were screened by q-PCR using mouse genomic DNA and were confirmed to produce a single product of the right size, as analysed by agarose gel electrophoresis and dissociation curve analysis. These primers also yielded no DNA bands in the no-template control. At least two sets of primers were selected for each region of interest.

2.14 Real-time quantitative PCR (q-PCR)

Q-PCR were carried out in a 10 µl or 20 µl reaction volume containing 0.45 µM primer pairs, at least 0.5 ng ChIP DNA or cDNA and 2X SYBR® Green Master Mix (Applied Biosystems, CA). 96 and 384-well clear plates and covers were purchased from Applied Biosystems, CA and qPCR were performed on the ABI ABI PRISM 7900HT Sequence Detection System or the 7500 Real-time PCR System using the SDS2.2 software. The parameters used were absolute quantitation, stage 1: 95°C, 10 min; stage 2: 40 cycles of 95°C, 30s, 60°C 30s, 72°C 1 min (data collection); melting curve stage 3: 95°C, 15s, 60°C, 15s, 2% ramp to 95°C, 15s. Relative occupancy values (ChIP fold enrichment) were calculated by determining the apparent immunoprecipitation efficiency (ratios of the amount of immunoprecipitated DNA to that of the input sample) and normalized to the level observed at a control region, which was set at 1.0. Relative expression levels (RNA transcript) were calculated by taking the ratios of the amount of DNA against that of the control sample and normalized to the endogenous β-actin control levels.

2.15 ChIP-PET (paired-end ditag) cloning and sequencing

The ChIP-PET analysis was done in collaboration with Dr Wei Chia Lin, Cloning and Sequencing Group, GIS, as described in Ng *et al.* (2005) and Wei *et al.* (2006). Briefly, the ChIP DNA fragments were blunted and ligated to the cloning vector pGIS3 containing two *Mme*I recognition sites. The ChIP DNA library was obtained by transformation of ligated products into electro-competent TOP10 bacterial cells. Purified plasmid prepared from the ChIP DNA library was digested with *Mme*I, end-polished with T4 DNA polymerase to remove the 3'-dinucleotide overhangs, and the resulting plasmids containing a signature tag from each terminal of the original ChIP DNA insert were self-ligated to form single-ditag plasmids. These were then transformed into TOP10 cells to form a "single-ditag library". Plasmid DNA extracted from this library was digested with *BamH*I to release the 50 bp paired end ditags. The PETs were PAGE-purified, then concatenated and separated on 4-20% gradient TBE-PAGE. Appropriate sized fractions (1kb - 2 kb) of the concatenated DNA were excised, extracted and cloned into *BamH*I-cut pZErO-1 (Invitrogen) to form the final ChIP-PET library for sequencing. This library was then subjected to large-scale sequencing. PET sequences containing 18 bp from 5' and 18 bp from 3' ends of the original ChIP DNA fragments were extracted from raw sequences obtained from the PET library, and mapped to the mouse genome (mm5 build) assembly. The mapping criteria are that both the 5' and 3' signatures must be present on the same chromosome, on the same strand, in the correct orientation (5' to 3'), with a minimal 17 bp match and within 4 kb of genomic distance. The locations of the ChIP-enriched DNA present in the library were visualized using the T2G browser (http://t2g.bii.a-star.edu.sg) developed by the Bioinformatics Institute (Singapore), which contains the location of the PETs based on the UCSC genome browser (mouse genomic sequences, build mm5).

2.15.1 Manual and computer-assisted *de novo* motif search

Sequence alignments of selected binding sites were carried out manually to identify conserved motifs. Genomic sequences from the product of peak primers generating the highest enrichment or

from PET clusters for selected loci were aligned using the multiple sequence alignment tool from VectorNTI. Consensus sequences resembling the Sox2 and Oct4 motifs were eye-balled from the alignment.

Computer-assisted search of enriched motifs within Oct4 and Sox2 ChIP-PET data was carried out by our Bioinformatics colleague, Vinsensius Vega. The masked sequences from clusters containing a minimum of 7 overlaps were fed into the motif discovery algorithm Weeder (Pavesi *et al.*, 2004), setting MM (*Mus musculus*, build mm5) as the background genome, searching both strands, allowing multiple motif occurrences in each sequence, and running the most thorough search (i.e. analysis type = "extra"). As the Weeder algorithm only allows for a maximal motif length of 12 bp, the motif was extended by two bp upstream and one nucleotide downstream to obtain one that resembles a full Sox2-Oct4 site. The good quality sites (i.e. >90% similarity to the main discovered motif) was extracted, as determined in the Weeder output, while at the same time extending several bp out from each arm of the main discovered motif. From these sequences, a Position Weight Matrix (PWM) M was built to model the joint Sox2-Oct4 binding sites. This PWM was used to predict the label of a sequence. If a given sequence contained a site scored at least T under PWM M, the sequence was labeled as positive (i.e. contained Sox2-Oct4 binding sites), otherwise it was labeled as negative. Next, we further refined the Sox2-Oct4 motif using a refinement strategy akin to the Expectation-Maximization optimization procedure (Dempster *et al.*, 1977). A collection of 1000 random promoter sequences and 1000 random coding sequences with an average length of about 1800 bp was used as a background set (sequence set B1). A positive set (sequence set P1) containing the clusters with 6 or more PET sequences overlaps (including 10bp flanking sequences) was constructed. The following steps were done iteratively: 1) Use the background set B1 to calculate the appropriate PWM scoring cutoff T such that the false discovery rate is at most $1e^{-3}$, 2) Scan the positive set P1 for occurrences of the Sox2-Oct4 motifs using the current PWM M and the cutoff score T, 3) Calculate the statistical significance (p-value) p of motif over-representation in the set P1 against B1, in terms of predicted

sites per nucleotide, using the current PWM M and cutoff score T, 4) Construct a new PWM M' using the discovered sites, 5) Calculate threshold T' for M' using background set B1 such that the false discovery rate is at most $1e^{-3}$ (similar to step 1), 6) Compute the p-value, p', of set P1 being enriched, in terms of predicted sites per nucleotides, for sites scoring better than T' under the matrix M', 7) If p' is smaller than p then use M' as M for the next iteration and go back to step 1, or else output M as the final matrix. Two other sets were created to test whether the refined matrix identified over-represented sites in the PET sequences. PET5 clusters sequences were used as the positive set P2 (which did not intersect with P1), while for the background/negative set B2, another 1000 random promoters and 1000 random coding sequences (non-overlapping to sequence set B1) were extracted. Varying the cutoff score T produced different sensitivity and specificity. The performance of the seed matrix and the refined matrix were compared by stratifying the threshold T and plotting the Receiver Operating Characteristic (ROC) curves. The refined matrix well outperformed the original seed matrix.

2.15.2 Computational co-motif enrichment analysis

To identify potential co-factors of Sox2-Oct4, a search for transcription factor DNA binding motifs was carried out on the Oct4 and Sox2 overlapping binding sites from a merged Oct4-Sox2 ChIP-PET dataset by our Bioinformatics colleagues, Vinsensius Vega and Bernard Leong. A list of regions which are highly probable to be bound by both Sox2 and Oct4 was generated. Sox2 ChIP-PET and Oct4 ChIP-PET overlap regions, each from the moPET2+ clusters, were merged. 1507 unique locations supported by presence of both Sox2 and Oct4 ChIP-PET cluster overlapping regions were identified.

500 bp regions centred on the 1507 high-confidence Sox2-Oct4 binding regions were extracted and scanned for other putative transcription factor binding motifs, based on the weight matrices provided in the TRANSFAC Professional v9.1 and its associated MATCH program and cutoffs criteria. TRANSFAC (Wingender *et al.*, 2000) is a database on transcription factors, their corresponding genomic binding sites and DNA-binding profiles. It contains more than 18000

transcription factor binding sites matrices. To assess the significance of each potential co-motif, a 1000-iteration Monte Carlo simulation (Zhou and Liu, 2004) to estimate the expected occurrences of each motif was done. In each iteration, a set of random background sequences of equal length as the input sequences were generated based on 3rd order Markov Chain model of mouse genome mm5. Following which, the random sequences were similarly scanned for putative binding motifs. To simplify the explanation, the observed count of putative binding sites from the positive dataset was obtained, after which, using randomly generated sequences from the specified background model, the expected rate of putative binding sites to occur per random bp was estimated. Then, using the binomial formula, the *p*-value was computed, where the probability of success p is the expected rate of putative binding sites occurrence per random bp, the number of trials N is the total bp in the positive dataset, and the number of success Y is the count of putative binding sites found in the positive dataset. More information on the binomial formula can be obtained from http://core.ecu.edu/psyc/wuenschk/docs30/Binomial.doc.

2.16 Sequential chromatin immunoprecipitation (seqChIP)

In this assay, chromatin extracts were subjected to ChIP twice (Geisberg and Struhl, 2004). 30 µl of protein G sepharose 4 fast flow beads (Amersham Biosciences) were incubated with the first antibody (10 µg) for 3 hr, nutating at room temperature. Subsequently, the beads were washed twice with low SDS FA lysis buffer, twice with Immunopure binding/wash buffer (Pierce) and crosslinking buffer (0.2 M triethanolamine pH 8.2). The antibodies were crosslinked to protein G sepharose beads using 20 mM dimethyl pimelimidate (DMP; PIERCE Biotechnology, #21666) in crosslinking buffer for 45 min, in the dark. The beads were then washed with blocking buffer (0.1 M ethanolamine pH 8.2), Immunopure elution buffer (PIERCE Biotechnology), twice with Immunopure binding/wash buffer (Pierce) and twice with low SDS FA lysis buffer. Subsequently,

500 ml chromatin extract which was pre-cleared (3 hr, nutating at 4°C in 100 µl protein G sepharose) was added to the antibody-bound beads and nutated at 4°C, overnight.

The next day, beads were washed as in ChIP protocol, with Tris component in all buffers replaced by 10 mM Hepes pH 7.9. The beads were incubated with 130 µl elution buffer (10 mM Hepes pH 7.5, 1 mM EDTA, 1% SDS) at 37°C, 1400 rpm for 45 min. 120 µl which was recovered after centrifugation and diluted into 1.1 ml FA lysis buffer (no SDS) was subjected to a second ChIP using another antibody. The steps for the second round of IP are the same as that described for the one step ChIP above. Washed beads were eluted with ChIP elution buffer at 68°C for 30 min. 260 µl supernatant was added to a reaction containing 220 µl TE buffer and 20 µl pronase and incubated at 42°C for 2 hr, and 68°C for 6 hr. Phenol:chloroform extraction and ethanol precipitation were carried out and the pellet was resuspended in 100 µl TE buffer to obtain IP DNA for qPCR. Significant enrichment reflected by qPCR after the second ChIP was indicative of co-occupancy.

2.17 ChIP on NimbleGen DNA Microarray

Sequential ChIP for Oct4-Sox2 as well as ChIPs for H3K4Me3, Stat3 and Smad1 were amplified and hybridized onto customized gene microarray chips, 2005-03-23_Hui_MM5_ES_Chip (designed from mouse genomic sequences build mm5 by Dr Ng Huck Hui and manufactured by NimbleGen Systems Inc.) following the manufacturer's ChIP-on-chip protocol. The chips (25 x 75 mm each) contain 50-mer probes (50 bp apart) spanning the loci of 200 mouse genes, miRNAs and other genes, 385000 features (16 µm x 16 µm) in an array size of 17.4 mm x 13 mm.

2.17.1 Ligation-mediated PCR

ChIP DNA was amplified following Nimblegen's ChIP-on-chip protocol. Briefly, 30 ng of ChIP DNA fragments were blunted using 0.6 U T4 DNA polymerase (New England Biolabs, Inc., MA) at 12°C, 20 min. The DNA was then purified by phenol:chloroform and ethanol precipitation.

Linkers were ligated to the ends of the DNA fragments in a reaction containing pre-annealed linkers (oJW102: 5'-GCGGTGACCCGGGAGATCTGAATTC-3', oJW103: 5'-GAATTCAGATC-3' (Proligo); the pair of oligos were boiled for 10 min in a beaker containing 3L of water and left overnight at room temperature to cool) and T4DNA ligase (New England Biolabs, Inc., MA) for 16 hr at 16°C. Ethanol precipitation was carried out, followed by PCR amplification in a 50 μl reaction containing *Taq* polymerase (Qiagen), *Pfu* Hotstar (Stratagene), dNTP, oligo oJW102 (Proligo). PCR program: 55°C 4 min, 72°C 5 min, 95°C 2 min; 35 cycles of 95°C 30s, 55°C 30s, 72°C 1 min; 72°C 1 min. The PCR product was then purified using the QIAquick PCR purification kit (Qiagen). After amplification, ChIP DNA and control ChIP DNA was run on 1.5% agarose gel to ensure similarity in size distribution and absence of degradation. DNA concentration was determined using the ND-1000 Spectrophotometer (NanoDrop Technologies, DE) and the samples were tested by q-PCR to ensure that the ChIP enrichment fold of DNA was maintained after the amplification process.

2.17.2 Labeling, hybridization and analyses

Samples were labeled using Klenow fragment (3'-> 5' exo$^-$) (New England Biolabs, Inc., MA) and the Cy3 (ChIP DNA) and Cy5 (control DNA) 9-mer Wobble primers (5'-Cy3/Cy5NNNNNNNNN-3') were purchased from Research Biolabs, following the Nimblegen protocol. The labeled samples were purified using isopropanol precipitation and an equal amount of labeled ChIP and control DNA were mixed and dried in a speedvac. Hybridization reaction containing Hybridization buffer (NimbleGen array hybridization kit), Hybridization component A (NimbleGen), Cy3 and Cy5 CPK 50-mers [IDT, 250 nmole HPLC purified; obtained from Lee Yew Kok (Prof. Edison Liu's group, GIS)] were added to the DNA and heated at 95°C 5 min before loading onto the array slide. Samples were hybridized at 42°C, 18 hr, mix mode: B, in the MAUI Mixer SL Low Temperature Hybridisation Chambers (BioMicro Systems #02-A008-03) and the MAUI (MicroArray User Interface) Hybridization System (BioMicro Systems, model #01-

A002-03). Slides were washed sequentially with the Nimblegen wash buffers I, II, III according to manufacturer's instructions and spun dry using the ArrayIt microarray highspeed centrifuge. Raw data were analyzed on GenePix analysis software version 4.0 on the GenePix 4000A scanner (Axon Instruments) at a 5 μm resolution. Thereafter, NimbleScan™ and SignalMap™ softwares from Nimblegen Systems Inc. were used for data analysis and visualization.

2.18 Dual-luciferase reporter assay

pGL3-mOct4 pp vector [an Oct4 minimal promoter driving luciferase; Oct4 minimal promoter upstream (*Bgl*II and *Nco*I sites) of the luciferase gene in the pGL3-basic vector (Promega)] was obtained from a colleague, Chen Xi (Wu *et al.*, 2006). Oct4 and Sox2 binding motif elements were ordered as oligo pairs from Proligo.

Oligo pairs were annealed at 95°C, 4 min, 70°C 10min, room temperature 1hr followed by addition of T_4 Polynucleotide Kinase (New England Biolabs, Inc., MA). The phosphorylated probes were then cloned upstream (*Mlu*I and *Bgl*II sites) into the pGL3-mOct4 pp vector. The plasmids were transformed into MAX Efficiency® Stbl2™ Competent Cells (Invitrogen) and isolated plasmids then transfected into E14 mESCs according to the Lipofectamine 2000 (Invitrogen) manual. In brief, 24000 mESCs were seeded per well in 96-well flat-bottom luminometer plates (Costar, Corning Inc., NY) 14 hr before transfection. 150 ng plasmid DNA was mixed with 5.5 ng of phRLSV40 (normalization control expressing *Renilla* luciferase, Promega), and incubated in 25 μl OptiMEM (GIBCO, Invitrogen Corp) for 5 min. 0.5 μl Lipofectamine 2000 was incubated with 25 μl OptiMEM for 5 min before adding to the plasmids mixture and incubated at room temperature for 20 min. The resulting solution was then added to ESCs. Transfections were done in triplicates. Luciferase activities were measured 36 hr after transfection with the Dual-Luciferase® Reporter Assay System (Promega) using Centro LB960 96-well luminometer (Berthold Technologies). Reporter activity was calculated as ratio of the average of experimental readings over the average of readings in vector-transfected controls.

Co-transfection of the plasmids together with 150 ng Oct4 and Sox2 RNAi plasmids, pSUPER.puro-Oct4 shRNA or pSUPER.puro-Sox2 shRNA were carried out as described above. Cells were treated with mESC media containing 1 μg/ml puromycin (Sigma) to select for cells carrying the RNAi plasmids. Luciferase activities were measured 3 days post-transfection. Probe sequences containing 3x elements are listed in Appendix E.

2.19 RNAi-mediated depletion of Oct 4 and Sox2 in mESCs

pSUPER.puro-Oct4 and pSUPER.puro-Sox2 which confers ampicillin and puromycin resistance plasmids for RNAi were obtained from a colleague, Loh Yuin Han. In the construction of RNAi plasmids, 19 bp gene-specific regions for RNA interference were designed based on the work of Reynolds *et al.* (2004) and Ui-Tei *et al.* (2004). Oligonucleotides were cloned into pSUPER.puro (Oligoengine) (*Bgl*II and *Hind*III sites), which contains the polymerase III H1-RNA gene promoter for directing the synthesis of 19-nucleotide hairpin-type short hairpin RNAs (shRNAs) with a 9-nucleotide loop. All sequences were analyzed by BLAST search to ensure that they did not have significant sequence similarity with other genes. For the *Oct4* RNAi target sequence, the Reynolds score and Ui-Tei class value were 6 and class Ib, respectively. For the *Sox2* RNAi target sequence, these were 5 and class Ia, respectively. The oligonucleotides used were as follows: for *Luciferase* control RNAi, 5'-GATCCCCGATGAAATGGGTAAGTACATTCAAGAGATGTACTTACCCATTTCATCTTTTTA and 5'-AGCTTAAAAAGATGAAATGGGTAAGTACATCTCTTGAATGTACTTACCCATTTCATCGGG-3' ; for *Oct4* RNAi, 5'-GATCCCCGAAGGATGTGGTTCGAGTATTCAAGAGATACTCGAACCACATCCTTCTTTTTA-3' and 5'-AGCTTAAAAAGAAGGATGTGGTTCGAGTATCTCTTGAATACTCGAACCACATCCTTCGGG-3'; for *Sox2* RNAi, 5'-

GATCCCCGAAGGAGCACCCGGATTATTTCAAGAGAATAATCCGGGTGCTCCTTCTTTT TA-3' and 5'-AGCTTAAAAAGAAGGAGCACCCGGATTATTCTCTTGAAATAATCCGGGTGCTCCTTCG GG-3'. For Oct4 and Sox2 RNAi, E14 mESCs at 50% confluency were transfected with RNAi and plasmids. Selection with 1 μg/ml puromycin (Sigma) was initiated 24 hr after transfection and continued for 48 hr. The specificity of Oct4 and Sox2 knockdown was reported in Chew *et al.* (2005). Cell morphology was observed and knockdown cells were harvested for protein, RNA and ChIP analyses.

2.20 Overexpression of Oct4 and Sox2 proteins in HEK293T cells

Plasmids pTri-Ex 1.1 which confers ampicillin resistance and contains the Oct4 or Sox2 open reading frame (ORF) obtained from colleague Zhang Wensheng and the expression vector pTri-Ex 1.1 were transfected into HEK293T cells using Lipofectamine 2000 (Invitrogen). HEK293T cells were seeded at ~1 x 10^6 cells onto 10 cm culture plates for 14 hr before transfection. These adherent cells at ~70% confluency were rinsed with PBS and transfected with 24 μg plasmid DNA containing Oct4 and Sox2 overexpression contructs or vector alone by liposome mediated transfection with Lipofectamine^{TM} 2000 reagent (Gibco-BRL). Plasmids were incubated in OptiMEM for 5 min. Lipofectamine 2000 was incubated with OptiMEM for 5 min before adding to the plasmids mixture. The mixture was incubated at room temperature for 20 min. The complex was then added to cells. At 4-6 hr post-treatment, the transfection reagent was aspirated and the cells were incubated for another 24-36 hr in complete growth medium before they were harvested. Whole cell lysates were prepared and run on Western blot to detect the Oct4 and Sox2 overexpressed proteins.

2.21 Electrophoretic mobility shift assay (EMSA)

1 μg double-stranded DNA oligonucleotides (Proligo) labeled with biotin at the 5' termini of the sense strands were annealed with reverse strands in an annealing buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM EDTA) and purified with an agarose gel DNA extraction kit (Roche). The gel shift assays were performed using a LightShift Chemiluminescent EMSA kit (Pierce). 10 μg of nuclear extract or overexpressed Oct4, Sox2 and vector control were added to a final 10 μl reaction mixture containing 3 ng of biotin-labeled oligonucleotide and 1 μg of poly(dG-dC) (Amersham Biosciences). The final binding buffer composition was 60% dialysis buffer. Binding reaction mixtures were incubated for 10 min at room temperature. Where specified, antibodies were added after the initial incubation for a further 20 min. For competitive studies, 200X (600 ng) unlabeled double-stranded competitor were added for a further 20 min after the initial incubation. For cooperativity binding studies, indicated amounts of overexpressed Oct4 or Sox2 were added for a further 20 min after the initial incubation. Binding reaction mixtures were resolved on pre-run 6% native polyacrylamide gels in 0.5X Tris-borate-EDTA (TBE). Gels were transferred to Biodyne B nylon membranes (Pierce Biotechnologies) using Western blot techniques and crosslinked using the Hoefer UVC 500 ultraviolet crosslinker (Amersham Biosciences) with energy setting: 120000 $\mu J/cm^2$, time setting: 10 s. The membranes were then subjected to chemiluminescence detection as described in the LightShift Chemiluminescent EMSA kit (Pierce). Probe sequences are listed in Chapter 6.

2.22 Co-Immunoprecipitation (Co-IP) of protein complexes

50 μl of protein G sepharose 4 fast flow beads (Amersham Biosciences, #17-0618-02) were washed twice with NP-40 lysis buffer (50 mM Tris pH8, 150 mM NaCl, 2 mM EDTA, 1% NP-40) with Complete, EDTA-free protease inhibitor (1 tablet/ 50 ml, Roche). 6 μg Oct4 antibodies (N-19, Santa Cruz) were incubated with the beads in 0.25% Triton X-100 in 1xPBS for 3 hr at room

temperature, nutating. 200 μg E14 mESC nuclear extract was precleared using 100 μl of protein G sepharose 4 fast flow beads (Amersham Biosciences, #17-0618-02) at 4°C, 3 hr, nutating. Antibody-bound beads were washed 3 times with NP-40 lysis buffer plus protease inhibitor and precleared nuclear extract was added to the antibody-bound beads and nutated overnight at 4°C. Beads were washed 5 times with NP-40 lysis buffer plus protease inhibitor before 100 μl 2x sample buffer was added. The beads were then heated at 95°C 15 min and the eluted samples were loaded onto SDS-PAGE gel for Western blot analysis to detect Smad1 and Stat3 in the Oct4 protein complex.

2.23 Error bars in figures

Error bars in figures represent either two technical replicates or two biological replicates, as stated. Technical replicates are duplicates included during qPCR experiments. Biological replicates are duplicates of the same experiment repeated twice with different cells in different experimental sittings.

2.24 Contribution of collaborators

All high-throughput bioinformatics analyses were done by collaborators from the Bioinformatics group in Genome Institute of Singapore. Their names have been added to the respective methods stated in this Chapter. Cloning and sequencing of ChIP DNA in the ChIP-PET approach was carried out by the Cloning and Sequencing group in Genome Institute of Singapore. Their names have been mentioned in this Chapter accordingly. All collaborators and contributors to reagents have also been acknowledged in the acknowledgement. Other than high-throughput bioinformatics analyses and cloning and sequencing of ChIP-PETs, other work in this thesis was carried out by the author.

CHAPTER 3

Establishing the Circuitry of Oct4, Sox2 and Nanog in Embryonic Stem Cells

3.1 Introduction

Oct4, Sox2 and Nanog are transcription factors essential for the formation of the ICM during mouse preimplantation development and for the pluripotent and self-renewing properties of ESCs (Avilion *et al.*, 2003; Chambers *et al.*, 2003; Mitsui *et al.*, 2003; Nichols *et al.*, 1998; Scholer *et al.*, 1990). These factors are highly expressed in both human and mouse ESCs, and its expression diminishes when these cells differentiate and lose pluripotency (Palmieri *et al.*, 1994).

Several targets of Oct4 were previously identified in mESCs, including *Fgf4*, *Utf1*, *Opn*, *Rex1*/*Zfp42*, *Fbx15*, and *Sox2* (Ben Shushan *et al.*, 1998; Botquin *et al.*, 1998; Catena *et al.*, 2004; Dailey *et al.*, 1994; Nishimoto *et al.*, 1999; Tokuzawa *et al.*, 2003; Tomioka *et al.*, 2002; Yuan *et al.*, 1995). The regulatory regions of these genes contain an octamer element capable of binding Oct4 *in vitro*. These sites have been shown to be important for transcriptional activity of their respective genes in reporter assays. The octamer elements within the enhancers of *Fgf4*, *Utf1*, *Opn*, *Fbx15*, and *Sox2* were also found in proximity to sox elements. Of these genes, all but *Opn* have the octamer and sox heptamer elements separated by either 0 or 3 bp. Such proximity suggested that Oct4 and Sox2 may interact with each other on genomic DNA. Moreover, two structures have been solved for a POU/HMG ternary complex bound to composite sox-oct elements where one of these is on an element separated by 3 bp (Remenyi *et al.*, 2003) and the other is on an element separated by 0 bp (Williams, Jr. *et al.*, 2004). Both reveal that the POU and HMG domains are involved in mediating specific protein-protein and DNA-protein interactions. In addition, it has also been demonstrated that Oct4 and Sox2 can interact in the absence of DNA and that the HMG and POU domains are involved in this interaction (Ambrosetti *et al.*, 2000). Therefore, Oct4 and Sox2 are capable of forming heterodimers regardless of whether they are bound to DNA.

Previous studies have identified regulatory elements of the *Oct4* and *Sox2* genes. For *Oct4*, its regulatory regions are important for driving expression in early stage mouse embryos and were defined through analysis of *LacZ* reporter genes regulated by different mouse *Oct4* genomic fragments (Yeom *et al.*, 1996). The *Oct4* regulatory regions identified were the core promoter, proximal enhancer and distal enhancer. The core promoter is located within the first 250 bp of the transcription initiation site. The proximal enhancer, located about 1.2 kb upstream, was shown to direct Oct4 expression in the epiblast, Oct4 downregulation in the anterior to posterior direction after gastrulation, and retinoic acid-dependent Oct4 downregulation in embryonic carcinoma (EC) cells (Okazawa *et al.*, 1991). The distal enhancer region (located about 2 kb upstream) was shown to regulate *Oct4* expression in preimplantation embryos (morula and ICM), primordial germ cells (PGCs), and in pluripotent ES, EC and embryonic germ (EG) cells. In another study, alignment of the upstream sequences of human, bovine, and mouse *Oct4* promoters revealed four conserved regions, CR1 to CR4 (Nordhoff *et al.*, 2001). Interestingly, CR4 overlaps with the distal enhancer, suggesting that evolutionarily conserved elements may be regulating the activity of the distal enhancer. However, the factors that bind and regulate these regulatory elements have not been identified.

Sox2 is transcriptionally regulated in ESCs by an enhancer containing a composite sox-oct element where Oct4 and Sox2 were reported to bind in a combinatorial interaction. Two regulatory regions (SRR1 and SRR2) in Sox2 were reported to confer ESC-specific expression (Tomioka *et al.*, 2002). SRR1 has been demonstrated to bind Oct4 at an octamer site by chromatin immunoprecipitation (ChIP) (Catena *et al.*, 2004). SRR2, located 1.2 kb downstream of the Sox2 transcription start site, contains the composite sox-oct element. Mutations within this composite element disrupted the *in vitro* formation of a DNA/protein complex and also resulted in the loss of SRR2 enhancer activity. More importantly, the reduction of Oct4 abolished the SRR2 activity, indicating that Oct4 can positively regulate SRR2. Further studies would be needed to clarify

whether Oct4 and Sox2 are bound to the SRR2 in ESCs and whether Sox2 binding to the SRR2 region is required for the transcriptional activity of Sox2.

Since Oct4 and Sox2 have been suggested to work together and that their binding elements were found in many ES-specific genes, Oct4 and Sox2 may also bind to genes encoding themselves as well as to the other key ESC gene, *Nanog*. Factors that bind to and regulate the transcription of *Nanog* gene have yet to be identified. This study employs chromatin immunoprecipitation (ChIP) in a small-scale gene-by-gene mapping of Oct4 and Sox2 binding sites in living mESCs and hESCs. Results showed that Oct4 and Sox2 bind to *Oct4*, *Sox2* and *Nanog* loci on the composite sox-oct elements, leading to the establishment of the core transcription factor network circuitry in mESCs.

3.2 Results

In this study, ChIP-qPCRs were carried out to investigate the binding of Oct4 and Sox2 on genomic *Oct4*, *Sox2* and *Nanog* DNA. Undifferentiated mESCs grown in feeder-free conditions were crosslinked with formaldehyde, and the fragmented chromatin lysates were subjected to immunoprecipitation with the Oct4 or Sox2 antibodies.

3.2.1 Optimisation of the Oct4 and Sox2 ChIP assays

Firstly, optimisation of ChIP assays using mESCs was conducted. A series of mESC chromatin sonication was carried out to find the optimal conditions for producing DNA fragment sizes of about 500 bp (data not shown). Optimal ESC chromatin fragment sizes were obtained with sonication conditions of: 40% amplitude, 15 pulses, 1 min rest in between pulses at 4°C. 1ml sonication buffer was used for every 100 µl chromatin pellet and 1ml glass beads were added to facilitate the fragmentation process. Minimum sonication volume in a 14 ml bacterial culture tube is 3 ml. For smaller sonication volumes, sonication was optimally carried out in a 2 ml volume in

an Eppendorf tube with 100 µl glass beads with sonication conditions of 40% amplitude, 4 pulses, 1 min rest in between pulses, in an ice-filled container. In this Chapter, protein G sepharose beads were used. At least two different antibodies each for Oct4 (N-19 goat polyclonal antibody from Santa Cruz and an in-house produced Oct4 rabbit polyclonal antibody) and Sox2 (Y-17 goat polyclonal antibody from Santa Cruz, an in-house produced Sox2 rabbit polyclonal antibody and Ab5306 rabbit polyclonal antibody from Chemicon) were tested by ChIP and yielded relatively similar enrichment results (data not shown). This was carried out to exclude the possibility of non-specific recognition of proteins and epitope masking that could reduce the efficiency of ChIP and preclude the identification of some targets. Western blots of mESC nuclear extracts showed that the two main antibodies used (anti-Oct4: N-19 sc-8628 and anti-Sox2: Y-17 sc-17320) detected proteins that had sizes that correspond to that of Oct4 (43kDa) and Sox2 (34kDa) (Figure 3.1).

3.2.2 Oct4 and Sox2 bind to the distal enhancer of *Oct4* in mESCs

Ten pairs of primers which were located sequentially along the entire conserved promoter region and the first exon were used to quantify ChIP-enriched DNA by real-time PCR (Figure 3.2A). A peak representing Oct4 and Sox2 binding was observed (approximately 30-fold above background) at the distal enhancer, indicating that this region was specifically bound by the two transcription factors (Figure 3.2B and C). A control antibody (glutathione-S-transferase; GST) showed no significant enrichment over the entire surveyed region (Figure 3.2D). Similar results were obtained when polyclonal antibodies against other epitopes of Oct4 and Sox2 were used, further confirming the specificity of this binding (data not shown). Thus, *Oct4* distal enhancer is a *bona fide* target of Oct4 and Sox2 in undifferentiated mESCs.

Retinoic-acid (RA) treatment overrides the self-renewal properties of mESCs and induces their differentiation. There is a corresponding decrease in endogenous levels of Oct4 and Sox2 (Figure 3.3A). The binding profiles of Oct4 and Sox2 on *Oct4* in this RA-induced differentiation

model were examined, as this also allowed us to determine the specificity of the antibodies by analyzing the binding in cells deficient in these factors. Mouse ESCs at different periods of RA treatment were crosslinked with formaldehyde, and the occupancies of Oct4 and Sox2 on the *Oct4* promoter were analyzed by ChIP. Upon differentiation, the binding of Oct4 and Sox2 to the *Oct4* distal enhancer was reduced in close correlation with the degree of differentiation (Figure 3.3B and C). By day 3 of RA treatment, the level of Oct4 binding was reduced by almost 50%, and by day 6, no significant enrichment was detected. A mock ChIP using a control antibody (myeloid/ lymphoid leukemia; MLL) did not result in enrichment of the enhancer sequences (Figure 3.3D). This indicated that the antibodies recognize specific complexes found in mESCs but not in their differentiated derivatives. The results indicate that Oct4 and Sox2 bind to the *Oct4* enhancer when this gene is being actively transcribed in undifferentiated, pluripotent mESCs.

3.2.3 Oct4 and Sox2 bind to the SRR2 of *Sox2* in mESCs

Previous studies have demonstrated the importance of the enhancer element at the 3' end of *Sox2* by reporter assays and mutagenesis (Tomioka *et al.*, 2002). This SRR2 (Sox regulatory region 2) region contains a composite sox-oct element. Although Oct6 and Oct4 have been implicated to bind to SRR2 along with Sox2, it is unclear if Oct4 and Sox2 are indeed bound to the SRR2 site in living ESCs. The ChIP results in Figure 3.4 shows Sox2 and Oct4 bound to the SRR2 region of *Sox2* in mESCs (Figure 3.4) and that these interactions were found specifically in undifferentiated ESCs. This result is similar to that for Oct4 and Sox2 binding on the *Oct4* gene, in which the levels of Oct4 and Sox2 bindings were reduced upon differentiation (Figure 3.4B and C). A control antibody (myeloid/ lymphoid leukemia; MLL) showed no significant enrichment for any of the amplicons from any of the ESC states analyzed (data not shown). Taken together, these results demonstrate that the SRR2 region in *Sox2* is a *bona fide* target for Oct4 and Sox2.

3.2.4 Oct4 and Sox2 bind to the *Nanog* promoter in mESCs

To investigate if Oct4 and Sox2 interact with the *Nanog* promoter *in vivo*, ChIP experiments were carried out using Oct4 and Sox2 antibodies and nuclear extracts from undifferentiated mESCs and in mESCs differentiated with RA. ChIP DNA fragments from undifferentiated mESCs were enriched up to 43- and 37-fold at the *Nanog* proximal promoter when immunoprecipitated with Oct4 and Sox2 antibodies, respectively (Figure 3.5B and C). Two neighbouring regions were not significantly enriched. Upon retinoic acid-induced differentiation of mESCs, this enrichment was reduced proportionally in response to the degree of differentiation. After 3 days of differentiation, enrichment only reached a maximum of 10-fold above background with both Oct4 and Sox2 antibodies, and after 6 days of differentiation, no significant enrichment was detectable (Figure 3.5B and C). Using an antibody against MLL as a negative control, no significant enrichment was found for any of the amplicons from the three ESC states analyzed (data not shown).

3.2.5 OCT4 and SOX2 bind to the CR4 region of *OCT4,* SRR2 region of *SOX2* and promoter region of *NANOG* in human ESCs

To determine if OCT4 and SOX2 bind to *Oct4*, *Sox2* and *Nanog* genes in hESCs, Oct4 and Sox2 ChIP-qPCR were carried out using HUES-6 hESCs propagated on inactivated mouse embryonic fibroblasts or grown in feeder-free conditions in the presence of matrigel. Both methods of culturing hESCs were tested using ChIP-qPCR and gave similar results. As fibroblast mouse cells do not express Oct4 or Sox2, the growth of hESCs on a mouse feeder cell layer did not affect the ChIP analysis. Moreover, the primers used to quantify the ChIP-enriched DNA were specific to human sequences (Figure 3.6A, C, E). Both the OCT4 and SOX2 ChIP assays on hESCs showed enrichment of DNA fragments in the distal enhancer of *OCT4* (Figure 3.6B). A control antibody did not show any enrichment of these enhancer sequences (Figure 3.6B). This shows the *in vivo* binding of OCT4 and SOX2 to the distal enhancer of *OCT4* in hESCs. Furthermore, ChIP-qPCR also showed that OCT4 and SOX2 binds to the SRR2 region 3' of *SOX2* in hESCs (Figure 3.6D).

The control antibody did not show any enrichment (Figure 3.6D). Thus, OCT4 and SOX2 also bind to the SRR2 region of *SOX2* in hESCs. To establish that OCT4 and SOX2 also interact with the *NANOG* promoter in hESCs, similar ChIP analysis was performed using HUES-6 hESCs. Six different amplicons (Figure 3.6E), all within close proximity to exon 1 of *NANOG* were used to detect for enrichment of OCT4 and SOX2 ChIP DNA. Three primer pairs showed significant enrichment with OCT4 and SOX2 antibodies (Figure 3.6F). Using the GST antibody as a negative control, no significant enrichment was found for any of the amplicons in these hESCs. In summary, these ChIP analyses demonstrate the *in vivo* occupancy of OCT4 and SOX2 on the *NANOG* promoter in undifferentiated hESCs.

3.2.6 Conserved elements in the CR4 region of *Oct4* promoter, SRR2 region of *Sox2* enhancer and promoter region of *Nanog*

Multiple sequence alignment of genomic regions with Oct4 and Sox2 binding (highest ChIP enrichment) were carried out. A consensus motif that resembles the composite sox-oct element which is involved in regulating the transcription of several other ESC-specific genes was identified in the CR4 region of *Oct4*, SRR2 region of *Sox2* and promoter region of *Nanog* of mESCs. The mouse elements are shown together with other Sox-Oct elements found in other studies in Figure 3.7. This element has also been identified in the conserved DNA regions in several organisms (Rodda *et al.*, 2005; Nordhoff *et al.*, 2001; Tomioka *et al.*, 2002). This indicates that Oct4 and Sox2 bind to a specific motif which is evolutionarily conserved, and thus implying that this motif has an important functional role in mediating the actions of Oct4 and Sox2.

3.3 Discussion

3.3.1 Oct4, Sox2 and Nanog circuitry in ESCs

During early embryogenesis, Oct4 and Sox2 are found to be co-expressed in several pluripotent cells such as the morula, ICM, epiblast, and germ cells. Gene knockout studies revealed that the primary defect for both the *Oct4*- and *Sox2*-null animals is in the pluripotent epiblast, though there are slight differences between the two null phenotypes. There is no epiblast development in the *Oct4*-null blastocyst, and the fate of all cell types is towards trophectoderm lineage (Nichols *et al.*, 1998). On the other hand, *Sox2*-null animals are capable of giving rise, at least transiently, to the epiblast, as epiblast-derived extraembryonic endoderm is detected (Avilion *et al.*, 2003; Ben Shushan *et al.*, 1995). However, this transient epiblast formation may be the result of maternally-derived Sox2 protein. Therefore, *Oct4*- and *Sox2*-null blastocysts are incapable of giving rise to pluripotent ESCs, indicating that they are both crucial regulators for these cells. Besides Oct4 and Sox2, the other key regulator of ESCs is Nanog. The removal of *Nanog* via gene targeting or RNAi leads to differentiation of mESCs (Chambers *et al.*, 2003; Mitsui *et al.*, 2003; Rodda *et al.*, 2005).

Results show that the enhancer elements of Oct4 and Sox2 are the direct targets of their respective gene products and are reciprocally bound by the other regulator. In addition, Oct4 and Sox2 also bind to the *Nanog* promoter region (Figure 3.8A). Within these binding sites, composite sox-oct elements were identified. This study expands the list of genes (*Fgf4*, *Utf1*, *Opn*, and *Fbx15*) which are potentially regulated by both Oct4 and Sox2, and is also the first to directly link these three transcription factors within the pluripotent cell genetic regulatory network. As the highest level of binding was detected in undifferentiated cells when these genes are known to be most transcriptionally active, they are likely associated with the transcriptional activation of these genes and may play a role in positively regulating expression. Moreover, gene-specific sequence

conservation within these sox-oct composite elements suggests functional significance. This was further confirmed by the knockdown of both *Sox2* and *Oct4* mRNAs by RNAi (Chew *et al.*, 2005; Rodda *et al.* 2005). Knockdown of Oct4 and Sox2 caused a reduction in the *Oct4* distal enhancer activity and in endogenous Oct4, Sox2 and Nanog transcript levels, and caused the cells to differentiate. This indicates that both Sox2 and Oct4 are positively regulating *Oct4, Sox2* and *Nanog*. Together with the binding results obtained from this study, a regulatory network containing Oct4, Sox2 and Nanog was established whereby Oct4 and Sox2 autoregulate, regulate reciprocally and regulate *Nanog*.

3.3.2 Network motifs in the Oct4, Sox2 and Nanog circuit

The interactions of Oct4 and Sox2 with the respective genes can be described as a transcriptional regulatory network consisting of autoregulatory, multi-component, feed-forward and multi-input motif loops (Figure 3.8B) (Lee *et al.*, 2002). A network motif (architecture) is a fundamental unit within a complex transcriptional regulatory network.

In an autoregulation model, the gene product binds to its own regulatory element (Figure 3.8B). This may allow for self-perpetuation and enhanced stability of gene expression. Autoregulation can be shut down by active mechanisms during differentiation. The promoter of *Oct4* contains negative regulatory elements which are required for repression when embryonal carcinoma cells differentiate (Ben Shushan *et al.*, 1995; Scholer, 1994; Schoorlemmer *et al.*, 1994; Sylvester and Pikarsky *et al.*, 1994). One example of a negative regulator is the germ cell nuclear factor (Gcnf) which has been shown to mediate repression of the *Oct4* proximal promoter (Fuhrmann *et al.*, 2001). The expression of *Gcnf* is inversely correlated with the *Oct4* expression in embryonal carcinoma cells (ECCs) and in *Gcnf*-knockout mouse embryos, the *Oct4* expression is no longer confined to the germ cell lineage.

An additional relationship between Oct4 and Sox2 is depicted by a multi-component loop motif whereby a regulator binds to the regulatory elements of another regulator in a closed loop (Figure 3.8B). Such a closed-circuit loop can efficiently generate a bi-stable system with the ability to switch between two different states. For ESCs, the two states may be the decision to undergo self-renewal or to differentiate and exit from symmetrical division. This model also requires that the concentrations of the two factors remain relatively constant, as any slight change in the abundance of one protein will destabilize the circuitry.

Besides that, Oct4 and Sox2 are depicted in a feedforward loop motif containing a regulator that controls a second regulator and has the additional feature that both regulators bind a common target gene (Figure 3.8B). The feedforward loop may act as a switch that is designed to be sensitive to sustained inputs. Feedforward loops have the potential to provide temporal control of a process, because expression of the ultimate target gene may depend on the accumulation of adequate levels of the master and secondary regulators. In this case, the expression of Nanog is proposed to be regulated by an appropriate level of Oct4 and a second regulator Sox2. Feedforward loops may provide a form of multistep ultrasensitivity, as small changes in the level or activity of the master regulator at the top of the loop might be amplified at the ultimate target gene because of the combined action of the master regulator and a second regulator that is under the control of the master regulator. Interestingly, it has been shown that mESCs are exquisitely sensitive to the level of Oct4 (Niwa *et al.*, 2000). Increasing the level of Oct4 by 50% is sufficient to induce differentiation of ESCs into primitive endoderm and mesoderm.

Lastly, Oct4 and Sox2 are involved in multi-input motifs consisting of a set of regulators (Oct4 and Sox2) that bind together to a set of genes (*Oct4*, *Sox2*, *Nanog*) (Figure 3.8B). In ESCs, this motif may offer the potential for coordinating gene expression for maintaining the ESC state.

3.3.3 Conjectures

It is apparent that Oct4 and Sox2 are important transcriptional regulators in ESCs. The binding of Oct4 and Sox2 may thus be instrumental in recruiting various other protein partners. However, it should be emphasized that not all Oct4 and Sox2 sites on the same regulatory region are synergistic in transcriptional activation. For example, in the *Osteopontin* intron, a sox site 39 bp away from an inverted pair of Oct4 sites acts antagonistically in transactivation by Oct4 (Botquin *et al.*, 1998). Repression by Sox2 was shown to require DNA binding and a carboxy-terminal transactivation domain. For *Fgf4*, *Utf1*, and *Fbx15*, the regulatory elements contain Oct4 and Sox2 sites in proximity (either 0- or 3-bp separation) and the Oct4/Sox2 complex is implicated in transactivation. This raises the interesting possibility that perhaps Oct4 and Sox2 collaborate to globally control ESC-specific gene expression through the sox-oct motifs. To address this possibility, the genome-wide targets of both regulators have to be identified.

It is also possible that the Oct4/Sox2 complex interacts with another factor(s) to activate the network of ESC-specific genes. For example, both Oct4 and Sox2 are present in the nucleus of cells of the morula and their precursors (Botquin *et al.*, 1998; Donovan and Gearhart, 2001) which do not express Nanog, indicating that other molecular signals, besides the appropriate cellular location of Oct4 and Sox2, are required for pluripotent-specific expression of *Nanog*. It will be interesting to identify other factors that act through other *cis*-elements within this conserved promoter and/or through coactivators interacting with the Oct4/Sox2 complex.

It is also important to note that there are other key regulators for maintenance of the undifferentiated state of ESCs. The LIF/Stat3 pathway is essential for self-renewal of mESCs (Matsuda *et al.*, 1999; Niwa *et al.*, 1998; Raz *et al.*, 1999). This was shown by the removal of LIF which leads to the inactivation of Stat3 and induces differentiation. Intriguingly, overexpression of Nanog is sufficient to bypass the LIF/Stat3 requirement. This may mean that Stat3 is regulating

Nanog. However, how Stat3 interacts with Nanog and the Oct4/Sox2 pathway remains to be studied. All these questions were examined in the following chapters.

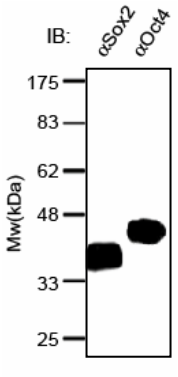


Figure 3.1: Specificity of αOct4 and αSox2 antibodies used in ChIP. Western blot analyses of mESC nuclear extracts using antibodies against Oct4 (N-19) and Sox2 (Y-17).

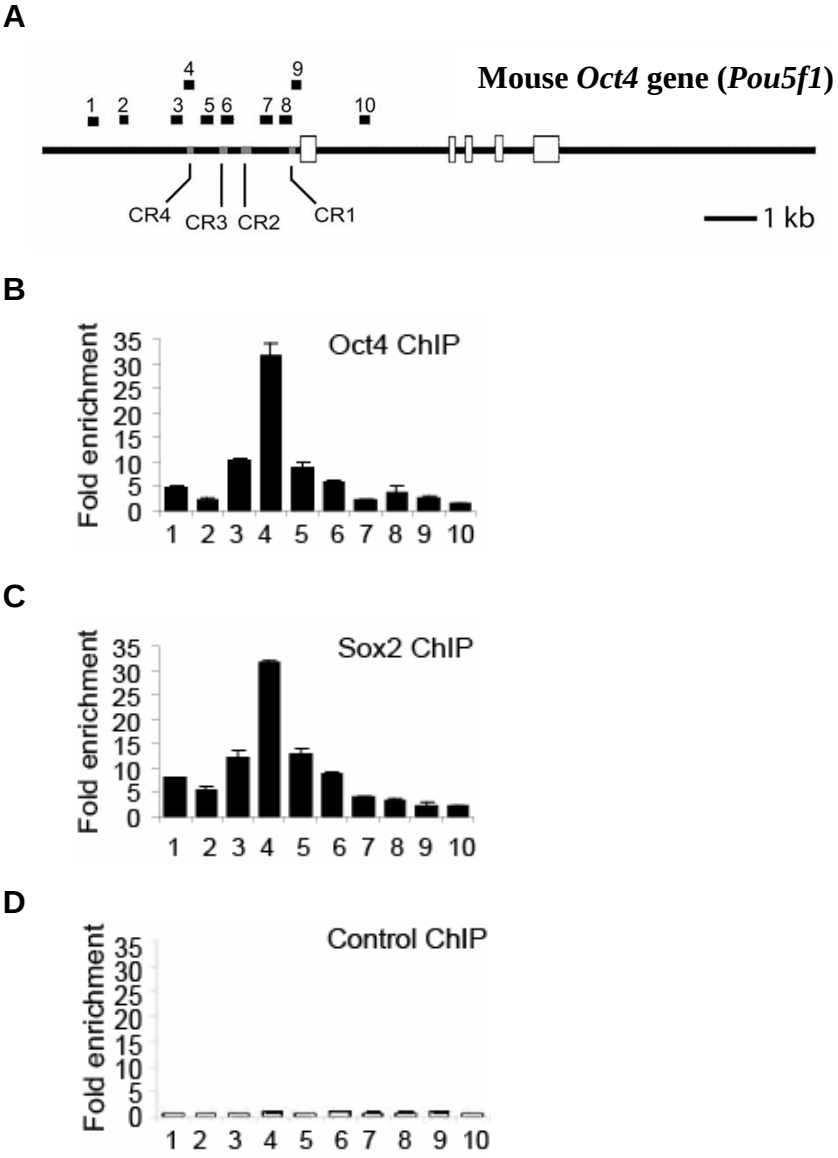


Figure 3.2: Oct4 and Sox2 binding to *Oct4* CR4 region in mESCs. (A) Locations of the amplified products (numbered black boxes) of ten primer pairs used in Oct4 ChIP-qPCR on the mouse *Oct4* gene. The locations of the conserved regions, CR1-4 are indicated. Open boxes represent exons. (B-D) High resolution ChIP-qPCR mapping of Oct4 (B), Sox2 (C), and control (glutathione S-transferase; GST) antibody (D) binding sites across the *Oct4* promoter in mESCs by ChIP analysis. Fold enrichment is the relative abundance of DNA fragments at the indicated regions (A) over a control region as quantified by real-time PCR.

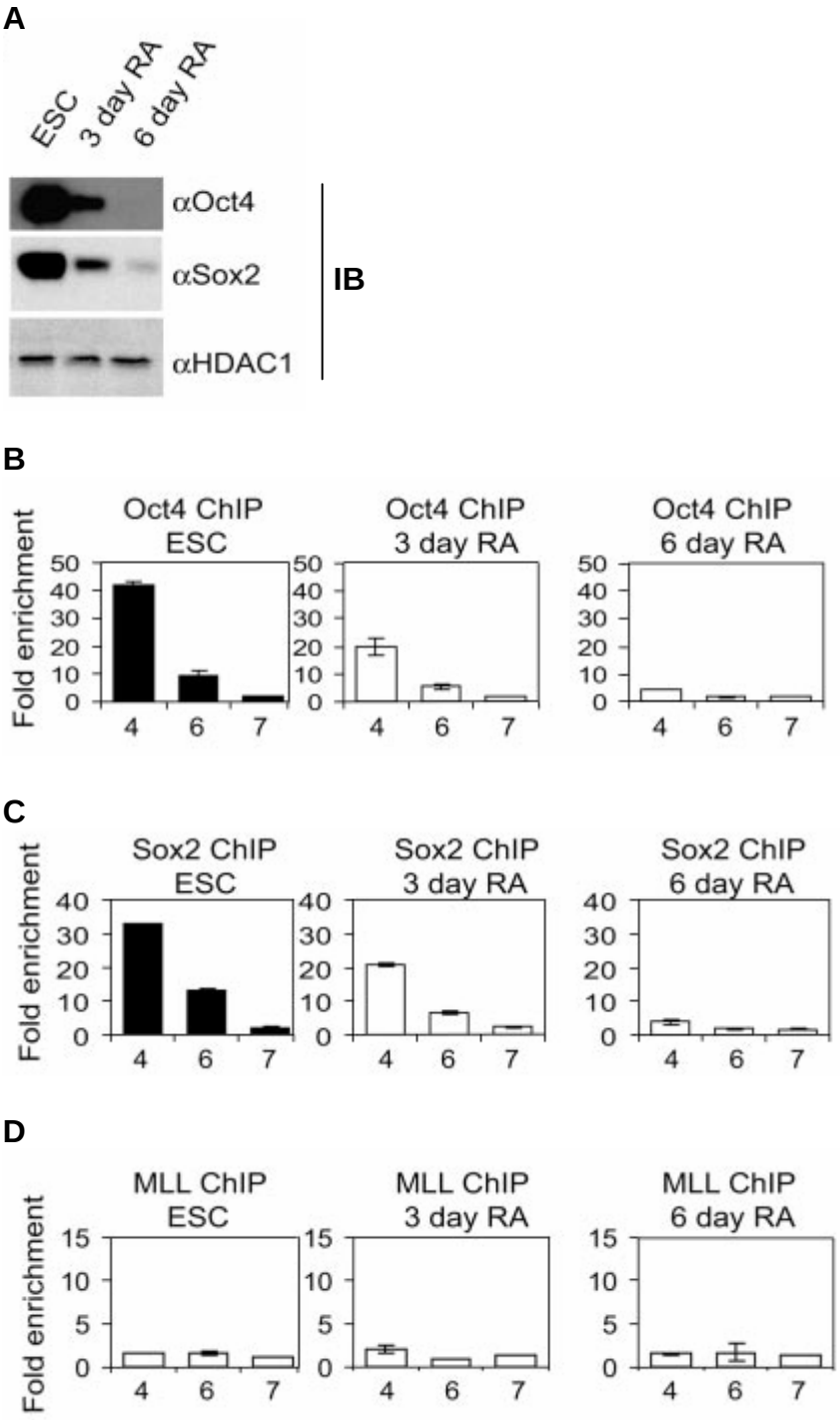


Figure 3.3: Oct4 and Sox2 binding reduces after retinoic acid differentiation of mESCs. (A) Oct4 and Sox2 levels decreased upon retinoic acid-induced differentiation of mESCs. Western blot analyses of chromatin extracts using antibodies against Oct4, Sox2 and a histone deacetylase 1 antibody (HDAC1) as a loading control in undifferentiated mESCs (ESC) and mESCs treated with retinoic acid for 3 and 6 days. A similar analysis of Oct4 (B) and Sox2 (C) occupancy on the *Oct4* promoter in undifferentiated mESCs and in mESCs induced to differentiate with retinoic acid for 3 and 6 days. ChIP analysis was used as a subset (4, 6, and 7) of the amplicons described in panel Figure 3.2. (D) A control ChIP assay using an anti-MLL antibody is also shown.

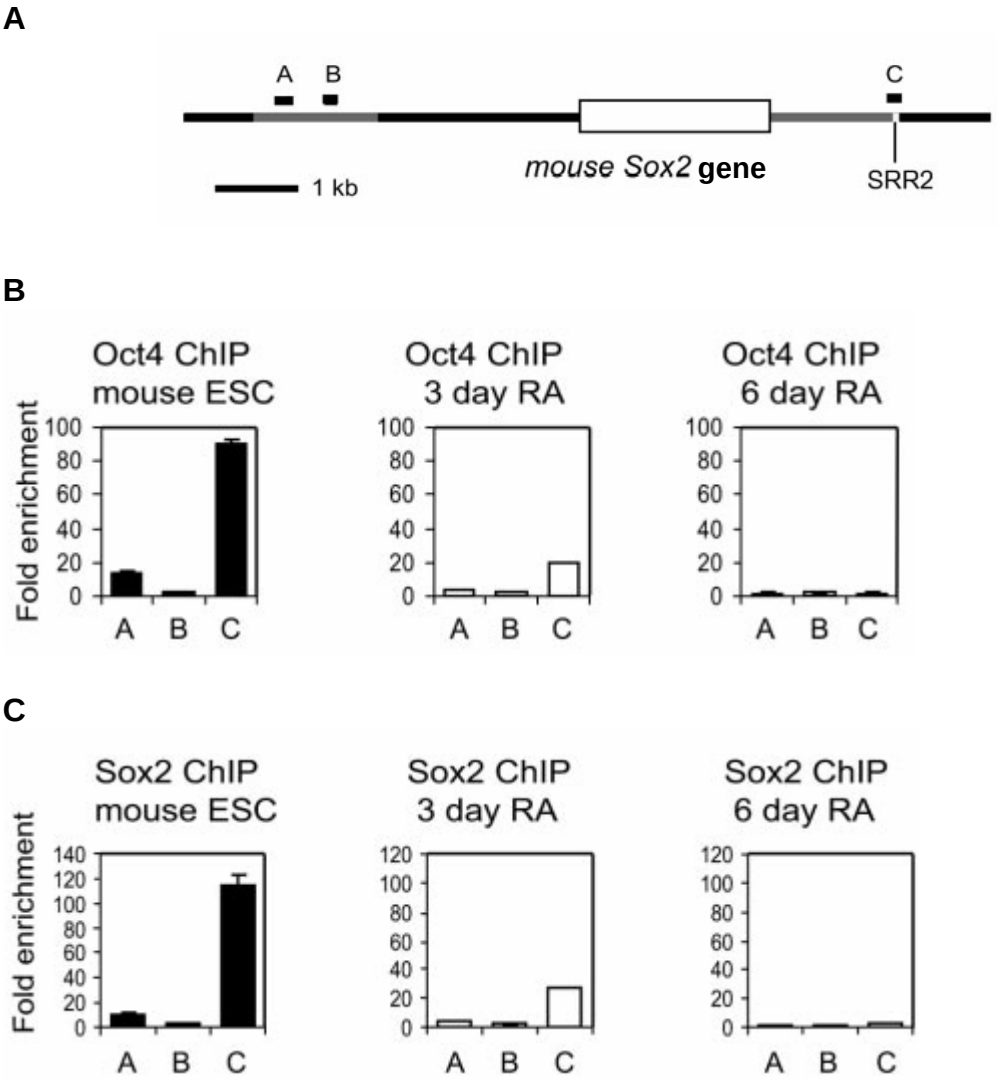


Figure 3.4: Oct4 and Sox2 bind to the SRR2 at the 3' enhancer of *Sox2* in mESCs. (A) Schematic diagram of mouse *Sox2* genomic locus with its single exon represented by an open box and the SRR2 region indicated. The relative locations of the amplicons used to detect enriched ChIP fragments are shown (A to C). (B) Measurement by ChIP analysis of Oct4 occupancy in regions of *Sox2* in undifferentiated mESCs and those induced to differentiate for 3 and 6 days by retinoic acid. Letters correspond to the amplicons indicated in (A). (C) Measurement by ChIP analysis of Sox2 occupancy in regions of *Sox2* in undifferentiated mESCs and those induced to differentiate for 3 and 6 days by retinoic acid.

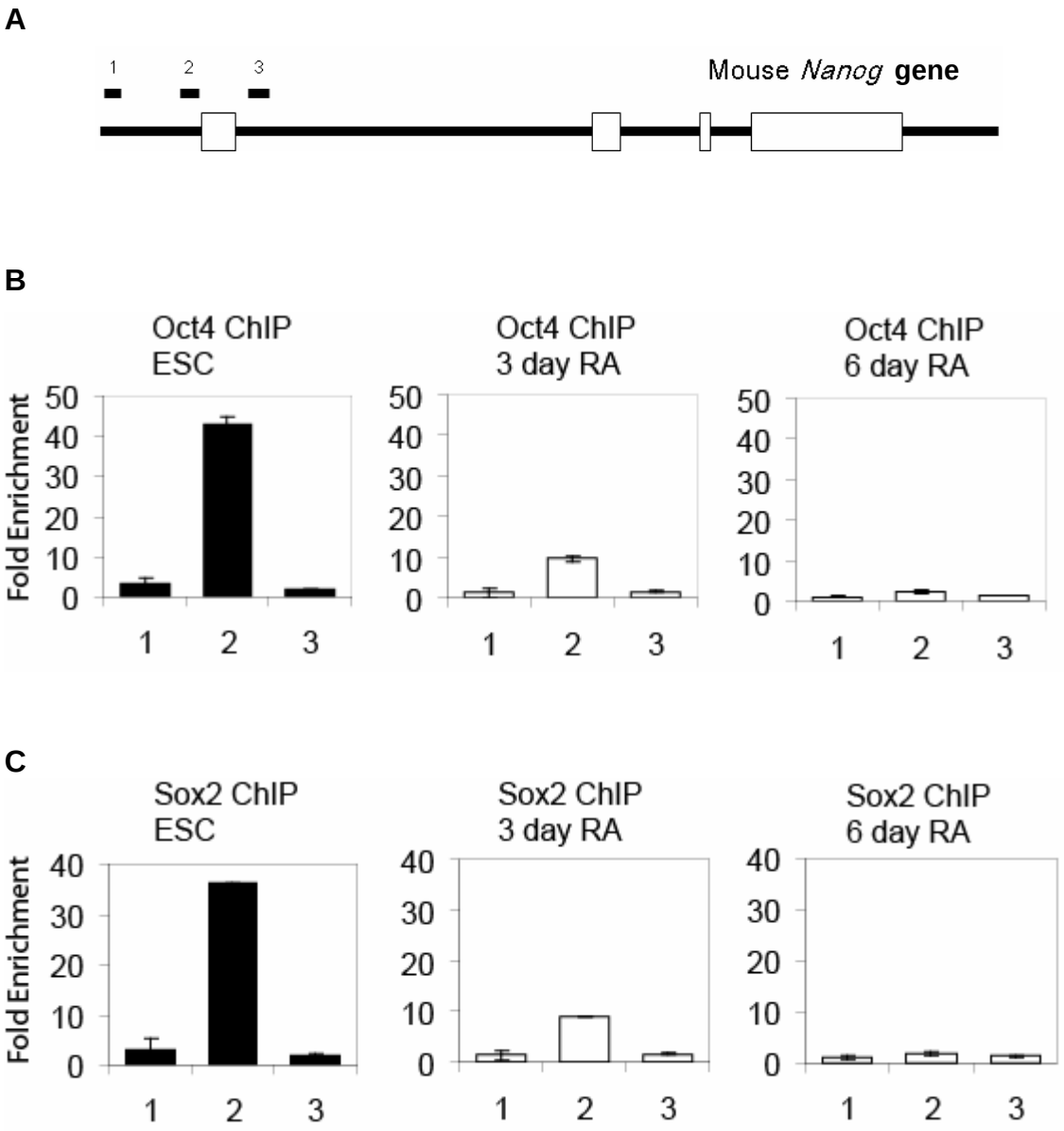
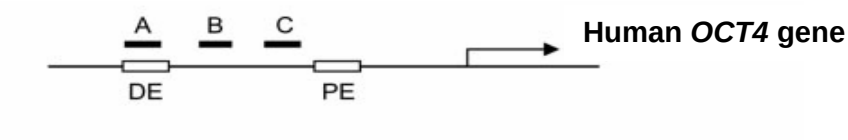
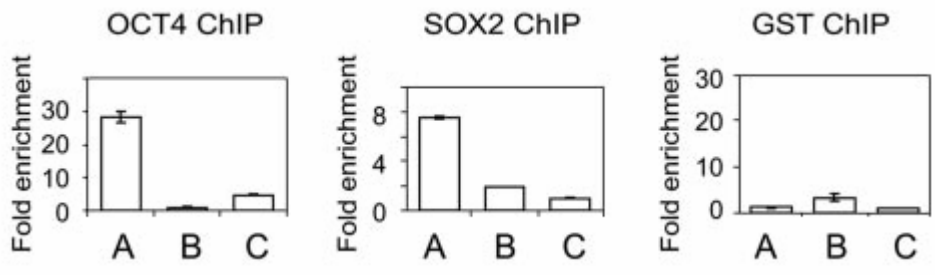


Figure 3.5: Oct4 and Sox2 bind to the *Nanog* promoter in mESCs. (A) Schematic diagram of mouse *Nanog* genomic locus with exons represented by open boxes. The relative locations of the amplicons used to detect enriched ChIP fragments are shown (1-3). (B) Measurement by ChIP analysis of Nanog occupancy in regions of *Nanog* in undifferentiated mESCs and those induced to differentiate for 3 and 6 days by retinoic acid. Letters correspond to the amplicons indicated in (A). (C) Measurement by ChIP analysis of Sox2 occupancy in regions of *Nanog* in undifferentiated mESCs and those induced to differentiate for 3 and 6 days by retinoic acid.

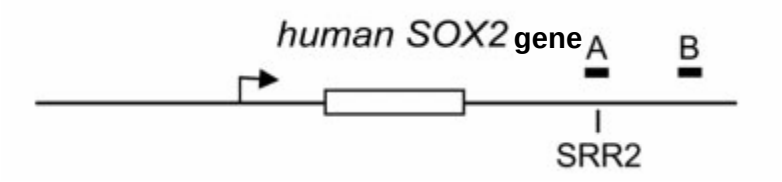
A


A
B
C
DE
PE
Human *OCT4* gene

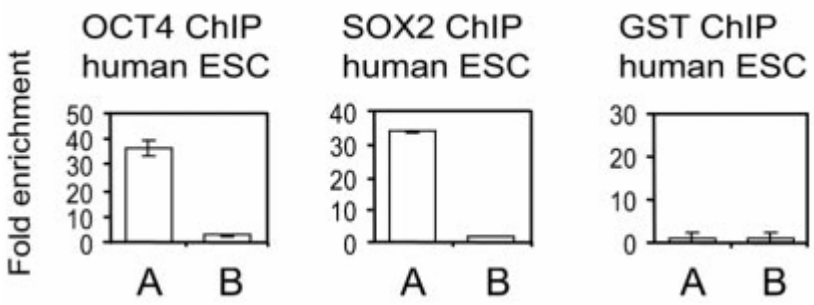
B


OCT4 ChIP
SOX2 ChIP
GST ChIP
Fold enrichment
30
20
10
0
8
4
A
B
C

C


human SOX2 gene
A
B
SRR2

D


OCT4 ChIP
human ESC
SOX2 ChIP
human ESC
GST ChIP
human ESC
Fold enrichment
50
40
30
20
10
0
A
B

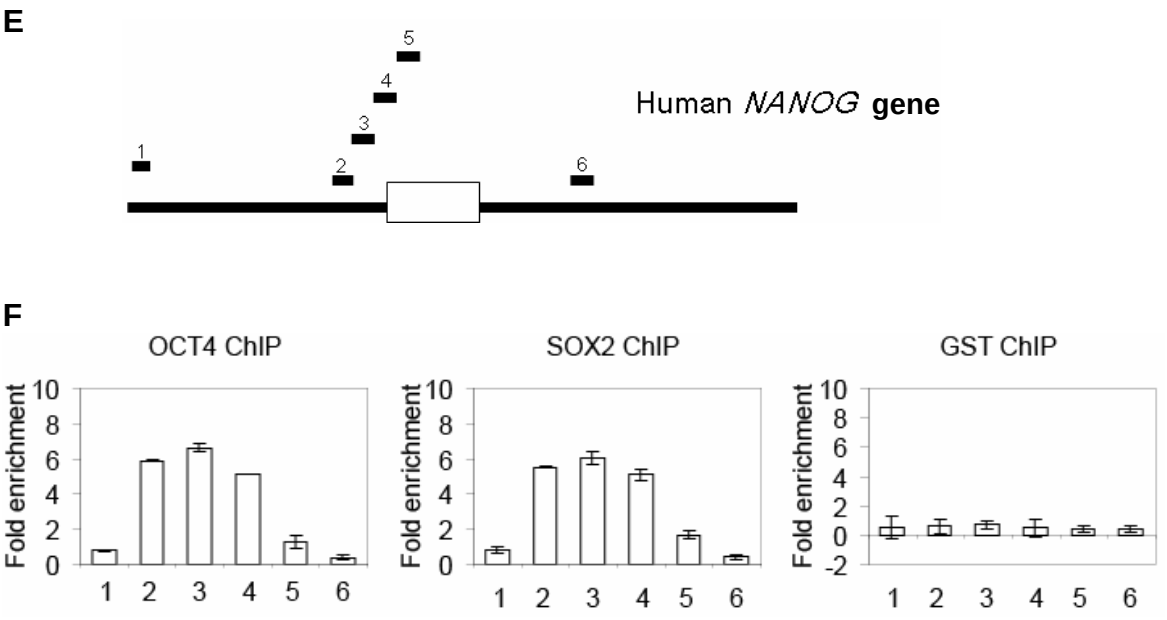


Figure 3.6: OCT4 and SOX2 bind to the distal enhancer (DE)/CR4 region of *OCT4*, the SRR2 region of the *SOX2* and promoter region of *NANOG* in living human ESCs. (A) Schematic diagram of the location of the amplicons (A, B, and C) used to detect ChIP-enriched fragments in *OCT4* shown relative to the distal enhancer (DE)/CR4 region, to the proximal enhancer (PE), to the transcription start site (arrow) and to the exon (box). (B) ChIP-qPCR using OCT4, SOX2, and a control glutathione S-transferase (GST) antibody. (C) Schematic diagram of the human *SOX2* genomic locus with the single exon represented by an open box and the SRR2 region indicated. The relative locations of the amplicons used to detect enriched ChIP fragments are shown (A and B). (D) Measurement by ChIP analysis of OCT4 and SOX2 occupancies on *SOX2* in living hESCs. A glutathione S-transferase antibody (GST) was used as a negative control. (E) Schematic diagram of the location of the amplicons (1 to 6) used to detect ChIP-enriched fragments in *NANOG* shown relative to the first exon (box). (F) Measurement by ChIP analysis of OCT4 and SOX2 occupancies on *NANOG* in living hESCs. Control glutathione S-transferase (GST) ChIP-qPCR is also shown.

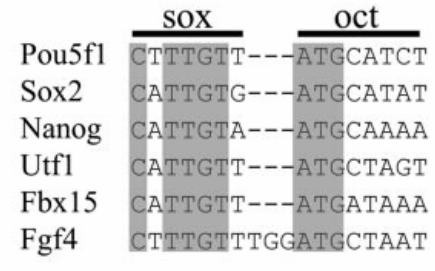


Figure 3.7: Multiple alignment analysis of Oct4 and Sox2 binding sites in mESCs from this study and previous studies (*Utf1*, *Fbx15*, *Fgf4*) identified the Sox-Oct composite element.

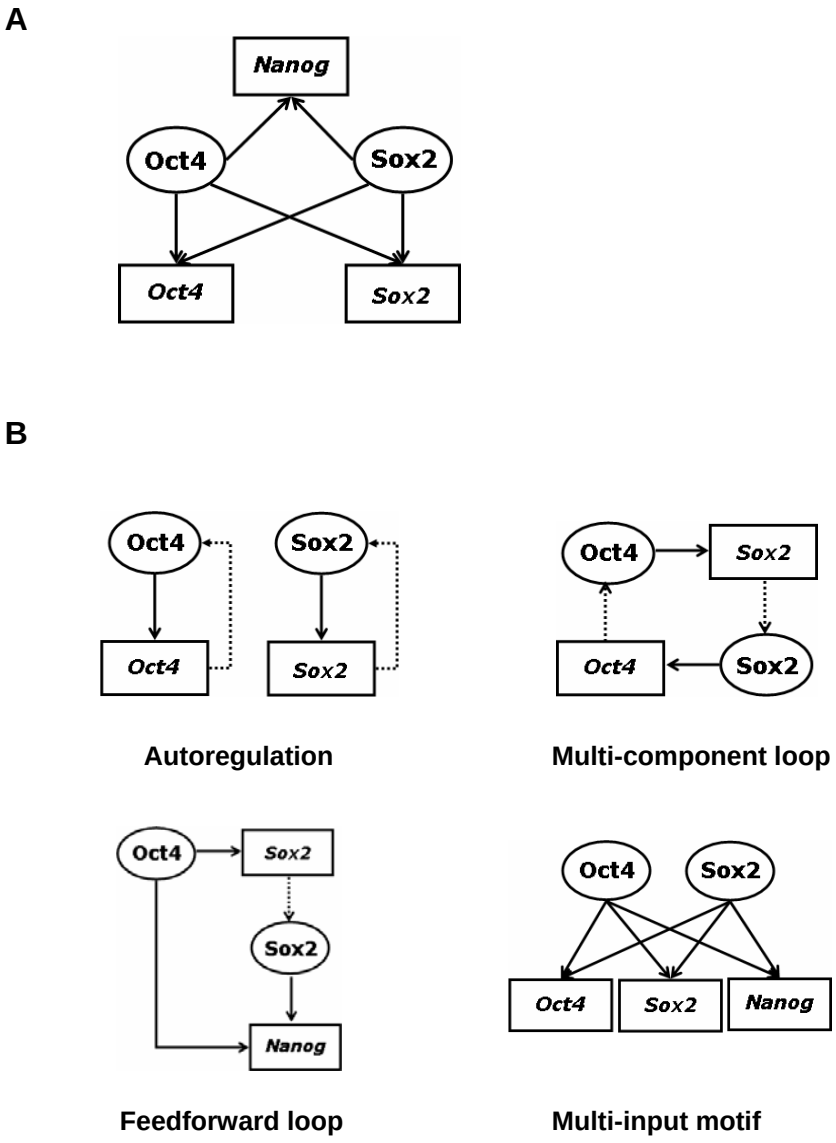


Figure 3.8: Oct4, Sox2 and Nanog circuitry and network motifs. (A) Mapping of the Oct4, Sox2, and Nanog circuitry in ESCs. The relationship between Oct4 and Sox2 proteins (circled) and *Oct4*, *Sox2* and *Nanog* genes (boxed). Solid arrows represent binding shown by ChIP in this study. (B) The network motifs in the Oct4, Sox2 and Nanog regulatory network includes autoregulation, multi-component loop, feed-forward loop and multi-input motif. Regulators are represented by circles; gene promoter/enhancers are represented by rectangles. Binding of a regulator to a promoter is indicated by a solid arrow. Genes encoding regulators are linked to their respective regulators by dashed arrows.

CHAPTER 4

Genome-wide Mapping of Oct4-DNA Interactions in Mouse Embryonic Stem Cells

4.1 Introduction

The previous chapter demonstrated that Oct4 and Sox2 bind and regulate transcription of *Oct4*, *Sox2* and *Nanog* (Boyer *et al.*, 2005; Chew *et al.*, 2005; Kuroda *et al.*, 2005; Okumura-Nakanishi *et al.*, 2005; Rodda *et al.*, 2005). Both Sox2 and Nanog are essential for maintaining the pluripotent phenotype, although the *in vivo* phenotype of both hints at a slightly later role in development as compared with Oct4 (Avilion *et al.*, 2003; Chambers *et al.*, 2003; Mitsui *et al.*, 2003). Results from the previous chapter suggest that Oct4 and Sox2 function through autoregulatory feedback loops (*Sox2* and *Oct4*) and feed-forward mechanisms (*Nanog*) to maintain the expression of transcription factors essential to the ESC phenotype.

Other targets of the Oct4-Sox2 complex in pluripotent cells include *Fgf4*, *Utf1*, and *Fbxo15*. Fgf4 is an extracellular signalling molecule synthesized by the epiblast and it plays an essential function in mediating signals to maintain the trophoblast stem cell (Tanaka *et al.*, 1998). Utf1 is a transcriptional co-activator that has been implicated in enhancing ESC proliferation, thereby associating Oct4 with a proliferative regulatory network (Nishimoto *et al.*, 1999; Nishimoto *et al.*, 2005). Fbx15 is a transcription factor with no known targets or function. The only other known targets of Oct4 are *Spp1* encoding the extracellular molecule osteopontin and *Zfp42* encoding the transcription factor also known as Rex1 (Ben Shushan *et al.*, 1998; Botquin *et al.*, 1998). Neither of these have any known function in the pluripotent cell.

The *cis*-regulatory element, to which the Sox2-Oct4 complex is bound, consists of neighbouring Sox2 (CATTGTA) and Oct4 (ATGCAAAT) binding sites. These elements are immediately adjacent to one another with the sox element in the 5' position in five out of the six

Sox2-Oct4 target genes. The sixth target gene (*Fgf4*) contains 3 intervening base pairs between the two binding sites (Ambrosetti *et al.*, 1997; Ambrosetti *et al.*, 2000). For *Spp1*, the Oct4 site is important in transcriptional activation while a Sox2 site 35 bp 5' to this is implicated in transcriptional repression (Botquin *et al.*, 1998). Sox2 binding to the *Zfp42* promoter has not been described while the binding by Oct4 is through a representative octamer motif (Ben Shushan *et al.*, 1998).

Based on the limited knowledge in the involvement of Oct4 in cellular functions such as ESC maintenance and proliferation, it is anticipated that there are many more Oct4 targets yet to be identified. Moreover, none of the known Oct4 targets display the early phenotypic defect shown by Oct4-null embryos, implying that additional Oct4 targets should exist. Also, given the central role that Oct4 plays in ESC biology, the identification of transcriptional targets of Oct4 would expand our understanding into its regulatory network.

In the previous chapter, ChIP-qPCR was used to discover binding sites of Oct4 and Sox2 in a gene-by-gene approach. While qPCR detection is effective and convenient, it is limited to loci that are selected for study. Here, this study uses chromatin immunoprecipitation coupled with a paired-end ditags (ChIP-PET) sequencing technique developed in the Genome Institute of Singapore (Wei *et al.*, 2006) to map the Oct4 binding sites in mESCs on a global manner. In brief, mESCs were fixed using formaldehyde to preserve *in vivo* protein-DNA interactions. Chromatin extracts were sonicated and subjected to immunoprecipitation using an antibody against Oct4. ChIP-enriched DNA was purified and cloned into a PET library. Subsequently, 36 bp sequence tags (PET) from both ends (18 bp from 5' and 18 bp from 3' end) of each individual ChIP-enriched DNA fragment were generated. These PETs were then concatenated to allow for efficient sequencing. Sequenced PETs were then mapped to the mouse genome to demarcate the boundaries of ChIP-enriched DNA fragments. Single PETs, called singletons, were observed throughout the

genome and they represent background noise due to random recovery of genomic DNA. Genome locations derived from multiple overlapping PETs indicate *bona fide* transcription factor binding sites.

In this study, 366,639 PETs mapping to unique genomic loci were generated from Oct4 ChIP-enriched DNA from undifferentiated mESCs. ChIP-qPCR validations done on these regions empirically defined PET overlaps of 4 and above as high confidence Oct4 transcription factor binding sites (TFBS). 1083 distinct Oct4 binding sites were identified from the clusters of multiple overlapping PETs. Along with known targets of Oct4 such as *Oct4* and *Nanog*, Oct4 was also shown to bind a variety of gene classes. A substantial number of these genes code for transcription factors implicated in a much broader range of cellular processes in ESCs. Genes associated with Oct4 binding encode products that function in an array of cellular processes, including cell proliferation (*Mycn*), transcription (*Rest*, *Tcf3*), DNA integrity (*Trp53*, *Trp53BP1*), metabolism (proteases and pumps), signalling (kinases) and translation (microRNAs).

The strategy used here synergises ChIP, high-throughput cloning and sequencing, computational analyses, and comparative analyses with previous observations by other groups on the functions of the target genes to define the role of Oct4. The results implicate Oct4 in triggering a cascade of pathways governing pluripotency, self-renewal and cell fate determination of ESCs.

4.2 Results

4.2.1 Optimisation of large-scale ChIP

MagnaBind™ goat anti-rabbit IgG (Pierce), Dynabeads M-450 Tosylactivated (Dynal), protein G and protein A magnetic beads (New England Biolabs), Dynabeads M-450 Epoxy beads (Dynal Biotech), peptide [Oct4 PS (N-19) sc-8628PS, Sox2 PS (Y-17) sc-17320PS, Santa Cruz)] elutions,

protein G sepharose (Amersham) and protein G Dynabeads (Dynal) were all tested in Oct4 ChIP-qPCR assays. Protein G sepharose beads were chosen for its high ChIP DNA recovery (approximately 50ng per ChIP reaction) and high enrichment. Feeder-free mESCs grown at a confluency of 70% were optimal for crosslinking and ChIP assays. Large-scale mESC cultures in 500cm^2 dishes (at least 8 dishes at each time) were done to obtain sufficient chromatin extracts. Large-scale Oct4 ChIP of 14 reactions was carried out at each time. Input DNA (decrosslinked chromatin extracts) was re-purified after RNase digestion, measured for enrichment by ChIP-qPCR and quantified before being sent for library construction by the GIS Cloning and Sequencing group.

4.2.2 Global mapping of Oct4 binding sites in mESCs

This study utilises a robust sequencing approach, termed the ChIP-PET technology developed by Dr Ruan Yi Jun's group in GIS, which efficiently identified and positioned Oct4 ChIP DNA fragments onto the mouse genome. Oct4 ChIP-enriched DNA fragments were cloned into a primary library that preserved the original representation of ChIP-enriched DNA. The clones were then converted into paired-end ditags that were concatenated and cloned into a plasmid vector to yield the final ChIP-PET library (please refer to Chapter 2 for more details). Approximately 80,000 randomly selected clones were sequenced, generating 10 to 15 PETs per sequence read for a total of 1,088,836 PET sequences. These PET sequences were then mapped to the mouse genome to define the boundaries of the Oct4 ChIP DNA fragments (termed as 'cluster'). Definition of the terms used in ChIP-PET is illustrated in Figure 4.1. The binding data was then visualized through a browser maintained by the Bioinformatics Institute (http://t2g.bii.a-star.edu.sg). With the help of large scale data prosessing by the Bioinfomatics group in GIS, 515,717 (47%) PETs were found to map to specific locations on the mouse genome (build mm5) and the rest of the PETs either did not map to the genome (469,032; 43%) or were mapped to multiple locations (109,087; 10%) as they were derived from repetitive sequences. The 515,717

PET-mapped locations were grouped into 366,639 unique genomic loci. It is highly unlikely to have two independently generated PETs mapping to identical locations due to the random nature of DNA shearing used in the ChIP protocol. Any redundantly mapped PETs are likely a result of amplification events during the ChIP-PET cloning process.

True Oct4 binding sites should be distinguishable from background noise by virtue of multiple overlapping PETs mapping to specific locations within the genome. Of the 366,639 PETs mapping to unique genome locations, the majority (276,006; 75%) were represented by single non-overlapping PETs (singletons). These singletons most likely represent background noise. The remainder of these PETs (90,633 or 24.7%) was found to overlap with others and comprised 37,623 distinct PET clusters. The total number of PETs in these clusters ranged from 2 PETs for 29,451 of these clusters to 17 PETs represented by one cluster. Some of these clusters, particularly those containing few PETs, may represent background noise from random sampling of genomic DNA.

Subsequently, the minimum number of overlapping PET sequences required within a cluster to distinguish a *bona fide* Oct4 binding site was determined through ChIP-qPCR. Clusters with 2, 3, 4 and 5 overlapping PETs were randomly selected and tested for Oct4 ChIP enrichment by the conventional ChIP-qPCR. Out of the clusters represented by 3 overlapping PETs, only 3 showed more than a 2-fold enrichment over background and were considered to represent background in the data analysis (Figure 4.2). In contrast, all clusters containing 4 overlapping PETs showed significant enrichment of Oct4 binding (greater than 2-fold). In addition, clusters with 5 or more PET overlaps were all enriched over background. Therefore, the majority of loci with 4 PET overlaps and above were deemed to represent authentic Oct4 binding sites. Based on this cut-off, 1083 distinct sites within the mouse genome were determined to be Oct4 high confident binding sites in undifferentiated mESCs.

4.2.3 Oct4 ChIP-PET experiment identifies known Oct4 binding targets

Prior to this study, there were eight reported Oct4 DNA binding sites in mESCs (five of which were validated by ChIP). To verify the reliability of the Oct4 ChIP-PET data, these known sites were searched for positive binding. Firstly, Oct4 bindings on three key transcription factors essential to the ESC phenotype, namely the *Oct4* gene itself, *Sox2*, and *Nanog* as previously characterized by this study and others (Chew *et al.*, 2005; Kuroda *et al.*, 2005; Okumura-Nakanishi *et al.*, 2005; Rodda *et al.*, 2005; Tokuzawa *et al.*, 2003; Tomioka *et al.*, 2002) were examined. As expected, the ChIP-PET method identified Oct4 binding to these three genes where nine PET overlaps mapped to *Oct4* gene (*Pou5f1*), five PET overlaps mapped to *Sox2*, and four PET overlaps mapped to *Nanog*. Furthermore, the PET overlap region mapping to each of these genes encompassed the known Oct4 binding sites as shown and confirmed by conventional ChIP-qPCR. The Oct4 ChIP with its corresponding PET overlaps showing the binding sites on *Oct4* and *Nanog* are depicted in Figure 4.3. Control ChIP using an antibody against yeast Ena-1 protein showed no enrichment with the same primers.

In addition, the ChIP-PET method also identified a fourth previously known Oct4 target gene, *Fbx15* (Tokuzawa *et al.*, 2003), with four PET overlaps spanning the known Oct4 binding site located 530 bp 5' to the transcription start site of this gene. Identification of these known Oct4 binding sites by the ChIP-PET method attested to the reliability of this approach for determining Oct4 binding targets in mESCs. However, the other four previously reported Oct4 binding sites were not identified by the ChIP-PET method. There were no PETs, singletons or otherwise, spanning the Oct4 binding sites in *Fgf4*, *Zfp42*, or *Spp1* and only a singleton spanning the known Oct4 site of *Utf1*. This omission may be due to the inadequate depth of sequencing. When conventional ChIP and a two step ChIP (sequential ChIP)-qPCR were carried out on these loci, there were indeed enrichment or positive binding on the *Zfp42*, *Fgf4*, *Spp1* and *Utf1* loci (data not shown). Therefore, the Oct4 ChIP-PET data may likely be an underestimate of the actual number

of Oct4 binding sites in the mouse genome. Nonetheless, the ChIP-PET method was still able to detect 1083 high confidence Oct4 binding loci, expanding the catalogue more than 100 times the number of sites previously known.

4.2.4 Annotation of Oct4 binding sites to the transcriptome of ESCs

To identify genes that are potentially regulated by these 1083 Oct4-DNA interactions, these loci were annotated together with Dr Leonard Lipovich to determine the nearest known gene (Appendix A). The ChIP-PET data identified 957 distinct Oct4-bound genes. Subsequently, these Oct4 binding loci were also annotated for their relative position to the respective gene (Appendix A). For distances of less than 100 kb, bound loci were annotated as 5' distal (10 to 100 kb upstream), 5' proximal (0 to 10 kb upstream), intragenic (contained within introns or exons of the respective genes), 3' proximal (0-10 kb downstream), and 3' distal (10 to 100 kb downstream) (Figure 4.4A). Loci mapping greater than 100 kb away from the nearest gene were annotated as gene deserts although the nearest genes were still listed.

Results showed that about 44% of the Oct4 binding sites mapped within a gene, with 437 mapping to introns and 25 to exons (Figure 4.4B). 197 loci (19%) mapped to the 5' proximal region of genes (eg. *Oct4*) and of these, 82 (42%) were within 1 kb of the 5' end of the gene (this includes *Nanog* and *Fbx15*). The binding within the 1 kb proximal promoter window represented 7.9% of all Oct4-bound loci which indicates a significant enrichment for this immediate proximal promoter region (this is compared with 13% of Oct4-bound loci mapping to the 90 kb region classified as 5' distal). The number of Oct4 binding sites mapped to the downstream of genes was 79 in the 3' proximal (including *Sox2*) and 104 in the 3' distal regions. Of the 957 distinct genes with Oct4 binding sites, the vast majority contained a single Oct4 ChIP-PET cluster mapping to it. However, there were also a number of genes that contained multiple Oct4 binding sites. The highest number of binding sites associated with a single gene was 5. This was associated with an

uncharacterized transcript (*2610042L04Rik*). A second gene, *Igfbpl1*, had 4 Oct4 binding sites. Five genes were associated with three Oct4 binding sites and 67 genes had two Oct4 binding sites. The remainder (883 genes) had only one Oct4 binding site. This suggests that clustering of Oct4 binding sites is not a common mechanism of action for transcriptional regulation by Oct4. With the help of Dr Paul Robson, all 957 distinct genes associated with Oct4 binding sites were further annotated for functional relevance with the Panther classification system (Mi *et al.*, 2005). 714 of these genes were recognized in the Panther system but only 527 of these were identified with a molecular function. In total, there were 430 genes of unknown function in close proximity to the Oct4 binding sites. Understanding the association of Oct4 and these genes will unravel the functions of these novel genes in transcriptional regulation. Within the genes of known molecular functions, transcription factors were the most significantly over-represented. Within a random list of 527 genes generated from the 25358 genes characterized in the Panther system, 105 transcription factors such as *Trp53*, *Cdx2* and *Mycn* were indicated to associate with Oct4 binding sites (Table 4.1). Kinases are a second group of genes that were significantly over-represented. A gene that was identified as having an Oct4 ChIP-PET cluster overlap of 5 mapping to its first intron was *Akp2*. This is an alkaline phosphatase routinely used as a marker for both mouse and human pluripotent cells. This suggests that Oct4 is activating the transcription of *Akp2* in pluripotent cells.

The 1083 Oct4 binding sites were also annotated for indication of expression by relating expression profiling results from a massively parallel signature sequencing (MPSS) study (Wei *et al.*, 2005) (Appendix A). MPSS generates millions of 20 bp signature sequence tags per sample which are quantified as transcripts per million (tpm). The frequencies of *Oct4* transcript in this MPSS study were 388 tpm and 21 tpm in ESCs and EBs, respectively. Utilizing MPSS data for the genes that map closest to each Oct4 binding locus allowed the comparison of their expression levels between ESCs and EBs with that of Oct4 itself. Approximately 8% of the genes associated

with these Oct4-bound loci had greater than 100 tpm in undifferentiated ESCs. Surprisingly, the majority (507 genes or 52%) did not show expression in ESCs, as indicated by zero MPSS tags. From the relative abundance, measured by calculating the ratio of ES to EB tpm, 296 genes showed higher expression in ESCs than in EBs. This ES/EB ratio provides a correlation between pluripotent specific expression and Oct4 binding, and serves to imply potential activation, repression or non-regulation of these genes by Oct4. Annotation for all of the Oct4 binding sites are presented in the Appendix A.

4.2.5 Identification of novel Oct4 bound genes and associated pathways

The Oct4 ChIP-PET data uncovered a broad spectrum of genes previously not associated with Oct4 binding or regulation. These include *Esrrb*, *Rest*, *Jarid2*, *Tcf3*, *Tcf7*, *Mycn*, *Zic3*, *Rif1*, *Ehmt1*, *Tcfcp2l1*, *FoxH1*, *Id3*, *Trp53*, *Trp53BP1*, *Dido1*, *Tdgf1*, *Sgk*, *RARa*, *RARg*, *Bmi1* and *Zfp64* (Figure 4.5A). Importantly, microRNA (miRNA) genes (*mir302* and *mir296*) were also identified among the candidate targets of Oct4 (Figure 4.5B). Oct4 bound to sites within 10 kb of both miRNA genes. For all the protein coding and miRNA genes mentioned above, Oct4 binding was confirmed (*i.e.* enrichment of at least 2-fold) by traditional ChIP with detection by real-time PCR as compared to that of a control ChIP using an unrelated antibody (Figure 4.5A, C). To test the specificity of Oct4 binding in mESCs, Oct4 ChIP-qPCR was carried out using retinoic acid (RA) -induced differentiated mESCs. Fold enrichment for these targets decreased when mESCs were treated with retinoic acid. ChIP using a control antibody showed no significant enrichment (Figure 4.6). This further confirmed that the detected Oct4 binding at these target sites in undifferentiated mESCs was authentic and that upon differentiation, Oct4 binding was reduced.

The majority of genes identified in this analysis has not been identified in previous studies and thus represent novel targets of Oct4. These candidate target genes encode for transcription factors, chromatin remodellers, miRNA and other proteins. Oct4 was found to bind to genes

involved in the DNA damage response pathway (*Trp53*, *Trp53BP1*, *Rif1*). In addition, Oct4 also bound to genes important in cell proliferation (*Mycn*). Importantly, a cohort of Oct4 bound genes (Table 1) were those that encode for factors important in inhibiting transcription of lineage specific genes, such as the Rest transcription factor which is an inhibitor of neuronal genes (Ballas *et al.*, 2005). The repertoire of Oct4 target genes and the roles they play in different pathways may contribute to the ESC self-renewing and pluripotent potential.

4.3 Discussion

A total of 1083 high confidence Oct4 binding sites were identified using the ChIP-PET method. Apart from identifying known targets of Oct4 in ESCs, such as *Oct4*, *Sox2* and *Nanog*, a cohort of Oct4-associated genes involved in diverse functions that include transcription regulation, cell fate determination pathways, cellular proliferation, metabolism, and genetic integrity were also discovered. Transcription factors were found to be the most enriched subset of Oct4 binding sites (Table 4.1), suggesting that Oct4 may be one of the most important transcription factors in the hierarchy of the transcriptional regulatory network in mESC.

When transcripts of some of the Oct4 bound genes were tested in Oct4 RNAi-depleted mESCs, *Esrrb*, *Rest*, *Jarid2*, *Tcf3*, *Tcf7*, *Mycn*, *Zic3*, *Rif1*, *Ehmt1*, *Tcfcp2l1*, *FoxH1*, *Trp53*, *Trp53BP1*, *Dido1*, *Tdgf1*, *Sgk*, *RARa* and *RARg* transcripts were significantly downregulated after *Oct4* knockdown (Loh *et al.*, 2005), indicating that Oct4 binding on these targets confer expression. Interestingly, the trophectoderm (TE) marker genes *Cdx2* and *Cldn4* which were bound by Oct4, were dramatically upregulated upon *Oct4* reduction; demonstrating that Oct4 could also repress genes in ESCs (Chew *et al.* 2005). Concurringly, Oct4 was reported to directly prevent differentiation towards trophectoderm by interacting with Cdx2 to form a repressor complex (Niwa *et al.*, 2005). This complex interferes with the autoregulation of these two factors, giving rise to a reciprocal inhibition system that establishes their mutually exclusive expression.

As such, the downregulation of Oct4 results in an upregulation of Cdx2, and *vice versa* - a mechanism that might account for the two different pathways that lead to pluripotent stem cells and to trophectoderm cells. Besides *Cdx2*, other genes important for proper TE development were also bound by Oct4. An Oct4 binding site was found at 88 kb proximal to *Eomes*, a transcription factor essential for proper TE development (Russ *et al.*, 2000). In addition, *Cldn4*, which is expressed in trophoblastic stem cells and in the blastocyst, was also found to be bound by Oct4. This gene product is a major player in tight junction function and therefore implicated in TE epithelial integrity. This suggests that Oct4 also acts as a repressor of the trophoblast lineage.

There are also many other genes with Oct4 binding sites which show very low or no expression in mouse ESCs but are upregulated in differentiating embryoid bodies as determined by MPSS data-mining. A number of these genes are related to lineage development such as *Mid1*, *RbPms*, *Myo10*, *Cspg2*, *Spn2* and *Nestin*. This suggests that Oct4 may not only repress genes for trophectoderm differentiation, but it also represses genes for other lineages as well. For example, this data suggests that Oct4 indirectly inhibits neuronal lineages by positively regulating the expression of *Rest*. *Rest* encodes for a transcriptional repressor that restricts neuronal gene expression in neural progenitors. Ballas and coworkers demonstrated that in neural progenitor cells, Rest is degraded to levels just sufficient to repress transcription of neuronal genes. As progenitors differentiate into neurons, Rest and its co-repressors dissociate from the binding site, triggering activation of neuronal genes (Ballas *et al.*, 2005). The sustained activation of inhibitors of lineage genes such as *Rest* by Oct4 suggests another mechanism utilized by ESCs to remain undifferentiated whereby Oct4 may repress differentiation genes.

Oct4 also binds to genes important in proliferation and cell cycle such as *Mycn*, *Esrrb*, *Rif1*, *Trp53* and *Trp53bp1*. Mycn has been reported to be among the key mediators in the self-renewal and proliferation of ESCs (Cartwright *et al.*, 2005). Esrrb belongs to the superfamily of

nuclear hormone receptors, and homozygous mutant embryos show abnormal trophoblast proliferation, precocious differentiation toward the giant cell lineage and reduction in primordial germ cells (Mitsunaga *et al.*, 2004; Tremblay *et al.*, 2001a). Rif1 is an ortholog of a yeast telomeric protein and is upregulated in mESC and germ cells (Adams and McLaren, 2004). In human cells, Rif1 associates with dysfunctional telomeres and has a role in DNA damage response (Adams and McLaren, 2004; Silverman *et al.*, 2004; Xu and Blackburn, 2004). p53 has been suggested to maintain the genetic stability of ESCs by eliminating DNA damaged cells (Sabapathy *et al.*, 1997; Xu, 2005). p53BP1 is a p53 binding protein and functions as a key transducer of the DNA damage checkpoint signal and monitoring of dysfunctional telomeres (d'Adda *et al.*, 2003; DiTullio, Jr. *et al.*, 2002; Takai *et al.*, 2003; Wang *et al.*, 2002). Besides genes of the cell cycle, Oct4 also binds to *Id3*, *FoxH1* and *Tcf3* genes which are implicated in the BMP and Wnt signalling pathways (Hollnagel *et al.*, 1999; Hoodless *et al.*, 2001; Kofron *et al.*, 2004; Merrill *et al.*, 2004; Norris *et al.*, 2002; von, I *et al.*, 2004;). In addition, Oct4 binding was also identified at microRNA genes *mir296* and *mir302*. These microRNAs have been isolated in ESCs and their expression was shown by Northern blot analyses (Houbaviy *et al.*, 2003). This finding suggests a role for miRNA in early embryonic development. In addition, a substantial number of Oct4-bound genes are related to metabolic pathways including solute carrier proteins (Slcs) and enzymes related to ubiquitination pathways (ubiquitin conjugating enzymes and ubiquitin specific proteases). These may contribute to the high proliferation rate and metabolic turnover of ESCs.

In many instances, Oct4 bound to genomic sites with no apparent upstream or downstream genes within the range of 500 kb. These may be non-functional Oct4 sites, sites where novel transcription units which have not yet been defined or sites of enhancer regions acting at a distance due to the conformational structure of the chromatin. This demonstrates that the ChIP-PET method identified Oct4 binding sites in a global manner. This global manner of mapping binding sites is particularly important as regulatory elements do not always fall within the 5' proximal region of

the first exon (Cawley *et al.*, 2004). The location map generated in this study will serve as a useful guide in identifying additional components in the regulatory network important for self-renewal, pluripotency and differentiation of ESCs. Moreover, the data will provide clues for the selection of genes in ESCs that may be targeted for a directed differentiation of mESCs.

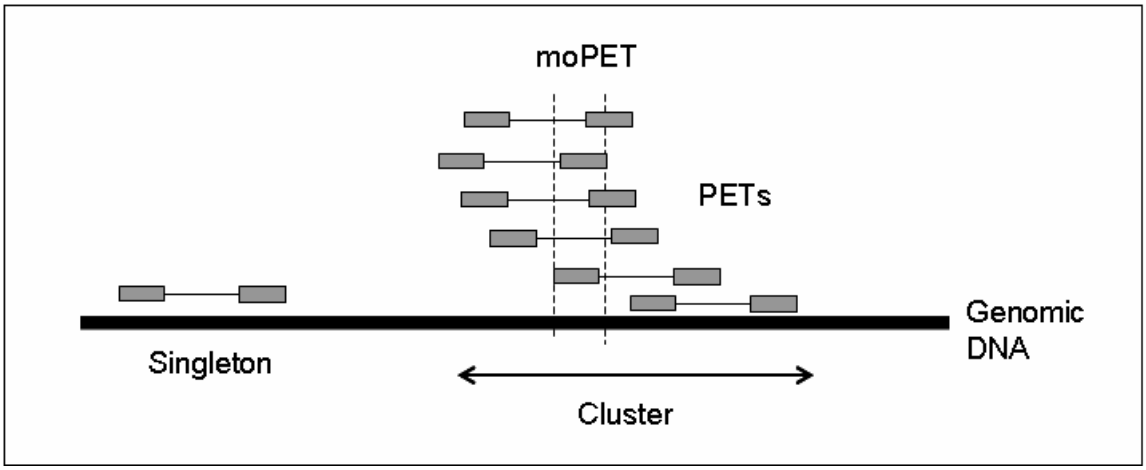


Figure 4.1: Diagram illustrating the features of the ChIP-PET readout. Paired-end ditags (PETs) mapped to the genome either appear as a single PET (termed singleton) which most likely represent background noise, or as multiple overlapping PETs forming a PET cluster demarcating the boundaries of the ChIP DNA fragments. Maximum overlapping PETs (moPET) are located at the region where the number of overlapping PETs is the highest.

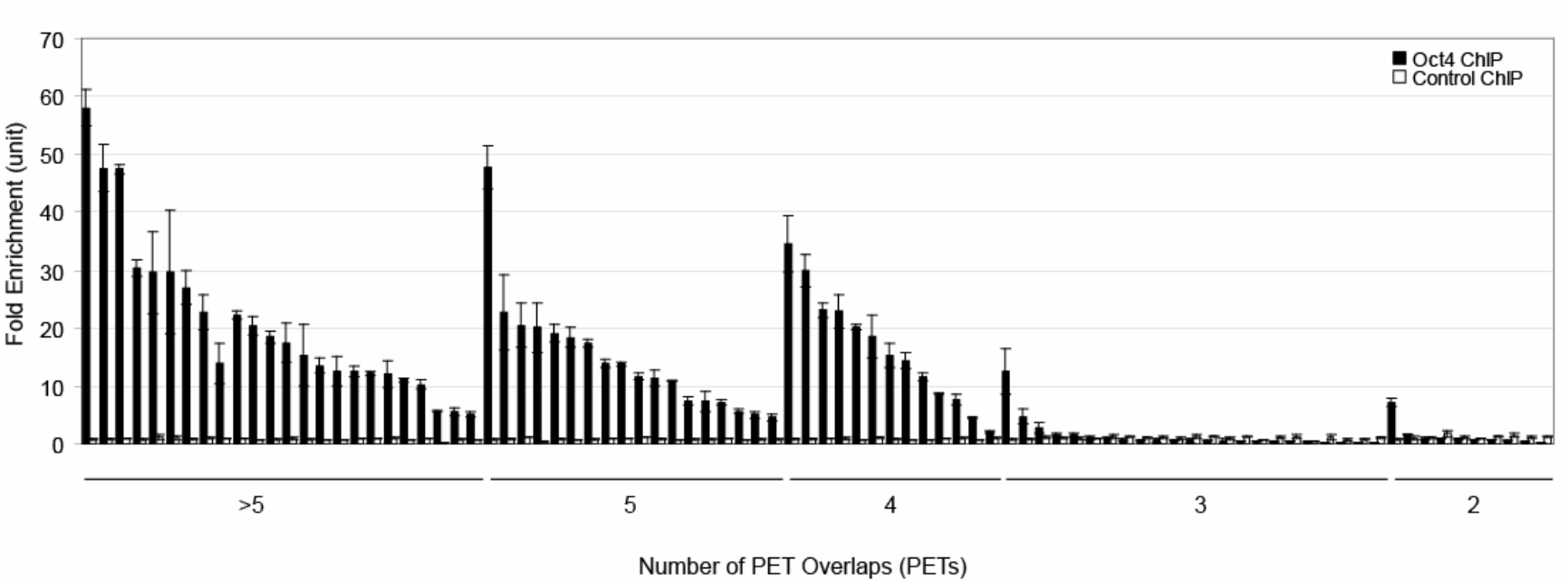
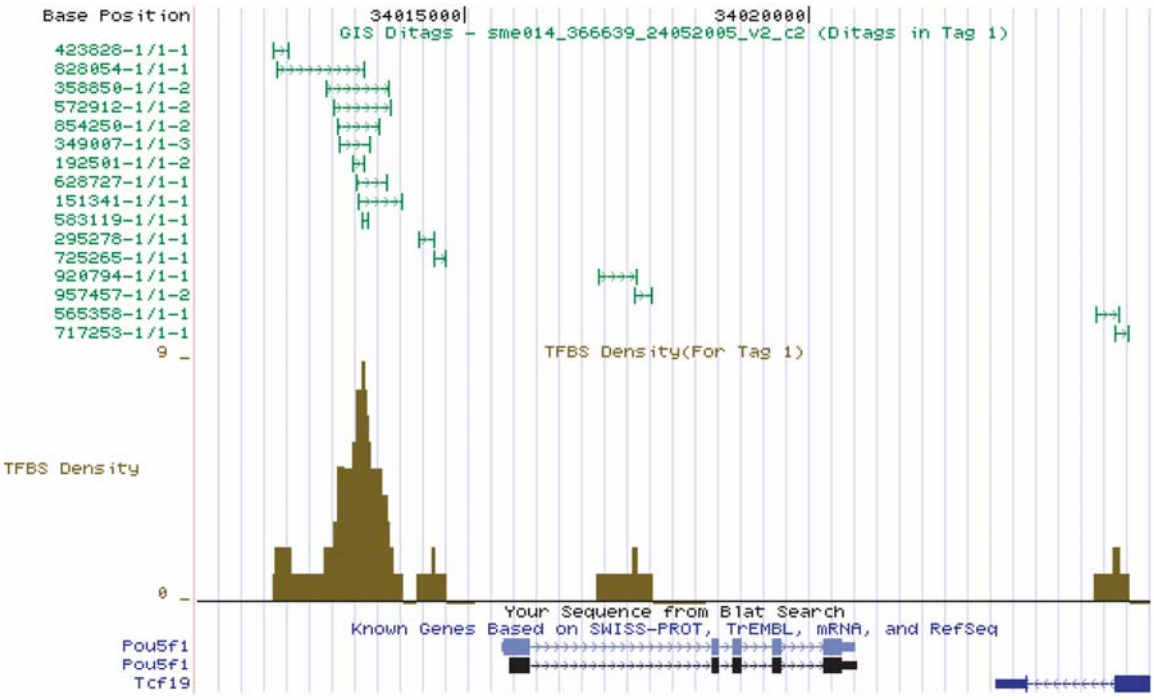
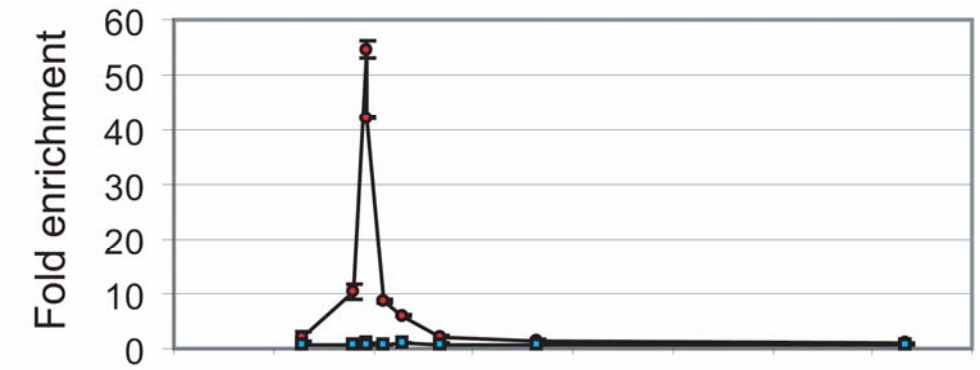


Figure 4.2: Determination of the minimum PET cluster size as Oct4 *bona fide* binding sites. Oct4 ChIP-qPCR enrichment at randomly selected regions containing different numbers of PETs are shown. Standard deviations and control ChIP are shown for all targets.

A





Oct4/ Pou5f1 gene
chr17:34,006,516-34,020,395

B

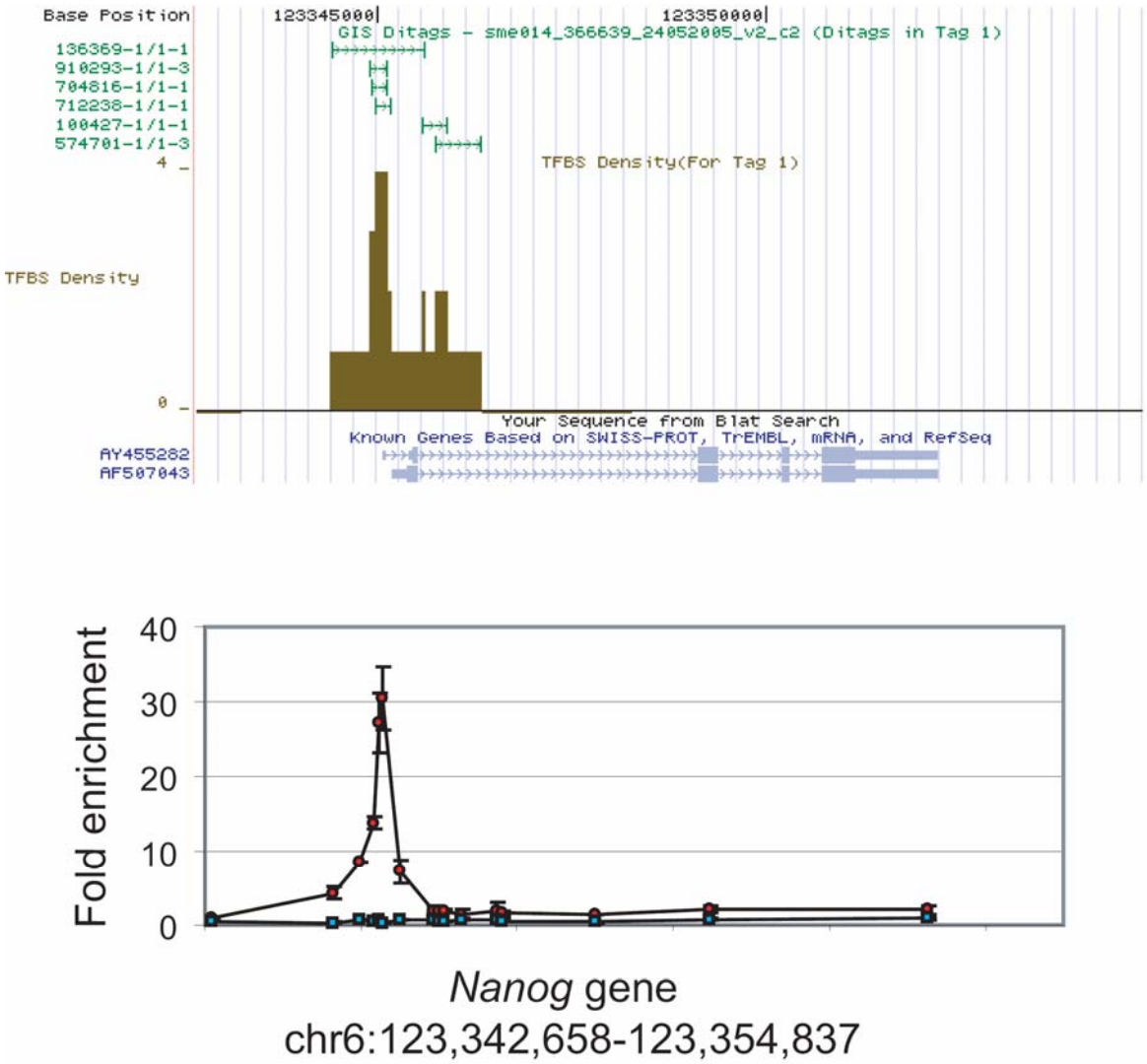
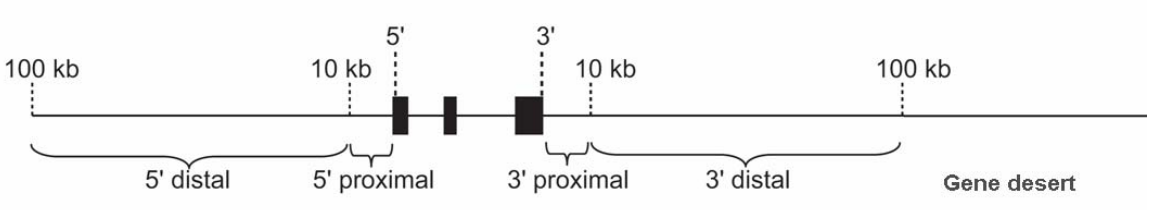


Figure 4.3: Profiles of Oct4 binding revealed by ChIP-PET. Validation of known Oct4 occupied genes in mESCs. An image of the T2G browser showing PET clusters at *Oct4/Pou5f1* (upper panel of A) and *Nanog* (upper panel of B) with 9 overlaps and 4 overlaps respectively. Each horizontal green line represents a DNA fragment mapped to the genome based on the UCSC mm5 build. TFBS density (coloured in brown) is a plot showing the profile of the transcription factor binding and is based on the number of overlaps of the DNA fragments. Conventional Oct4 ChIP-qPCR was carried out using *Oct4* and *Nanog* genes specific primers to confirm the Oct4 binding sites shown by the ChIP-PET analysis. Graphs depicting the mapping of the Oct4 (red circles) and control (Ena-1 antibody, light blue circles) binding sites across the chromosomal locations of PET clusters are shown (lower panels of A and B). Fold enrichment is the relative abundance of DNA fragments at the regions shown (relative to the position of the respective gene as illustrated in the browser) over a control region as quantified by realtime PCR.

A



B

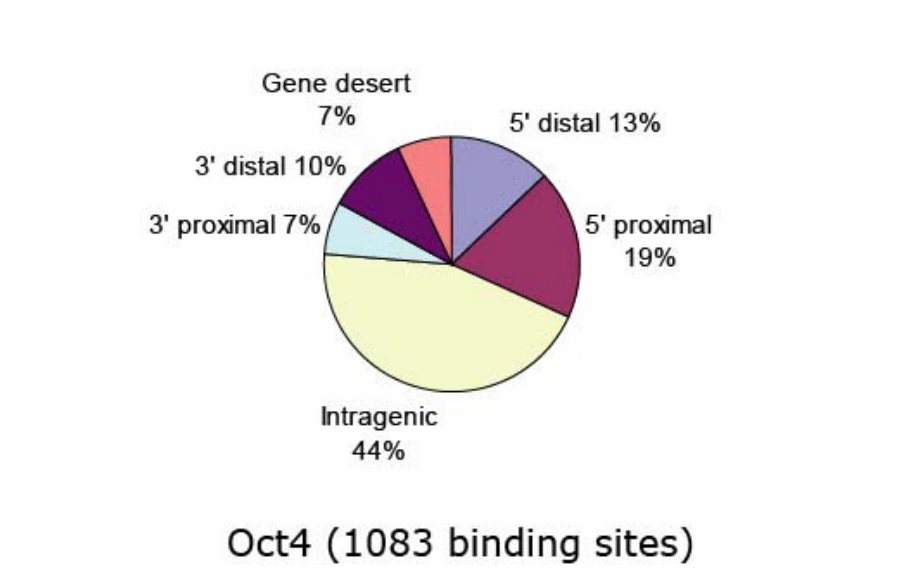


Figure 4.4: Annotation of Oct4 binding sites in relation to genomic locations. (A) Schematic diagram illustrating the definition of the location of a binding site in relation to a transcription unit. 5' distal, 5' proximal, 3' proximal and 3' distal regions are depicted within 100kb upstream and 100kb downstream of the transcriptional unit. (B) Locations of Oct4 binding sites relative to the nearest transcription units. The percentages of binding sites at the respective locations are shown.

Table 4.1: Molecular function of Oct4 target genes

	Total in	527 genes with known molecular function		
Molecular Function	Genome	expected	actual	Representative Gene
Transcription factors	2092	58.9	105	*Trp53, Esrrb, Jarid2, Edr1, Zic3, FoxH1*
Homeobox transcription factor	243	6.84	15	*Nanog, Pou5f1, Otx2, Cdx2*
Basic helix-loop-helix transcription factor	88	2.48	6	*Nmyc*
Zinc finger transcription factor	768	21.62	11	*Myst2, Mtf2, Sall1*
HMG box transcription factor	83	2.33	5	*Sox2, Tcf3*
Transcription cofactor	159	4.48	15	*Trp53bp1, Trim24*
Kinases	729	20.53	44	*Brd2, Yes1*
Tyrosine protein kinase receptor	72	2.03	9	*Tyro3*
Non-receptor Ser/Thr protein kinase	288	8.1	12	*Sgk, Mast2*
Kinase activator/modulator	138	3.88	5	*Tcl1*
G-protein modulator	317	8.93	16	*Rapgef1*
Cell adhesion molecule	501	14.11	12	*Clstn2*
Signaling molecule	799	22.5	21	*Spry4, Cd63*
Membrane-bound signaling molecule	129	3.63	9	*Dll1, Jag1*
Other receptor	310	8.73	18	*Ptch1*
Other membrane traffic protein	79	2.22	5	*Sec61g*
Cytokine receptor	100	2.82	3	*Il6st, Lifr, Cntfr*
Ligase	232	6.53	10	*Sars1*
Glycosidase	47	1.32	3	*Hexb*
Esterase	93	2.79	7	*Sulf1*
Glycosyltransferase	206	5.8	10	*Ext1*
Phosphorylase	10	0.28	2	*Upp1*
Phosphatase	276	7.77	10	*Ppp2r2d, Akp2*
Reductase	105	3.15	5	*Akr1b8*
Annexin	106	2.98	5	*Prkcb1*
Chaperone	157	4.71	4	*Cdc37l1*
G-protein coupled receptor	1827	51.44	16	*Bai1*
Ribosomal protein	556	15.66	0	-
Structural protein	146	4.11	3	*Col18a1*
Protease inhibitor	180	5.07	1	*Itih3*
TGF-beta superfamily member	34	0.96	3	*Lefty1*

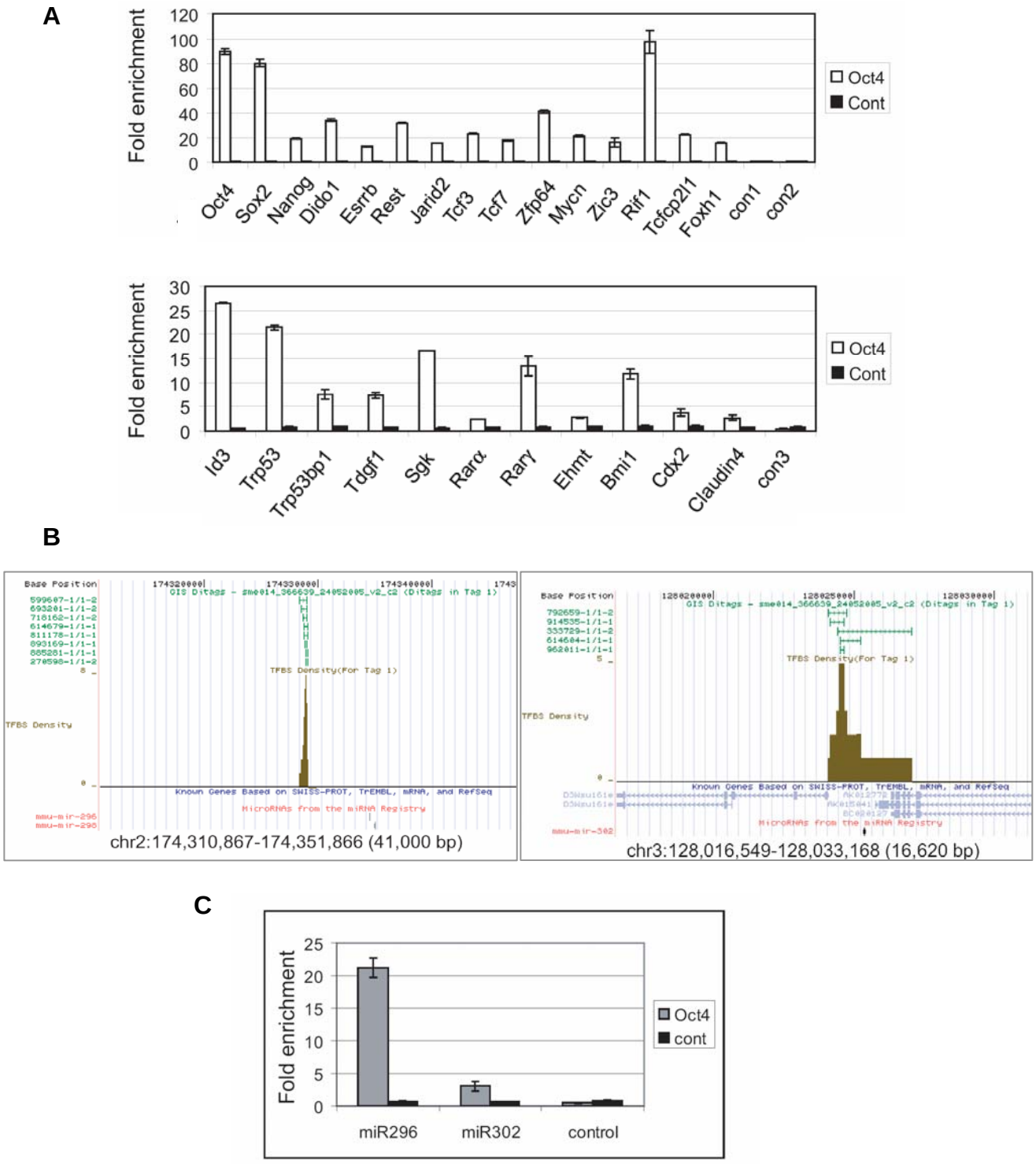


Figure 4.5: Novel genes bound and potentially regulated by Oct4 in mESCs. (A) Novel and known Oct4-bound targets were validated by conventional ChIP-qPCR using Oct4 and control (Ena-1) antibodies. Con1 to con3 are control regions. (B) Image of the T2G browser illustrating the binding of Oct4 on *miR296* and *miR302*. (C) Verification of Oct4 occupancy by conventional ChIP-qPCR. ChIP using Oct4 (N-19) and control Ena-1 antibodies were carried out to validate Oct4 occupancy on *miR296* and *miR302*.

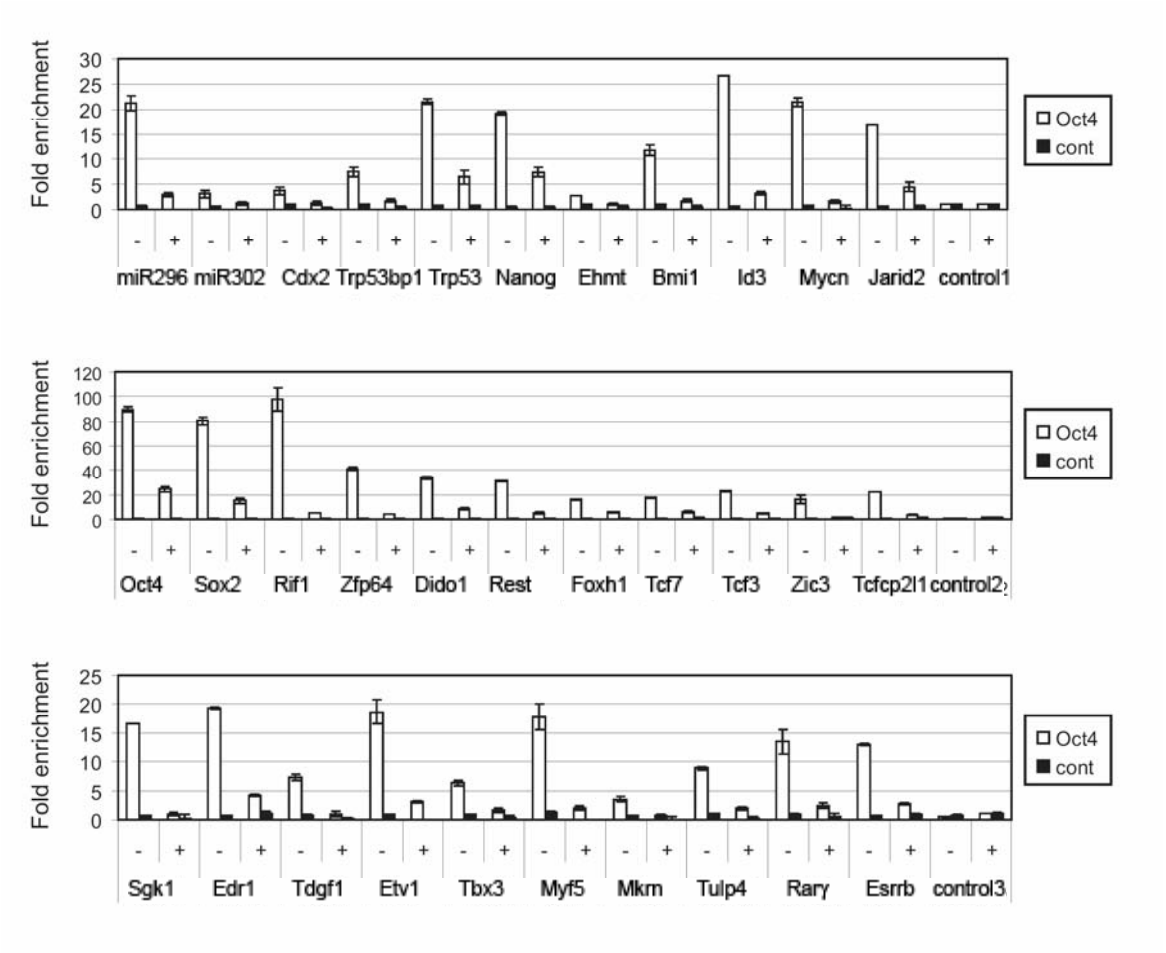


Figure 4.6: Retinoic-acid induced differentiation reduces Oct4 binding levels on targets in mESCs. Oct4 (N-19) and control (Ena-1) ChIP were carried out using formaldehyde-crosslinked mESCs treated with retinoic acid (+) and non-treated ESCs (-). Fold enrichment represents the abundance of enriched DNA fragments over a control region not enriched for the respective targets. Standard deviations representing technical replicates are shown.

CHAPTER 5

Genome-wide Identification of Sox2-DNA Interactions in Mouse Embryonic Stem Cells

5.1 Introduction

Similar to the previous chapter, the Sox2 genome-wide binding sites in mESCs were also mapped using the ChIP-PET approach. The essential function of Sox2 in ESCs is mediated through the genes that it transcriptionally regulates; several which have been identified. In Chapter 3, Sox2 was shown to bind to the regulatory regions of *Sox2* itself, *Oct4* and *Nanog*. Other reports also corroborated with this finding (Boyer *et al.*, 2005; Kuroda *et al.*, 2005; Okumura-Nakanishi *et al.*, 2005). The autoregulatory loop and feed-forward mechanisms of Sox2 action implicated its function in the regulation of these three transcription factors that are essential to the pluripotent phenotype. In addition, Sox2 is known to bind regulatory regions of *Fgf4*, *Utf1* and *Fbxo15* (Ambrosetti *et al.*, 1997; Botquin *et al.*, 1998; Nishimoto *et al.*, 1999; Yuan *et al.*, 1995) to drive their pluripotent-specific expression. In all these six known target genes, Sox2 binds synergistically with Oct4 to neighbouring sox and octamer elements, implying that Sox2 may also bind to many of Oct4 target genes. To investigate this possibility, the genome-wide binding sites of Sox2 in mESCs (and the genes associated with them) were identified to establish the extent in which these binding sites overlap with those of Oct4.

In this study, 1133 high confidence Sox2 binding sites and numerous other lower affinity sites were identified using the ChIP-PET method. Of these high affinity sites, 49% are located within 20 kb of a gene whereas a surprisingly high number (15%) are more than 100 kb away. A number of sites mapped to the proximal promoters of uncharacterized transcripts. As with Oct4, transcription factors such as *FoxD3*, *Mycn*, *Rif1*, *Rest*, and *Tcf3*, were the most significantly represented targets of Sox2 binding. Novel binding sites such as those from the *Oct4*-bound loci, which was not detected in the gene-by-gene examination in Chapter 3, was identified here. Results

show that Sox2 binding sites associate with a diverse array of genes encoding transcription factors, chromatin remodelling factors, proteins involved in cell cycle control, cytokine signalling and miRNAs.

5.2 Results

5.2.1 Optimisation of ChIP and global mapping of Sox2 binding sites in mESCs

Single-step Sox2 ChIP did not yield high enrichments. Therefore, a two step ChIP using the Sox2 antibody was carried out to obtain good quality ChIP DNA for cloning and sequencing. Since sequential ChIP only gave approximately 5 ng of DNA per ChIP reaction, multiple reactions were carried out to acquire enough DNA (at least 500 ng for ChIP-PET and minimum validations). Purified Sox2 ChIP-enriched DNA was cloned into a primary library that preserved the original representation of the ChIP-enriched DNA by the GIS Cloning and Sequencing group led by Dr Ruan Yijun. As with the Oct4 ChIP-PET process, sequence tags (18 bp) from both ends of each cloned DNA fragment were generated and ligated together to form a paired-end ditag (PET). Subsequently, these PETs were concatenated and cloned into a second library in which each clone contained 5-6 PETs in tandem. Clones from this library were randomly picked and single-pass sequenced by the Cloning and Sequencing group, GIS.

A total of 115,443 clones were sequenced from the Sox2 ChIP-PET library, generating 616,299 PETs. These PET sequences were mapped to the mouse genome (build mm5) to demarcate Sox2 binding sites. Of these PETs, 50.85% (313,418) were successfully mapped, 40.68% did not match any sequence in the genome, and 8.46% hit repetitive sequences and was therefore omitted. Multiple identical PETs are most likely the result of library amplifications of one original genomic fragment. 203,987 unique PETs were successfully mapped to the mouse

genome. The majority of these mapped genomic loci (91.8%) were represented by single non-overlapping PETs (singletons) and considered to be background.

Out of the 16,729 non-singleton loci, the number of PET overlaps ranged from those containing two overlaps (PET2) (13,798) to a single cluster containing twenty six (PET26). Those clusters with a lower number of overlapping PETs most likely contained Sox2 binding sites of a lower affinity and therefore, such DNA fragments were pulled down less frequently than those sites of greater affinity for Sox2. To determine the minimum number of PET overlaps (cutoff) that definitively identified a *bona fide* Sox2 binding site, 25 PET3 and 29 PET4 and above clusters were randomly selected to test for enrichment using conventional ChIP-qPCR (Figure 5.1). The majority of PET3 clusters produced no more than a 2-fold enrichment whereas all PET4 and above clusters showed significant enrichment (*i.e.* greater than 2-fold) over background for Sox2 binding (Figure 5.1). Sampling of clusters with 5 or more PETs (PET5+) indicated that the number of PETs within a cluster was generally proportional to the fold-enrichment as detected by ChIP-qPCR.

Clusters with 4 and more overlapping PETs (PET4+) were taken to be high-confidence Sox2 binding sites. However, there is likely to be a significant number of true Sox2 binding loci in clusters containing PET2 and 3 as 32% of the PET3 clusters tested by ChIP-PCR had significant enrichment. From this percentage, it is assumed that 575 out of the 1799 total PET3 clusters are likely to be Sox2 binding sites. Additional binding sites may also be represented by some of the 13,798 PET2 clusters. From the set of high-confidence Sox2 binding sites represented by PET4+ clusters, of which 83% fall within the PET4-6 range, 1133 distinct Sox2 binding sites were identified in the mESC genome, which is a substantial increase over the 8 previously known Sox2 binding sites.

5.2.2 Sox2 ChIP-PET experiment identifies known Sox2 targets

To investigate the reliability of the Sox2 ChIP-PET data, the data were analysed for known Sox2 targets. In Chapter 3, *in vivo* binding of Sox2 was demonstrated on *Oct4*, *Nanog* and *Sox2* itself (Chew *et al.*, 2005; Rodda *et al.*, 2005) and these loci were detected in the Sox2 ChIP-PET data. For the *Oct4* loci, a cluster containing 11 PET overlaps mapped to the known Sox2 binding site located 1977 bp upstream of the transcription start site (TSS). Interestingly, a second Sox2 binding site was detected 862 bp downstream of this first site in the CR2 region and is represented by a cluster of 5 PET overlaps (Figure 5.2A). Conventional ChIP-qPCR using primer sets scanning this entire region reiterated a similar profile as detected by the ChIP-PET data (Figure 5.2B), demonstrating that the ChIP-PET method is more efficient and comprehensive in identifying binding sites in a global manner. At the known Sox2 binding sites of *Sox2* and *Nanog*, there were 5 and 3 Sox2 PET overlaps, respectively (Figure 5.3A, B). These regions have previously been shown to bind Sox2 by ChIP-qPCR in Chapter 3. Hence, the identification of these 3 known binding sites confirmed, to some extent, the reliability of the ChIP-PET data that contained Sox2 binding targets.

5.2.3 Linking the Sox2 binding sites to the transcriptome of mESCs

As with the Oct4 ChIP-PET data, the 1133 Sox2 binding sites were annotated to the nearest genes that they may potentially regulate (with the help of attachment students, Amanda Ang and Yong Chun). 957 distinct Sox2 associated genes were identified. Binding sites were then categorized by their location to their respective genes (Figure 5.4A). Surprisingly, a significant number of Sox2 binding loci were further than 100 kb away from the nearest gene. However, the majority of sites fall within 20 kb of the nearest gene (Figure 5.4B). These results signify the global nature of the ChIP-PET approach based on its ability to detect Sox2 binding irrespective of gene location.

With the help of Dr Paul Robson, the genes were classified according to the Panther classification system (Mi *et al.*, 2005) to identify their corresponding molecular function (Appendix B). 550 of 717 genes have recognizable molecular functions. Similar to Oct4, the most significantly over-represented molecular function (with $p=2.99\text{x}10^{-10}$) in the Sox2 binding dataset was transcription factors which include *Otx2*, *Rarb*, *Tcf3*, *Phc1*, *JunB*, *Mycn*, *Sall4*, *Id2*, *Id4*, *Ets1*, *Smad7*, *Gbx2*, *Sp1*, *Dido1*, and *Lef1*.

To investigate if the Sox2-associated genes are transcriptionally active or inactive in mESCs, expression data from massively parallel signature sequencing (MPSS) was integrated into the Sox2 ChIP-PET analysis annotation. The Sox2 annotations were carried out together with attachment students, Yong Chun and Amanda Ang (GIS). MPSS generates millions of 20 bp sequence tags per sample and these are expressed as transcripts per million (tpm). Data from the previously profiled transcriptomes of mESC were used (Wei *et al.*, 2005). *Sox2*, *Oct4* and *Nanog* transcripts have tpm counts of 105, 388, and 112 in mESCs, respectively. 67.4% of the Sox2 associated genes had 0-3 tpm in ESCs (Appendix B). Only the MPSS tags with a single genome hit were included in the computation of the percentage of Sox2 associated genes that are not expressed in ESCs to ensure specificity. A large number of Sox2-associated genes appear not to be expressed in ESCs, suggesting that Sox2 is negatively regulating these genes. There were 54 genes with tpm of 100 or more in ESCs, suggesting that these relatively abundant genes were positively regulated by Sox2. Genes that fall into this category include *Rest*, *Rcor2*, *Otx2*, *Rif1*, *Lefty1*, *Lefty2*, *Dppa5*, *Spp1* and *Esrrb*.

In addition to the above mentioned genes, other classes of protein-coding genes that were identified proximal to Sox2 binding include chromatin remodellers (*Pcaf*), phospholipases (*Pla2g1b*), phosphatases (*Ppm1b*), kinases (*Stk38*), cell surface receptors (*Notch1*) and cell adhesion molecules (*Ncam2*). Pathways which were implicated to involve Sox2 regulation

included cell proliferation (*Mycn*), apoptosis (*Cradd*), ubiquitination (*Uble1a*), FGF (*Fgf1*) and Wnt (*Lef1*) signalling pathways. Besides protein coding genes, Sox2 binding sites were also located proximally to genes encoding microRNAs (miRNAs) (Figure 5.5).

5.2.4 Comparative location analyses of Sox2 and Oct4

In Chapter 4, the ChIP-PET method was used to map the DNA binding sites of Oct4 in mESCs. Interestingly, many of Sox2 binding sites identified in this study are located near the Oct4 binding sites identified in Chapter 4. Among the genes which had DNA regions demonstrating Oct4 and Sox2 binding are *Rest*, *Mycn*, *Oct4*, *Sox2* and *Esrrb* (Figure 5.6).

5.3 Discussion

In this study, 1133 high-affinity Sox2 binding sites (PET4+) were identified in the mESC genome. This is likely an under-representation of all Sox2 binding sites in ESCs as there might be false negative targets that have been omitted amidst the noise in overlaps of 2 or 3 PETs. An example of a true target in this lower range of PET overlaps is *Nanog*, a known Sox2 binding target (Chapter 3; Rodda *et al.*, 2005) that is represented by a 3 PET overlap in this study. Nonetheless, this expanded list of Sox2 binding sites represented a more than 100-fold expansion of the previously known targets of Sox2. This expanded list would be a good starting point to elucidate the mechanism of Sox2 action in its role in maintaining the pluripotent cell.

The distribution of Sox2 binding sites showed a preference for proximity to genes with a substantial number of these sites falling within 20 kb of a known gene. The actual number of Sox2-bound genes may be even higher as many more genes would be identified upon the completion of the fully annotated mouse genome. Although the 5' proximal regions of many genes are enriched for Sox2 binding, there were more Sox2 binding sites within intragenic regions of genes. However, not all Sox2 binding sites are located proximally to genes, as some of them were

also found to be greater than 100 kb away from the nearest gene (gene desert). This suggests that a *cis*-regulatory module associating with Sox2 can function over large linear genomic distances to regulate gene expression.

In addition, a subset of the Sox2 binding sites was found proximal to genes of unknown function. An example of such a gene is 1700019D03Rik (PET cluster ID chr1.53325798). Indeed, there were 5 and 9 PET sequence overlaps in Sox2 and Oct4 PET clusters, respectively within the 1 kb proximal promoter region of this full-length transcript. Such associations between these transcription factor binding sites on genes of unknown function should aid in the functional characterizations of these genes. These associations as well as those binding sites associated with gene deserts highlight the advantage of using a global approach for transcription factor binding site discovery.

Besides *Oct4*, *Sox2* and *Nanog*, *FoxD3* which encodes another transcription factor essential in the maintenance of the pluripotent phenotype in ESCs, was also found to have a Sox2 binding site with PET4 located at the 5' region (Hanna *et al.*, 2002). This indicates that Sox2 is involved in the regulation of these transcription factors. In addition, a large cohort of Sox2-associated genes has been implicated in many ESC functions such as self-renewal and pluripotency. These include those in the Wnt/β-catenin/Tcf signalling pathway (*eg. Lef1*, *Tcf3*, *Tcf7*, and *Ctnnbl1*) (Reya and Clevers, 2005). In addition, several genes associated with Sox2 binding have been established to be functionally important in early development (*i.e.* in the egg cylinder) and include those that have a role in pattern formation (*Nodal*, *Lefty1*, *Lefty2*) and members of the fragilis family (*Ifitm1*, *Ifitm2*, *Ifitm6*, and *Ifitm7*) which are thought to be involved in germ cell specification (Saitou *et al.*, 2002).

Sox2 was also shown to bind genes encoding for miRNAs. Studies have indicated that miRNAs have important roles in gene regulation as more than a third of mammalian protein-encoding genes are conserved miRNA targets (Bartel, 2004; Lewis *et al.*, 2005). Moreover, ESCs that lack the machinery that processes miRNA transcripts are unable to differentiate (Kanellopoulou *et al.*, 2005).

Sox2 binding was identified at genes that play important roles in cell cycle and proliferation. A PET7 Sox2 binding site located 2 kb 5' to *Mycn* was identified in this study. *Mycn* is functionally equivalent to *Myc* in ESCs which possesses a proliferative function (Malynn *et al.*, 2000). In addition, a binding site located 592 bp 5' to *Rif1* was identified as one of the most highly represented Sox2 binding loci (PET16). Oct4 was also previously found to associate highly with this gene, with an Oct4 PET overlap of 17. Rif1 has been implicated in the DNA damage response at the intra S-phase checkpoint (Silverman *et al.*, 2004). Other Sox2-associated genes involved in maintaining DNA integrity either through DNA repair or cell cycle control included *Msh6*, *Pms1*, *Xpa*, *Gadd45g*, *Cdc37l1*, *Ches1*, *Cdk6*, *Syk38*, and *Terf1*. *Terf1* has been shown to be essential for development beyond the peri-implantation stage (Karlseder *et al.*, 2003). Earlier data had singled out *Trp53* and *Trp53bp1* as Oct4 target genes. Together, this indicates that Sox2 and Oct4 are involved in DNA-damage response ensuring only DNA of high integrity is passed on from the pluripotent cells to the somatic cell lineages and to the next generation.

As with Oct4, several Sox2 targets have been identified to be involved in the regulation of ESC differentiation and commitment. One example is the Rest transcriptional repressor that acts at the terminal stage of the neuronal differentiation pathway and blocks the transcription of several differentiation genes. Ballas and coworkers (2005) demonstrated that regulation of the Rest transcriptional repressor plays a fundamental role in the progression of pluripotent cells to lineage-restricted neural progenitors. Sox2 (PET14) and Oct4 (PET5) were found to bind within a 100 bp

of each other at a genomic region located 3.1 kb 5' to *Rest*. In addition, Oct4 and Sox2 also bind to the Rest corepressor *Rcor2*.

Sox2 binding was also identified on genes encoding chromatin remodeling factors such as PCAF, Ehmt1, Phc1 and Nsd1. p300/CBP-associating factor (PCAF), a ligand-dependent coactivator, belongs to the histone acetyltransferase (HAT) group of enzymes that are critical in the regulation of chromatin structure and gene expression (Santos-Rosa *et al.*, 2003). The polyhomeotic-like 1 (Phc1)/Rae28-like protein helps in maintaining transcription states upon initiation by generating heritable higher-order chromatin structures. Less is known of the euchromatic histone methyltransferase 1 (Ehmt1) but it has been suggested to play a role in histone modification (Okazaki *et al.*, 2002). Nsd1 has been found to be essential for early post-implantation development and possesses intrinsic methyltransferase activity (Rayasam *et al.*, 2003). Thus, Sox2 is implicated in the control of numerous proteins involved in chromatin remodeling.

From this study, it is evident that many of the Sox2 binding sites co-localise with those of Oct4, which was previously described in Chapter 4. It will be noteworthy to examine the extent in which Sox2 interacts with Oct4 on DNA in ESCs and to characterize the Oct4 and Sox2 binding elements that are crucial in mediating the binding of these transcription factors to DNA.

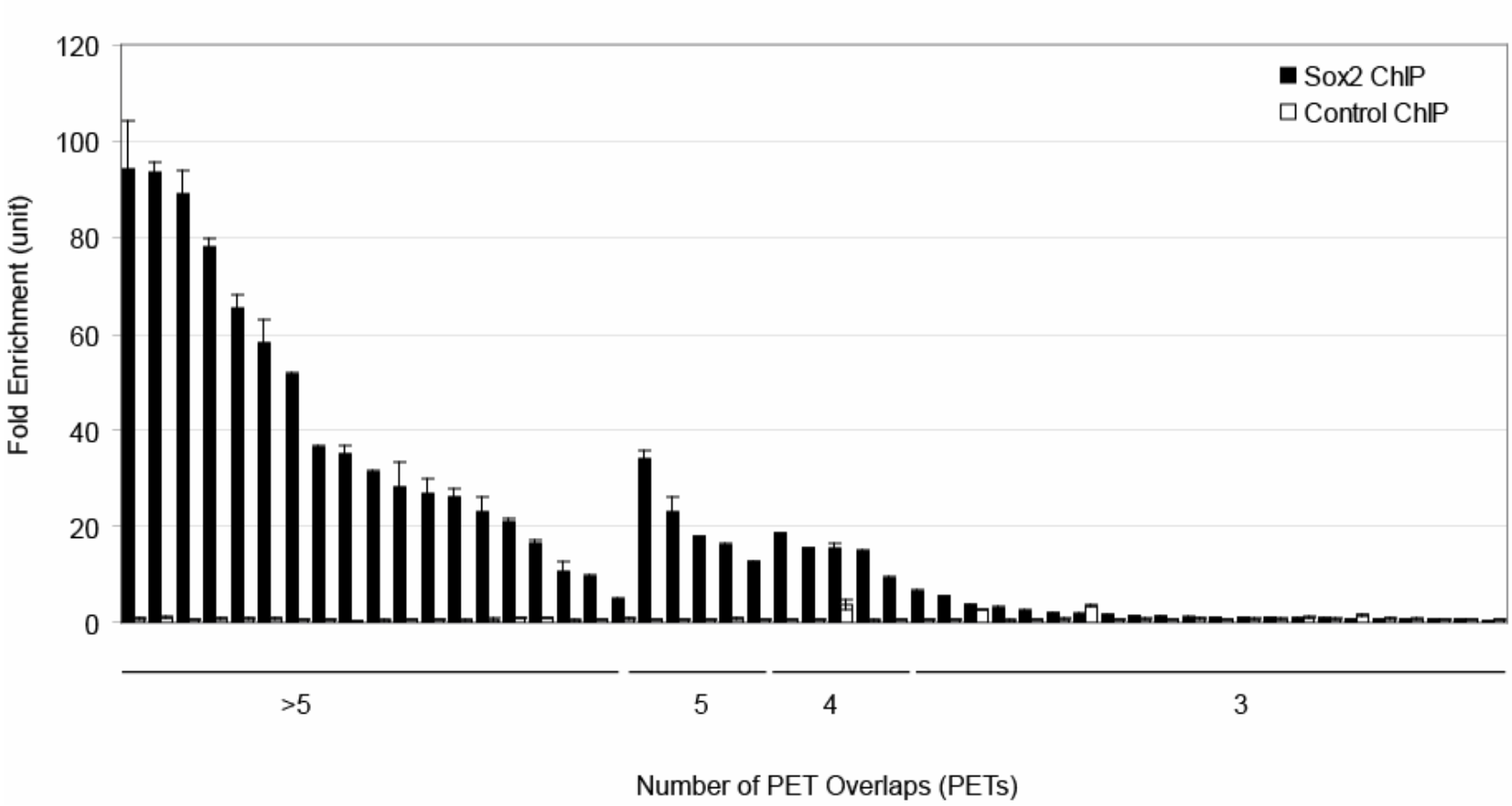


Figure 5.1: Determination of PET cluster size as Sox2 *bona fide* binding sites. Sox2 ChIP-qPCR enrichment at randomly selected regions containing different numbers of PETs are shown. Standard deviation represents technical replicates.

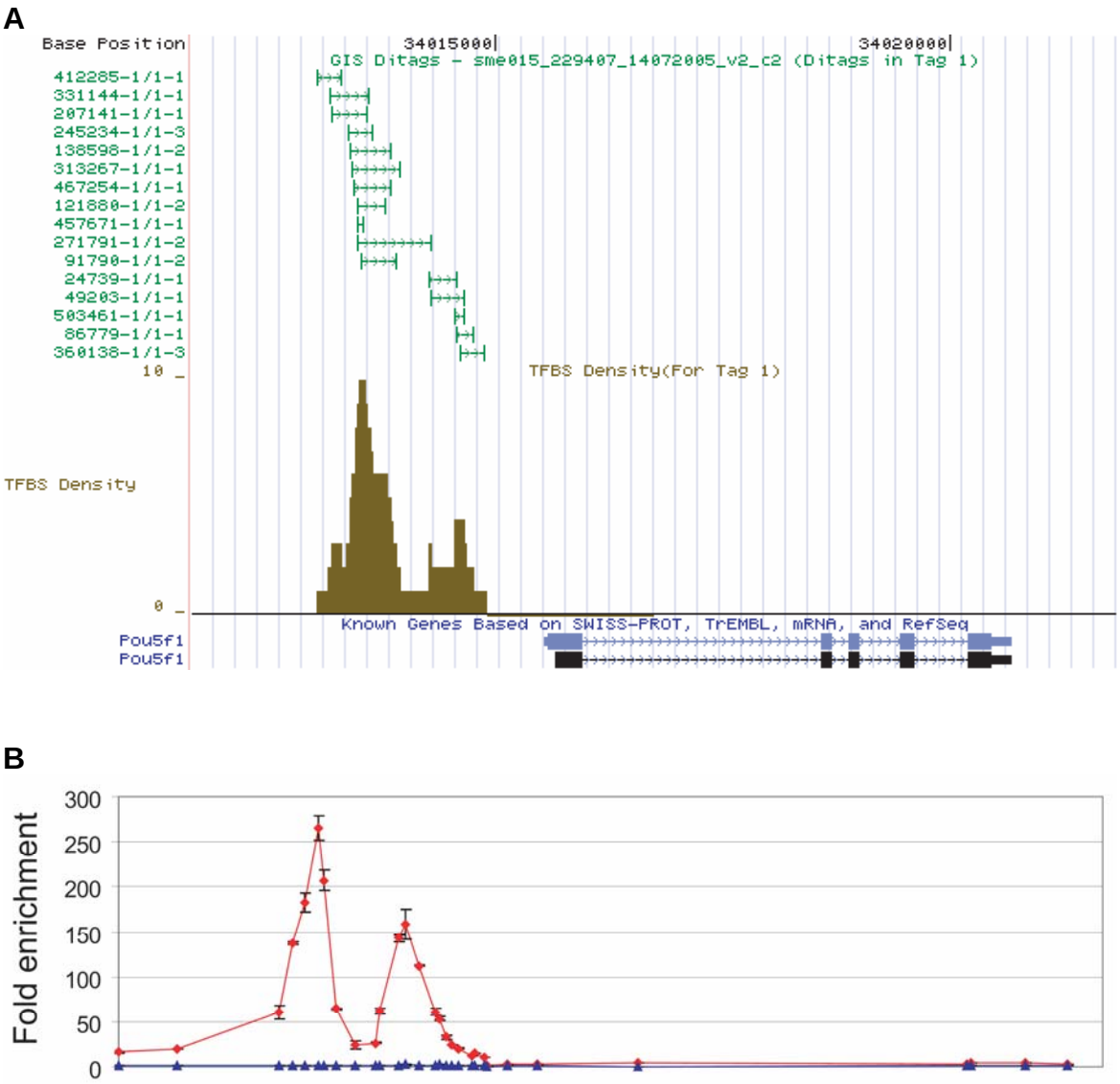


Figure 5.2: Validation of Sox2 binding profiles at the *Oct4* (*Pou5f1*) upstream regulatory regions. (A) An image capturing the T2G browser showing PET clusters at *Oct4*. Each horizontal green line represents a DNA fragment mapped to the genome. TFBS density (coloured in brown) shows the profile of the transcription factor binding and is based on the number of overlaps of the DNA fragments. (B) Conventional ChIP-qPCR was carried out using Oct4 gene primers to confirm the Sox2 binding sites shown by the ChIP-PET analysis. The locations of the amplified products of the primer sets used to detect the ChIP-enriched fragments are shown in the context of the genomic structure of *Oct4*. Fold enrichment is the relative abundance of DNA fragments at the regions shown over a control region (red plots represent Sox2 ChIP and blue plots represent control ChIP).

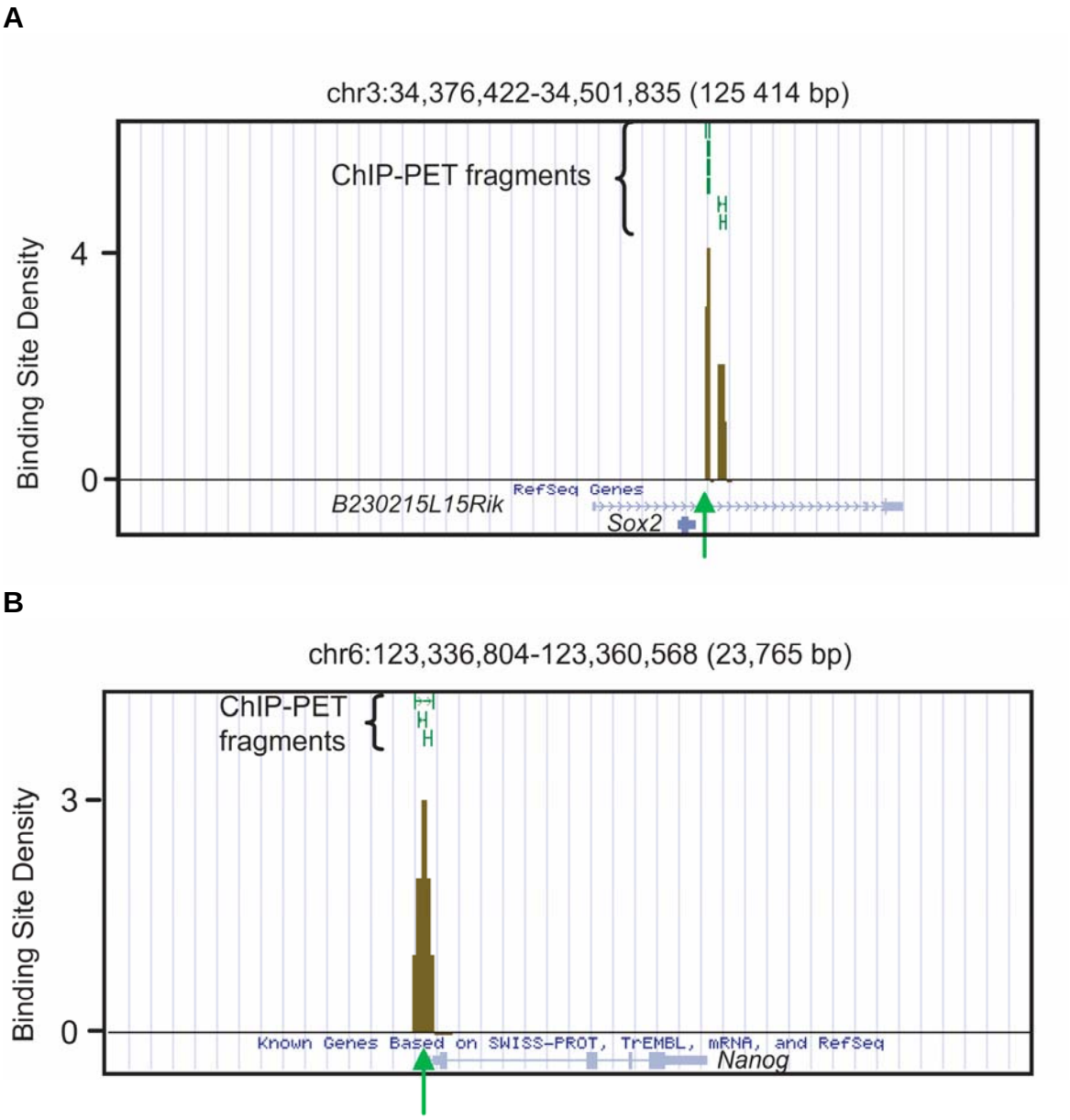
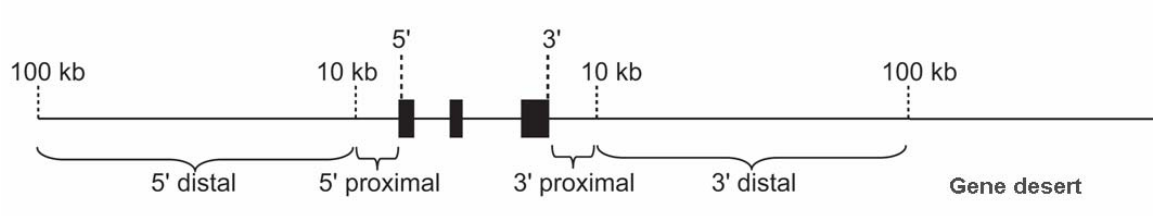


Figure 5.3: Sox2 binding on *Sox2* and *Nanog* from the ChIP-PET data. Images from the T2G browser showing Sox2 PETs at the previously known and validated regions of (A) *Sox2* and (B) *Nanog*.

A



B

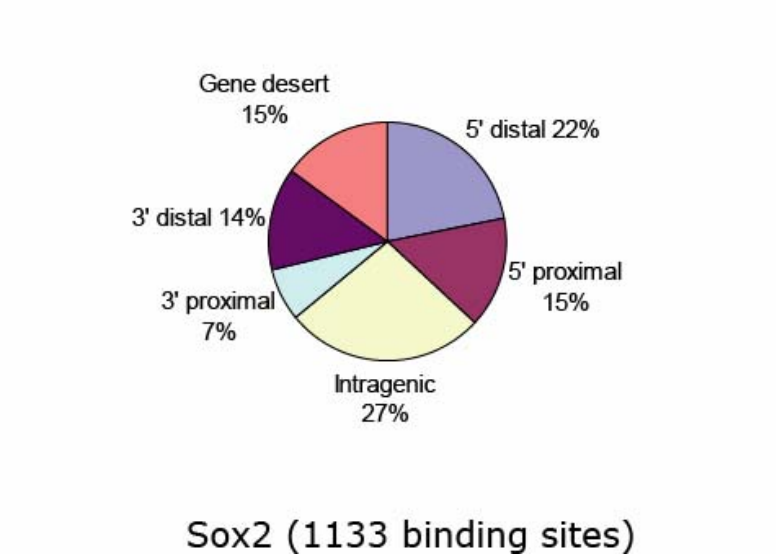
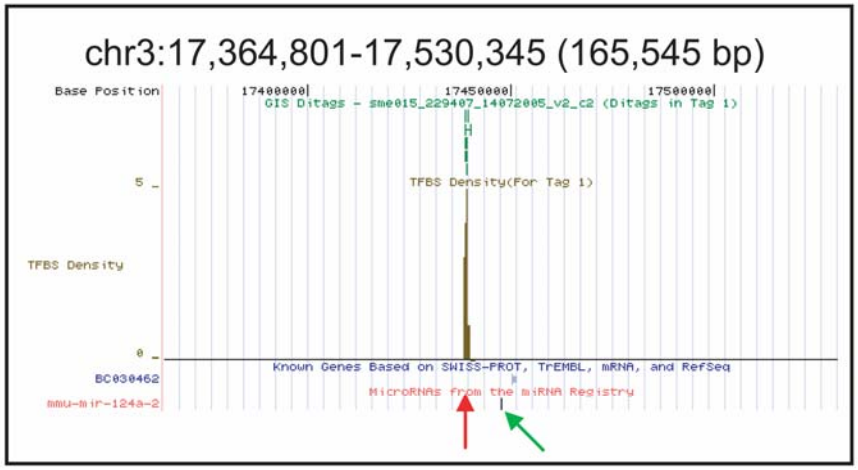
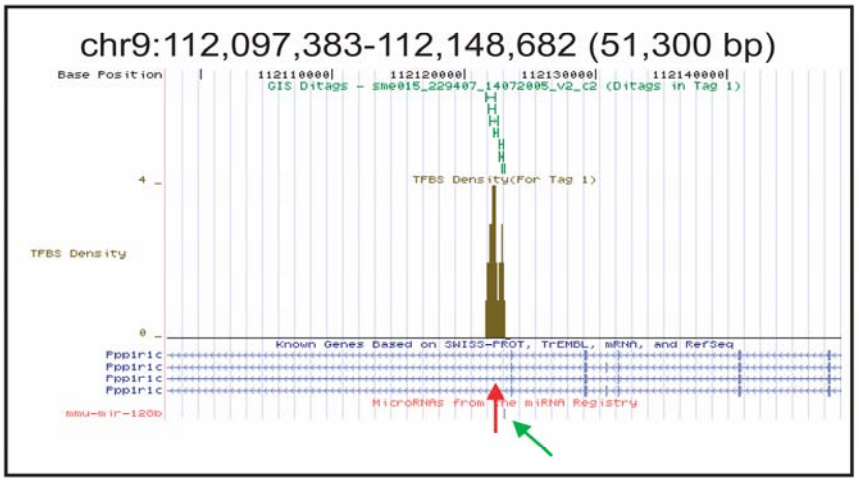


Figure 5.4: Annotation of Sox2 binding sites in relation to genomic locations. (A) Schematic diagram illustrating the definition of the location of a binding site in relation to a transcription unit. 5' distal, 5' proximal, 3' proximal and 3' distal regions are depicted within 100kb upstream and 100kb downstream of the transcriptional unit. (B) Locations of Sox2 binding sites relative to the nearest transcription units. The percentages of binding sites at the respective locations are shown.

A



B



C

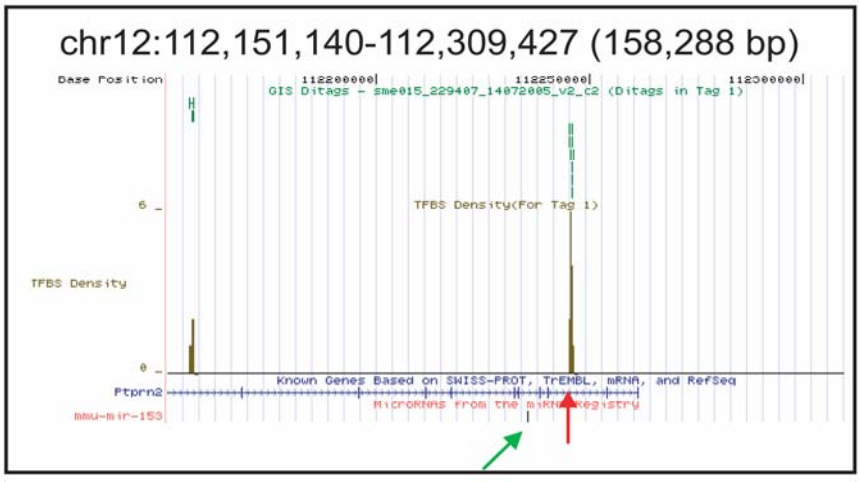
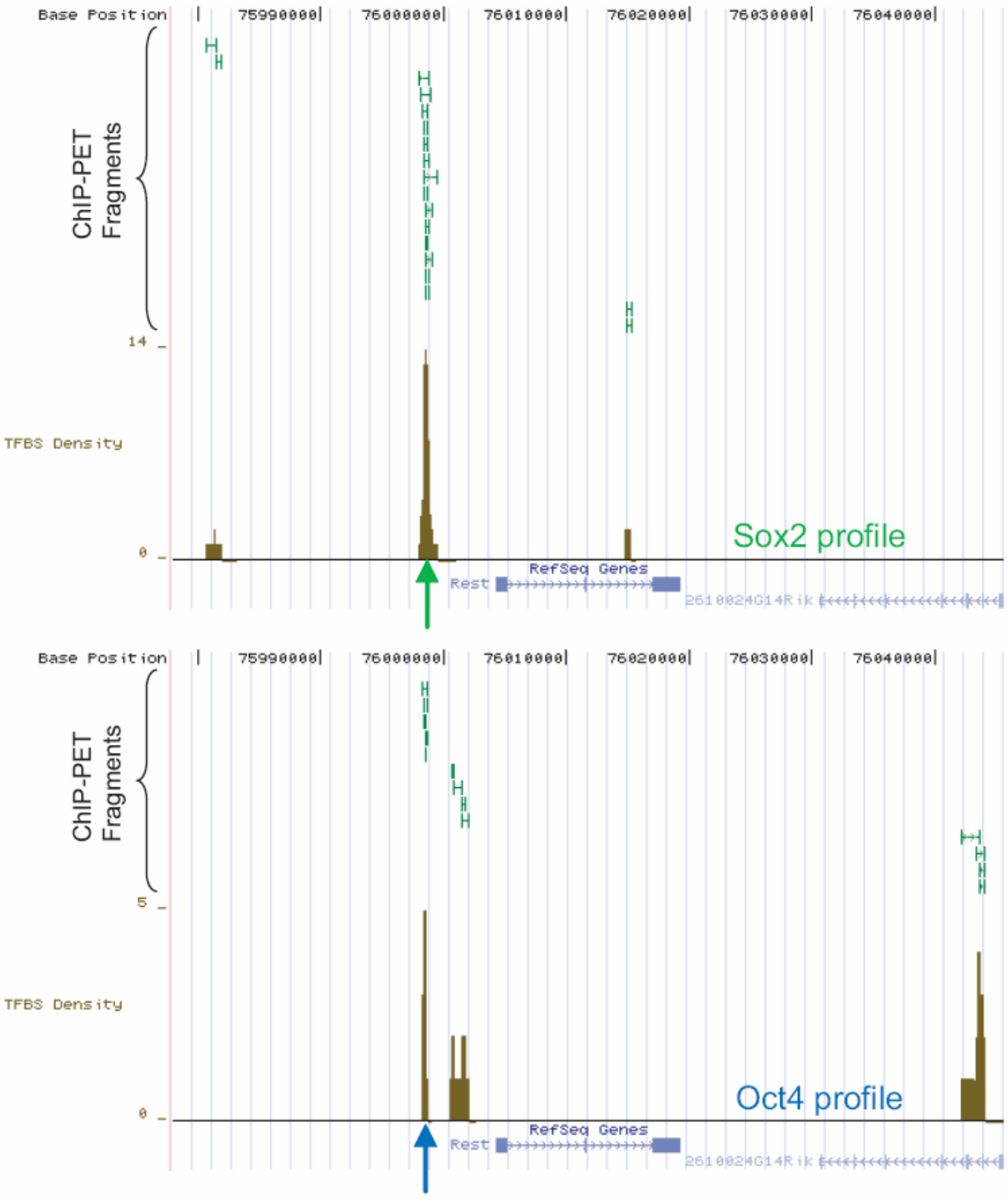
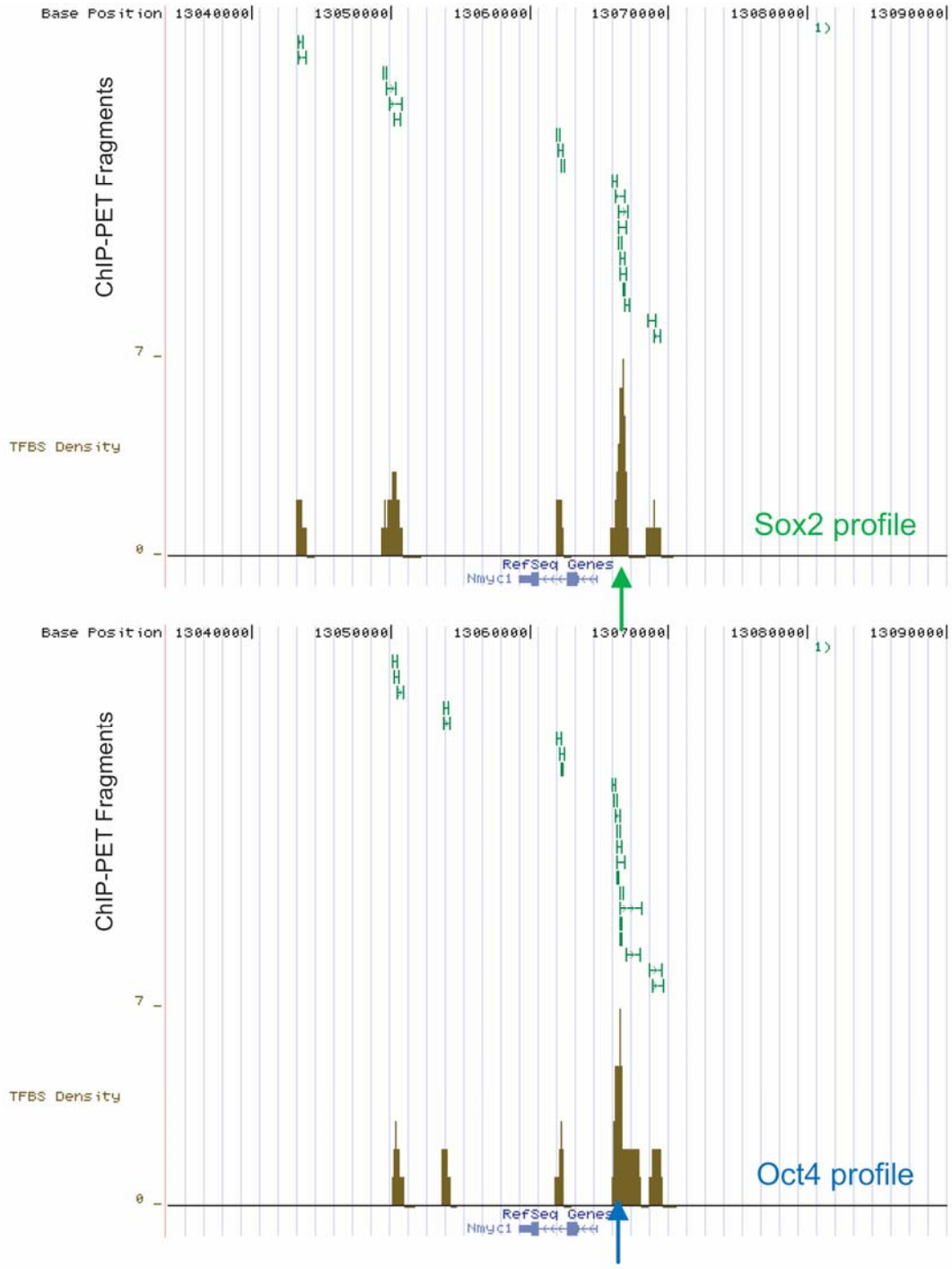


Figure 5.5: Binding of Sox2 on microRNA genes. Sox2 ChIP-PET data for (A) *miR124a*, (B) *miR128b*, (C) *miR153* are shown. Vertical arrow represents Sox2 binding loci and diagonal arrow represents the position of the microRNA.

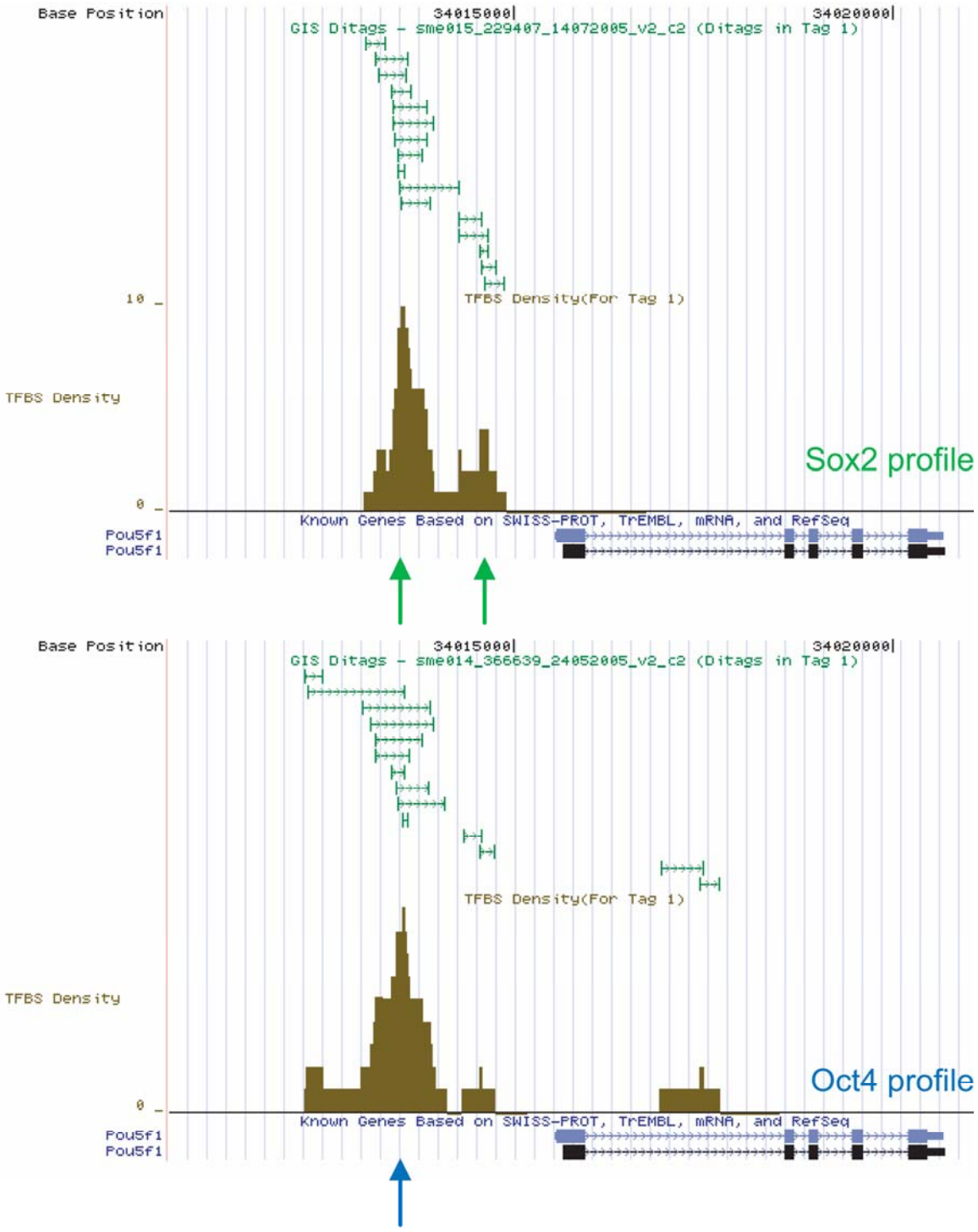
A **Sox2 and Oct4 binding profiles at *Rest***
chr5: 75,977,921-76,045,618 (67,698 bp)



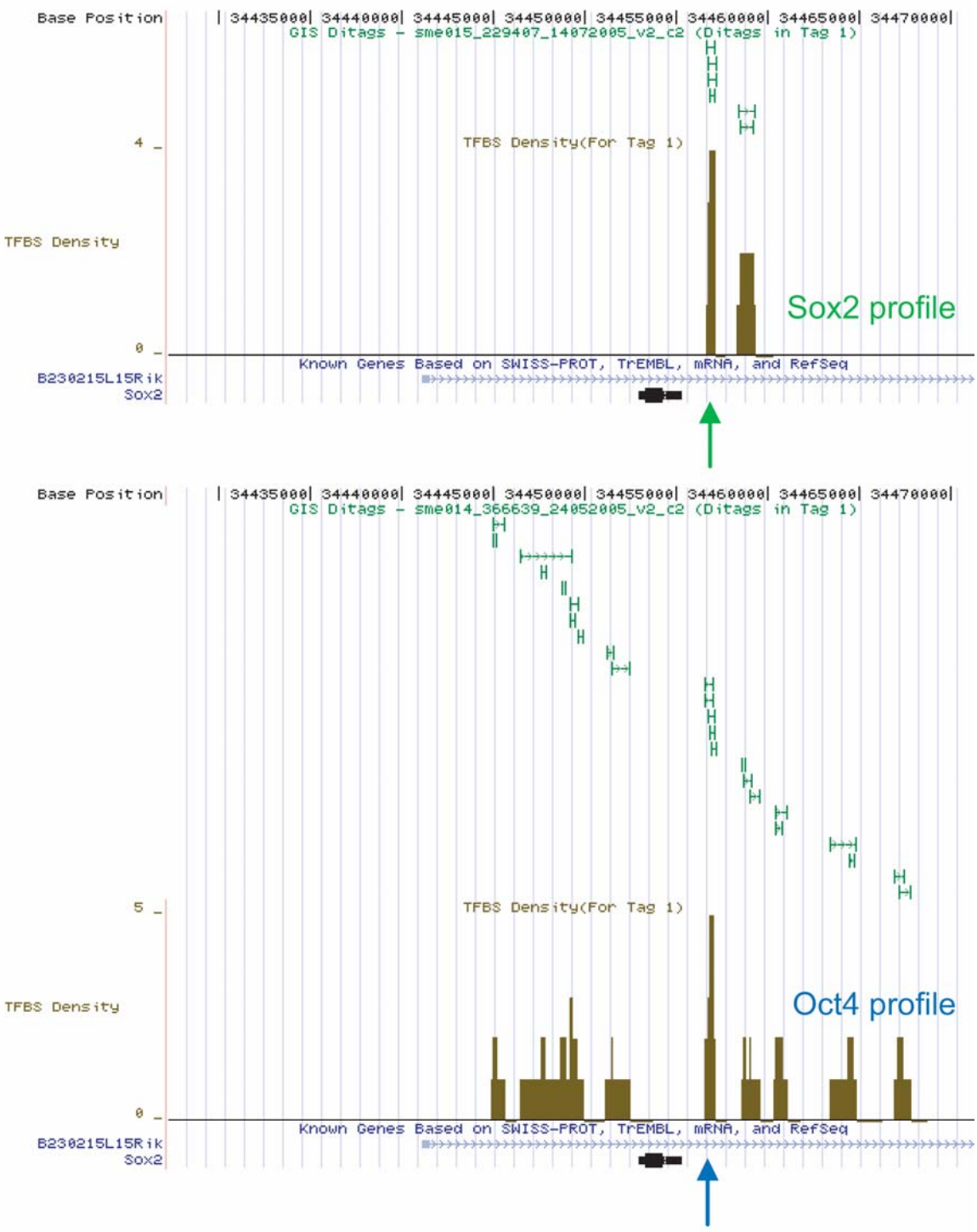
B **Sox2 and Oct4 binding profiles at *Mycn***
chr12: 13,033,905-13,090,054 (56,150 bp)



C **Sox2 and Oct4 binding profiles at *Oct4***
chr17: 34,010,465-34,021,124



D **Sox2 and Oct4 binding profiles at *Sox2***
chr3: 34,427,243-34,471,442



E **Sox2 and Oct4 binding profiles on *Esrrb***
chr12:81,786,235-81,858,459 (72,225 bp)

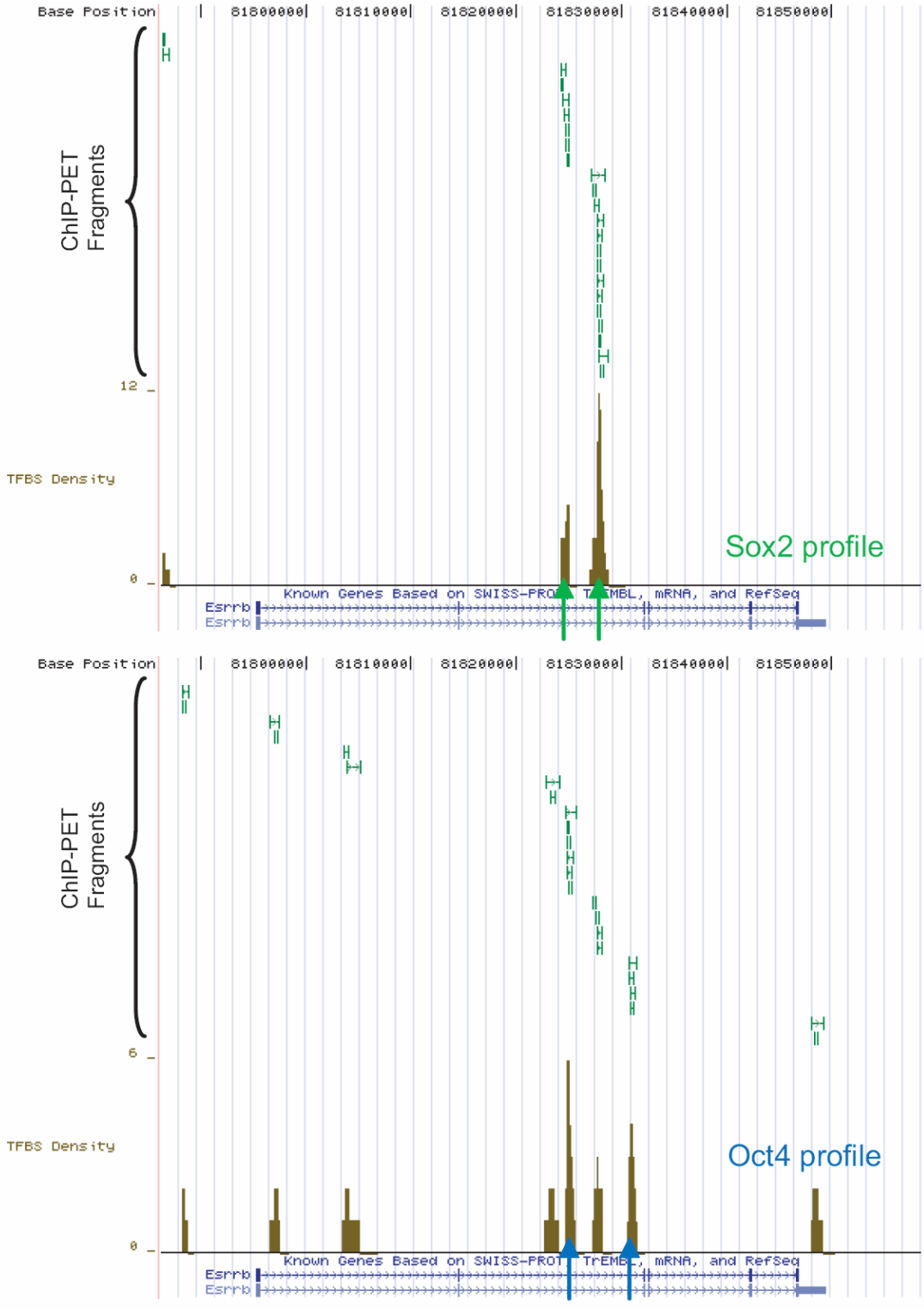


Figure 5.6: Sox2 and Oct4 ChIP-PET binding profiles at (A) *Rest* and (B) *Mycn* (C) *Oct4*, (D) *Sox2* and (E) *Esrrb*.

CHAPTER 6

Analyses of the Combined Oct4 and Sox2 DNA Binding Sites

6.1 Introduction

Oct and Sox proteins have been shown to selectively interact with each other and bind to DNA via their conserved domains, POU and HMG, respectively (Herr and Cleary, 1995; Wegner, 1999). Their functional partnership on regulatory elements have been shown to exist in several species including that of the human and mouse (Dailey and Basilico, 2001; Kuhlbrodt *et al.*, 1998). Moreover, during early development and in ESC lines, *Oct4* and *Sox2* genes are co-expressed and their combinatorial function is critical in specifying the first three lineages in the mammalian embryo (Avilion *et al.*, 2003; Nichols *et al.*, 1998; Niwa *et al.*, 2000).

The synergistic partnership between Oct4 and Sox2 on DNA has been demonstrated on *Fgf4* (Ambrosetti *et al.*, 1997; Yuan *et al.*, 1995), *Nanog* (Kuroda *et al.*, 2005; Rodda *et al.*, 2005), *Oct4* (Chew *et al.*, 2005) and *Sox2* (Chew *et al.* 2005) gene loci. In addition, *Oct4* and *Sox2* elements have been found in the enhancers of *Opn* (Botquin *et al.*, 1998), *Utf1* (Nishimoto *et al.*, 1999) and *Lefty1* (Nakatake *et al.*, 2006). The data from the ChIP-PET study described in Chapters 4 & 5 has also facilitated the recent characterisation of two composite sox-oct binding sites within the first intron of *Zfp206* where Oct4 and Sox2 are involved in activating its transcription in a synergistic manner (Wang *et al.*, 2007). With the use of homology modelling tools to analyze the crystal structure of Oct1-Sox2/*Fgf4* ternary protein-enhancer complex, Remenyi *et al.* (2003) derived a structural model of the Oct4-Sox2 heterodimer interaction on the *Fgf4* and *Utf1* enhancers. The specific assemblies of these factors are reported to rely on the *cis*-regulatory elements on the DNA (Remenyi *et al.*, 2003).

In this study, a significant number of the same genes were found in the Oct4 and Sox2 ChIP-PET datasets, indicating that Oct4 and Sox2 may target genomic sites through a joint *cis* element. Here, the approach taken was to merge both sets of Oct4 and Sox2 binding site data in order to obtain a list of genes that possess both Oct4 and Sox2 binding sites. A selection of overlapping genes was then experimentally verified for co-occupancy of Oct4 and Sox2. Subsequently, sequences from the Oct4 and Sox2 *bona fide* DNA binding sites were search for the presence of Oct4 and Sox2 DNA binding motifs. The DNA motifs were then characterised for specificity in conferring Oct4 and Sox2 DNA binding ability and in providing functional activity *in vitro*.

6.2 Results

6.2.1 Oct4 and Sox2 co-occupy shared binding sites

6.2.1.1 Comparative location analyses of Oct4 and Sox2

The Oct4 and Sox2 ChIP-PET datasets were used to investigate the relationship between Sox2 and Oct4 occupancies on a global scale in mESCs (Chapters 4 & 5). A list of genes (based on known, Refseq and MGC) associated with Sox2 and/or Oct4 PET4+ clusters and within 50 kb upstream and downstream of the transcription start site was generated by a bioinformatics colleague, Vinsensius Vega. While Oct4 and Sox2 were predicted to bind to genes independently of each other (540 genes for Oct4, 508 genes for Sox2), a significant proportion of the genes (approximately 30% or 242 genes of Oct4-associated and Sox2-associated genes) were also predicted to be bound by both Oct4 and Sox2 (Figure 6.1). These included *Rest*, *Dido1*, *Phc1*, *Id3*, *Mycn*, *Tcf3*, *Esrrb*, *Otx2*, *Lefty1*, *Rif1*, as well as *Sox2*, *Oct4*, and *Nanog*. Although the bioinformatic search showed that Oct4 and Sox2 may bind to the same genes, further verifications are needed to show that they bind to *cis* elements in close proximity with one another. This is

because Oct4 and Sox2 may be binding individually at the same gene loci but on different DNA molecules or different populations of cells.

6.2.1.2 Oct4 and Sox2 co-occupy on the same DNA molecules

Six genes were selected from the 242 genes possessing both Oct4 and Sox2 binding sites in Figure 6.1 to test for Oct4 and Sox2 co-occupancy. A two-step ChIP (termed sequential ChIP; seqChIP) was used to pull down DNA fragments (approximately 500 bp) bound by both Oct4 and Sox2. Crosslinked chromatin extracts were first immunoprecipitated with either the Oct4 or Sox2 antibody. Recovered material from this first ChIP was then subjected to a second ChIP using the other antibody (Figure 6.2A). Subsequently, qPCR using primers amplifying 150-250 bp products specific to binding sites on selected genes in the Oct4 and Sox2 overlap data were carried out to measure ChIP fold enrichments. Significant enrichment after the second ChIP was indicative of Oct4 and Sox2 co-occupancy.

Oct4 or Sox2 binding was still detectable after this first round of ChIP, although enrichment was reduced due to the process of crosslinking of the antibody to the protein G sepharose using dimethyl pimelimidate (DMP). The fold enrichments after the 2nd ChIP (Oct4 followed by Sox2 and vice versa) were significantly higher than Oct4 or Sox2 single ChIPs and control ChIPs using a non-specific antibody (Ena-1) (Figure 6.2B). This demonstrates that Oct4 and Sox2 co-occupy the same DNA molecules of *Oct4*, *Sox2*, *Nanog*, *Tcf3*, *Trp53* and *Mycn* in very close proximity to each other.

6.2.2 Regulation of target genes by Oct4 and Sox2

Selected gene loci that were co-bound by Oct4 and Sox2 were shown to be active in mESCs by the presence of histone 3 trimethylated lysine 4 (H3K4Me3) marks which denote active regions of the chromosome. To examine genes that are bound by both Oct4 and Sox2, seqChIP using Oct4

followed by Sox2 antibodies were carried out. Oct4-Sox2 sequential ChIP and GST mock ChIP DNA were amplified, labelled with Cy dyes, and hybridised onto NimbleGen gene microarrays designed by Dr Ng Huck Hui. The chips (25 x 75 mm each) contain 50-mer probes (50 bp apart) spanning about 600 genes represented by 385000 features (16 μm x 16 μm) in an array size of 17.4 mm x 13 mm. The same procedure was repeated using the H3K4Me3 ChIP DNA and mock ChIP DNA.

SeqChIP-on-chip data showed that Oct4 and Sox2 co-occupy *Oct4*, *Sox2*, *Nanog*, *Tcf3*, *Mycn*, *Trp53*, *Rest* and *Tcf7* (Figure 6.3, blue peaks) while mock ChIP-on-chip did not show any peaks at these regions (Appendix F). These ChIP-on-chip results confirm the seqChIP-qPCR results in section 6.2.1. In addition, the seqChIP-on-chip results showed that Oct4 and Sox2 co-occupy at the 5' and 3' enhancer regions of Sox2 as well as both the promoter and 5' enhancer region of *Nanog* demonstrating the sensitivity of using sequential ChIP coupled to microarray. Moreover, these genes are also bound by H3K4Me3 at the 5' or throughout the transcribed regions (Figure 6.3, green peaks), indicating that the bound genes are active. In contrast, Oct4-Sox2 and H3K4Me3 did not show occupancy on *Hoxa9* and *Hoxa10*, which are part of the repressed Hox cluster genes in ESCs. H3K4Me3 and Oct4-Sox2 binding do not always occur together. One example is the binding on the *Sox15* loci. Sox15 was reported to be expressed in mESCs but is not essential for ESC self-renewal (Maruyama *et al.*, 2005). Concurringly, H3K4Me3 peaks were detected on the *Sox15* loci. However, Oct4-Sox2 binding was not detected, suggesting that *Sox15* is not a downstream target of Oct4 and Sox2.

These data indicates that Oct4 and Sox2 regulate the genes that are bound. Gene activity increases when histone 3 lysine 4 (H3K4) is converted from a di- to a tri-methyl state (Santos-Rosa *et al.*, 2002). However, recent studies have suggested that both active as well as repressive chromatin marks may be closely juxtaposed at promoters of highly conserved genes in ESCs

(Bernstein *et al.*, 2006), hence other methods of inferring regulation have been utilized. Firstly, MPSS data correlated with those of ChIP-PET binding as shown in the previous chapters as well as other studies have confirmed that these genes are highly expressed in undifferentiated mESCs (Pritsker *et al.*, 2006). Moreover, RNAi-mediated depletion of Oct4 or Sox2 was found to also perturb the expression of these genes (Loh *et al.*, 2006 & Loh *et al.* unpublished data). These data collectively suggest that Oct4 and Sox2 DNA binding may be important for gene regulation in mESCs. The next step was to identify and characterize the *cis* elements that determine the binding of both Oct4 and Sox2.

6.2.3 Identification of the joint Sox2-Oct4 DNA binding motif

To look for consensus sequences that may be present in the Oct4 and Sox2 bound genes, DNA sequences of enriched regions on selected genes were aligned using the VectorNTI multiple sequence alignment tool together with previously identified elements from *Oct4*, *Sox2*, *Nanog*, *Utf1* and *Fbx15*. A 15 nucleotide consensus sequence CATTGTTATGCAAAT, which resembles the Sox2 and Oct4 motif joint in tandem, was identified (Figure 6.3A, Chew *et al.*, 2005).

Subsequently, a comprehensive motif search was carried out by our collaborator, Vinsensius Vega, using the *de novo* motif discovery algorithms *Weeder* (Dempster *et al.*, 1977; Pavesi *et al.*, 2004). Sequences from PET7+ clusters of the Oct4 and Sox2 ChIP-PET datasets (total of 90 Oct4 and 191 Sox2 clusters) were extracted and their repeat regions (as annotated by the USCS mm5 mouse genome browser) were masked. The CATTGTTATGCAAAT sequence was the predominant motif found in both the Oct4 and Sox2 datasets (Figure 6.3B). This motif is similar to the Sox2-Oct4 composite consensus element derived earlier (Figure 6.3A). 69% of PET6+ Oct4 binding loci and 62% of PET9+ Sox2 binding loci contained this motif, indicating that Sox2 and Oct4 may act together as a complex in the mESC genome.

6.2.4 Characterization of the Sox2-Oct4 DNA binding motif

ChIP-qPCR, ChIP-PET and ChIP-on-chip experiments can only identify DNA binding regions of about 100-300 bp. Therefore, in order to investigate the specificity of the 15 nucleotide Sox2-Oct4 joint motif and to examine its functionality in conferring binding and activity, selected motifs from the ChIP-PET datasets were characterized using a series of EMSA and luciferase reporter assays.

6.2.4.1 Interactions of Sox2 and Oct4 with the Sox2-Oct4 joint motifs

6.2.4.1.1 Sox2 and Oct4 bind to the Sox2-Oct4 DNA motif *in vitro*

Transcription factors are capable of binding to a basic consensus sequence of various permutations. EMSA relies on the ability of a protein to bind to a DNA probe *in vitro*, followed by electrophoretic separation of DNA-protein complexes from the unbound DNA (free probes) on non-denaturing polyacrylamide gels.

To test whether Sox2 and Oct4 proteins bind to the Sox2-Oct4 joint element, EMSA was carried out using Sox2-Oct4 motifs found at *Ebf1*, *Rest*, *Rif1* and *Tcf7* loci. A 33 bp biotin-labelled double-stranded DNA containing the 15 nucleotide composite Sox2-Oct4 elements and 18 nucleotide flanking sequences (Figure 6.5A) was incubated with Sox2 or Oct4 over-expressed protein from HEK293T cells that do not express endogenous Oct4 and Sox2. When the Oct4 protein was incubated with the probes, a single Oct4/DNA complex was detected (Figure 6.5B, lane 2). This complex was supershifted (retardation in mobility) upon addition of anti-Oct4 antibody (Figure 6.5B, lane3). Anti-Sox2 and anti-JunB (control) antibodies did not affect the mobility of the complex (Figure 6.5B, lane 4 & 5). This demonstrates that Oct4 binds to the dsDNA probe. Similar observations were noted when Sox2 protein was added to the probe, demonstrating that Sox2 also binds to the same dsDNA probe (Figure 6.5B, lanes 6-9). Control lysates from HEK293T cells transfected with the vector alone did not form any specific complex

with the DNA probes (Figure 6.5B, lanes 10-13). For all the probes, except *Rif1*, the anti-Sox2 supershifted complex showed weaker bands than that formed by the anti-Oct4 antibody, suggesting that the Sox2 antibody may have interfered with the protein-DNA interaction.

The *Ebf1* probe, which contains the exact consensus sequence, was also used to test for binding with native proteins extracted from the nuclei of mESCs (nuclear extract). A major complex between the probe and native proteins was detected as shown in Figure 6.6, lane 2. The addition of anti-Oct4 or -Sox2 antibodies led to a supershift of this complex (Figure 6.6, lanes 3 & 4). Conversely, the control anti-JunB antibody did not affect the mobility of this complex. Therefore, this indicated that the DNA probe was bound specifically by endogenous Oct4 and Sox2. The addition of a 200-fold excess unlabelled probe successfully competed for binding to the Sox2/Oct4 complex (Figure 6.6, lane 6), while addition of an unrelated probe had no effect (Figure 6.6, lane 7). When an excess of unlabelled probe was added, the DNA binding protein will bind to both the unlabelled probe and biotin-probe, resulting in a decrease in the amount of factor available for binding to the probe by competition from the unlabelled site, and subsequent reduction in the intensity of the retarded band. When an unrelated probe was used, there was no difference in the intensity of the retarded band because the competing probe is unable to bind to the factor.

Thus, *in vitro*, Oct4 and Sox2 have been demonstrated to bind specifically to the Sox2-Oct4 joint motifs of these targets as a ternary complex.

6.2.4.1.2 Mutations of Sox2 and Oct4 DNA motif sequences abolished binding

Subsequently, the importance of each DNA sequence within the Sox2-Oct4 element in mediating the formation of the DNA-protein complex was examined. EMSA was carried out using the wild-type *Rest* probe and *Rest* probes containing mutations within the Sox2 and Oct4 elements (Figure

6.7A). Mutations within the Oct4 element prevented the formation of the Oct4-DNA complex, indicating that these sequences are crucial for Oct4 binding. Mutations within the Sox2 element also abolished the formation of the Sox2-DNA complex, except for mutation E at the 7th position (Figure 6.7B), which is the last nucleotide of the Sox2 element. Instead, more Sox2-DNA complex was formed when wild type G was mutated to A at that position. Upon inspection of the consensus motif matrix generated from the ChIP-PET datasets (Figure 6.7C), the 7th position of the consensus motif contained nucleotides T and A (square box in Figure 6.7C). This showed that changing the consensus sequence from G to A increased the efficiency of forming the Sox2-DNA complex, further validating the joint Sox2-Oct4 consensus motif. However, it does not rule out the possibility that the 7th position may be dispensable for Sox2 (or Oct4) binding. Nonetheless, these results indicate that all but one of the sequences in the motif is crucial for Oct4 or Sox2 binding and that Oct4 and Sox2 are capable of binding on this element *in vitro*.

6.2.4.1.3 Sequences flanking the Sox2-Oct4 DNA motif are not essential for binding

Mutation of the motif sequences investigates the requirement of the particular sequence in the formation of DNA-protein complex. However, differences in the sequence of flanking DNA outside the core consensus binding site may also affect binding of the protein onto DNA. To examine whether the flanking sequences of the Sox2-Oct4 motifs affect Sox2 and Oct4 binding, EMSA was carried out using dsDNA probes containing the joint motifs with mutations in the flanking sequences (Figure 6.8A). All of the probes carrying flanking sequence mutations could form Oct4/DNA and Sox2/DNA complexes (Figure 6.8B). This demonstrates that the flanking sequences of the Sox2-Oct4 motif are not essential for Oct4 and Sox2 binding.

6.2.4.2 The Sox2-Oct4 joint motif sequences are functional

6.2.4.2.1 The Sox2-Oct4 motifs confer reporter activities which are Oct4 and Sox2-dependent

To study the role of these specific sequence elements in the activation of pluripotent cell transcription, Sox2-Oct4 DNA motifs from *Ebf1*, *Tcf7*, *Rest* and *Rif1* were cloned in three tandem repeats upstream of the pGL3-*Oct4* pp vector (Wu *et al*. 2006) which contains the ESC-specific *Oct4* proximal promoter upstream of the luciferase gene (pGL3-3x element-*Oct4*pp, Figure 6.10A). The constructs were then transfected into mESC and assayed for luciferase activity after 3 days. All the four Sox2-Oct4 elements caused a two to seven-fold increase in reporter activity compared to the empty pGL3-*Oct4* pp vector control, indicating that the presence of the joint DNA motifs can augment the activity of the *Oct4* promoter in driving the expression of the luciferase gene in mESCs (Figure 6.9B). The luciferase reporter assay was also carried out using only one copy of each motif. Luciferase activity was measured to be two-fold above that of the empty pGL3-*Oct4* pp vector (data not shown). However, this is less compared to that which was observed when the motifs were present in three tandem repeats. The proportional increase in activity observed with increasing number of copies of motifs indicated that the activity measured was specific, i.e. luciferase activity is induced by the presence of the DNA motifs. The Sox2-Oct4 DNA motifs are therefore functional.

To further evaluate whether the activity conferred by the motifs were caused by the presence (and therefore binding as shown by EMSA) of Oct4 and Sox2 proteins, mESCs were co-transfected with pGL3-3x element-Oct4pp plasmids together with *Oct4* or *Sox2* RNAi plasmids. In the cells, *Oct4* and *Sox2* shRNAs mediate depletion of endogenous Oct4 and Sox2, respectively. The specificity of these shRNAs has been previously confirmed (Chew *et al.*, 2005), showing that the *Oct4* and *Sox2* shRNA specifically depletes the transcript and protein levels (also shown in Chapter 7) of Oct4 and Sox2, respectively. The constructs were transfected into mESC and after

12 hr, the cells were cultured in media containing puromycin for two additional days before assaying for luciferase activity. The use of puromycin was to selectively remove non-transfected mESCs. All the non-transfected cells (control) were killed after 2 days of puromycin selection. Activities of all the four reporter plasmids carrying the joint motifs were reduced upon Oct4 and Sox2 knockdown (Figure 6.9B). *Oct4* RNAi significantly reduced reporter activities to almost basal levels, whereas *Sox2* RNAi reduced reporter activities by approximately 20% (*Ebf1*) to 60% (*Rif1*), suggesting that the activities conferred by the motifs were dependent on Oct4 and Sox2 proteins.

The cells were harvested for luciferase activity at a relatively short time to reduce the indirect effects caused by differentiation. This is because silencing of Oct4 or Sox2 lead to differentiation of mESCs that will subsequently cause endogenous Oct4 and Sox2 levels to decrease. In addition, the effects of RNAi on the reporters may also constitute indirect effects as Oct4 and Sox2 reciprocally regulate each other, and thus the depletion of one factor may deplete the other. These are inherent disadvantages of the RNAi approach. However, since the specificity of the shRNAs was confirmed, the reporter activities conferred by the motifs are at least dependent on the presence of the Oct4 and Sox2 proteins. Moreover, since Oct4 and Sox2 have been shown to bind to these Sox2-Oct4 joint motifs by both EMSA and ChIP, the results indicate that the functionality of the DNA elements is dependent on Oct4 and Sox2.

6.2.4.2.2 The orientation of the Sox2-Oct4 DNA motif is important in conferring reporter activity

To test whether the orientation of the Sox2-Oct4 motif is important, the order of the Sox2 and Oct4 elements were swapped and tested for reporter activity. Oligos containing Oct4-Sox2 (instead of Sox2-Oct4) motifs were cloned into the pGL3-Oct4 pp in three tandem repeats and transfected into mESCs for luciferase reporter assay. The Oct4-Sox2 motifs showed a significantly

lower reporter activity compared to the original Sox2-Oct4 motif (Figure 6.10). This implies that the binding of Sox2 and Oct4 on their respective consensus motifs may be context-dependent or that the Sox2-Oct4 motif configuration is entailed by the heterodimer structure of the Sox2-Oct4 protein complex in mESCs. Therefore, these results indicate that the Sox2-Oct4 joint motif has to be in this particular orientation to be functional.

6.3 Discussion

Merging of the Oct4- and Sox2-associated genes (50 kb downstream to 50 kb upstream) obtained from individual ChIP-PET datasets revealed a core set of 242 candidate genes that are bound by both Oct4 and Sox2. Sequential ChIP-qPCR carried out on selected loci showed that Oct4 and Sox2 indeed co-occupied the same DNA molecule at close proximity. This suggests that Oct4 and Sox2 work in tandem to regulate gene expression of a majority of their target genes. However, there are still exceptions where some Oct4 and Sox2 binding sites uncovered by ChIP-PET did not contain any Sox2-Oct4 motif. This may be due to indirect Oct4 or Sox2 recruitment to genomic DNA independent of sequence-specific DNA recognition or indirect binding through a protein complex loop (further explained in Chapter 7). In addition, the Oct4 binding motif does not always exist joint to the Sox2 motif. Although the sox and oct elements in the enhancer regions of known targets *Fgf4*, *Utf1*, *Oct4*, *Nanog*, *Fbx15*, and *Sox2* are <3 bp apart (Botquin *et al.* 1998; Catena *et al.*, 2004; Dailey *et al.*, 1994; Nishimoti *et al.*, 1999; Tomioka *et al.*, 2002; Tokuzawa *et al.*, 2003; Yuan *et al.*, 1995), a recent study found that oct and sox elements in Zfp206 are separated by 11 bp (Wang *et al.*, 2007).

Subsequently, ChIP-on-chip using Oct4-Sox2 sequential ChIP also demonstrated that Oct4 and Sox2 co-occupy the same location on many genes. In addition, the ChIP-on-chip approach could detect Oct4/Sox2 co-bound sites that were not previously detected using the ChIP-PET method, such as *Myc*, *Jarid2* and *Sgk*, which were found in one library but not the other. This is

because ChIP-on-chip is a more sensitive method that can detect as low as 2-fold ChIP enrichment (NimbleGen), while many targets may be missed out using the ChIP-PET approach due to low ChIP enrichment or insufficient sequencing of PETs. Therefore, the ChIP-on-chip approach was used as a complementary approach to further examine the Oct4-Sox2 co-occupancy. The Oct4 and Sox2 co-occupied targets identified using ChIP-PET and ChIP-on-chip included those that are important in the maintenance of pluripotency (eg. *Zic3*, *Oct4*, *Nanog*, *Sox2*), cell cycle regulation (eg. *Trp53*, *Rif1*, *Myc*) and transcriptional regulation (eg. *Esrrb*). Furthermore, many of these target genes are transcriptionally active as indicated by H3K4Me3 marks shown by H3K4Me3 ChIP-on-chip on selected gene loci. The wide range of Oct4 and Sox2 targets, which are crucial for cell maintenance and survival, further implicates them as key factors of the ESC transcriptional regulatory network.

The ChIP-PET method provides a resolution of binding site location of approximately 100 bp (based on the Oct4 and Sox2 PET cluster sizes), thus allowing the use of *de novo* discovery algorithms to identify specific *cis*-elements enriched in the ChIP fragments. Both multiple alignment of selected binding site sequences and computerised *de novo* motif discovery from both ChIP-PET datasets identified a predominant 15 nucleotide consensus sequence, CATTGTTATGCAAAT. This sequence contains the Sox binding site motif, CATTGTT, followed immediately by the Oct binding motif, ATGCAAAT. Since this consensus motif does not contain any spacing between the Sox and Oct elements, it is postulated that there may be many more permutations of sox-oct binding sites present such as the sox-oct binding site with 3 bp separation found in *Fgf4*. The identification of a high frequency of Sox2-Oct4 joint motif in the genome from both the Oct4 and Sox2 ChIP-PET datasets, in addition to the sequential ChIP evidence, suggest that one of the main mechanisms for targeting Oct4 and Sox2 to their genomic sites is through the Sox2-Oct4 motif via a cooperative interaction between Oct4 and Sox2.

In vitro characterisation by EMSA and reporter assay also verified that the computationally-derived Sox2-Oct4 joint motif was bound by both Sox2 and Oct4. Moreover, the motif was found to be specific and transcriptionally functional. These results corroborated with previous studies showing that maximal transcriptional activity of the *Oct4* promoter in mESCs requires the cooperative binding of both Oct4 and Sox2 to the composite joint element (Chew *et al.*, 2005). The joint motif may enable Oct4 and Sox2 to interact with each other while binding onto DNA. The DNA-binding domains of these transcriptional regulators may play a key role in gene regulation by tethering activation domains and coordinating the assembly of the factors on the promoter or enhancer regions.

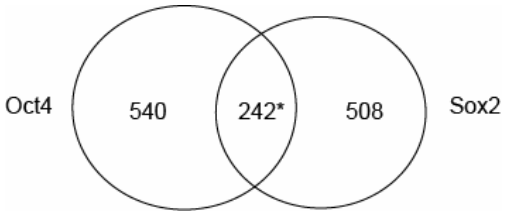


Figure 6.1: Venn diagram indicating the extent of overlap between genes associated with Sox2 and Oct4 binding in mESCs. Common targets (overlap) between Oct4 and Sox2-bound genes covered 50 kb upstream to 50 kb downstream of each gene. *Significant at Binomial Test with *p*-value < 1.55E-192.

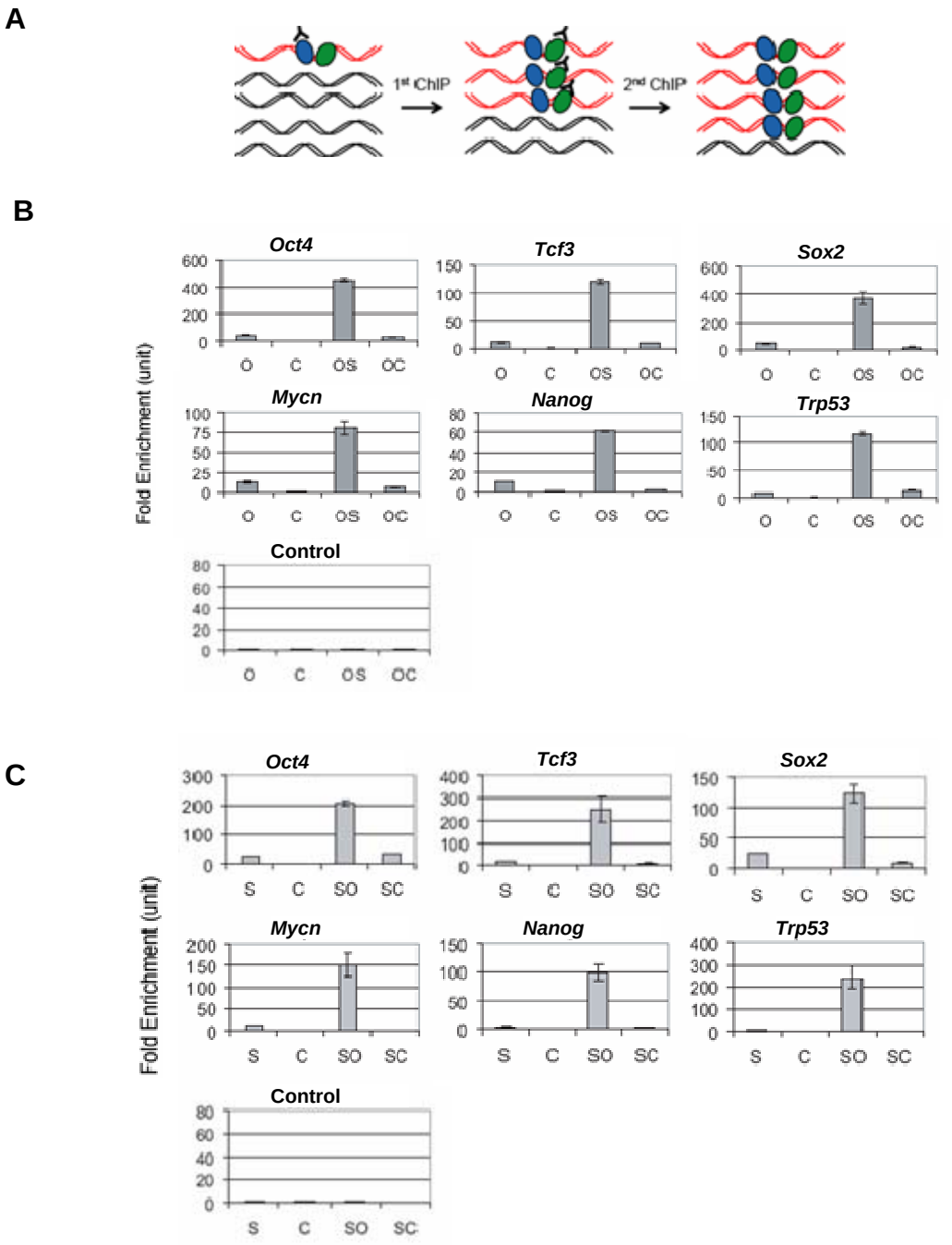
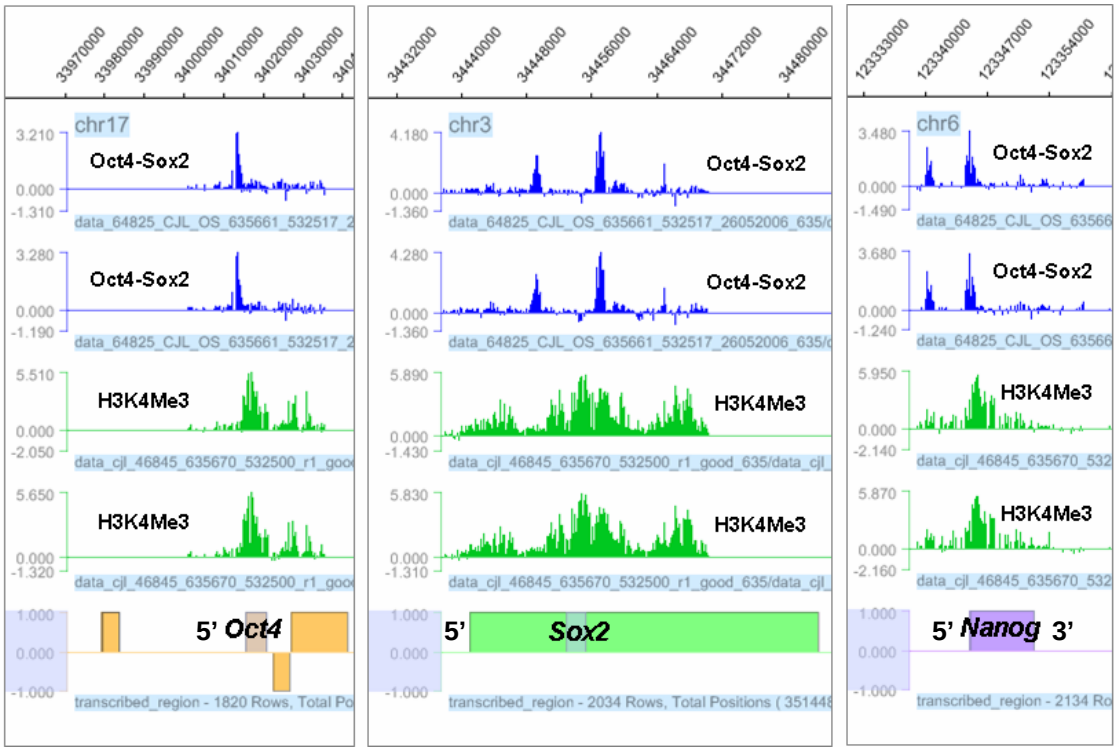
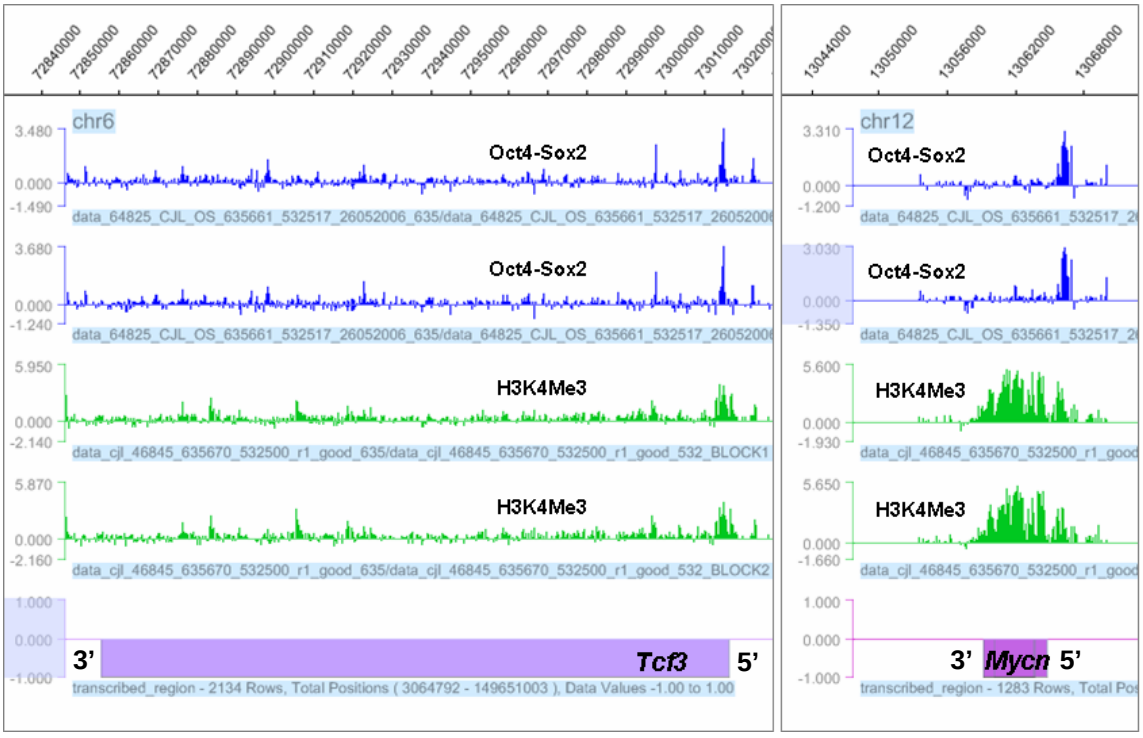


Figure 6.2: Co-occupancy of Oct4 and Sox2 on target sites. (A) Schematic diagram illustrating Oct4 and Sox2 co-occupancy as carried out in sequential ChIP. First ChIP using Oct4 antibody will purify for fragments bound by Oct4, as well as non-specific DNA. A second ChIP using the Sox2 antibody will further purify the pool of DNA to yield DNA fragments bound by both Oct4 as well as Sox2. Sequential ChIP using (B) Oct4 antibody followed by Sox2 antibody and (C) Sox2 antibody followed by Oct4 antibody. Fold enrichment represents the abundance of enriched DNA fragments over a control region not enriched for the respective targets. O: Oct4 ChIP, C: control Ena-1 ChIP, S: Sox2 ChIP, OS: Oct4 ChIP followed by Sox2 ChIP, OC: Oct4 ChIP followed by control Ena-1 ChIP, SO: Sox2 ChIP followed by Oct4 ChIP, SC: Sox2 ChIP followed by control Ena-1 ChIP. Standard deviations representing technical replicates are shown.


chr17
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
5' Oct4
transcribed_region - 1820 Rows, Total Po
chr3
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
5' Sox2
transcribed_region - 2034 Rows, Total Positions (351448
chr6
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
5' Nanog 3'
transcribed_region - 2134 Ro


chr6
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
3' Tcf3 5'
transcribed_region - 2134 Rows, Total Positions (3064792 - 149651003), Data Values -1.00 to 1.00
chr12
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
3' Mycn 5'
transcribed_region - 1283 Rows, Total Pos

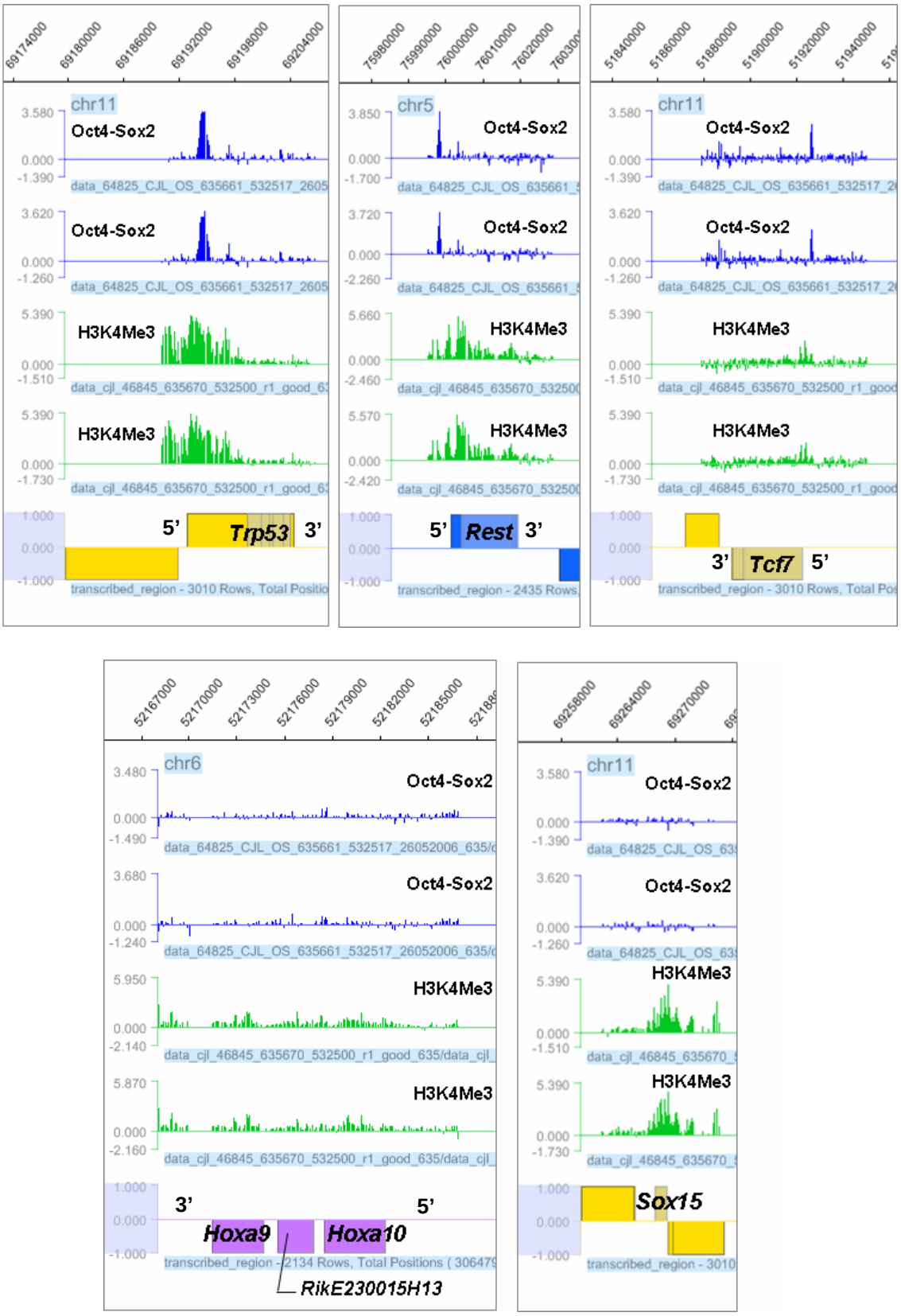

chr11
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
5' Trp53 3'
transcribed_region - 3010 Rows, Total Positio
chr5
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
5' Rest 3'
transcribed_region - 2435 Rows
chr11
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
3' Tcf7 5'
transcribed_region - 3010 Rows, Total Pos
chr6
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
3' 5'
Hoxa9 Hoxa10
RikE230015H13
transcribed_region - 2134 Rows, Total Positions (306479
chr11
Oct4-Sox2
Oct4-Sox2
H3K4Me3
H3K4Me3
Sox15
transcribed_region - 3010

Figure 6.3: ChIP-on-chip data showing occupancy of Oct4 and Sox2 on genes marked by H3K4Me3. Oct4-sox2 sequential ChIP (seqChIP) and H3K4Me3 ChIP were carried out and hybridised on Nimblegen customised DNA ChIP. NimbleGen ChIP-on-chip arrays are run as two colour experiments, one channel representing the experimental ChIP sample, the other representing the control non-specific GFP antibody ChIP sample. SignalMap readout showed peaks representing binding on *Oct4*, *Sox2*, *Nanog*, *Tcf3*, *Mycn*, *Trp53*, *Rest*, and *Tcf7* loci. Negative controls are also shown. No Oct4-Sox2 and H3K4Me3 peaks were detected at the *Hoxa9* and *Hoxa10* loci, while the *Sox15* loci showed H3K4Me3 peaks but no Oct4-Sox2 peak. Y axis represents enrichment which is the ratio between the intensities of ChIP over control probe. Bottom track of every panel shows transcribed regions of genes, with blocks above the x axis (+) representing the sense strand, and blocks below the x axis (-) representing antisense strand. Replicates shown are on-chip technical replicates of probes separated into physically distinct blocks on the arrays.

A

```
Oct4          CTTTGTTATGCATCT
Sox2          CATTGTGATGCATAT
Nanog         CATTGTAATGCAAAA
Utf1          CATTGTTATGCTAGT
Fbx15         CATTGTTATGATAAA
Zic3          CACTGTTTTGCAGAT
Jarid2        CTGTATTGTGCAAAA
Zfp57         CATTGAAATATTAGG
Phc1          TATTGTTATGCAAAT
Ctnnbl1       TATTGTCTCCTTGTT

                Sox2    Oct4
Consensus     CATTGTTATGCAAAT
```

B

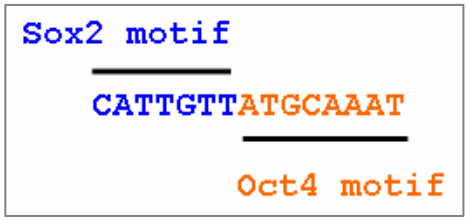


Figure 6.4: Identification of the 15 nucleotide Sox2-Oct4 joint consensus motif in the Oct4 and Sox2 ChIP-PET datasets. (A) The Sox2-Oct4 joint motifs were found by multiple alignment analysis of sequences from Oct4 and Sox2 binding sites on the *Oct4*, *Sox2*, *Nanog*, *Utf1*, *Fbx15*, *Zic3*, *Jarid2*, *Zfp57*, *Phc1* and *Ctnnbl1* loci. (B) *De novo* identification of Sox2-Oct4 consensus motif found in both the Oct4 and Sox2 ChIP-PET libraries.

A

	Sox2 Oct4
Ebf1	5′-gcaacaccg**CATTGTTATGCAAAT**tgttcctat-3′
Rest	5′-agtccttcg**TATTGTGATGCAAAT**aggccaatg-3′
Rif1	5′-tttccaggc**CTTTGTTATGCACGC**catgggtct-3′
Tcf7	5′-tggtaggcc**TATTGTTATGCAAAT**gagcccggt-3′

B

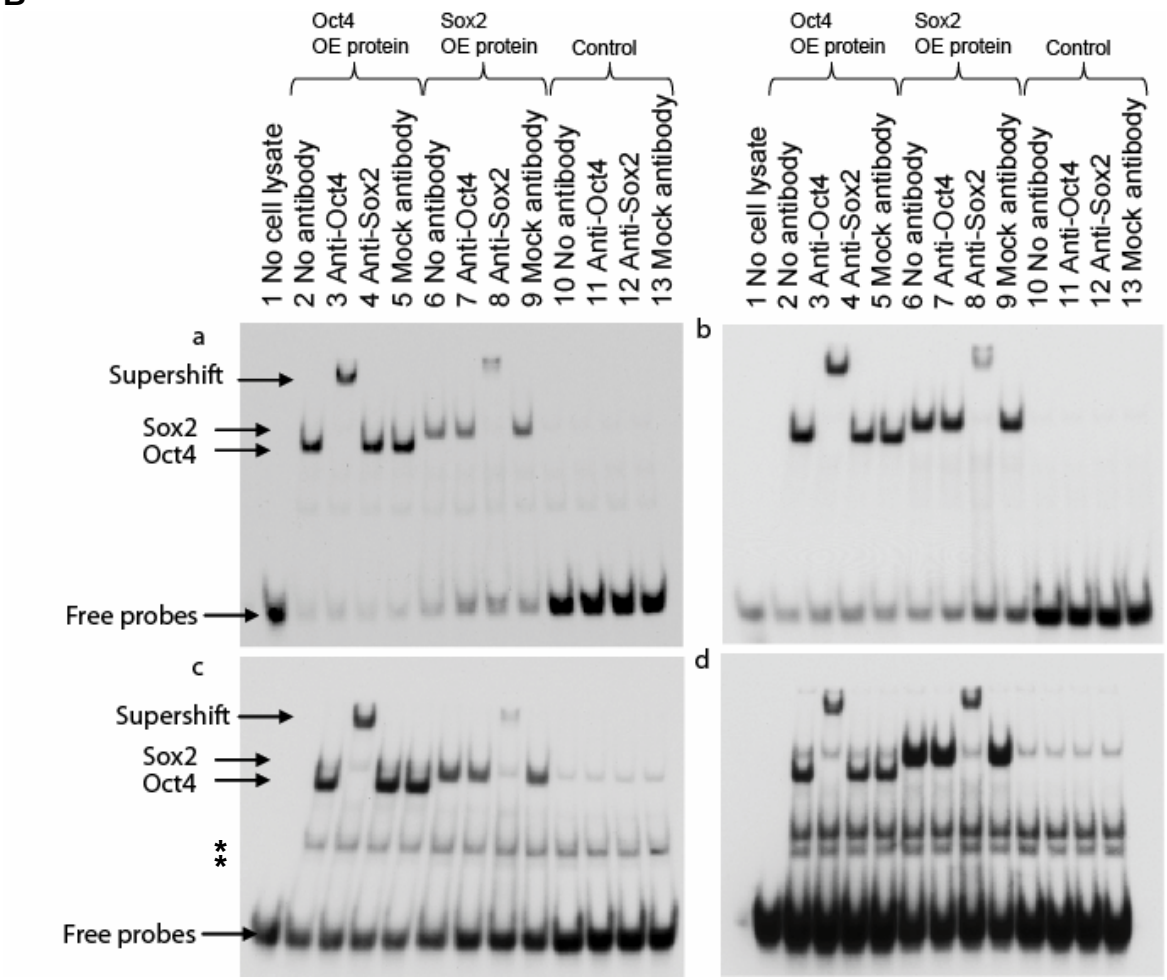


Figure 6.5: Binding of Oct4 and Sox2 overexpressed (OE) proteins on biotin-labelled DNA probes containing Oct4 and Sox2 binding sites using EMSA. (A) Sequences of corresponding probes containing the Sox2-Oct4 joint motifs. (B) Oct4 and Sox2 containing expression plasmids as well as an empty plasmid were transfected into HEK293T cells and the cell lysates were harvested. Specific anti-Oct4, anti-Sox2 and JunB (control) antibodies were used to obtain supershifts in reactions containing a) *Ebf1*, b) *Tcf7*, c) *Rest*, and d) *Rif1* probes. * denotes non-specific bands.

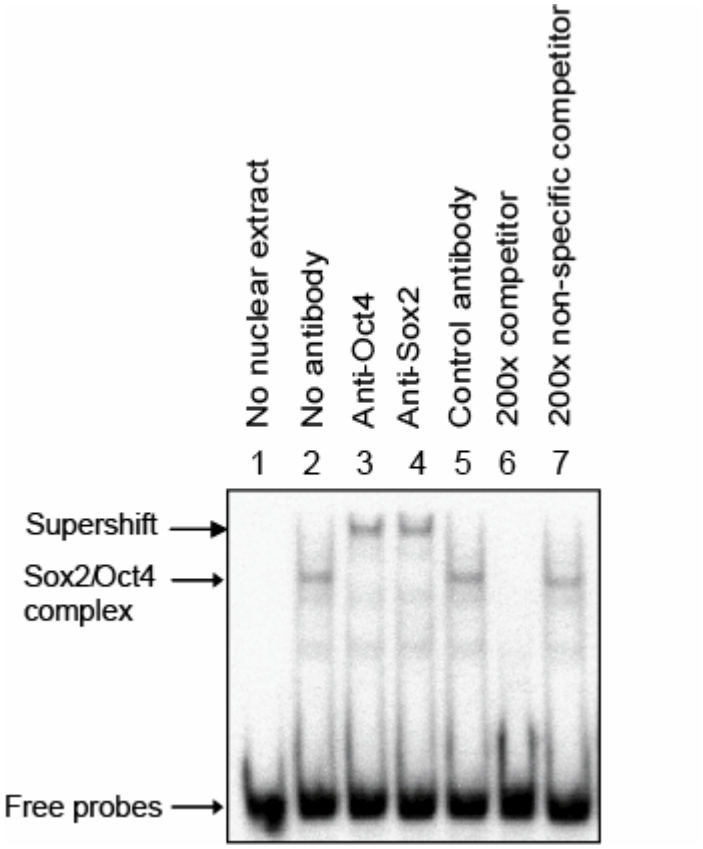
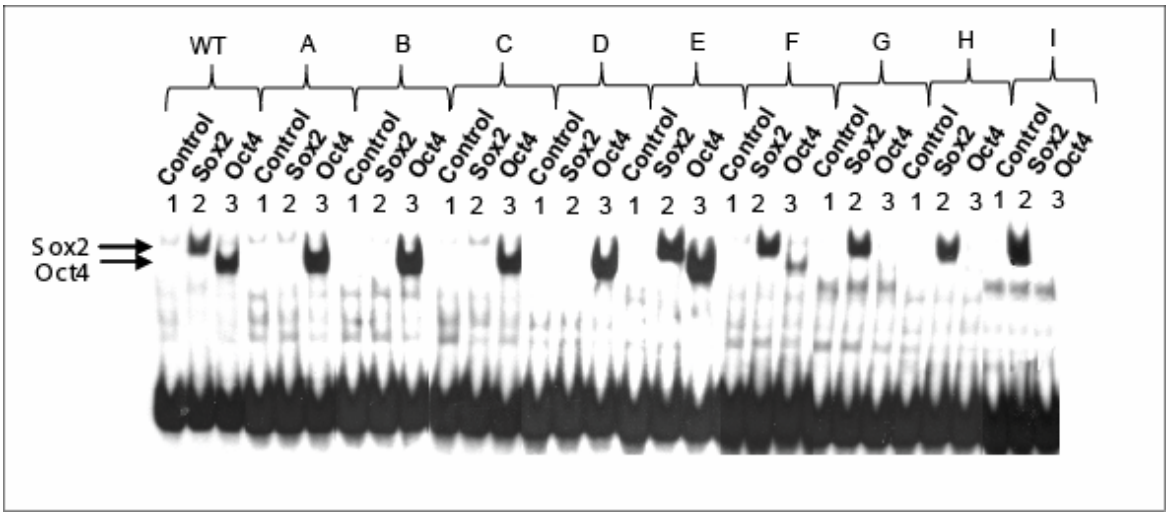


Figure 6.6: Endogenous native Oct4 and Sox2 bind to the composite Sox2-Oct4 joint motif of *Ebf1*. Oct4 and Sox2 binding on the *Ebf1* probe using 10 μg mESC nuclear extract.

A

		Sox2 Oct4
Rest WT	5′	agtccttcg**TATTGTGATGCAAAT**aggccaatg 3′
Sox2 motif mutations:		
A	5′	agtccttcg**TATaGTGATGCAAAT**aggccaatg 3′
B	5′	agtccttcg**TAaTGTGATGCAAAT**aggccaatg 3′
C	5′	agtccttcg**ggTTGTGATGCAAAT**aggccaatg 3′
D	5′	agtccttcg**TATTtaGATGCAAAT**aggccaatg 3’
E	5′	agtccttcg**TATTGTaATGCAAAT**aggccaatg 3′
Oct4 motif mutations:		
F	5′	agtccttcg**TATTGTGATGCAAta**aggccaatg 3′
G	5′	agtccttcg**TATTGTGATGCttAT**aggccaatg 3′
H	5′	agtccttcg**TATTGTGATcgAAAT**aggccaatg 3′
I	5′	agtccttcg**TATTGTGTaGCAAAT**aggccaatg 3′

B



C

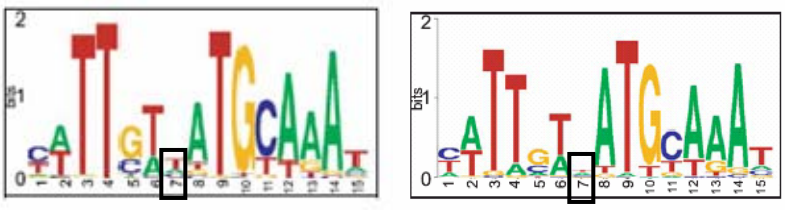


Figure 6.7: Mutations within the Sox2-Oct4 element of *Rest* abolished the Sox2/Oct4-DNA complex. (A) Sequence of the *Rest* composite element and corresponding mutations (lower case and shaded) used in this study. (B) EMSA using 10 μg cell lysate containing Sox2 and Oct4 over-expressed proteins from HEK293 cells with 3 ng wild type (WT) and corresponding probes with mutation. Control denotes 10 μg lysate from HEK293 cell transfected with vector control. EMSA with WT probe detected Sox2-DNA and Oct4/DNA complexes, while probes with Sox2 mutations A, B, C, D did not show the Sox2/DNA complex. EMSA with Sox2 mutation E detected both Sox2/DNA and Oct4/DNA complexes. EMSA with Oct4 mutations F, G, H, I detected low or no Oct4/DNA complex. (C) *De novo* motif discovery of the Sox2-Oct4 element from the Oct4 (left) and Sox2 (right) ChIP-PET data. The height of the alphabets represents frequency of sequence occurrences. Sequences that occur at position 7 are boxed.

A

f-*Ebf1*	5′	**gggggggggg**CATTGTTATGCAAAT**gggggggggg**	3′
f-*Tcf7*	5′	**gggggggggg**TATTGTTATGCAAAT**gggggggggg**	3′
f-*Rest*	5′	**gggggggggg**TATTGTGATGCAAAT**gggggggggg**	3′
f-*Rif1*	5′	**gggggggggg**CTTTGTTATGCACGC**gggggggggg**	3′

B

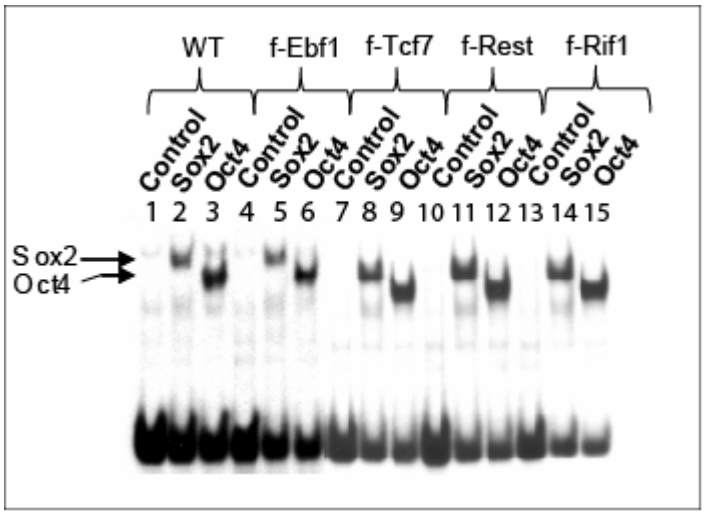


Figure 6.8: Flanking sequences of the Sox2-Oct4 joint motif do not affect Sox2 or Oct4 binding. (A) DNA probes containing mutations (lower case and shaded) in the 9 bp flanking sequences at both sides of the Sox2-Oct4 motifs of *Ebf1, Tcf7, Rest* and *Rif1*. (B) EMSA using 10 μg cell lysate containing Sox2 and Oct4 over-expressed proteins in HEK293 cells with 3 ng wild type (WT) probe (shown here as an example is *Ebf1*) or *Ebf1, Tcf7, Rest, Rif1* probes carrying mutation in the flanking sequences (represented with a prefix f-). Control denotes 10 μg lysate from HEK293 cell transfected with vector control. EMSA with WT probe and probes with flanking sequence mutation detected Sox2-DNA and Oct4-DNA complexes.

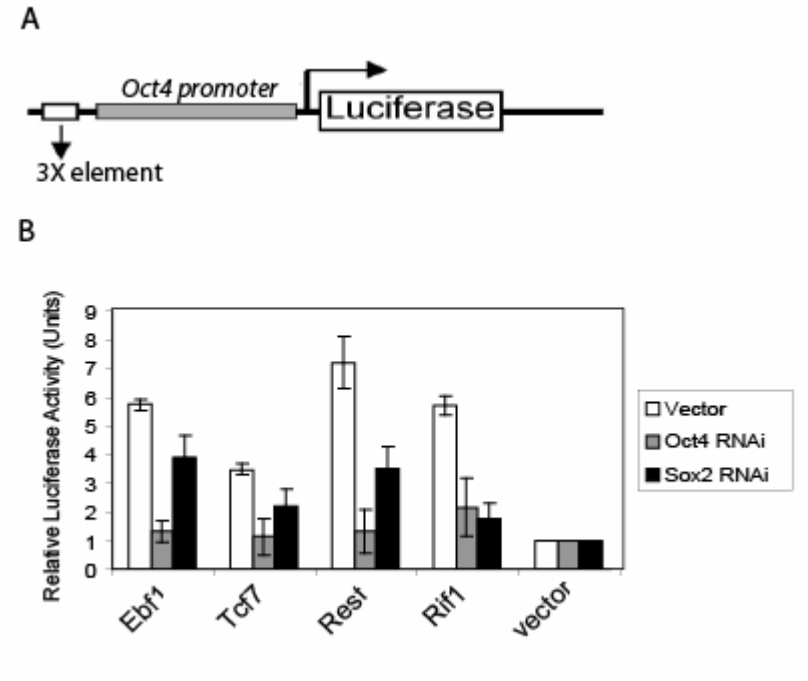


Figure 6.9: Increased reporter activity conferred by the Sox2-Oct4 elements. Three copies of Sox2-Oct4 elements were cloned upstream of the Oct4 promoter in the luciferase plasmid system (A) and transfected into mESCs. Increased reporter activity in Sox2-Oct4 elements-containing plasmids compared to control is shown (B). The enhancer activities are reduced upon knockdown of *Oct4* and *Sox2* demonstrating that the activity is Oct4 and Sox2 specific.

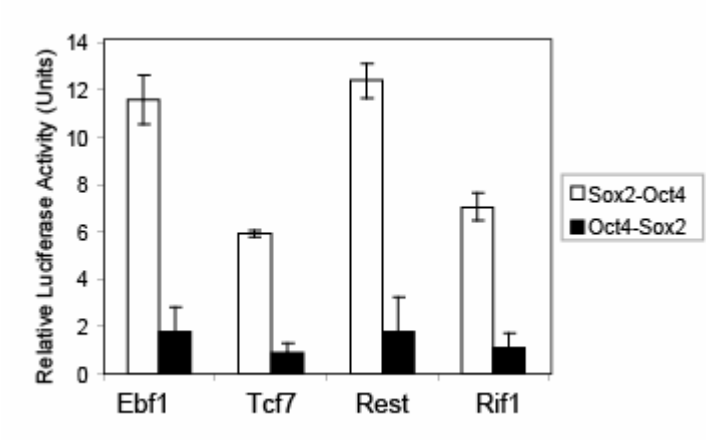


Figure 6.10: Swapping the orientation of Sox2-Oct4 motifs to Oct4-Sox2 abolished enhancer activity. Luciferase activity conferred by Sox2-Oct4 wild type element and Oct4-Sox2 swapped element.

CHAPTER 7

Discovery of Oct4 and Sox2 Collaborating Factors and Demonstrating a Link between Different Pathways in Mouse Embryonic Stem Cells

7.1 Introduction

Many questions regarding the mechanism of pluripotency remain unanswered, such as how are the transcriptional pathways of Oct4, Stat3 and Nanog regulated, and whether there is crosstalk between these pathways. Is there a link between the extrinsic signalling pathways and intrinsic transcription factors? This chapter attempts to answer these questions by utilizing the Oct4 and Sox2 binding site datasets obtained in Chapters 4 and 5.

Nanog is an essential determinant of pluripotency that induces activation of ES state genes, represses visceral-parietal state genes, or both (Figure 7.1). Overexpression of Nanog in mESCs enables maintenance of pluripotency in the absence of LIF (Chambers *et al.*, 2003), suggesting that Nanog may be a downstream target of the LIF signalling pathway. Oct4 is unable to prevent differentiation of mESCs upon LIF withdrawal. This implies that Oct4 and LIF probably activate two different pathways of gene activation, with the latter relying on Stat3 for downstream signalling. On the other hand, both LIF withdrawal and Oct4 up-regulation lead to the same pattern of ESC differentiation, suggesting that there may be crosstalk between the two pathways (Niwa *et al.*, 2000) (Figure 7.1). Niwa *et al.* (2001) proposed that increase in Oct4 expression or LIF withdrawal induces downregulation of a group of genes either by squelching of the co-activators lying downstream of the Stat3 pathway or by downregulation of a Stat3-induced transcriptional programme, which results in a differentiation into the mesoderm-endoderm lineage. Stat3 is hypothesized to either activate ESC state genes, to suppress endodermal-mesodermal genes, or both. Activated Stat3 and subtle changes in Oct4 expression go hand in hand in the maintenance of a pluripotent ESC fate. In this model proposed by Niwa *et al.* (2001), Stat3 may

activate the expression of an Oct4 partner to maintain ESC self-renewal. This model is supported by the existence of the E1A-like activities postulated to exist in ESCs (La Thangue and Rigby, 1987), which may likely represent the mentioned coactivators.

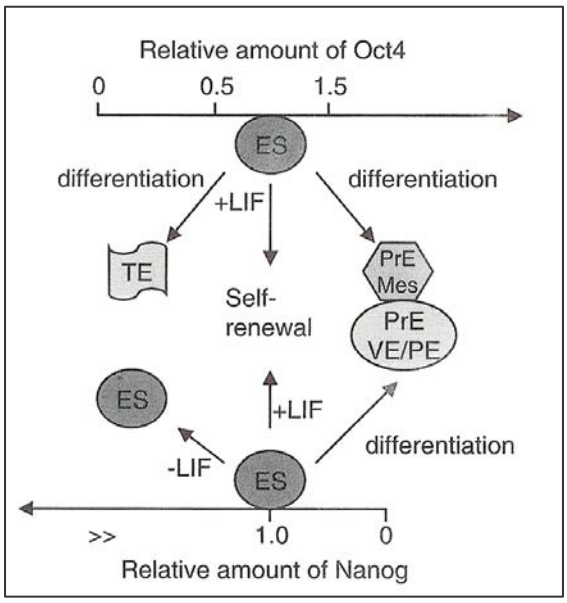


Figure 7.1: Model of the integrated roles of Oct4, Nanog and LIF (Stat3) on embryonic stem cell fate specification, according to different Oct4 and Nanog levels. (Mes: mesoderm, PE: parietal endoderm, PrE: primitive endoderm, TE: trophectoderm, VE: visceral endoderm). Figure was adapted from Lanza (2004).

Transcription factors are known to form multi-protein complexes on DNA, thereby orchestrating the precise temporal and spatial expression of specific genes. Therefore, Oct4 and Sox2 are likely to work in tandem with other co-factors in maintaining mESC pluripotency and self-renewal. To study these questions, a computational method was used to search for potential co-motifs that are present in the vicinity of Sox2-Oct4 overlapping binding sites. Two extrinsic signalling pathway proteins, Stat3 and Smad1, were identified as Oct4 and Sox2 collaborative factors by various methods: ChIP-qPCR, ChIP-on-chip, seqChIP, co-IP, RA induction of cell differentiation as well as RNAi-mediated depletion of Oct4 and Sox2. Subsequently, Stat3-

depletion via ligand removal in a serum-free culture system was carried out to study the roles of Stat3 on the regulation of *Oct4*, *Stat3* itself and *Nanog*. The collective results suggest a mechanistic model that links the Oct4, Stat3 and Nanog transcriptional pathways in governing pluripotency and self-renewal in mESCs.

7.2 Results

7.2.1 Stat3 and Smad1 as Oct4 and Sox2 collaborative factors

7.2.1.1 Expansion of the combined Oct4 and Sox2 ChIP-PET binding data

Previously, the Oct4 and Sox2 ChIP-PET datasets only included binding sites with clusters containing maximum overlap (mo) PET 4 and above as high confidence binding sites. This cutoff threshold was stringent and many true binding sites may have been omitted. Therefore, the Oct4 and Sox2 ChIP-PET datasets were expanded to include moPET of 2 and 3, and were subsequently merged to obtain the dataset of Oct4/Sox2 overlapping binding sites with the help of our Bioinformatics colleague, Vinsensius Vega. A total of 37625 PET clusters contained 2 or more Oct4 moPETs, 16794 clusters contained 2 or more Sox2 moPETs and 1507 clusters contained both Oct4 and Sox2 overlapping moPETs (Figure 7.2). These 1507 clusters represent binding sites shared by both Oct4 and Sox2.

Several binding sites that contained low moPETs from the overlapping 1507 clusters were validated by ChIP-qPCR. As expected, many sites that contain low moPETs also tested positive for binding (Figure 7.3). For example, the cluster chr10.45045909 which contained two Oct4 and two Sox2 moPETs yielded good enrichment of 10-fold for Sox2 ChIP and 30-fold for Oct4 ChIP. The enriched bindings for both Oct4 and Sox2, including those with moPET2 in both datasets

provide confidence that these 1507 overlapping binding sites contain both Oct4 and Sox2 bindings.

7.2.1.2 Matching of binding site sequences against TRANSFAC database identified putative co-motifs

To identify putative transcription factor partners for Oct4 and Sox2, masked sequences (sequences with masked repetitive sequences) from the 1507 overlapping clusters were matched against TRANSFAC matrices (motif sequences) by our Bioinformatics colleague, Vinsensius Vega. TRANFAC is a database containing more than 18000 transcription factor binding site matrices. 500 bp sequences centered on the 1507 high-confidence Sox2-Oct4 binding regions were extracted and scanned for putative transcription factor binding matrices that were provided by TRANSFAC. This scan generated 67 matching matrices (p=0) representing putative motifs present in the Sox2-Oct4 binding site loci (Table 7.1). From this list of putative Oct4 and Sox2 collaborative factors, it was interesting to find that two transcription factors, Stat3 and Smad1, which are activated by signalling pathways, may be working in collaboration with intrinsic factors Oct4 and Sox2.

7.2.1.3 Co-localisation of Stat3 and Smad1 to Oct4 and Sox2 binding sites

To examine whether Stat3 and Smad1 bind to DNA within the vicinity of Oct4-Sox2 binding sites, Stat3 and Smad1 ChIPs were carried out. Both protein G sepharose and Dynal magnetic beads were used to optimize Stat3 and Smad1 ChIPs. Protein G Dynal beads worked more efficiently in these ChIPs (data not shown) and were used throughout this study. Western blots using antibodies against Stat3 and Smad1 detected a band of the correct size (92 kDa and 56 kDa, respectively) in mESC chromatin extracts (Figure 7.4), indicating that these antibodies are specific.

7.2.1.4 ChIP-on-chip

Stat3 and Smad1 ChIP DNA were hybridized onto Nimblegen customized DNA arrays designed by Dr Ng Huck Hui based on mouse genomic sequences build mm5. The chips (25 x 75 mm each) contain 50-mer probes, 50 bp apart spanning the loci of 200 mESC genes, miRNAs and house keeping genes, 385000 features (16 μm x 16 μm) with an array size of 17.4 mm x 13 mm. The results were compared with that of Oct4-Sox2 sequential ChIP-on-chip. Results demonstrated that Oct4, Sox2, Stat3 and Smad1 bind to the same genomic regions on the *Mycn* and *Sgk* loci as shown by the ChIP-on chip results (Figure 7.5). ChIP-on-chip using a control antibody showed no peaks for these regions (Appendix F).

7.2.1.5 Scanning ChIP-qPCR

Subsequently, Stat3 and Smad1 ChIP-qPCR were carried out on loci containing Oct4 and Sox2 binding sites in order to obtain a scanning profile of the respective binding. Six loci (*Rest, Tbx3, Dido1*, *Mycn*, *Nanog* and *Sgk*), which contained Oct4 and Sox2 binding sites as identified by ChIP-PET and ChIP-on-chIP, were selected. Stat3 and Smad1 were shown to bind to all these six Oct4 and Sox2 binding sites (Figure 7.6), with peaks corresponding to Oct4 and Sox2 binding peaks. All loci showed enrichment of more than 20-fold except for Stat3 binding on *Tbx3* and *Nanog*, which showed lower enrichment. This may be due to a lower binding affinity.

7.2.1.6 ChIP-qPCR of Oct4, Sox2, Stat3 and Smad1 on 25 loci

Using primers that gave the highest ChIP enrichments (binding peaks), 25 gene loci were tested for Oct4, Sox2, Stat3 and Smad1 binding using ChIP-qPCR. A colour-coded heat map is shown in Figure 7.7, where the darker regions represent higher ChIP enrichments. Two independent antibodies for Stat3 (Stat3a,b) and Smad1 (Smad1a,b) were used in the ChIP assays. Stat3 ChIP showed positive enrichment on all loci bound by Oct4 and Sox2, while Smad1 showed enrichment for all with the exception of *Etv1*, *Tulp4* and *Zfp64*.

Smad1 has been implicated in influencing the expression of *Id* genes (Ying *et al.*, 2003) whereas *Zfp57* and *Myc* have previously been identified as candidates of Stat3 targets (Akagi *et al.*, 2005; Cartwright *et al.*, 2005). This further validates the results here which included *Id3* as a Smad1 binding target and *Zfp57* as a Stat3 binding target. Control ChIPs using mock antibodies (GFP, cFos, NFκB, JunB) and the p53 antibody did not yield any significant enrichment for all the 25 loci. Control genomic regions tested for binding of these factors (Test regions 1-5 and *p21* loci) showed that Oct4, Sox2, Stat3 and Smad1 did not bind to these regions. p53 has been shown to be highly expressed in mESCs (Sabapathy *et al.*, 1997). As *p21* is a known target of p53, it was used as a positive control locus to ensure that the p53 ChIP worked.

The genes associated with binding of all the four factors (Oct4, Sox2, Stat3 and Smad1) are involved in various functions including governing pluripotency, self-renewal, cell cycle and cell survival (Table 7.2). This suggests that Oct4 and Sox2 collaborate with Stat3 and Smad1 to maintain pluripotency and self-renewal of mESCs.

When all these 25 loci were scanned for the presence of Stat3 and Smad1 motifs with the help of our Bioinformatics colleague Vinsensius Vega, Stat3 motifs were found in the *Atbf1*, *Esrrb* and *Trp53* loci while Smad1 motifs were present in the *Ctnnbl1*, *Esrrb*, *Id3*, *Jarid2*, *mir296*, *Mycn*, *Nanog*, *Rest*, *Rif1* and *Sgk* loci. All loci that contained the motifs tested positive for the binding of the respective factors. However, control regions *p21*, test regions 1, 3 and 4 were also found to contain the Smad1 motif. This may be due to the possibility that the TRANSFAC motif sequence is not very defined (AGACNBCNN for V$SMAD_Q6 and NNNTTCCN for V$STAT3_02), increasing the probability of identifying false positives in the computational search. Thus, the computational method only provides a rough guide in the search for putative co-motifs. Another possibility is that Oct4 and Sox2 binding may be a prerequisite for Stat3 and Smad1 to bind, as none of the control regions contain Oct4, Sox2, Stat3 and Smad1 binding although some may

contain the Smad1 motif. Experiments involving the differentiation of mESCs and depletion of Oct4 and Sox2 were subsequently carried out to further support the notion of Stat3 and Smad1 binding dependency on Oct4 and Sox2.

7.2.1.7 Co-occupancy of Oct4 with Stat3 and Smad1 on the same DNA molecule

To further examine if Stat3 and Smad1 are binding onto the same DNA molecule as Oct4, sequential ChIPs (seqChIPs) using Oct4 followed by Stat3 or Smad1 antibodies were carried out. Significant enrichment was observed after Oct4-Stat3 and Oct4-Smad1 seqChIPs, as compared to low enrichment in the Oct4-control seq ChIPs (Figure 7.8). This indicates that Stat3 and Smad1 are binding to the same molecule as Oct4. This experiment omits the possibility that Oct4 and Sox2 may bind independently on the same DNA region of different DNA molecules.

7.2.1.8 Retinoic acid differentiation affects Stat3 and Smad1 binding

To examine the effect of Stat3 and Smad1 binding upon induction of differentiation, mESCs were treated with retinoic acid (RA) as described in Chapter 2. Upon RA induction, the cells begin to differentiate. Oct4 and Sox2 protein could not be detected in RA-treated cells (Figure 7.9A). Concurrently, binding of Oct4 and Sox2 on the *Nanog* locus was also reduced (Figure 7.9B). In mESCs treated with RA, Stat3 and Smad1 protein levels remain relatively unchanged (Figure 7.9C). However, Stat3 and Smad1 bindings on *Nanog*, *Mycn*, *Sgk*, *Dido1*, *Tbx3* and *Rest* loci were abolished (Figure 7.9D,E). Reduction in Stat3 and Smad1 binding were also observed in ChIP experiments using embryoid bodies and mESCs grown without LIF (data not shown). These results indicate that Stat3 and Smad1 binding on these loci are specific in undifferentiated mESCs. In addition, it also suggests that Stat3 and Smad1 could no longer bind to these loci due to the absence of Oct4 and Sox2 binding.

7.2.1.9 RNAi-mediated depletion of Oct4 and Sox2 affects Stat3 and Smad1 binding

To investigate whether Stat3 and Smad1 bindings are dependent on Oct4 and Sox2, mESCs were depleted of Oct4 and Sox2 proteins and were assayed for Stat3 and Smad1 binding. Cells were transfected with RNAi plasmids (shRNA) that targeted Oct4 and Sox2. The cells were then cultured in media containing puromycin to remove non-transfected cells. Cells transfected with vector plasmid and plasmid carrying the luciferase gene did not show any significant differentiation (Figure 7.10a,b). Knockdown of Oct4 and Sox2 caused the cells to flatten and change morphology into distinct individual epithelial-like cells (Figure 7.10Ac,d). All non-transfected cells died after 2 days of puromycin selection (Figure 7.10Ae). Oct4 and Sox2 protein levels were also shown to be reduced significantly in RNAi-treated cells (Figure 7.10B). A slight reduction in Sox2 protein levels was also observed in Oct4 RNAi-treated cells (Figure 7.10B). Similarly, there was also a slight reduction in Oct4 protein levels in Sox2 RNAi-treated cells (Figure 7.10B). These observations can be attributed to the indirect effects of RNAi as Oct4 and Sox2 reciprocally regulates each other in a positive manner. However, Oct4 and Sox2 RNAi-mediated depletion can still be used as a model for study as the primary effects caused by the specific shRNAs were much more significant. As expected, ChIP-qPCR showed that upon Oct4 and Sox2 RNAi, Oct4 and Sox2 binding reduced significantly on the *Mycn* and *Nanog* loci as compared to cells treated with luciferase shRNA (Figure 7.10C).

The protein levels of Stat3 and Smad1 were noted to remain relatively unchanged in both Oct4 and Sox2 knockdown cells (Figure 7.11A). The same shRNA-treated cells that were selected for 2 days were crosslinked and harvested for ChIP. Results show that although their protein levels remain unchanged, Stat3 and Smad1 binding on the *Mycn* and *Nanog* loci were reduced significantly in Oct4 and Sox2-depleted cells, as compared to the cells treated with control *luciferase* shRNA (Figure 7.11B). The RNAi-ChIP experiment was repeated for cells selected for 4 days in puromycin-containing media. Similarly, Stat3 and Smad1 binding were reduced

significantly on the *Mycn* and *Nanog* loci in Oct4 and Sox2 depleted mESCs (Figure 7.11C). Stat3 and Smad1 were not present in the Oct4 and Sox2 ChIP-PET datasets, indicating that they are not regulated by Oct4 and Sox2. These data suggests that Stat3 and Smad1 binding may be dependent on the presence of Oct4 and Sox2.

7.2.1.10 Stat3 and Smad1 are Oct4 protein partners

To examine whether Stat3 and Smad1 are Oct4 protein partners, co-immunoprecipitation (co-IP) was carried out using nuclear extracts (input) from undifferentiated mESCs. IP carried out using Oct4 antibodies could pull down all protein complexes containing Oct4, and subsequently these complexes can be analysed by Western blotting. As control, the Oct4 protein was detected in the Oct4 immunoprecipitate (Figure 7.12A). Smad1, Stat3 and their phosphorylated forms were also detected in Oct4 immunoprecipitates, while none of these proteins were detected in GST immunoprecipitates (control IP) (Figure 7.12B). This indicates that Stat3 and Smad1 are present in Oct4 protein complexes, and may likely be Oct4 partners in mESCs. A limitation of the co-IP experiment is that it does not discriminate between direct and indirect protein-protein interactions.

7.2.2 The connection between Stat3, Oct4 and Nanog pathways

7.2.2.1 Culturing mESCs in defined media containing BMP4 and LIF

To investigate the role of Stat3 binding, depletion of Stat3 via ligand removal was carried out. Cells were first acclimatized in serum-free media containing BMP4 and LIF (Clonal media from Chemicon; as optimsed by Ying *et al.* 2003). After 3 passages, mESCs form densely packed colonies (Figure 7.13Ab). A low density colony forming assay was performed using $1x10^3$ cells in a 10 cm culture dish. After 5 days of culture, densely packed colonies emerged (Figure 7.13Aa), verifying the quality of the mESCs grown in this serum-free media. The mESCs could be cultured in Clonal media for more than 6 passages.

7.2.2.2 Stat3 depletion in mESCs

Cells cultured in serum-free clonal media were then treated with basal media containing LIF, BMP4 or both. The morphology of cells cultured in different media for 24 hr is shown in Figure 7.13B. Some differentiation was observed in cells cultured in Basal media, while cells cultured in LIF, BMP4, LIF and BMP or Clonal media still formed densely packed colonies and showed no signs of differentiation after 24 hr. Cells grown in LIF, BMP4 and basal media all differentiated after 36 hr (data not shown).

Cell lysates were obtained from mESCs grown in Clonal media, Basal media and Basal media containing 1000 U/ml LIF, 30 ng/ml BMP4, 50 ng/ml BMP4 or a combination of 1000 U/ml LIF and 30 ng/ml BMP4 at 15 min, 2 hr, 6 hr and 12 hr. Western blot using these lysates were carried out to detect phosphorylated Stat3 (p-Stat3) proteins (Figure 7.14). Significant reduction of p-Stat3 proteins was observed in cells cultured without LIF, that is in media with BMP4 alone and Basal media (Figure 7.14, lanes 3, 4, 6, 7, 10, 12, 13, 15, 16). Stat3 is a substrate in the LIF-Stat3 signalling pathway. According to the LIF/STAT3 pathway model, removal of the LIF ligand may not allow Stat3 to be phosphorylated or translocated into the nucleus, and thus reduced the levels of phosphorylated Stat3 in cells cultured without LIF.

7.2.2.3 Stat3 binds to but does not regulate *Oct4*

Using mESC cultured in defined media for 24 hr, the effects of Stat3 depletion on *Oct4* was examined. Stat3 ChIP was carried out using crosslinked chromatin extracts from cells treated with (i) Clonal, (ii) Basal, (iii) Basal media with 50 ng/ml BMP4, (iv) Basal media with 1000 U/ml LIF and (v) Basal media with 30 ng/ml BMP4 + 1000 U/ml LIF. ChIP-qPCR of Stat3 on the *Oct4* loci showed high enrichment (15-25 fold) using extracts from Clonal, LIF + BMP4, and LIF alone, indicating that Stat3 binds to the *Oct4* loci in these cells (Figure 7.15A). In cells grown without LIF (BMP and Basal media), Stat3 binding was significantly lower (about 5-fold), indicating that

Stat3 binds to the *Oct4* loci in the presence of LIF (Figure 7.15A). However, the expression level of Oct4 remains relatively unchanged in all these cells as shown by RT-PCR analysis of Oct4 transcript levels (Figure 7.15B). Only after 24 hr, Oct4 transcript level in cells grown in Basal media started to show some reduction compared to the other samples. This is probably due to differentiation of mESCs. These results collectively indicate that although Stat3 binds to the *Oct4* loci, it does not regulate Oct4 transcription.

7.2.2.4 Stat3 binds to its own gene, providing a model for autoregulation

Next, to examine the effects of Stat3 depletion on its own regulation, ChIP-qPCR of Stat3 on the *Stat3* loci was carried out. Enrichment of 150-250 folds were detected in Clonal, LIF+BMP and LIF samples, indicating that Stat3 binds to the *Stat3* loci in cells grown in the presence of LIF. In cells grown without LIF (BMP4 only and Basal media), enrichment for Stat3 binding was comparatively lower (about 25-fold) (Figure 7.16A). Stat3 transcript levels were then measured. The expression of Stat3 remained relatively unchanged for Clonal, LIF+BMP and LIF samples (Figure 7.16B). On the other hand, *Stat3* transcript levels were significantly reduced in cells grown in BMP4 and basal media after 1hr (Figure 7.16B). Therefore, Stat3 binding and expression levels reduced concurrently in mESCs grown without LIF, suggesting that Stat3 regulates its own gene.

7.2.2.5 Stat3 regulates *Nanog*

Oct4 ChIP-qPCR on the *Nanog* loci did not show significant difference in fold enrichment for all treated samples, indicating that Oct4 binds to *Nanog* in the presence or absence of LIF (Figure 7.17A). To test the effects of Stat3 depletion on *Nanog*, Stat3 ChIP-qPCR on the *Nanog* loci was carried out. ChIP enrichment of 25-40 fold was detected in Clonal, LIF+BMP and LIF samples, indicating that Stat3 binds to the *Nanog* loci in cells grown in media containing LIF (Figure 7.17B). Enrichment was low (<10-fold) for Stat3 on the *Nanog* loci in cells grown without LIF (BMP4 and Basal media) (Figure 7.17B). RNA Polymerase II ChIP-qPCR on the *Nanog* loci was

also carried out and results showed that ChIP enrichment was approximately 16-24 fold in Clonal, LIF+BMP and LIF samples, and less than 3 fold in BMP4 and Basal media samples (Figure 7.17C). RNA Polymerase II binding pattern correlated with the Stat3 binding pattern, indicating that RNA Polymerase II and Stat3 were present together in cells grown in the same condition. The presence of RNA Polymerase II on *Nanog* in Clonal, LIF+BMP and LIF cells indicates that *Nanog* is being actively transcribed in these cells as the occupancy of RNA polymerase II is tightly correlated to the level of transcription activity and is a strong indicator of *in vivo* transcription elongation (Knight *et al.*, 2003).

Subsequently, the kinetics of Nanog expression in these cells were measured. There was a significant drop in the *Nanog* transcript levels at 6 hr and 24 hr (Figure 7.17D). The reduction in Nanog transcripts is correlated with the reduction in RNA polymerase II binding in these cells. These results suggest that Stat3 binds to and regulates the expression of *Nanog*. Oct4 and Sox2 have also been shown to partly regulate *Nanog* (Kuroda *et al.*, 2005; Rodda *et al.*, 2005); hence Stat3, Oct4 and Sox2 may work in collaboration to regulate *Nanog*. Oct4 and Sox2 binding may be a pre-requisite for Stat3 binding, and thus, regulation of *Nanog*.

7.3 Discussion

7.3.1 Cofactors collaborating with Oct4 and Sox2 in *cis*-regulatory modules

Oct4 and Sox2 may carry out their functions by collaborating with other factors. Motif search through TRANSFAC identified that two transcription factor binding sites, Stat3 and Smad1 were significantly enriched in the subset of Oct4-Sox2 overlapping binding sites. These extrinsic signalling pathway factors were distinguished after screening of the Oct4-Sox2 overlapping PET2+ clusters from mouse transcription factor motifs in the TRANSFAC database (version 9.1) (Wingender *et al.*, 2000). Results revealed 67 putative transcription factor matrices (p=0). The

large number of potential co-motifs identified using computational methods can only serve as a guide for forming biological hypotheses.

Stat3 and Smad1 are transcription factor substrates of extrinsic signalling pathways which are important in mESC self-renewal, while Oct4 and Sox2 are intrinsic key transcription factors essential in maintaining the pluripotency of mESC. Results from ChIP-qPCR and ChIP-on-chip indicated that Stat3 and Smad1 may co-localise to the same binding regions of Oct4 and Sox2, as shown on 25 gene loci. The genes targeted by the four factors play important roles in maintaining both self-renewal and pluripotency of mESCs. Further evidence showed that Oct4 co-occupied the same ChIP DNA fragment as Stat3 and Smad1, and that Stat3 and Smad1 are partners of Oct4 in mESCs. Subsequently, RA-induced differentiation and RNAi-depletion of Oct4 and Sox2 assays provided evidence that Stat3 and Smad1 bindings may be dependent on the presence of Oct4 and Sox2 on the genomic sites. Collectively, these results strongly suggest that Stat3 and Smad1 are cofactors collaborating with Oct4 and Sox2 in *cis*-regulatory modules.

7.3.2 Molecular mechanisms in the maintenance and differentiation of mESCs

This study presented strong evidence that Stat3 regulates itself. Autoregulation is one of the key features of transcription factors positioned at the top in the hierarchy of the transcriptional regulatory network. They act to regulate other transcription factors as well as auto-regulate themselves in order to maintain their own transcript level. Stat3, being a downstream factor of extrinsic LIF signalling pathway, may act to regulate intrinsic factors that are crucial in maintaining undifferentiated mESCs. Concurringly, a LIF dose-response study indicated the presence of a positive feedback loop in mESCs, whereby Stat3 activation may control the expression of Stat3, gp130 and LIFR (Davey *et al.*, 2007). Stat3 autoregulation was also shown in murine myeloid leukemic cells (Ichiba *et al.*, 1998).

Depletion of Stat3 (or LIF) and Nanog, as well as overexpression of Oct4, led to the differentiation of mESC into the endodermal lineage (Figure 7.1). Results from this study showed that Stat3 co-binds with Oct4 on many ESC genes, including *Nanog*. This study also demonstrated that depletion of Stat3 down-regulated *Nanog* expression. In concurrence, previous studies have also showed that over-expression of Nanog can maintain mESCs in the absence of LIF, which is a ligand for activation of Stat3 in the LIF-Stat3 signalling pathway. Therefore, a model linking Oct4, Stat3 and Nanog is proposed whereby Stat3 regulates itself and binds together with Oct4 to regulate *Nanog*.

Since both the inhibition of Stat3 activity and the overexpression of Oct4 stimulate mESCs to differentiate into primitive endoderm-like cells (Figure 7.1) (Niwa *et al.*, 1998; Niwa *et al.*, 2000), the function of Stat3 could be disrupted by an excess level of Oct4, which might act to block Stat3 from binding to the neighbouring genomic site via the saturation of protein interactions. This will then disrupt the functions of the ternary complex (consisting of Oct4, Stat3 and a general transcription unit, which activates target genes). This mechanism is substantiated by the evidence that the over-dosage effect of Oct4 on mESC differentiation does not require Oct4 DNA-binding activity (Niwa *et al.*, 2002).

Nanog has been reported to prevent mESCs from differentiating into primitive endoderm by repressing the trigger factor, *Gata6* (Chambers *et al.*, 2003; Mitsui *et al.*, 2003). Moreover, *Gata6* promoter sequences are bound by Nanog (Wang *et al.*, 2006) and *Nanog*-null mESCs was shown to differentiate into Gata6-positive parietal endoderm-like cells, which have a morphology that is similar to that of Gata6-induced cells (Mitsui *et al.*, 2003). In this model, depletion of Stat3 down-regulates Nanog which then may allow the cells to differentiate towards the endodermal lineage. This study provides evidence of crosstalk and interdependence between the Oct4, Stat3 and Nanog pathways in mESCs.

Another group showed similar conclusion of Stat3 regulating *Nanog*, but they identified Stat3 binding and Stat3 motif at about 5 kb upstream from the transcription start site (Suzuki *et al.*, 2006) (Figure 7.18A). In contrast, results from this chapter demonstrated Oct4, Sox2, Stat3 and Smad1 binding near the Sox2-Oct4 motif, which is about 500bp upstream from the transcription start site. Integrating these findings with evidence that Stat3 binding may be dependent on Oct4 (and Sox2 binding); a conformational model is derived whereby a loop containing these factors may be formed during transcription of this gene (Figure 7.18B). These results indicate that Oct4 and Sox2 have multiple DNA binding sites for a single transcription factor molecule and work in collaboration with other transcription factors in gene regulation.

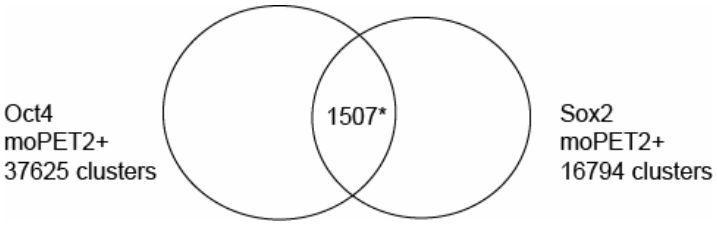


Figure 7.2: Expansion of the combined ChIP-PET data. Clusters containing maximum overlapping PET (moPET) 2 or more from both Oct4 and Sox2 binding datasets were merged by Vinsensius Vega (Bioinformatics, GIS). An overlap of 1507 clusters was obtained assuming that the maximum overlapping span is ~10 kb. *The p-value of observing 1507 Oct4/Sox2 overlapping clusters is 1.687927e-75.

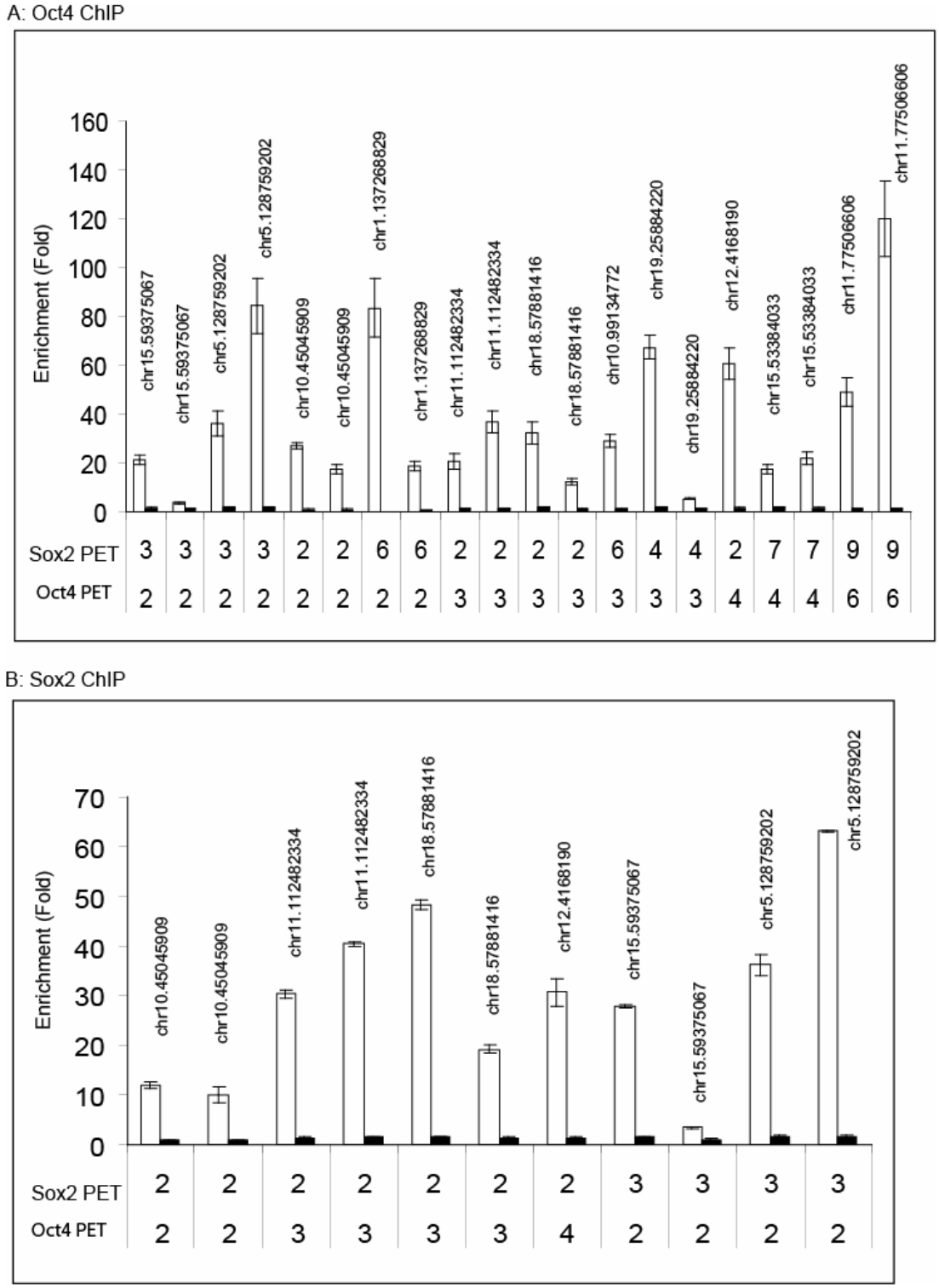


Figure 7.3: Validation of Oct4 and Sox2 overlapping binding sites containing low PET overlaps. Oct4 (A) and Sox2 (B) ChIP followed by qPCR at sites containing 2 or more PET overlaps were carried out. Open (white) bars represent Oct4 or Sox2 ChIP and solid (black) bars represent control GST ChIP. Standard deviations represent technical replicates.

Table 7.1: List of putative Oct4 and Sox2 co-occurring motifs denoted by their TRANSFAC matrix ID. Oct4 and Sox2 overlapping ChIP-PET masked sequences were matched against the TRANSFAC database. 67 potential co-occurring motifs (matrix ID) with *p*-values representing the expected rate of occurrence per random bp calculated as zero are listed. The list represents binding motif of transcription factor and its variation according to the TRANSFAC database.

1	KROX_Q6	35	MYOGENIN_Q6
2	GC_01	36	HNF3_Q6_01
3	SP1_Q6	37	CACBINDINGPROTEIN_Q6
4	SP1_Q4_01	38	LEF1_Q2
5	SP1_Q2_01	39	OCSBF1_01
6	VDR_Q3	40	TEF1_Q6
7	MAZR_01	41	SREBP1_Q6
8	SP1_Q6_01	42	HNF4_DR1_Q3
9	PAX4_03	43	DR1_Q3
10	MAZ_Q6	44	SGF3_Q6
11	SP1_01	45	ZIC3_01
12	Alfin1_Q2	46	SMAD_Q6
13	ZF5_01	47	HNF4_01
14	AP2_Q3	48	HNF3ALPHA_Q6
15	OCT1_Q6	49	PPAR_DR1_Q2
16	E2F_Q2	50	TST1_01
17	OCT_Q6	51	FOXD3_01
18	OCT1_04	52	MSX1_01
19	HSF_01	53	HNF3_Q6
20	SPZ1_01	54	MYC_Q2
21	OCT1_Q5_01	55	AP2REP_01
22	CHCH_01	56	HNF4_Q6_01
23	AP2_Q6_01	57	MAF_Q6_01
24	SRY_01	58	STAT3_02
25	ZF5_B	59	FOX_Q2
26	ADR1_01	60	GEN_INI2_B
27	MUSCLE_INI_B	61	CPRF2_01
28	MZF1_02	62	EVE_Q6
29	GEN_INI3_B	63	GAGAFACTOR_Q6
30	HSF_01	64	USF2_Q6
31	EGR_Q6	65	FAC1_01
32	SF1_Q6	66	NRF1_Q6
33	GEN_INI_B	67	NRSF_Q4
34	E2F1_Q3		

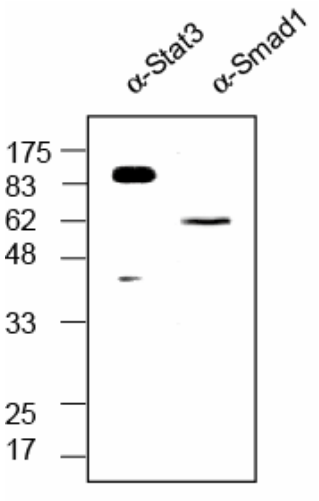


Figure 7.4: Specificity of Stat3 and Smad1 antibodies used. Western blot analysis of mESC chromatin extracts using antibodies against Stat3 (92kDa, Santa Cruz C-20) and Smad1 (56kDa, Santa Cruz A-4). Markers are shown in kDa.

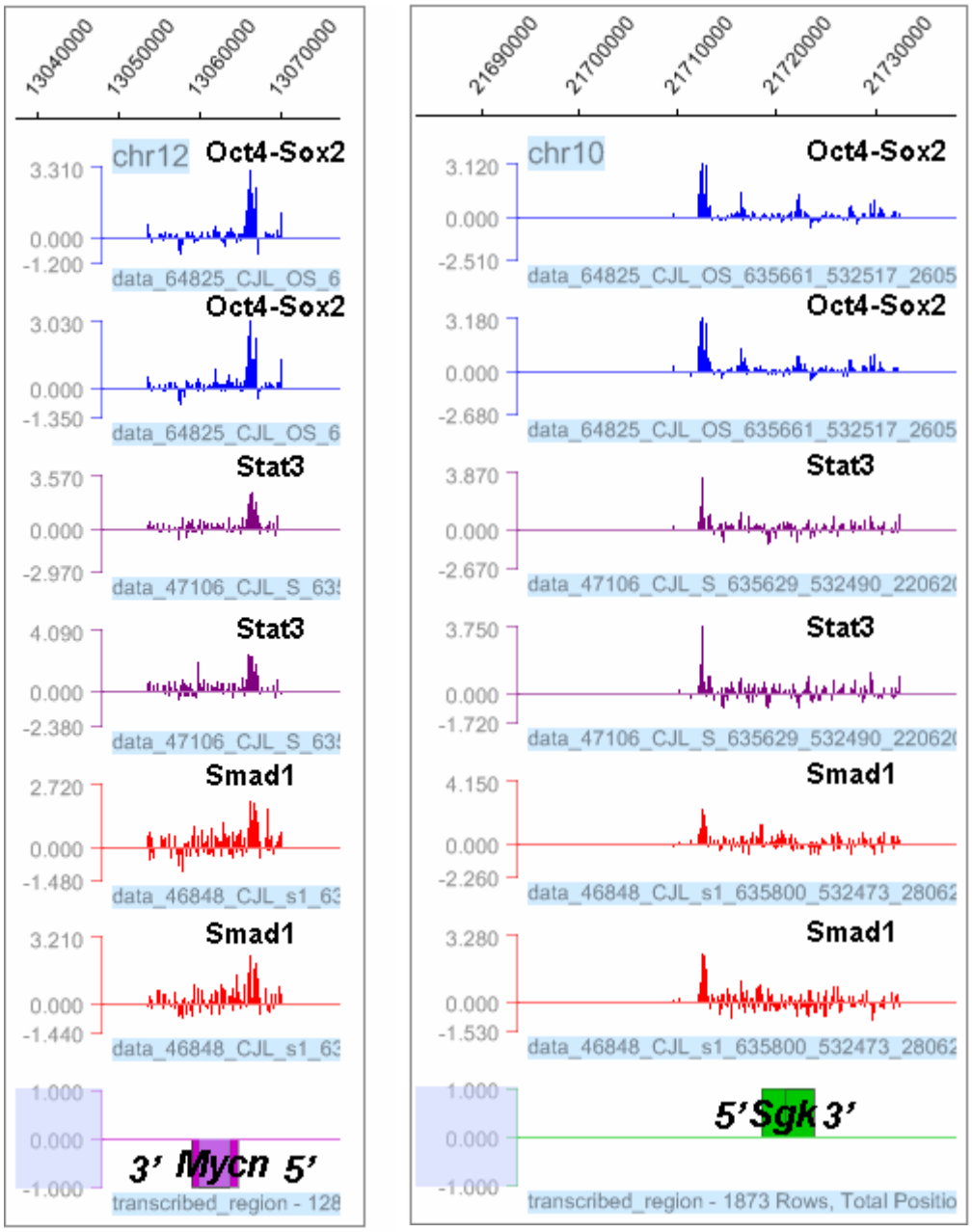
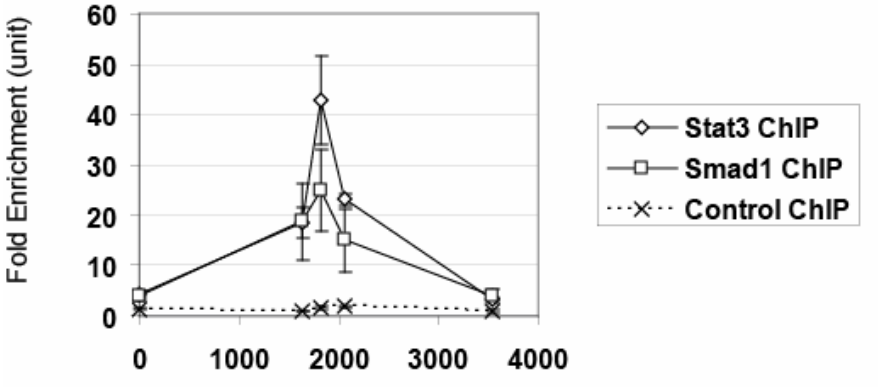
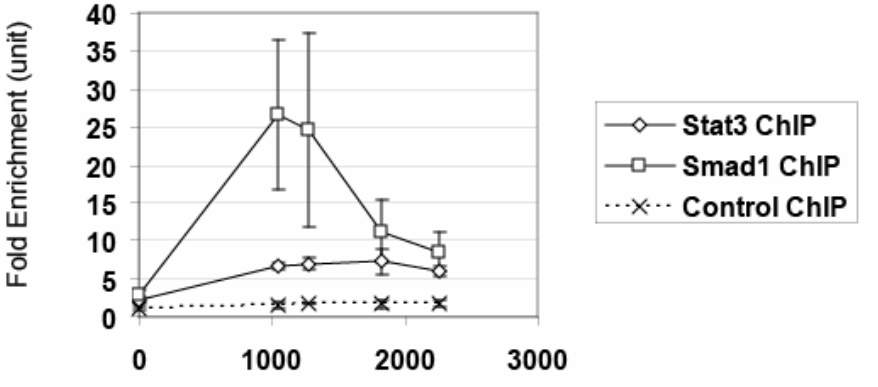


Figure 7.5: Co-localisation of Oct4-Sox2, Stat3 and Smad1. ChIP-on-chip SignalMap diagram showing co-localisation of Oct4-Sox2, Stat3 and Smad1 on the *Mycn* and *Sgk* loci. Oct4 and Sox2 sequential ChIP, Stat3 and Smad1 ChIP were hybridized on the Nimblegen DNA microarray. Peaks represent enrichment of ChIP DNA on the loci (transcription factor binding on that particular region). Block replicates for each ChIP-on-chip are shown. Control ChIP-on-chip is shown in Appendix F.

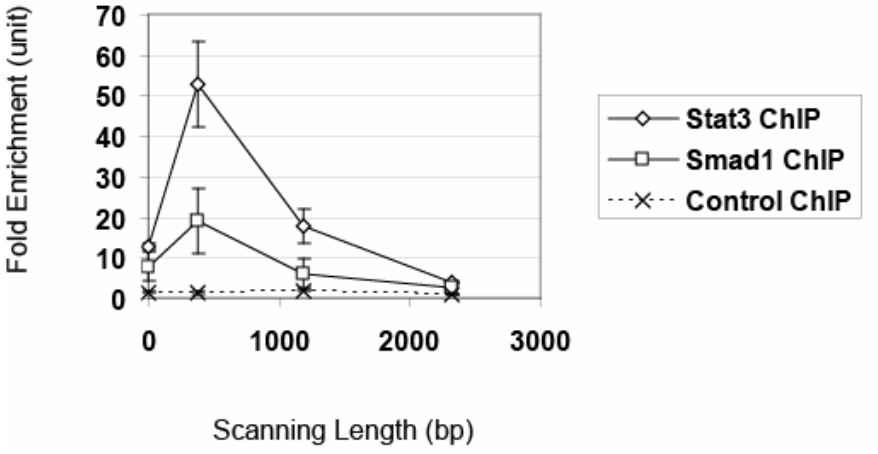
A: *Rest*



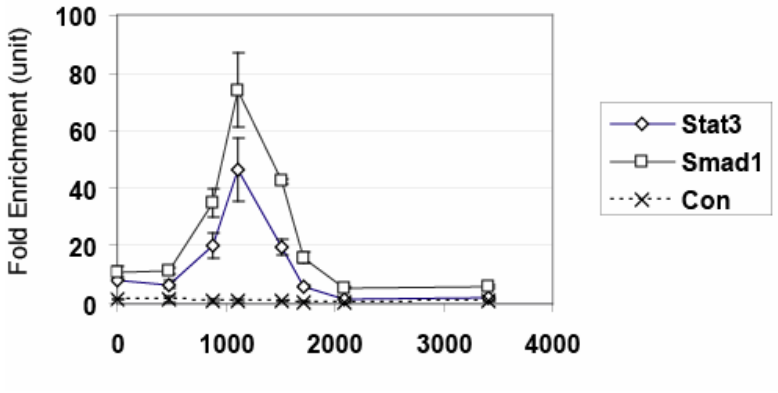
B: *Tbx3*



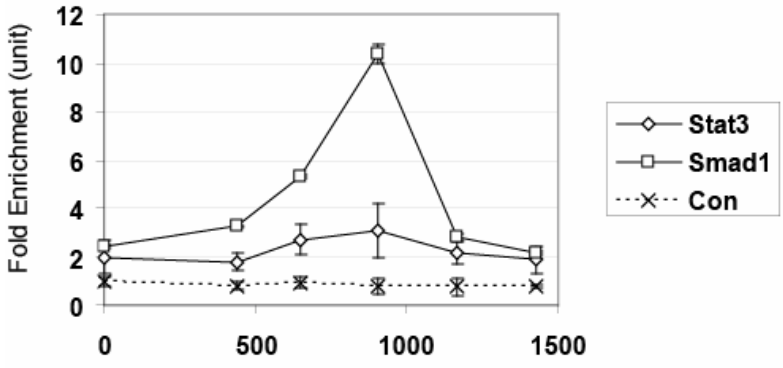
C: *Dido1*



D: *Mycn*



E: *Nanog*



F: *Sgk*

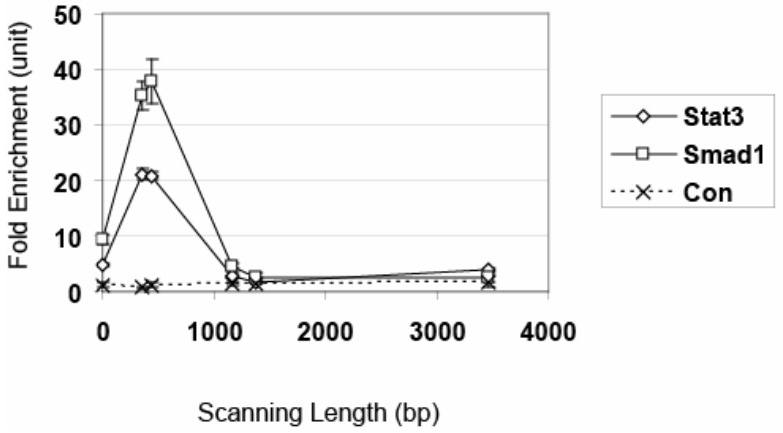


Figure 7.6: Co-localisation of Stat3 and Smad1 on Oct4 and Sox2 overlapping binding sites. ChIP-qPCR scanning of putative collaborating transcription factors at loci bound by Oct4 and Sox2. Binding profiles of Stat3 (open diamond) and Smad1 (open square) on A: *Rest*, B: *Tbx3*, C: *Dido1*, D: *Mycn*, E: *Nanog*, F: *Sgk* loci. Control ChIP (cross) using the GST antibody is also shown. Standard deviations represent biological replicates. Axis y represents coordinates relative to the most upstream amplicon. All coordinates are shown in Appendix G.

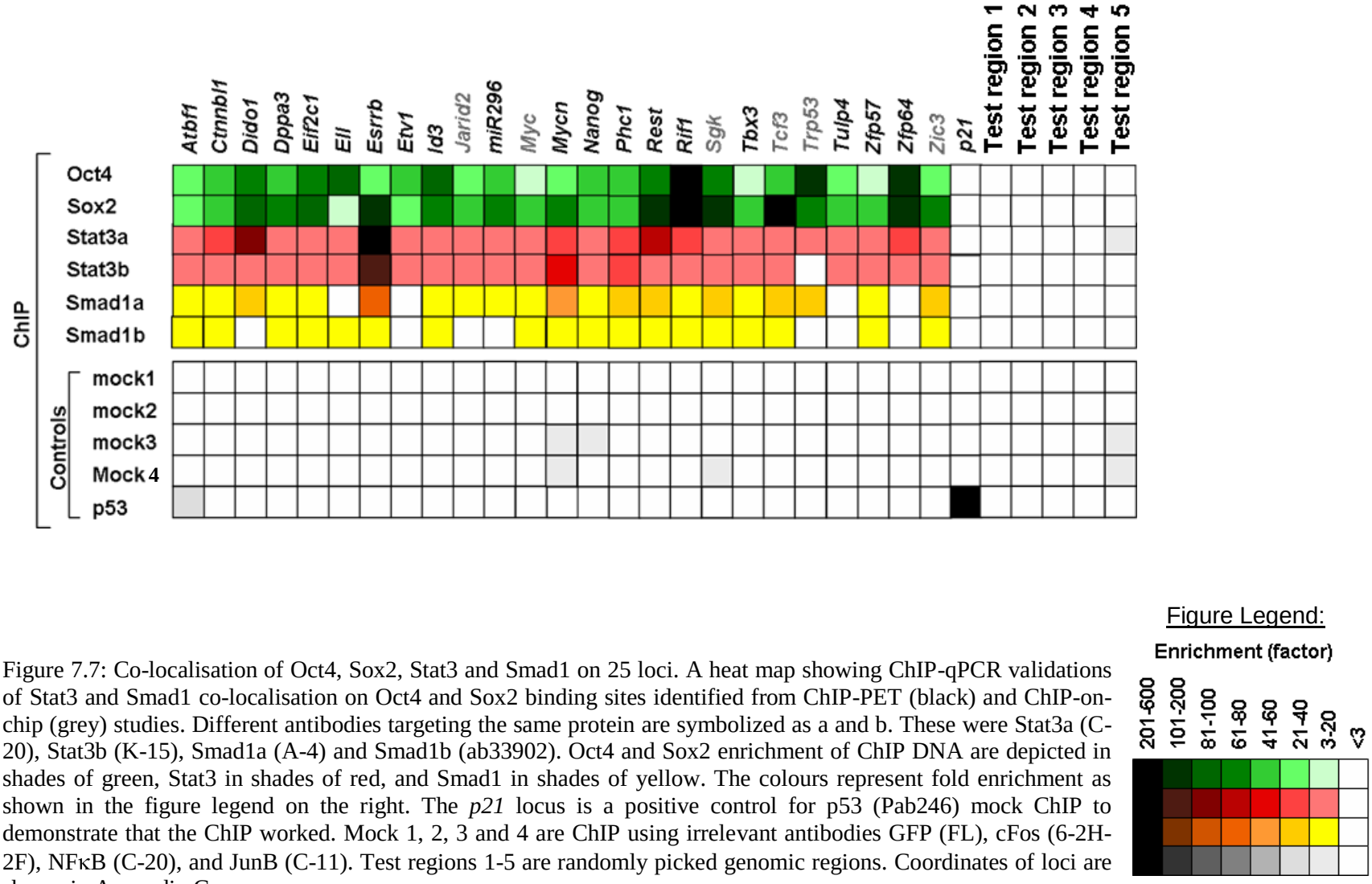


Figure 7.7: Co-localisation of Oct4, Sox2, Stat3 and Smad1 on 25 loci. A heat map showing ChIP-qPCR validations of Stat3 and Smad1 co-localisation on Oct4 and Sox2 binding sites identified from ChIP-PET (black) and ChIP-on-chip (grey) studies. Different antibodies targeting the same protein are symbolized as a and b. These were Stat3a (C-20), Stat3b (K-15), Smad1a (A-4) and Smad1b (ab33902). Oct4 and Sox2 enrichment of ChIP DNA are depicted in shades of green, Stat3 in shades of red, and Smad1 in shades of yellow. The colours represent fold enrichment as shown in the figure legend on the right. The *p21* locus is a positive control for p53 (Pab246) mock ChIP to demonstrate that the ChIP worked. Mock 1, 2, 3 and 4 are ChIP using irrelevant antibodies GFP (FL), cFos (6-2H-2F), NFκB (C-20), and JunB (C-11). Test regions 1-5 are randomly picked genomic regions. Coordinates of loci are shown in Appendix G.

Table 7.2: Function of genes associated with the binding of Oct4, Sox2, Stat3 and Smad1 in mouse embryonic stem cells as shown by ChIP-qPCR and annotated in NCBI.

Symbol	Gene full name	mRNA	Unigene	Function
Atf1	activating transcription factor 1	NM_007497	Mm.676	regulation of transcription, DNA-dependent; DNA binding; transcription factor activity; cysteine protease inhibitor activity; protein binding; nucleus; transcription factor complex, neural differentiation
Ctnnbl1	catenin, beta like 1	NM_025680	Mm.45193	molecular_function unknown; nucleus; apoptosis; induction of apoptosis
Dido1	death inducer-obliterator 1	NM_011805	Mm.253836	Apoptosis, regulation of transcription
Dppa3	developmental pluripotency-associated 3	NM_139218	Mm.27982	germ cell marker of pluripotency
Eif2c1	eukaryotic translation initiation factor 2C, 1	NM_153403	Mm.30800	Translation initiation activity
Ell	elongation factor RNA polymerase II	NM_007924	Mm.271973	nucleus; regulation of transcription, DNA-dependent
Esrrb	estrogen related receptor, beta	NM_011934	Mm.235550	DNA binding; transcription factor activity; steroid hormone receptor activity; receptor activity; ligand-dependent nuclear receptor activity; steroid binding; nucleus; transcription; regulation of transcription, DNA-dependent; protein targeting
Etv1	ets variant gene 1	NM_007960	Mm.4866	DNA binding; transcription factor activity; nucleus; regulation of transcription, DNA-dependent; axon guidance; muscle development; mechanosensory behavior
Id3	inhibitor of DNA binding 3	NM_008321	Mm.110	nucleus; protein domain specific binding; protein binding; negative regulation of transcription from Pol II promoter
Jarid2	jumonji, AT rich interactive domain 2	NM_021878	Mm.25059	DNA binding; DNA binding; intracellular; nucleus; nucleus; development; transcriptional repressor activity; negative regulation of transcription, DNA-dependent
mir296	microRNA 296	MI0000394	mmu-mir-296	regulation of development, gene regulation, potential mESC-specific microRNA
Myc	myelocytomatosis oncogene	NM_010849	Mm.2444	induction of apoptosis by intracellular signals; activation of pro-apoptotic gene products; activation of pro-apoptotic gene products; negative regulation of survival gene product activity; negative regulation of survival gene product activity; response to radiation; response to radiation; regulation of cell proliferation; regulation of apoptosis; cellular physiological process; regulation of cell cycle; release of cytochrome c from mitochondria; DNA binding; transcription factor activity; protein binding; nucleus; spindle; DNA fragmentation during apoptosis; regulation of transcription, DNA-dependent; caspase activation; induction of apoptosis by intracellular signals
Mycn	neuroblastoma myc-related oncogene 1	NM_008709	Mm.16469	regulation of cell cycle; DNA binding; transcription factor activity; protein binding; intracellular; nucleus; regulation of transcription, DNA-dependent; cellular physiological process
Nanog	Nanog homeobox	NM_028016	Mm.440503	DNA binding; nucleus; stem cell division
Phc1	polyhomeotic-like 1 (Drosophila)	NM_007905	Mm.6822	protein binding; nucleus; development; nuclear body; DNA binding
Rest	RE1-silencing transcription factor	NM_011263	Mm.28840	DNA binding; protein binding; transcriptional repressor activity; transcriptional repressor complex; negative regulation of transcription, DNA-dependent
Rif1	Rap1 interacting factor 1 homolog (yeast)	NM_175238	Mm.254530	Cell cycle, response to DNA damage stimulus, telomere-associated
Sgk	serum/glucocorticoid regulated kinase	NM_011361	Mm.28405	protein kinase activity; protein serine/threonine kinase activity; ATP binding; nucleus; protein amino acid phosphorylation; apoptosis; kinase activity; transferase activity
Tbx3	T-box 3	NM_011535	Mm.219139	DNA binding; transcription factor activity; nucleus; regulation of transcription, DNA-dependent; development; cell aging; negative regulation of transcription; transcriptional repressor activity
Tcf3	transcription factor 3	NM_009332	Mm.440067	Regulation of transcription, DNA-dependent; DNA binding; transcription factor activity; Wnt signalling pathway
Trp53	transformation related protein 53	NM_011640	Mm.222	protein-nucleus import, translocation; DNA binding; transcription factor activity; protein binding; nucleus; cytoplasm; cytosol; transcription; regulation of transcription, DNA-dependent; apoptosis; response to DNA damage stimulus; cell cycle; negative regulation of DNA replication; negative regulation of DNA replication; response to UV; response to X-ray; DNA damage response, signal transduction by p53 class mediator; regulation of cell proliferation; DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis; DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis; negative regulation of apoptosis; negative regulation of cell cycle; negative regulation of fibroblast proliferation
Tulp4	tubby like protein 4	NM_054040	Mm.28251	intracellular signaling cascade
Zfp57	zinc finger protein 57	NM_009559	Mm.305561	nuclear heterochromatin; regulation of transcription, DNA-dependent; negative regulation of transcription from Pol II promoter; nucleic acid binding; nucleus
Zfp64	zinc finger protein 64	NM_009564	Mm.2095	nucleic acid binding; DNA binding; nucleus; regulation of transcription, DNA-dependent; development; zinc ion binding
Zic3	zinc finger protein of the cerebellum 3	NM_009575	Mm.255890	nucleic acid binding; DNA binding; nucleus; regulation of transcription, DNA-dependent; pattern specification; zinc ion binding

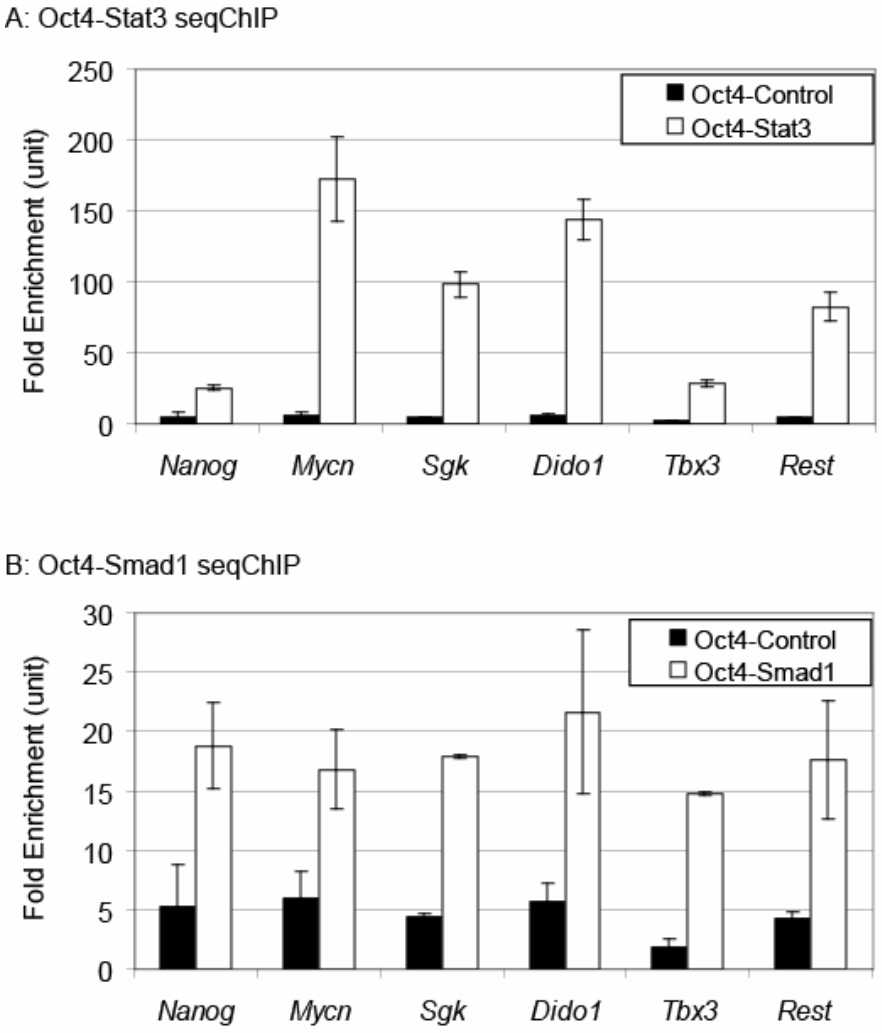
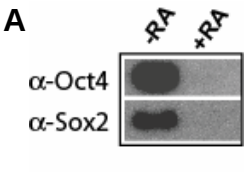
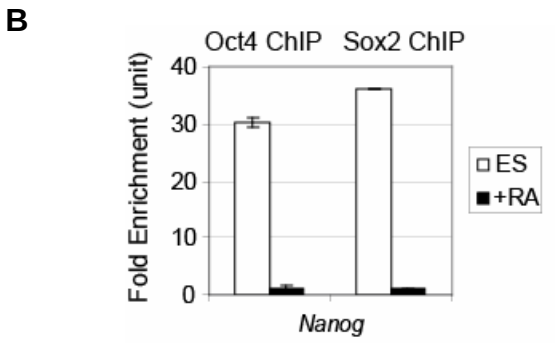
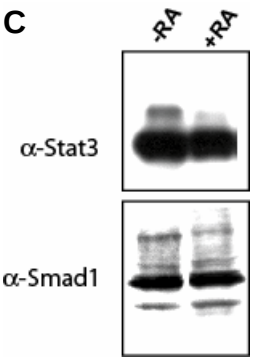


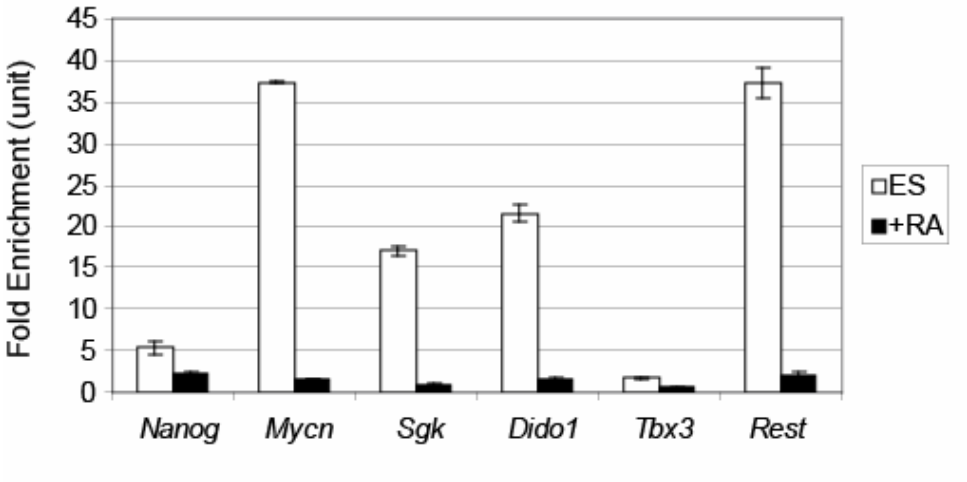
Figure 7.8: Co-occupancy of Oct4 with Stat3 and Smad1 on the same DNA molecule. Sequential chromatin immunoprecipitation (seqChIP) using antibodies against Oct4 followed by antibodies against (A) Stat3 and (B) Smad1 followed by qPCR on the respective gene loci. Control seqChIP using α-Oct4 followed by α-GST antibody are also presented. Biological replicates are shown as standard deviation.







D: Stat3 ChIP



E: Smad1 ChIP

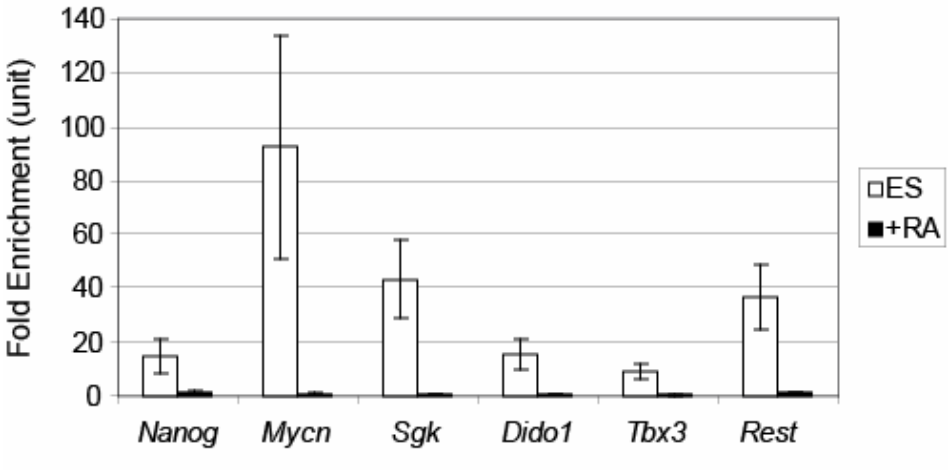
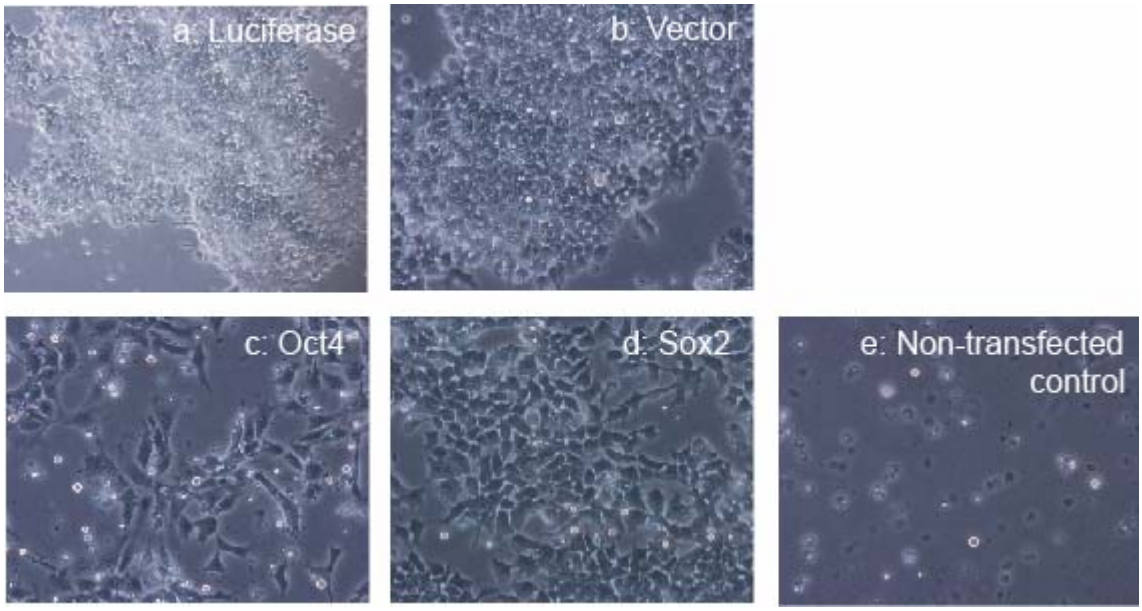
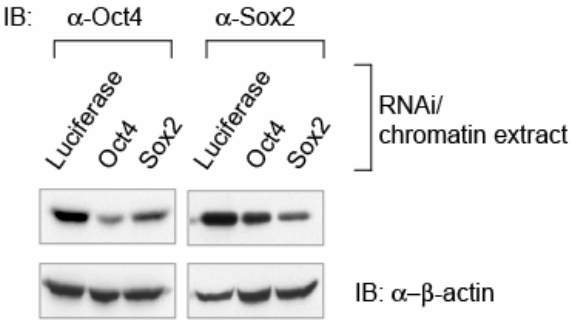


Figure 7.9: Retinoic acid differentiation of mESCs reduces Oct4 and Sox2 levels and binding, and abolishes Stat3 and Smad1 binding. (A) Oct4 and Sox2 protein levels reduced significantly after retinoic acid treatment. Western blot analysis of retinoic acid-treated (+RA) and untreated (-RA) E14 mESC chromatin extract using α-Oct4 and α-Sox2 antibodies. (B) Oct4 and Sox2 bindings on the *Nanog* loci were abolished after RA treatment. Oct4 and Sox2 ChIP-qPCR on the *Nanog* loci using chromatin extracts from non-treated (ES) and retinoic acid-treated (+RA) mESC. (C) Minimal changes in Stat3 and Smad1 protein levels after retinoic acid treatment. Western blot analysis of retinoic acid-treated (+RA) and untreated (-RA) mESC chromatin extract using α-Stat3 and α-Smad1. (D) Stat3 and (E) Smad1 binding abolished after RA treatment. ChIP using antibodies against Stat3 and Smad1, followed by qPCR on the respective gene loci were carried out. Standard deviation represents biological replicates.

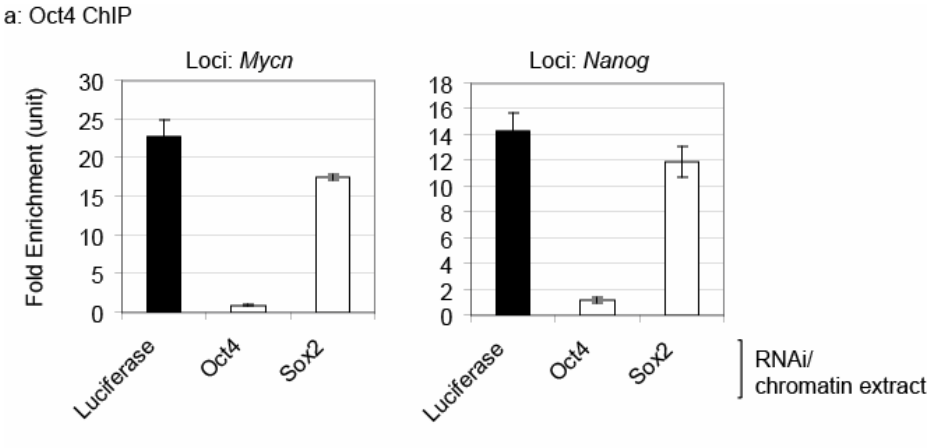
A


a: Luciferase
b: Vector
c: Oct4
d: Sox2
e: Non-transfected control

B


IB:
α-Oct4
α-Sox2
Luciferase
Oct4
Sox2
Luciferase
Oct4
Sox2
RNAi/
chromatin extract
IB: α–β-actin

C


a: Oct4 ChIP
Loci: Mycn
Loci: Nanog
Fold Enrichment (unit)
30
25
20
15
10
5
0
18
16
14
12
10
8
6
4
2
0
Luciferase
Oct4
Sox2
Luciferase
Oct4
Sox2
RNAi/
chromatin extract

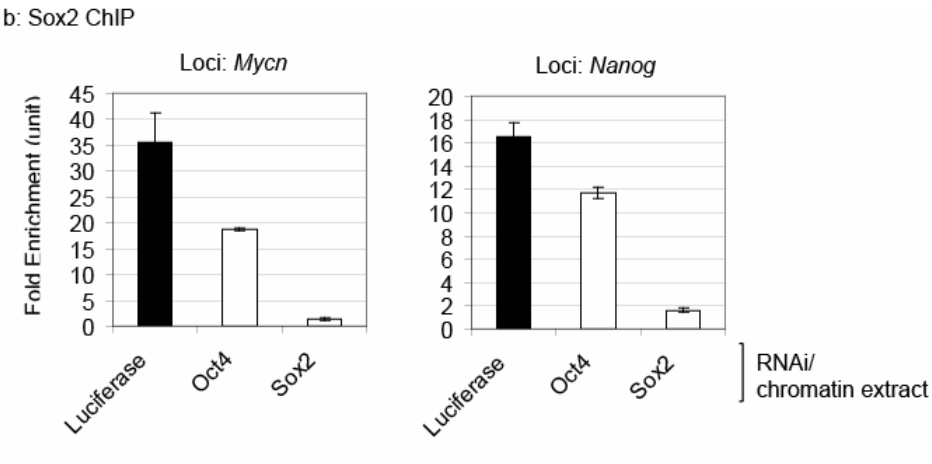
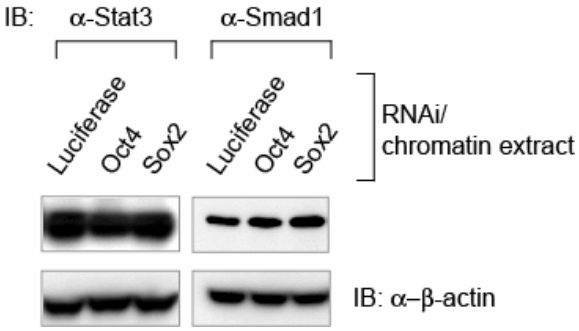
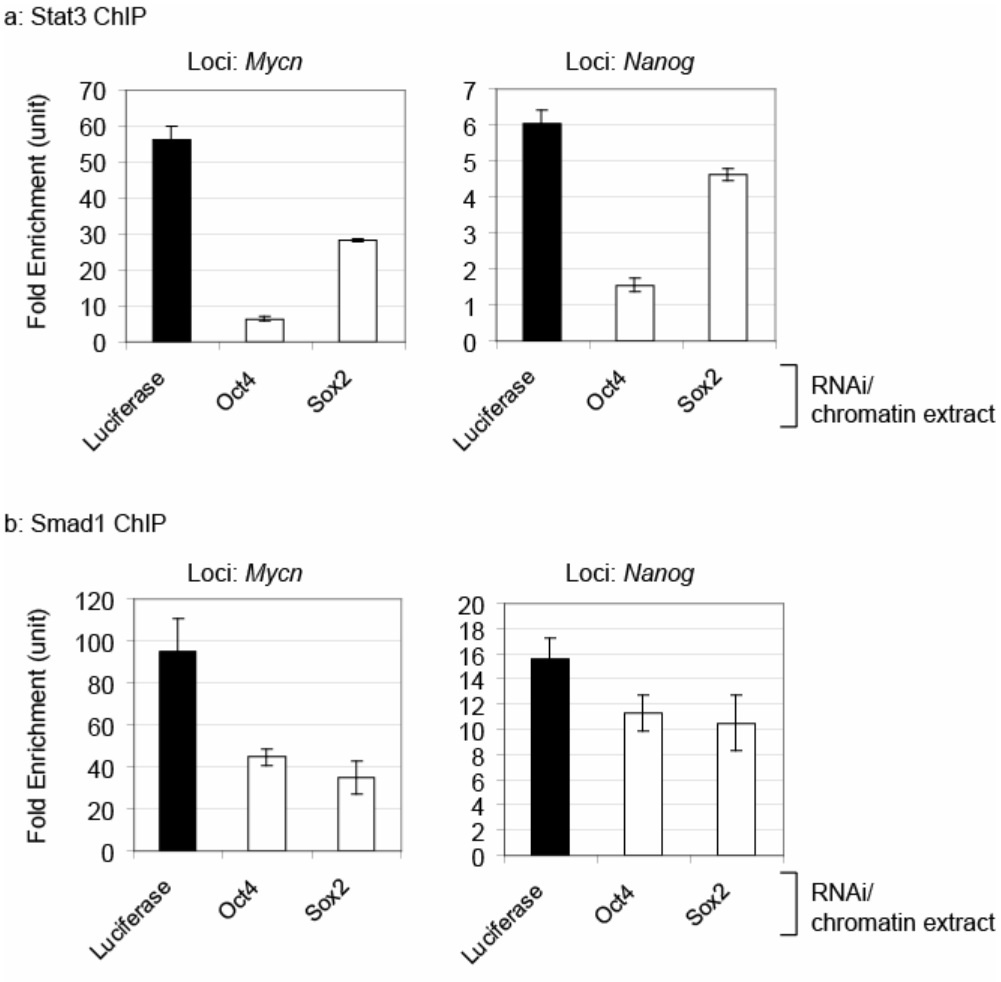


Figure 7.10: Knockdown of Oct4 and Sox2 in mESCs differentiates the cells, significantly reduces Oct4 and Sox2 protein levels and concurrently reduces Oct4 and Sox2 binding on the *Mycn* and *Nanog* loci. (A) Morphology of the cells after RNAi-mediated depletion of Oct4 and Sox2. Mouse ESCs were transfected with shRNA plasmids targeting a) luciferase as control, b) empty pSUPER vector as control, c) Oct4 and d) Sox2. Non-transfected mESCs died upon puromycin selection (e). Images were taken at 20x10x magnification, 3 days post-transfection. (B) Western blot analysis using chromatin extracts from Oct4 and Sox2 knockdown cells after 2 days of selection. Mouse ESCs were transfected with shRNA plasmids targeting luciferase as control, Oct4 and Sox2. The chromatin extracts were analysed by Western blot for changes in Oct4 and Sox2 protein levels. The same blots were detected for β-actin as loading control. (C) ChIP-qPCR analysis using chromatin extracts from Oct4 and Sox2 knockdown cells after 2 days of selection. Mouse ESCs were transfected with shRNA plasmids targeting luciferase as control, Oct4 and Sox2, followed by puromycin selection for 2 days. Chromatin extracts were analysed by ChIP-qPCR for changes in a: Oct4 and b: Sox2 fold enrichment on the *Mycn* and *Nanog* loci.

A


IB:
α-Stat3
α-Smad1
Luciferase
Oct4
Sox2
Luciferase
Oct4
Sox2
RNAi/
chromatin extract
IB: α–β-actin

B


a: Stat3 ChIP
Loci: Mycn
Loci: Nanog
Fold Enrichment (unit)
70
60
50
40
30
20
10
0
7
6
5
4
3
2
1
0
Luciferase
Oct4
Sox2
RNAi/
chromatin extract
b: Smad1 ChIP
Loci: Mycn
Loci: Nanog
Fold Enrichment (unit)
120
100
80
60
40
20
0
20
18
16
14
12
10
8
6
4
2
0
Luciferase
Oct4
Sox2
RNAi/
chromatin extract

C

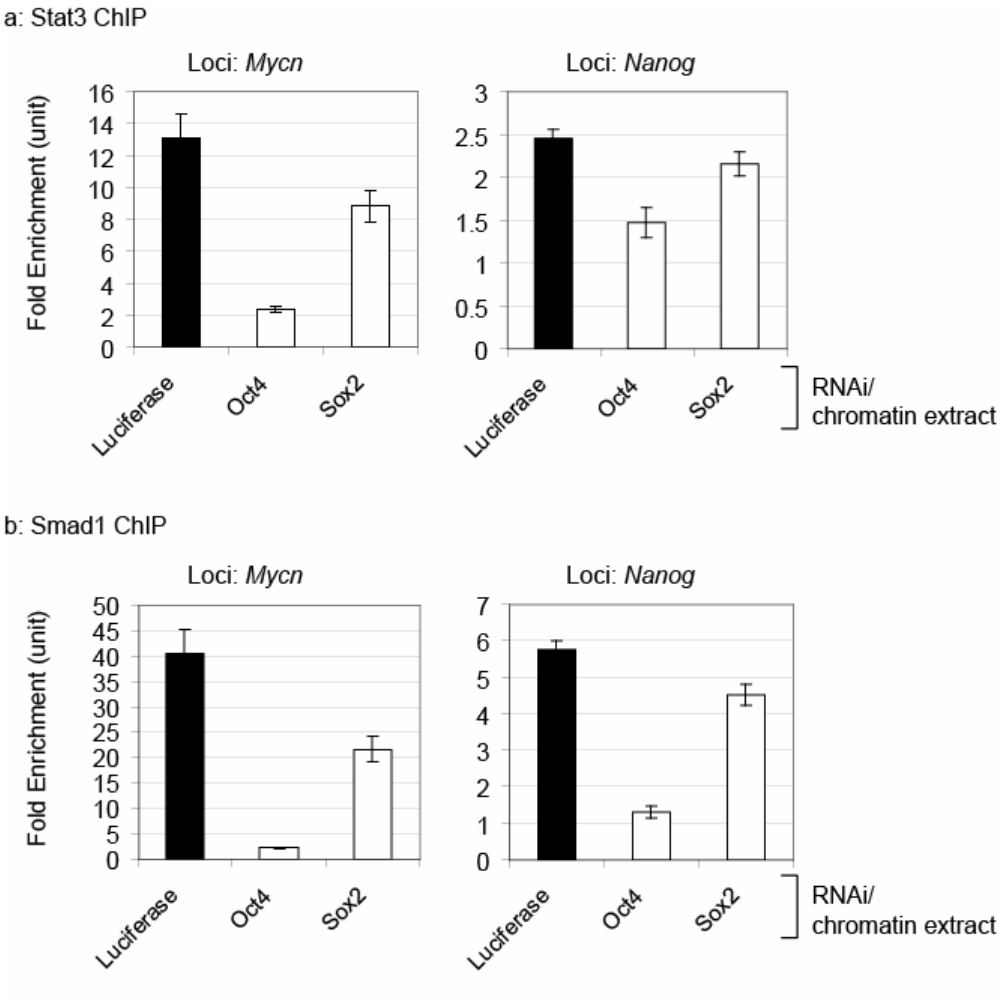


Figure 7.11: Knockdown of Oct4 and Sox2 in mESCs does not significantly affect Stat3 and Smad1 protein levels but reduces Stat3 and Smad1 binding on the *Mycn* and *Nanog* loci. (A) Western blot analysis using chromatin extracts from Oct4 and Sox2 knockdown cells after 2 days-selection. Mouse ESCs were transfected with shRNA plasmids targeting luciferase as control, Oct4 and Sox2 and the chromatin extracts were analysed by Western blot for changes in Stat3 and Smad1 protein levels. The same blots were detected for β-actin as loading control. (B&C) ChIP-qPCR analysis using chromatin extracts from Oct4 and Sox2 knockdown cells. Mouse ESCs were transfected with shRNA plasmids targeting luciferase as control, Oct4 and Sox2, followed by puromycin selection for (B) 2 days and (C) 4 days. The chromatin extracts were analysed by ChIP-qPCR for changes in a: Stat3 and b: Smad1 fold enrichment on the *Mycn* and *Nanog* loci.

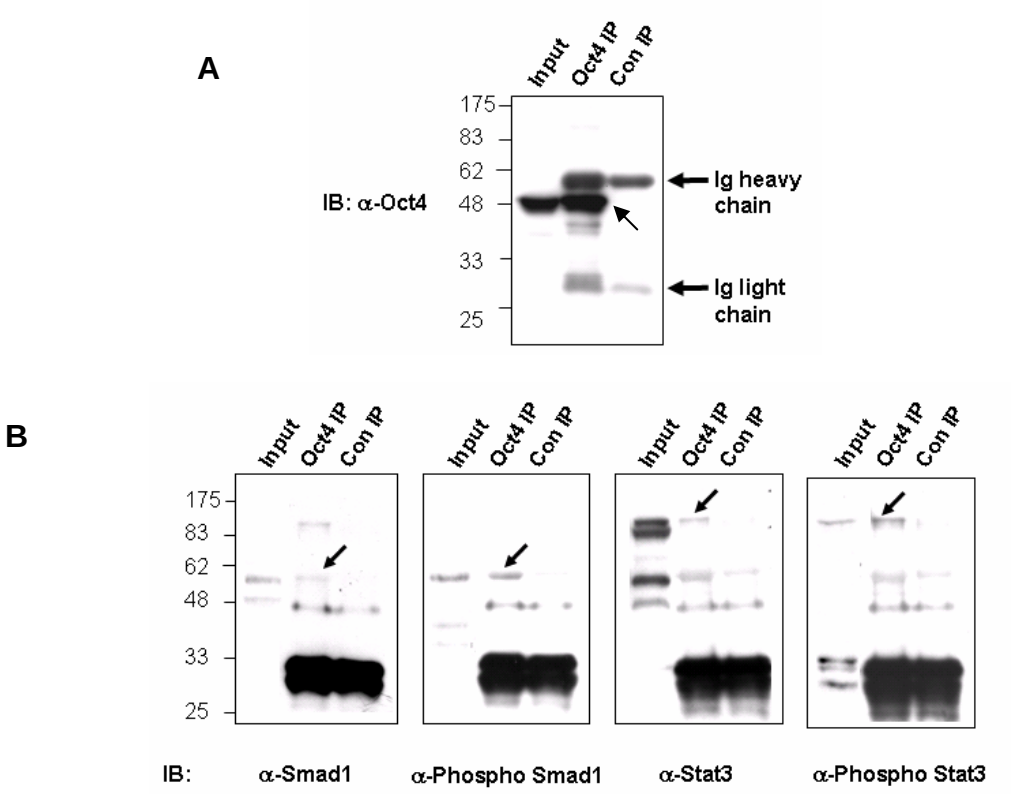
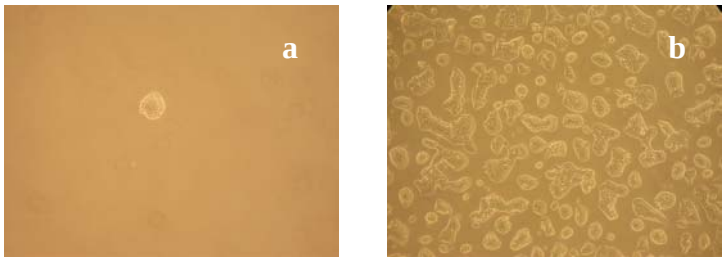


Figure 7.12: Stat3 and Smad1 are Oct4 partners. Co-immunoprecipitation (Co-IP) of Oct4 complex. Immunoprecipitation (IP) of 200 ug mESC nuclear extract was carried out using α-Oct4 and control α-GST (con) antibodies. Input represents mESC nuclear extract. IP was followed by Western immunoblot (IB) using (A) α-Oct4 and (panel B) α-Smad1, α-phosphorylated Smad1, α-Stat3, α-phosphorylated Stat3 antibodies. Arrows point to bands representing proteins detected by the respective antibodies. Molecular weight markers are presented in kDa.

A



B

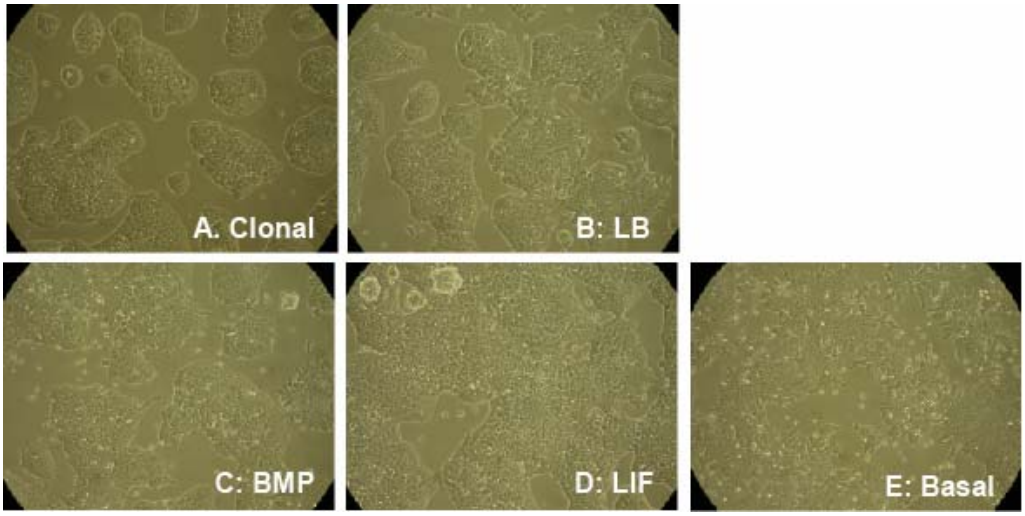


Figure 7.13: Feeder-free serum-free mESCs. (A) Culturing mESC in serum-free media containing LIF and BMP4 (Clonal media, Chemicon). a) Low density colony forming assay at 1000 cells/10cm dish and b) culture after 3 passages at 1:5 plating ratio using Clonal media. Images taken at 20X 10X magnification. (B) Morphology of mESCs cultured in defined media. Cells were acclimatized using clonal media for 2 passages, plated in clonal media for 24 hr, washed with PBS, and cultured with A: clonal, B: 1000 U/ml LIF and 30 ng/ml BMP4 (LB), C: 50 ng/ml BMP4, D: 1000 U/ml LIF and E: pure basal media for 24 hr.

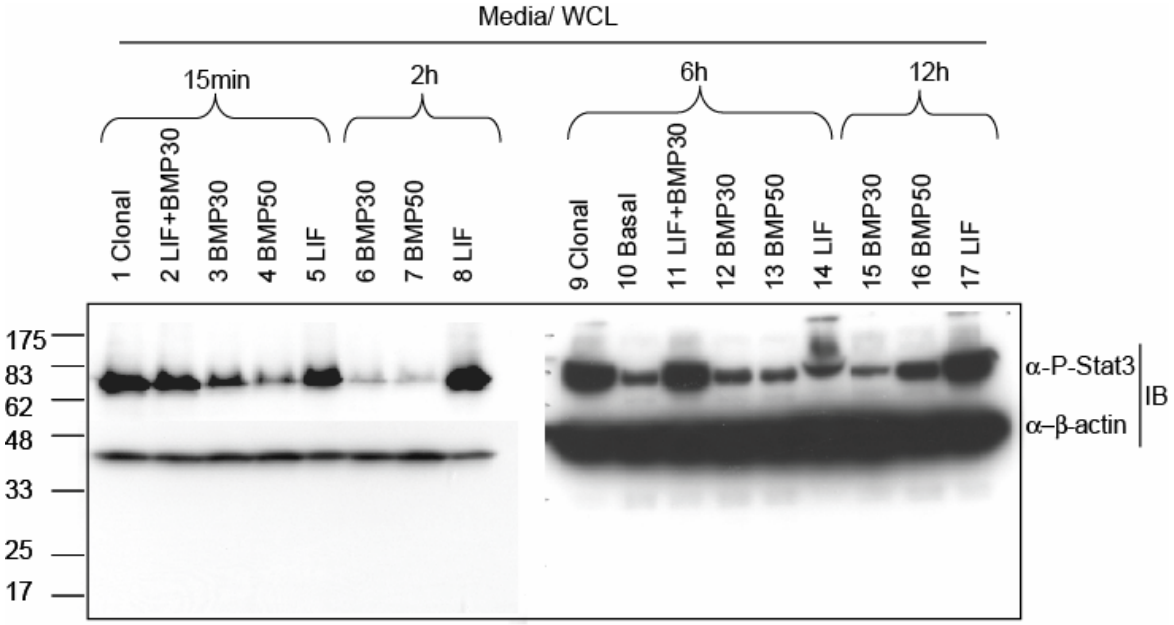
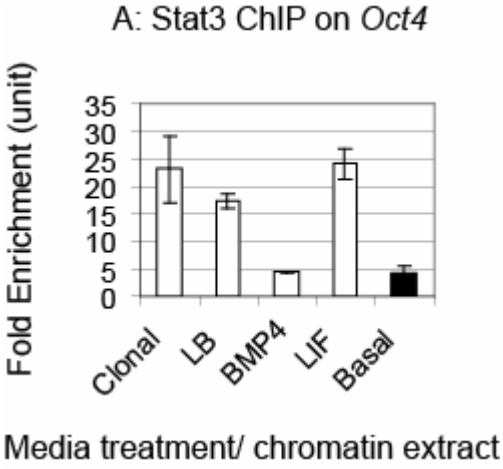


Figure 7.14: Stat3 and Smad1 depletion in mESCs cultured in defined media. Cells were grown in clonal media (Clonal), basal media (Basal), and basal media containing 1000 U/ml LIF and 30 ng/ml BMP4 (LIF+BMP30), 30 ng/ml BMP4 (BMP30), 50 ng/ml BMP4 (BMP50) and 1000 U/ml LIF for 15min, 2hr, 6hr and 12hr. Western blot analysis of the mESC cell lysates using antibodies against phosphorylated Stat3. The same blots were stripped and detected for β-actin as loading control.



B: *Oct4* transcript level

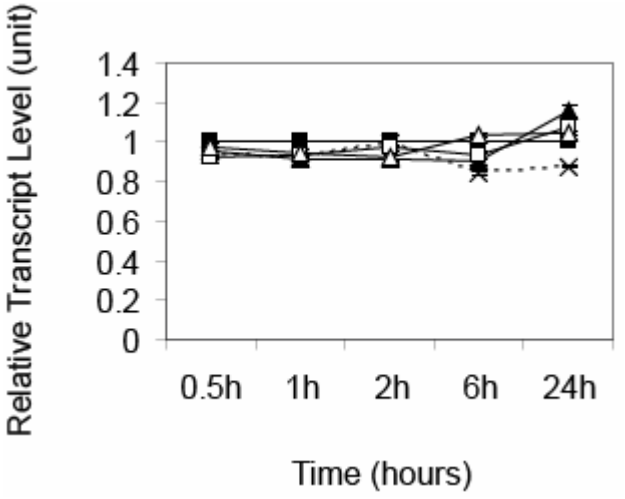
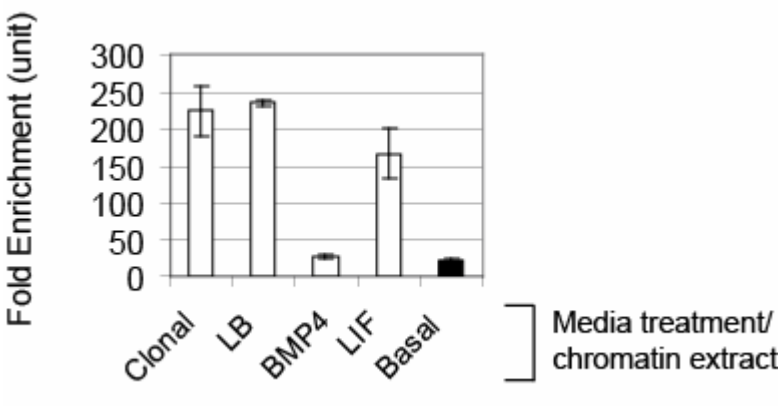


Figure 7.15: Stat3 does not regulate *Oct4*. (A) Stat3 binding on the *Oct4* loci. Stat3 ChIP using chromatin extracts from mESCs cultured in clonal media (Clonal) from Chemicon, basal media (Basal), and basal media containing LIF and BMP4 (LB), BMP4 (BMP4) and LIF (LIF) for 24 hr. ChIP is followed by qPCR on the *Oct4* locus. Standard deviations represent technical replicates. Fold enrichment represents the abundance of enriched DNA fragments over a control region not enriched for the respective targets. (B) Kinetics of *Oct4* expression in mESC cultured in basal, clonal, LIF and BMP4, BMP4 and LIF containing media over a time course of 30 min to 24 hr. Standard deviations represent technical replicates. The levels of transcripts were normalized against values derived from clonal media cultured cells.

A: Stat3 ChIP on *Stat3*



B: *Stat3* transcript level

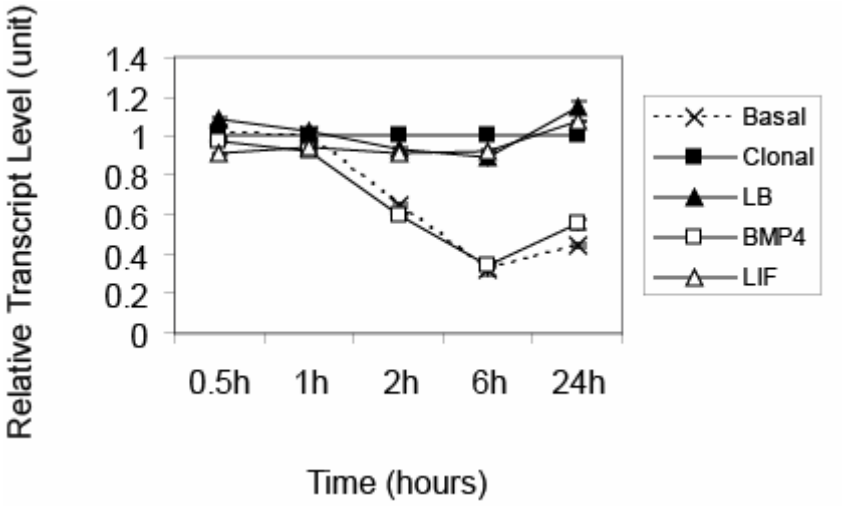
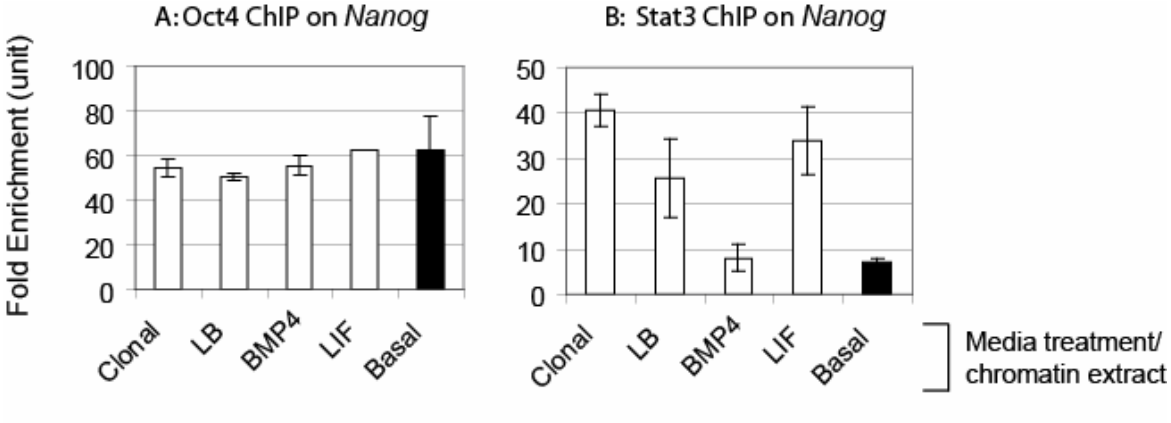
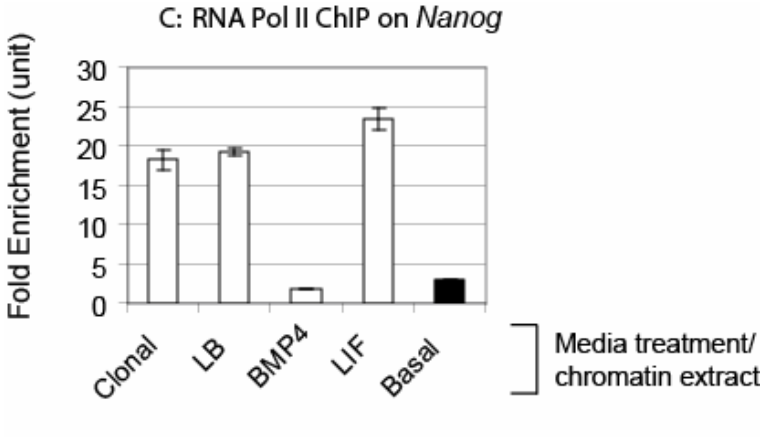


Figure 7.16: Stat3 binds to its own gene and autoregulates. (A) Stat3 ChIP using chromatin extracts from mESCs cultured in clonal media (Clonal) from Chemicon, basal media (Basal), and basal media containing LIF and BMP4 (LB), BMP4 (BMP4) and LIF (LIF) for 24 hr. ChIP is followed by qPCR on the *Stat3* locus. Standard deviations represent technical replicates. Fold enrichment represents the abundance of enriched DNA fragments over a control region not enriched for the respective targets. (B) Kinetics of *Stat3* expression in mESCs cultured in basal, clonal, LIF and BMP4, BMP4 and LIF containing media over a time course of 30 min to 24 hr. Standard deviations represent technical replicates. The levels of transcripts were normalized against values derived from clonal media cultured cells.


A: Oct4 ChIP on *Nanog*
B: Stat3 ChIP on *Nanog*
Fold Enrichment (unit)
100
80
60
40
20
0
50
40
30
20
10
0
Clonal
LB
BMP4
LIF
Basal
Media treatment/
chromatin extract


C: RNA Pol II ChIP on *Nanog*
Fold Enrichment (unit)
30
25
20
15
10
5
0
Clonal
LB
BMP4
LIF
Basal
Media treatment/
chromatin extract

D

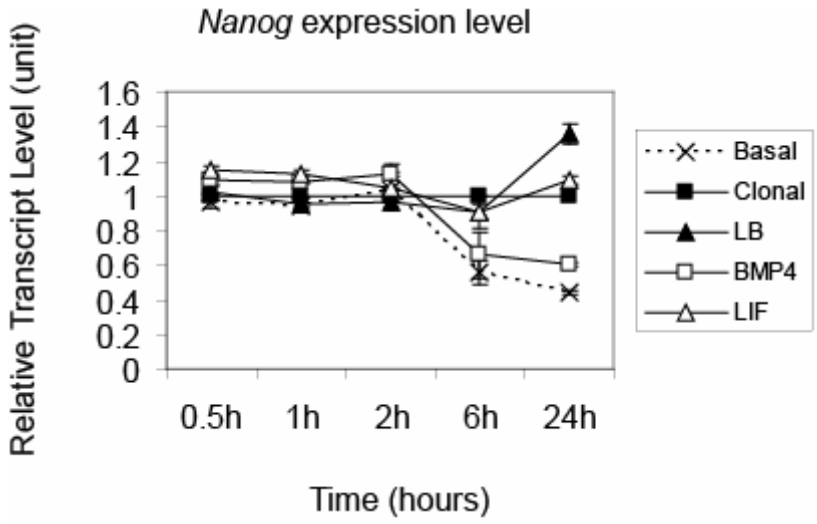

Nanog expression level
Relative Transcript Level (unit)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5h
1h
2h
6h
24h
Time (hours)
Basal
Clonal
LB
BMP4
LIF

Figure 7.17: Stat3 regulates *Nanog*. (A) Stat3 ChIP using chromatin extracts from mESCs cultured in clonal media (Clonal) from Chemicon, basal media (Basal), and basal media containing LIF and BMP4 (LB), BMP4 (BMP4) and LIF (LIF) for 24 hr. Oct4 (A), Stat3 (B) and RNA polymerase II (C) ChIP-qPCR on the *Nanog* locus. Standard deviations represent technical replicates. Fold enrichment represents the abundance of enriched DNA fragments over a control region not enriched for the respective targets. (D) Kinetics of *Nanog* expression levels in mESC cultured in basal, clonal, LIF and BMP4, BMP4 and LIF containing media over a time course of 30 min to 24 hr. Standard deviations represent technical replicates. The levels of transcripts were normalized against values derived from clonal media cultured cells.

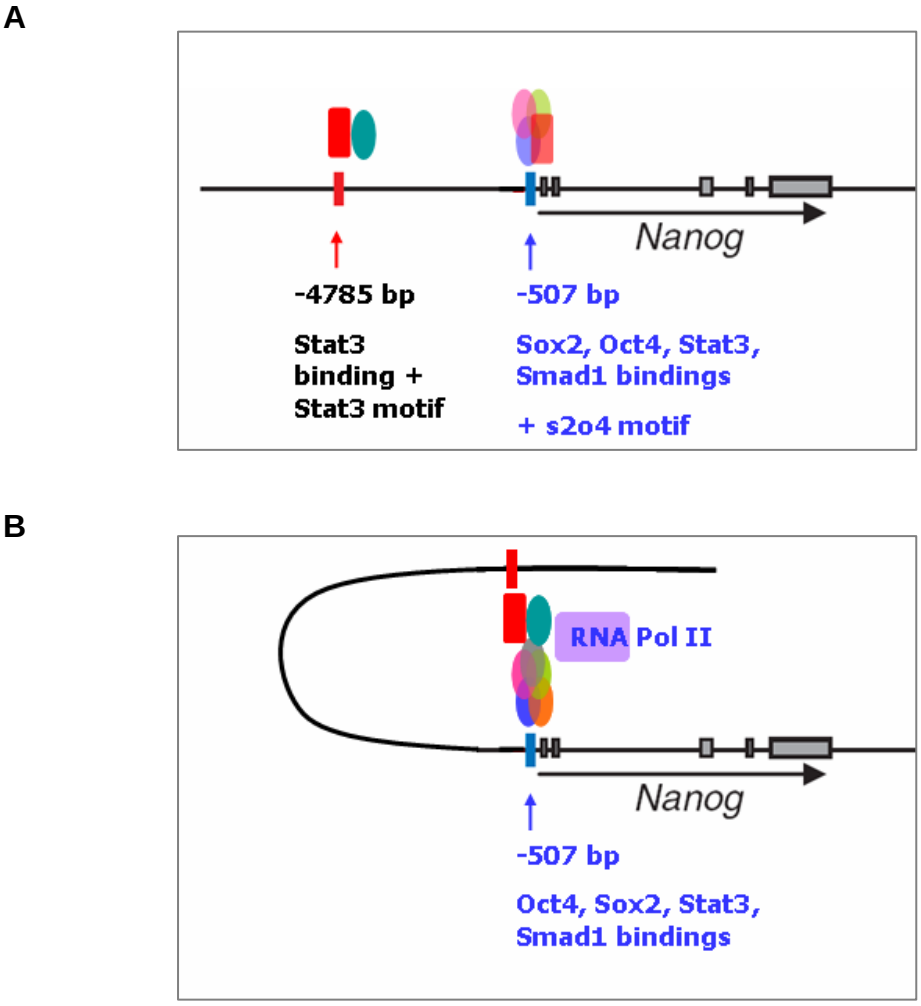


Figure 7.18: Looping mechanism of *Nanog* regulation by collaborating transcription factors. A conjecture presenting how a loop may form linking DNA elements bound by a protein complex is shown. Stat3 binding and Stat3 motif were identified at -4785 bp from the *Nanog* transcription start site by Suzuki *et al.* (2006) (shown in black fonts), whereas this present study identified Sox2, Oct4, Stat3 and Smad1 bindings as well as the Sox2-Oct4 (s2o4) motif at -507 bp from the transcription start site (shown in blue fonts).

CHAPTER 8

GENERAL DISCUSSION

8.1 Implications of the study

In this study, the DNA binding sites of Oct4 and Sox2 were identified. Transcriptional network involving Oct4, Sox2 and Nanog was established (Chapter 3). In addition, Oct4 and Sox2 were shown to bind to novel targets including transcription factors important in maintaining mESC pluripotency and self-renewal (Chapters 4 and 5). A joint Sox2-Oct4 motif was identified and characterized in Chapter 6, suggesting that these two factors collaborate to globally control mESC gene expression through the Sox2-Oct4 motif. Subsequently, Oct4 and Sox2 were found to collaborate with extrinsic signalling transcription factors Stat3 and Smad1 to activate the network of ESC specific genes. Finally, this study demonstrated the relationship between the Stat3, Nanog and Oct4/Sox2 pathways in mESCs (Chapter 7).

The results showed that a fraction of Oct4, Sox2, Stat3 and Smad1 binding sites did not contain consensus motifs. These binding sites could represent a novel DNA binding motif or the transcription factors could be tethered to the DNA independent of their DNA binding properties. In addition, this study also discovered Oct4 and Sox2 binding sites which were located far away from known genes. The absence of DNA motifs and the observation of distal binding sites may be explained by the looping model described in Chapter 7. Due to the chromosomal conformation, these sites may be indirect binding sites that associate with direct binding sites via transcription factor complexes. All genomic fragments that are associated with a particular protein, whether bound directly or indirectly through a complex are isolated by ChIP assays through formaldehyde crosslinking. Further studies need to be carried out in order to validate this model.

The presence of Oct4 and Sox2 may play a role in recruiting other factors that directly regulate the gene as shown in Chapter 7. This may be the reason why some of the Oct4 and Sox2 bindings do not seem to affect the regulation of the bound genes (as described in Chapters 4 and 5). It may also explain a recent study by Masui *et al.* (2007) which reported that Sox2 is dispensable for the activation of the Sox-Oct enhancers.

In addition, this study identified Stat3 and Smad1 as factors collaborating with Oct4 and Sox2 on *cis*-regulating modules (Chapter 7). Stat3 and Smad1 are downstream effectors of the LIF and BMP signalling pathways, respectively. This suggests that extrinsic signalling factors LIF and BMP4 activate Stat3 and Smad1, which then work together with Oct4 and Sox2 to regulate certain genes important in maintaining pluripotency and self-renewal. Indeed, Stat3 was shown to bind with Oct4 to regulate *Nanog*, providing evidence of a crosstalk between the Oct4, Stat3 and Nanog transcriptional pathways in mESCs. In addition, this study showed that Oct4, Sox2 and Stat3 autoregulate, thus confirming their importance as key regulatory factors in mESCs.

There are other studies that reported genome-wide mapping of key transcription factor binding sites in ESCs. A colleague, Wu Qiang, mapped the binding sites of Nanog in mESCs using ChIP-PET (Loh *et al.,* 2006). When the ChIP-PET datasets from Oct4, Sox2 and Nanog were compared (Appendix C), the three transcription factors were shown to bind a set of shared genes. The three transcription factors co-occupied genes in various arrangements. An example is where both Sox2 and Oct4 bind 3' of the *Sox2* gene whereas Nanog binds 3.8 kb 5' to its transcription start site. Further investigation into these different binding configurations may offer clues into regulatory mechanisms in addition to perhaps uncovering differential transcriptional responses. Triple sequential ChIP validated co-occupancy of the three factors on 7 genes (Appendix D). It appears that pluripotency and self-renewal of mESCs is achieved through the combination of a few key transcription factors that control the activation and repression of specific

genes. Boyer *et al.* (2005) also mapped the binding sites of Oct4, Sox2 and Nanog in hESCs using promoter microarrays. However, the binding sites generated by ChIP-PET/mESCs (this study) and ChIP-on-promoter chip/hESCs (Boyer *et al.* 2005) only overlapped by 6.9% for Oct4 and 9.5% for Sox2 when comparative analysis was done using both datasets by our Bioinformatics collaborator, Guillaume Bourque. The large discrepancies from these two binding sites study may be due to differences in (1) biological factors (mouse versus human ESCs) and (2) approaches used (ChIP-PET versus ChIP-on-promoter chip). Despite the small overlap, these sites may regulate core ES genes that are conserved between both mouse and human, and may represent the most important genes in controlling ESCs. Further studies on these genes will be important in discovering more transcriptional switches in the control of the ESC states. Identifying targets of key factors and investigation on the roles of these genes using mouse ESCs as a model system would be of importance to eventually develop switches in differentiating human ESCs, and hopefully bring us a step closer to achieve control of the human ESC fate for use in regenerative medicine.

8.2 Future studies

This study presents many interesting questions and future directions to be explored. Studies should be conducted to characterize the functions of identified Oct4 and Sox2 target genes. Core Oct4, Sox2 and Nanog shared genes which are conserved between mouse and human are likely to have important roles in ESCs. It will also be intriguing to dissect the roles of microRNAs which are targeted by these transcription factors in ESCs. Further examination of the interactions among all these crucial genes will provide a more thorough understanding of cell fate determination.

Besides that, further studies need to be done to verify long range interactions and the looping model proposed in Chapter 7. Transcriptional activation involves physical association of genes and their regulatory elements. Previous studies have demonstrated that active β-globin genes physically interact with multiple *cis*-regulatory elements (Tolhuis *et al.*, 2002) and this has led to

the proposal of an active chromatin hub (de Laat and Grosveld, 2003) (Figure 8.1). Moreover, long-range chromosomal interactions between genomic elements have been shown to regulate active gene expression (Levings *et al.*, 2006; Vernimmen *et al.*, 2007). Therefore, a higher chromosome organization may be a major determinant of gene regulation (Levings *et al.*, 2006). It would be interesting to find out how these distant activators interact with their target genes.

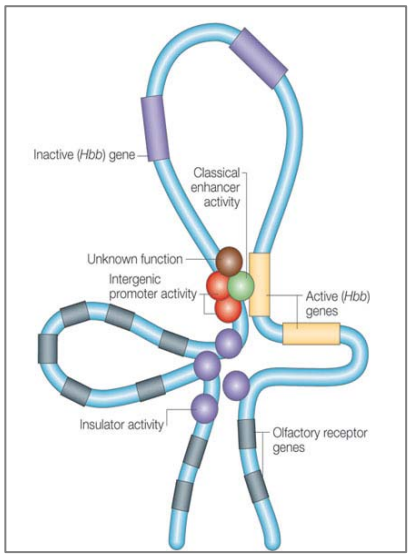


Figure 8.1: A chromatin hub formed by long-range interactions between the *haemoglobin β-chain complex* (*Hbb*) genes and the locus control region. Figure was adapted from Chakalova *et al.* (2005).

One of the methods that can be used to analyze the spatial organization of chromosomes is the Chromosome Conformation Capture (3C) assay (Figure 8.2a). In the 3C methodology, ChIP is carried out, followed by ligation to fuse nearby DNA fragments. Primer pairs consisting of one primer in the promoter region and the other at the enhancer region are then used to PCR-amplify potentially ligated fragments of DNA. Many variations of the 3C method, including coupling with microarray detection or high throughput DNA sequencing, have been employed in various studies (Dostie and Dekker, 2007; Simonis *et al.*, 2006; Zhao *et al.*, 2006). They have used this method to discover long range interactions such as interactions between the estrogen receptor and Forkhead protein Fox1A (Carroll *et al.*, 2005), an enhancer on mouse chromosome 14 and multiple olfactory receptor genes (Lomvardas *et al.*, 2006), and (Igf2)/H19 regulation (Ling *et al.*, 2006). In mESCs, Wurtele and Chartrand (2006) identified HoxB1-associated loci throughout the genome.

One of the disadvantages of the 3C method is that sites separated by less than 8 kb may not be detected (Dekker *et al.*, 2002). In these cases, the RNA TRAP (tagging and recovery of associated proteins) assay may be used to detect long-range interactions indirectly by identifying loci that are in the vicinity of nascent RNA (Dekker, 2003). Horseradish peroxidase which is targeted to a specific nascent RNA, catalyzes local deposition of biotin tags on proteins. DNA segments cross-linked to biotinylated proteins are then purified on streptavidin-agarose and detected by PCR (Figure 8.2b).

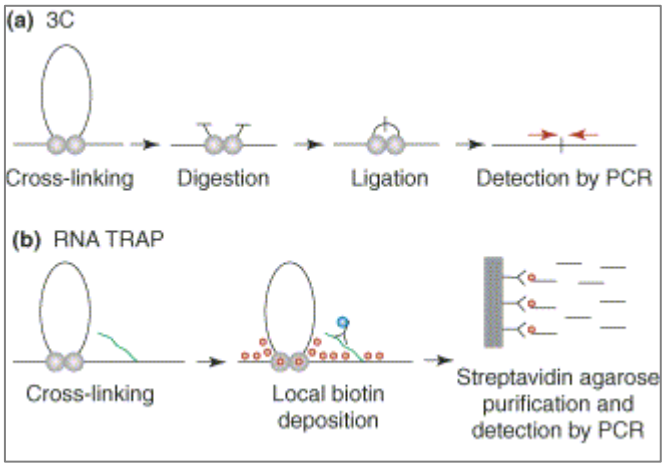


Figure 8.2: Approaches used for detecting long range protein-DNA interactions are (a) the Chromosome Conformation Capture (3C) method and (b) the RNA tagging and recovery of associated proteins (RNA TRAP) method. Figure was obtained from Dekker (2003).

Besides transcription factors, maintaining ESC identity ultimately depends on many other aspects such as the accurate replication of the specific covalent modifications to the histones and DNA as well as associated transcription factors that structure the chromatin and define the ESC. To date, the involvement of the epigenome in the maintenance and establishment of ESC pluripotency remain unclear. Further studies will be necessary to find out how epigenetic mechanisms determine the pluripotent epigenome and how it functions to maintain the ESC transcription factor network.

The advent of new sequencing technologies will be able to help expedite the identification of transcription factor-DNA or modified histone-DNA binding sites. New sequencing platforms such as the Genome Analyzer from Illumina and SOLiD from Applied Biosystems are now able to sequence short DNA fragments, such as ChIP DNA and micro RNAs. In addition, development to sequence up to a single DNA molecule will be the future challenge. Currently, a company called Helicos is developing this new platform technology and if successful, it will be a powerful technology to study DNA at a single cell level.

REFERENCES

Adams,I.R. and McLaren,A. (2004). Identification and characterisation of mRif1: a mouse telomere-associated protein highly expressed in germ cells and embryo-derived pluripotent stem cells. Dev. Dyn. *229*, 733-744.

Akagi,T., Usuda,M., Matsuda,T., Ko,M.S., Niwa,H., Asano,M., Koide,H., and Yokota,T. (2005). Identification of Zfp-57 as a downstream molecule of STAT3 and Oct-3/4 in embryonic stem cells. Biochem. Biophys. Res. Commun. *331*, 23-30.

Ambrosetti,D.C., Basilico,C., and Dailey,L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol. Cell Biol. *17*, 6321-6329.

Ambrosetti,D.C., Scholer,H.R., Dailey,L., and Basilico,C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J Biol. Chem. *275*, 23387-23397.

Avilion,A.A., Nicolis,S.K., Pevny,L.H., 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., Jorgensen,H.F., John,R.M., Gouti,M., Casanova,M., Warnes,G., Merkenschlager,M., and Fisher,A.G. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532-538.

Bain,G., Kitchens,D., Yao,M., Huettner,J.E., and Gottlieb,D.I. (1995). Embryonic stem cells express neuronal properties in vitro. Dev. Biol. *168*, 342-357.

Balciunaite,E., Spektor,A., Lents,N.H., Cam,H., Te,R.H., Scime,A., Rudnicki,M.A., Young,R., and Dynlacht,B.D. (2005). Pocket protein complexes are recruited to distinct targets in quiescent and proliferating cells. Mol. Cell Biol. *25*, 8166-8178.

Ballas,N., Grunseich,C., Lu,D.D., Speh,J.C., and Mandel,G. (2005). REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell *121*, 645-657.

Bartel,D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Ben Shushan,E., Sharir,H., Pikarsky,E., and Bergman,Y. (1995). A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells. Mol. Cell Biol. *15*, 1034-1048.

Ben Shushan,E., Thompson,J.R., Gudas,L.J., and Bergman,Y. (1998). Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. Mol. Cell Biol. *18*, 1866-1878.

Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K., Jaenisch,R., Wagschal,A., Feil,R., Schreiber,S.L., and Lander,E.S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Blackwood,E.M. and Kadonaga,J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Boeuf,H., Hauss,C., Graeve,F.D., Baran,N., and Kedinger,C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. J Cell Biol. *138*, 1207-1217.

Boiani,M. and Scholer,H.R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. Nat. Rev. Mol. Cell Biol. *6*, 872-884.

Bongso,A., Fong,C.Y., Ng,S.C., and Ratnam,S. (1994). Isolation and culture of inner cell mass cells from human blastocysts. Hum. Reprod. *9*, 2110-2117.

Botquin,V., Hess,H., Fuhrmann,G., Anastassiadis,C., Gross,M.K., Vriend,G., and Scholer,H.R. (1998). New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. Genes Dev. *12*, 2073-2090.

Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zucker,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L., Jenner,R.G., Gifford,D.K., Melton,D.A., Jaenisch,R., and Young,R.A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyer,L.A., Mathur,D., and Jaenisch,R. (2006a). Molecular control of pluripotency. Curr. Opin. Genet. Dev. *16*, 455-462.

Boyer,L.A., Plath,K., Zeitlinger,J., Brambrink,T., Medeiros,L.A., Lee,T.I., Levine,S.S., Wernig,M., Tajonar,A., Ray,M.K., Bell,G.W., Otte,A.P., Vidal,M., Gifford,D.K., Young,R.A., and Jaenisch,R. (2006b). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J., Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R., Lebkowski,J., and Stanton,L.W. (2004). Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. Nat. Biotechnol. *22*, 707-716.

Brehm,A., Ohbo,K., and Scholer,H. (1997). The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain. Mol. Cell Biol. *17*, 154-162.

Brehm,A., Ohbo,K., Zwerschke,W., Botquin,V., Jansen-Durr,P., and Scholer,H.R. (1999). Synergism with germ line transcription factor Oct-4: viral oncoproteins share the ability to mimic a stem cell-specific activity. Mol. Cell Biol. *19*, 2635-2643.

Brenner,S., Johnson,M., Bridgham,J., Golda,G., Lloyd,D.H., Johnson,D., Luo,S., McCurdy,S., Foy,M., Ewan,M., Roth,R., George,D., Eletr,S., Albrecht,G., Vermaas,E., Williams,S.R., Moon,K., Burcham,T., Pallas,M., DuBridge,R.B., Kirchner,J., Fearon,K., Mao,J., and Corcoran,K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. *18*, 630-634.

Brenner,S. and Livak,K.J. (1989). DNA fingerprinting by sampled sequencing. Proc. Natl. Acad. Sci. U. S. A *86*, 8902-8906.

Buck,M.J. and Lieb,J.D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics *83*, 349-360.

Burdon,T., Chambers,I., Stracey,C., Niwa,H., and Smith,A. (1999). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. Cells Tissues. Organs *165*, 131-143.

Burdon,T., Smith,A., and Savatier,P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. Trends Cell Biol. *12*, 432-438.

Carey,M. and Smale,S.T. (2000). Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques. Cold Spring Harbor Laboratory Press).

Carroll,J.S., Liu,X.S., Brodsky,A.S., Li,W., Meyer,C.A., Szary,A.J., Eeckhoute,J., Shao,W., Hestermann,E.V., Geistlinger,T.R., Fox,E.A., Silver,P.A., and Brown,M. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33-43.

Cartwright,P., McLean,C., Sheppard,A., Rivett,D., Jones,K., and Dalton,S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885-896.

Catena,R., Tiveron,C., Ronchi,A., Porta,S., Ferri,A., Tatangelo,L., Cavallaro,M., Favaro,R., Ottolenghi,S., Reinbold,R., Scholer,H., and Nicolis,S.K. (2004). Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. J Biol. Chem. *279*, 41846-41857.

Cawley,S., Bekiranov,S., Ng,H.H., Kapranov,P., Sekinger,E.A., Kampa,D., Piccolboni,A., Sementchenko,V., Cheng,J., Williams,A.J., Wheeler,R., Wong,B., Drenkow,J., Yamanaka,M., Patel,S., Brubaker,S., Tammana,H., Helt,G., Struhl,K., and Gingeras,T.R. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.

Chakalova,L., Debrand,E., Mitchell,J.A., Osborne,C.S., and Fraser,P. (2005). Replication and transcription: shaping the landscape of the genome. Nat. Rev. Genet. *6*, 669-677.

Chambers,I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells. Cloning Stem Cells *6*, 386-391.

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. and Smith,A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene *23*, 7150-7160.

Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A. (2007) Nanog safeguards pluripotency and mediates germline development. Nature *450(7173)*:1230-4.

Chew,J.L., Loh,Y.H., Zhang,W., Chen,X., Tam,W.L., Yeap,L.S., Li,P., Ang,Y.S., Lim,B., Robson,P., and Ng,H.H. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol. Cell Biol. *25*, 6031-6046.

Chung,A.C. and Cooney,A.J. (2001). Germ cell nuclear factor. Int. J Biochem. Cell Biol. *33*, 1141-1146.

Cosma,M.P. (2002). Ordered recruitment: gene-specific mechanism of transcription activation. Mol. Cell *10*, 227-236.

Coucouvanis,E. and Martin,G.R. (1995). Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. Cell *83*, 279-287.

Cowan,C.A., Klimanskaya,I., McMahon,J., Atienza,J., Witmyer,J., Zucker,J.P., Wang,S., Morton,C.C., McMahon,A.P., Powers,D., and Melton,D.A. (2004). Derivation of embryonic stem-cell lines from human blastocysts. N. Engl. J Med. *350*, 1353-1356.

d'Adda,d.F., Reaper,P.M., Clay-Farrace,L., Fiegler,H., Carr,P., Von Zglinicki,T., Saretzki,G., Carter,N.P., and Jackson,S.P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. Nature *426*, 194-198.

Dailey,L. and Basilico,C. (2001). Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J Cell Physiol *186*, 315-328.

Dailey,L., Yuan,H., and Basilico,C. (1994). Interaction between a novel F9-specific factor and octamer-binding proteins is required for cell-type-restricted activity of the fibroblast growth factor 4 enhancer. Mol. Cell Biol. *14*, 7758-7769.

Dani,C., Smith,A.G., Dessolin,S., Leroy,P., Staccini,L., Villageois,P., Darimont,C., and Ailhaud,G. (1997). Differentiation of embryonic stem cells into adipocytes in vitro. J Cell Sci. *110 (Pt 11)*, 1279-1285.

Davey,R.E., Onishi,K., Mahdavi,A., and Zandstra,P.W. (2007). LIF-mediated control of embryonic stem cell self-renewal emerges due to an autoregulatory loop. FASEB J *21*, 2020-2032.

Davuluri,R.V., Grosse,I., and Zhang,M.Q. (2001). Computational identification of promoters and first exons in the human genome. Nat. Genet. *29*, 412-417.

de Laat,W. and Grosveld,F. (2003). Spatial organization of gene expression: the active chromatin hub. Chromosome. Res. *11*, 447-459.

Dekker,J. (2003). A closer look at long-range chromosomal interactions. Trends Biochem. Sci. *28*, 277-280.

Dekker,J., Rippe,K., Dekker,M., and Kleckner,N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Dempster,A., Laird,N., and Rubin,D. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B *39*, 1-38.

Derynck,R. and Zhang,Y.E. (2003). Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature *425*, 577-584.

Dignam,J.D., Lebovitz,R.M., and Roeder,R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475-1489.

Ding,L. and Buchholz,F. (2006). RNAi in embryonic stem cells. Stem Cell Rev. *2*, 11-18.

DiTullio,R.A., Jr., Mochan,T.A., Venere,M., Bartkova,J., Sehested,M., Bartek,J., and Halazonetis,T.D. (2002). 53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer. Nat. Cell Biol. *4*, 998-1002.

Doetschman,T.C., Eistetter,H., Katz,M., Schmidt,W., and 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.

Donovan,P.J. and Gearhart,J. (2001). The end of the beginning for pluripotent stem cells. Nature *414*, 92-97.

Dostie,J. and Dekker,J. (2007). Mapping networks of physical interactions between genomic elements using 5C technology. Nat. Protoc. *2*, 988-1002.

Dunn,N.R., Winnier,G.E., Hargett,L.K., Schrick,J.J., Fogo,A.B., and Hogan,B.L. (1997). Haploinsufficient phenotypes in Bmp4 heterozygous null mice and modification by mutations in Gli3 and Alx4. Dev. Biol. *188*, 235-247.

Dynan,W.S. and Tjian,R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature *316*, 774-778.

Euskirchen,G., Royce,T.E., Bertone,P., Martone,R., Rinn,J.L., Nelson,F.K., Sayward,F., Luscombe,N.M., Miller,P., Gerstein,M., Weissman,S., and Snyder,M. (2004). CREB binds to multiple loci on human chromosome 22. Mol. Cell Biol. *24*, 3804-3814.

Evans,M.J. and Kaufman,M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Faherty,S., Kane,M.T., and Quinlan,L.R. (2005). Self-renewal and differentiation of mouse embryonic stem cells as measured by Oct 4 gene expression: effects of lif, serum-free medium, retinoic acid, and dbcAMP. In Vitro Cell Dev. Biol. Anim *41*, 356-363.

Finley,M.F., Devata,S., and Huettner,J.E. (1999). BMP-4 inhibits neural differentiation of murine embryonic stem cells. J Neurobiol. *40*, 271-287.

Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E., and Mello,C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Fuhrmann,G., Chung,A.C., Jackson,K.J., Hummelke,G., Baniahmad,A., Sutter,J., Sylvester,I., Scholer,H.R., and Cooney,A.J. (2001). Mouse germline restriction of Oct4 expression by germ cell nuclear factor. Dev. Cell *1*, 377-387.

Fujikura,J., Yamato,E., Yonemura,S., Hosoda,K., Masui,S., Nakao,K., Miyazaki,J.J., and Niwa,H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev. *16*, 784-789.

Furusawa,T., Ikeda,M., Inoue,F., Ohkoshi,K., Hamano,T., and Tokunaga,T. (2006). Gene expression profiling of mouse embryonic stem cell subpopulations. Biol. Reprod. *75*, 555-561.

Galan-Caridad,J.M., Harel,S., Arenzana,T.L., Hou,Z.E., Doetsch,F.K., Mirny,L.A., and Reizis,B. (2007). Zfx controls the self-renewal of embryonic and hematopoietic stem cells. Cell *129*, 345-357.

Geisberg,J.V. and Struhl,K. (2004). Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. Nucleic Acids Res. *32*, e151.

Gu,P., Goodwin,B., Chung,A.C., Xu,X., Wheeler,D.A., Price,R.R., Galardi,C., Peng,L., Latour,A.M., Koller,B.H., Gossen,J., Kliewer,S.A., and Cooney,A.J. (2005a). Orphan

nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol. Cell Biol. *25*, 3492-3505.

Gu,P., LeMenuet,D., Chung,A.C., Mancini,M., Wheeler,D.A., and Cooney,A.J. (2005b). Orphan nuclear receptor GCNF is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation. Mol. Cell Biol. *25*, 8507-8519.

Hanna,L.A., Foreman,R.K., Tarasenko,I.A., Kessler,D.S., and Labosky,P.A. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. Genes Dev. *16*, 2650-2661.

Hao,J., Li,T.G., Qi,X., Zhao,D.F., and Zhao,G.Q. (2006). WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. Dev. Biol. *290*, 81-91.

Hay,D.C., Sutherland,L., Clark,J., and Burdon,T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells *22*, 225-235.

Herr,W. and Cleary,M.A. (1995). The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. Genes Dev. *9*, 1679-1693.

Hocke,G.M., Cui,M.Z., and Fey,G.H. (1995). The LIF response element of the alpha 2 macroglobulin gene confers LIF-induced transcriptional activation in embryonal stem cells. Cytokine *7*, 491-502.

Hollnagel,A., Oehlmann,V., Heymer,J., Ruther,U., and Nordheim,A. (1999). Id genes are direct targets of bone morphogenic protein induction in embryonic stem cells. J Biol. Chem. *274*, 19838-19845.

Hoodless,P.A., Pye,M., Chazaud,C., Labbe,E., Attisano,L., Rossant,J., and Wrana,J.L. (2001). FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. Genes Dev. *15*, 1257-1271.

Hooker,C.W. and Hurlin,P.J. (2006). Of Myc and Mnt. J Cell Sci. *119*, 208-216.

Houbaviy,H.B., Murray,M.F., and Sharp,P.A. (2003). Embryonic stem cell-specific MicroRNAs. Dev. Cell *5*, 351-358.

Hough,S.R., Clements,I., Welch,P.J., and Wiederholt,K.A. (2006a). Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. Stem Cells *24*, 1467-1475.

Hough,S.R., Clements,I., Welch,P.J., and Wiederholt,K.A. (2006b). Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. Stem Cells *24*, 1467-1475.

Ichiba,M., Nakajima,K., Yamanaka,Y., Kiuchi,N., and Hirano,T. (1998). Autoregulation of the Stat3 gene through cooperation with a cAMP-responsive element-binding protein. J Biol. Chem. *273*, 6132-6138.

Imagawa,M., Miyamoto,A., Shirakawa,M., Hamada,H., and Muramatsu,M. (1991). Stringent integrity requirements for both trans-activation and DNA-binding in a trans-activator, Oct3. Nucleic Acids Res. *19*, 4503-4508.

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

Ivanova,N.B., Dimos,J.T., Schaniel,C., Hackney,J.A., Moore,K.A., and Lemischka,I.R. (2002). A stem cell molecular signature. Science *298*, 601-604.

Iwai,N., Kitajima,K., Sakai,K., Kimura,T., and Nakano,T. (2001). Alteration of cell adhesion and cell cycle properties of ES cells by an inducible dominant interfering Myb mutant. Oncogene *20*, 1425-1434.

Iyer,V.R., Horak,C.E., Scafe,C.S., Botstein,D., Snyder,M., and Brown,P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.

Kamachi,Y., Uchikawa,M., and Kondoh,H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet. *16*, 182-187.

Kanellopoulou,C., Muljo,S.A., Kung,A.L., Ganesan,S., Drapkin,R., Jenuwein,T., Livingston,D.M., and Rajewsky,K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. *19*, 489-501.

Karlseder,J., Kachatrian,L., Takai,H., Mercer,K., Hingorani,S., Jacks,T., and de Lange,T. (2003). Targeted deletion reveals an essential function for the telomere length regulator Trf1. Mol. Cell Biol. *23*, 6533-6541.

Keller,G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes Dev. *19*, 1129-1155.

Kim,T.H., Barrera,L.O., Zheng,M., Qu,C., Singer,M.A., Richmond,T.A., Wu,Y., Green,R.D., and Ren,B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

Knight,J.C., Keating,B.J., Rockett,K.A., and Kwiatkowski,D.P. (2003). In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat. Genet. *33*, 469-475.

Kofron,M., Puck,H., Standley,H., Wylie,C., Old,R., Whitman,M., and Heasman,J. (2004). New roles for FoxH1 in patterning the early embryo. Development *131*, 5065-5078.

Kolch,W. (2000). Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. Biochem. J *351 Pt 2*, 289-305.

Kuhlbrodt,K., Herbarth,B., Sock,E., Enderich,J., Hermans-Borgmeyer,I., and Wegner,M. (1998). Cooperative function of POU proteins and SOX proteins in glial cells. J Biol. Chem. *273*, 16050-16057.

Kuroda,T., Tada,M., Kubota,H., Kimura,H., Hatano,S.Y., Suemori,H., Nakatsuji,N., and Tada,T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol. Cell Biol. *25*, 2475-2485.

La Thangue,N.B. and Rigby,P.W. (1987). An adenovirus E1A-like transcription factor is regulated during the differentiation of murine embryonal carcinoma stem cells. Cell *49*, 507-513.

Lan,Z.J., Chung,A.C., Xu,X., DeMayo,F.J., and Cooney,A.J. (2002). The embryonic function of germ cell nuclear factor is dependent on the DNA binding domain. J Biol. Chem. *277*, 50660-50667.

Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., Funke,R., Gage,D., Harris,K., Heaford,A., Howland,J., Kann,L., Lehoczky,J., Levine,R., McEwan,P., McKernan,K., Meldrim,J., Mesirov,J.P., Miranda,C., Morris,W., Naylor,J., Raymond,C., Rosetti,M., Santos,R., Sheridan,A., Sougnez,C., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Wyman,D., Rogers,J., Sulston,J., Ainscough,R., Beck,S., Bentley,D., Burton,J., Clee,C., Carter,N., Coulson,A., Deadman,R., Deloukas,P., Dunham,A., Dunham,I., Durbin,R., French,L., Grafham,D., Gregory,S., Hubbard,T., Humphray,S., Hunt,A., Jones,M., Lloyd,C., McMurray,A., Matthews,L., Mercer,S., Milne,S., Mullikin,J.C., Mungall,A., Plumb,R., Ross,M., Shownkeen,R., Sims,S., Waterston,R.H., Wilson,R.K., Hillier,L.W., McPherson,J.D., Marra,M.A., Mardis,E.R., Fulton,L.A., Chinwalla,A.T., Pepin,K.H., Gish,W.R., Chissoe,S.L., Wendl,M.C., Delehaunty,K.D., Miner,T.L., Delehaunty,A., Kramer,J.B., Cook,L.L., Fulton,R.S., Johnson,D.L., Minx,P.J., Clifton,S.W., Hawkins,T., Branscomb,E., Predki,P., Richardson,P., Wenning,S., Slezak,T., Doggett,N., Cheng,J.F., Olsen,A., Lucas,S., Elkin,C., Uberbacher,E., Frazier,M., Gibbs,R.A., Muzny,D.M., Scherer,S.E., Bouck,J.B., Sodergren,E.J., Worley,K.C., Rives,C.M., Gorrell,J.H., Metzker,M.L., Naylor,S.L., Kucherlapati,R.S., Nelson,D.L., Weinstock,G.M., Sakaki,Y., Fujiyama,A., Hattori,M., Yada,T., Toyoda,A., Itoh,T., Kawagoe,C., Watanabe,H., Totoki,Y., Taylor,T., Weissenbach,J., Heilig,R., Saurin,W., Artiguenave,F., Brottier,P., Bruls,T., Pelletier,E., Robert,C., Wincker,P., Smith,D.R., Doucette-Stamm,L., Rubenfield,M., Weinstock,K., Lee,H.M., Dubois,J., Rosenthal,A., Platzer,M., Nyakatura,G., Taudien,S., Rump,A., Yang,H., Yu,J., Wang,J., Huang,G., Gu,J., Hood,L., Rowen,L., Madan,A., Qin,S., Davis,R.W., Federspiel,N.A., Abola,A.P., Proctor,M.J., Myers,R.M., Schmutz,J., Dickson,M., Grimwood,J., Cox,D.R., Olson,M.V., Kaul,R., Raymond,C., Shimizu,N., Kawasaki,K., Minoshima,S., Evans,G.A., Athanasiou,M., Schultz,R., Roe,B.A., Chen,F., Pan,H., Ramser,J., Lehrach,H., Reinhardt,R., McCombie,W.R., de la,B.M., Dedhia,N., Blocker,H., Hornischer,K., Nordsiek,G., Agarwala,R., Aravind,L., Bailey,J.A., Bateman,A., Batzoglou,S., Birney,E., Bork,P.,

Brown,D.G., Burge,C.B., Cerutti,L., Chen,H.C., Church,D., Clamp,M., Copley,R.R., Doerks,T., Eddy,S.R., Eichler,E.E., Furey,T.S., Galagan,J., Gilbert,J.G., Harmon,C., Hayashizaki,Y., Haussler,D., Hermjakob,H., Hokamp,K., Jang,W., Johnson,L.S., Jones,T.A., Kasif,S., Kaspryzk,A., Kennedy,S., Kent,W.J., Kitts,P., Koonin,E.V., Korf,I., Kulp,D., Lancet,D., Lowe,T.M., McLysaght,A., Mikkelsen,T., Moran,J.V., Mulder,N., Pollara,V.J., Ponting,C.P., Schuler,G., Schultz,J., Slater,G., Smit,A.F., Stupka,E., Szustakowski,J., Thierry-Mieg,D., Thierry-Mieg,J., Wagner,L., Wallis,J., Wheeler,R., Williams,A., Wolf,Y.I., Wolfe,K.H., Yang,S.P., Yeh,R.F., Collins,F., Guyer,M.S., Peterson,J., Felsenfeld,A., Wetterstrand,K.A., Patrinos,A., Morgan,M.J., de Jong,P., Catanese,J.J., Osoegawa,K., Shizuya,H., Choi,S., and Chen,Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lanza,R. (2004). Handbook of Stem Cells. Elsevier Academic Press).

Lawson,K.A., Dunn,N.R., Roelen,B.A., Zeinstra,L.M., Davis,A.M., Wright,C.V., Korving,J.P., and Hogan,B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev. *13*, 424-436.

Lee,T.I., Jenner,R.G., Boyer,L.A., Guenther,M.G., Levine,S.S., Kumar,R.M., Chevalier,B., Johnstone,S.E., Cole,M.F., Isono,K., Koseki,H., Fuchikami,T., Abe,K., Murray,H.L., Zucker,J.P., Yuan,B., Bell,G.W., Herbolsheimer,E., Hannett,N.M., Sun,K., Odom,D.T., Otte,A.P., Volkert,T.L., Bartel,D.P., Melton,D.A., Gifford,D.K., Jaenisch,R., and Young,R.A. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lee,T.I., Rinaldi,N.J., Robert,F., Odom,D.T., Bar-Joseph,Z., Gerber,G.K., Hannett,N.M., Harbison,C.T., Thompson,C.M., Simon,I., Zeitlinger,J., Jennings,E.G., Murray,H.L., Gordon,D.B., Ren,B., Wyrick,J.J., Tagne,J.B., Volkert,T.L., Fraenkel,E., Gifford,D.K., and Young,R.A. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.

Levings,P.P., Zhou,Z., Vieira,K.F., Crusselle-Davis,V.J., and Bungert,J. (2006). Recruitment of transcription complexes to the beta-globin locus control region and transcription of hypersensitive site 3 prior to erythroid differentiation of murine embryonic stem cells. FEBS J *273*, 746-755.

Lewis,B.P., Burge,C.B., and Bartel,D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Li,M., Sendtner,M., and Smith,A. (1995). Essential function of LIF receptor in motor neurons. Nature *378*, 724-727.

Lim,C.A. and Ng,H.H. (2007). Application of advanced technologies in ageing research. Mech. Ageing Dev. *128*, 149-160.

Lim,L.S., Loh,Y.H., Zhang,W., Li,Y., Chen,X., Wang,Y., Bakre,M., Ng,H.H., and Stanton,L.W. (2007). Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol. Biol. Cell *18*, 1348-1358.

Lin,T., Chao,C., Saito,S., Mazur,S.J., Murphy,M.E., Appella,E., and Xu,Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nat. Cell Biol. *7*, 165-171.

Ling,J.Q., Li,T., Hu,J.F., Vu,T.H., Chen,H.L., Qiu,X.W., Cherry,A.M., and Hoffman,A.R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science *312*, 269-272.

Liu,E.T. (2005). Genomic technologies and the interrogation of the transcriptome. Mech. Ageing Dev. *126*, 153-159.

Loh,Y.H., Wu,Q., Chew,J.L., Vega,V.B., Zhang,W., Chen,X., Bourque,G., George,J., Leong,B., Liu,J., Wong,K.Y., Sung,K.W., Lee,C.W., Zhao,X.D., Chiu,K.P., Lipovich,L., Kuznetsov,V.A., Robson,P., Stanton,L.W., Wei,C.L., Ruan,Y., Lim,B., and Ng,H.H. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat. Genet. *38*, 431-440.

Lomvardas,S., Barnea,G., Pisapia,D.J., Mendelsohn,M., Kirkland,J., and Axel,R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell *126*, 403-413.

Maltsev,V.A., Rohwedel,J., Hescheler,J., and Wobus,A.M. (1993). Embryonic stem cells differentiate in vitro into cardiomyocytes representing sinusnodal, atrial and ventricular cell types. Mech. Dev. *44*, 41-50.

Malynn,B.A., de Alboran,I.M., O'Hagan,R.C., Bronson,R., Davidson,L., DePinho,R.A., and Alt,F.W. (2000). N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. Genes Dev. *14*, 1390-1399.

Martin,G.R. (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*, 7634-7638.

Martone,R., Euskirchen,G., Bertone,P., Hartman,S., Royce,T.E., Luscombe,N.M., Rinn,J.L., Nelson,F.K., Miller,P., Gerstein,M., Weissman,S., and Snyder,M. (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proc. Natl. Acad. Sci. U. S. A *100*, 12247-12252.

Maruyama,M., Ichisaka,T., Nakagawa,M., and Yamanaka,S. (2005). Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. J Biol. Chem. *280*, 24371-24379.

Masui,S., Nakatake,Y., Toyooka,Y., Shimosato,D., Yagi,R., Takahashi,K., Okochi,H., Okuda,A., Matoba,R., Sharov,A.A., Ko,M.S., and Niwa,H. (2007). Pluripotency governed

by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat. Cell Biol. *9*, 625-635.

Matoba,R., Niwa,H., Masui,S., Ohtsuka,S., Carter,M.G., Sharov,A.A., and Ko,M.S. (2006). Dissecting oct3/4-regulated gene networks in embryonic stem cells by expression profiling. PLoS. ONE. *1*, e26.

Matsuda,T., Nakamura,T., Nakao,K., Arai,T., Katsuki,M., Heike,T., and Yokota,T. (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. EMBO J *18*, 4261-4269.

Matsui,Y., Zsebo,K., and Hogan,B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.

McKnight,S. and Tjian,R. (1986). Transcriptional selectivity of viral genes in mammalian cells. Cell *46*, 795-805.

Merrill,B.J., Pasolli,H.A., Polak,L., Rendl,M., Garcia-Garcia,M.J., Anderson,K.V., and Fuchs,E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. Development *131*, 263-274.

Mi,H., Lazareva-Ulitsky,B., Loo,R., Kejariwal,A., Vandergriff,J., Rabkin,S., Guo,N., Muruganujan,A., Doremieux,O., Campbell,M.J., Kitano,H., and Thomas,P.D. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. *33*, D284-D288.

Miller-Hance,W.C., LaCorbiere,M., Fuller,S.J., Evans,S.M., Lyons,G., Schmidt,C., Robbins,J., and Chien,K.R. (1993). In vitro chamber specification during embryonic stem cell cardiogenesis. Expression of the ventricular myosin light chain-2 gene is independent of heart tube formation. J Biol. Chem. *268*, 25244-25252.

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.

Mitsunaga,K., Araki,K., Mizusaki,H., Morohashi,K., Haruna,K., Nakagata,N., Giguere,V., Yamamura,K., and Abe,K. (2004). Loss of PGC-specific expression of the orphan nuclear receptor ERR-beta results in reduction of germ cell number in mouse embryos. Mech. Dev. *121*, 237-246.

Miyazono,K., Maeda,S., and Imamura,T. (2005). BMP receptor signaling: transcriptional targets, regulation of signals, and signaling cross-talk. Cytokine Growth Factor Rev. *16*, 251-263.

Moore,K.A. and Lemischka,I.R. (2006). Stem cells and their niches. Science *311*, 1880-1885.

Moustakas,A., Souchelnytskyi,S., and Heldin,C.H. (2001). Smad regulation in TGF-beta signal transduction. J Cell Sci. *114*, 4359-4369.

Nakashima,K., Wiese,S., Yanagisawa,M., Arakawa,H., Kimura,N., Hisatsune,T., Yoshida,K., Kishimoto,T., Sendtner,M., and Taga,T. (1999). Developmental requirement of gp130 signaling in neuronal survival and astrocyte differentiation. J Neurosci. *19*, 5429-5434.

Nakatake,Y., Fukui,N., Iwamatsu,Y., Masui,S., Takahashi,K., Yagi,R., Yagi,K., Miyazaki,J., Matoba,R., Ko,M.S., and Niwa,H. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. Mol. Cell Biol. *26*, 7772-7782.

Negre,N., Lavrov,S., Hennetin,J., Bellis,M., and Cavalli,G. (2006). Mapping the distribution of chromatin proteins by ChIP on chip. Methods Enzymol. *410*, 316-341.

Ng,P., Wei,C.L., Sung,W.K., Chiu,K.P., Lipovich,L., Ang,C.C., Gupta,S., Shahab,A., Ridwan,A., Wong,C.H., Liu,E.T., and Ruan,Y. (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. Nat. Methods *2*, 105-111.

Nichols,J., Evans,E.P., and Smith,A.G. (1990). Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. Development *110*, 1341-1348.

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

Nishimoto,M., Fukushima,A., Okuda,A., and Muramatsu,M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol. Cell Biol. *19*, 5453-5465.

Nishimoto,M., Miyagi,S., Yamagishi,T., Sakaguchi,T., Niwa,H., Muramatsu,M., and Okuda,A. (2005). Oct-3/4 maintains the proliferative embryonic stem cell state via specific binding to a variant octamer sequence in the regulatory region of the UTF1 locus. Mol. Cell Biol. *25*, 5084-5094.

Niwa,H. (2001). Molecular mechanism to maintain stem cell renewal of ES cells. Cell Struct. Funct. *26*, 137-148.

Niwa,H. (2007). How is pluripotency determined and maintained? Development *134*, 635-646.

Niwa,H., Burdon,T., Chambers,I., and Smith,A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. *12*, 2048-2060.

Niwa,H., Masui,S., Chambers,I., Smith,A.G., and Miyazaki,J. (2002). Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. Mol. Cell Biol. *22*, 1526-1536.

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

Niwa,H., Toyooka,Y., Shimosato,D., Strumpf,D., Takahashi,K., Yagi,R., and Rossant,J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Nordhoff,V., Hubner,K., Bauer,A., Orlova,I., Malapetsa,A., and Scholer,H.R. (2001). Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. Mamm. Genome *12*, 309-317.

Norris,D.P., Brennan,J., Bikoff,E.K., and Robertson,E.J. (2002). The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo. Development *129*, 3455-3468.

Okamoto,K., Okazawa,H., Okuda,A., Sakai,M., Muramatsu,M., and Hamada,H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell *60*, 461-472.

Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E., and Hayashizaki,Y. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature *420*, 563-573.

Okazawa,H., Okamoto,K., Ishino,F., Ishino-Kaneko,T., Takeda,S., Toyoda,Y., Muramatsu,M., and Hamada,H. (1991). The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. EMBO J *10*, 2997-3005.

Okuda,A., Fukushima,A., Nishimoto,M., Orimo,A., Yamagishi,T., Nabeshima,Y., Kuro-o M, Nabeshima,Y., Boon,K., Keaveney,M., Stunnenberg,H.G., and Muramatsu,M. (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. EMBO J *17*, 2019-2032.

Okumura-Nakanishi,S., Saito,M., Niwa,H., and Ishikawa,F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol. Chem. *280*, 5307-5317.

Orlando,V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem. Sci. *25*, 99-104.

Paddison,P.J., Silva,J.M., Conklin,D.S., Schlabach,M., Li,M., Aruleba,S., Balija,V., O'Shaughnessy,A., Gnoj,L., Scobie,K., Chang,K., Westbrook,T., Cleary,M., Sachidanandam,R., McCombie,W.R., Elledge,S.J., and Hannon,G.J. (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature *428*, 427-431.

Palmieri,S.L., Peter,W., Hess,H., and Scholer,H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Dev. Biol. *166*, 259-267.

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

Pan,G., Li,J., Zhou,Y., Zheng,H., and Pei,D. (2006). A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. FASEB J *20*, 1730-1732.

Pavesi,G., Mereghetti,P., Mauri,G., and Pesole,G. (2004). Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res. *32*, W199-W203.

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

Pera,M.F., Reubinoff,B., and Trounson,A. (2000). Human embryonic stem cells. J Cell Sci. *113 (Pt 1)*, 5-10.

Pereira,L., Yi,F., and Merrill,B.J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. Mol. Cell Biol. *26*, 7479-7491.

Petersen,B.E. and Terada,N. (2001). Stem cells: a journey into a new frontier. J Am. Soc. Nephrol. *12*, 1773-1780.

Pierce GB, Arechaga J, Muro C, Wells RS. (1988) Differentiation of ICM cells into trophectoderm. Am J Pathol. *132(2)*:356-64.

Pikarsky,E., Sharir,H., Ben Shushan,E., and Bergman,Y. (1994). Retinoic acid represses Oct-3/4 gene expression through several retinoic acid-responsive elements located in the promoter-enhancer region. Mol. Cell Biol. *14*, 1026-1038.

Pritsker,M., Ford,N.R., Jenq,H.T., and Lemischka,I.R. (2006). Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A *103*, 6946-6951.

Prost,S., Bellamy,C.O., Clarke,A.R., Wyllie,A.H., and Harrison,D.J. (1998). p53-independent DNA repair and cell cycle arrest in embryonic stem cells. FEBS Lett. *425*, 499-504.

Qi,X., Li,T.G., Hao,J., Hu,J., Wang,J., Simmons,H., Miura,S., Mishina,Y., and Zhao,G.Q. (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. Proc. Natl. Acad. Sci. U. S. A *101*, 6027-6032.

Rajewsky,K., Gu,H., Kuhn,R., Betz,U.A., Muller,W., Roes,J., and Schwenk,F. (1996). Conditional gene targeting. J Clin. Invest *98*, 600-603.

Ramalho-Santos,M., Yoon,S., Matsuzaki,Y., Mulligan,R.C., and Melton,D.A. (2002). "Stemness": transcriptional profiling of embryonic and adult stem cells. Science *298*, 597-600.

Rao,R.R. and Stice,S.L. (2004). Gene expression profiling of embryonic stem cells leads to greater understanding of pluripotency and early developmental events. Biol. Reprod. *71*, 1772-1778.

Rathjen,P.D., Toth,S., Willis,A., Heath,J.K., and Smith,A.G. (1990). Differentiation inhibiting activity is produced in matrix-associated and diffusible forms that are generated by alternate promoter usage. Cell *62*, 1105-1114.

Rayasam,G.V., Wendling,O., Angrand,P.O., Mark,M., Niederreither,K., Song,L., Lerouge,T., Hager,G.L., Chambon,P., and Losson,R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J *22*, 3153-3163.

Raz,R., Lee,C.K., Cannizzaro,L.A., d'Eustachio,P., and Levy,D.E. (1999). Essential role of STAT3 for embryonic stem cell pluripotency. Proc. Natl. Acad. Sci. U. S. A *96*, 2846-2851.

Remenyi,A., Lins,K., Nissen,L.J., Reinbold,R., Scholer,H.R., and Wilmanns,M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev. *17*, 2048-2059.

Ren,B., Robert,F., Wyrick,J.J., Aparicio,O., Jennings,E.G., Simon,I., Zeitlinger,J., Schreiber,J., Hannett,N., Kanin,E., Volkert,T.L., Wilson,C.J., Bell,S.P., and Young,R.A. (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Reubinoff,B.E., Pera, M.F., Fong, C., Trounson, A., and Bongso, A. (2000) Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. Nature Biotechnol. *18*, 399–404.

Reya,T. and Clevers,H. (2005). Wnt signalling in stem cells and cancer. Nature *434*, 843-850.

Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S., and Khvorova,A. (2004). Rational siRNA design for RNA interference. Nat. Biotechnol. *22*, 326-330.

Richards,M., Tan,S.P., Tan,J.H., Chan,W.K., and Bongso,A. (2004). The transcriptome profile of human embryonic stem cells as defined by SAGE. Stem Cells *22*, 51-64.

Rodda,D.J., Chew,J.L., Lim,L.H., Loh,Y.H., Wang,B., Ng,H.H., and Robson,P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol. Chem. *280*, 24731-24737.

Roh,T.Y., Ngau,W.C., Cui,K., Landsman,D., and Zhao,K. (2004). High-resolution genome-wide mapping of histone modifications. Nat. Biotechnol. *22*, 1013-1016.

Rosner,M.H., Vigano,M.A., Ozato,K., Timmons,P.M., Poirier,F., Rigby,P.W., 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.

Ruan,Y., Le Ber,P., Ng,H.H., and Liu,E.T. (2004). Interrogating the transcriptome. Trends Biotechnol. *22*, 23-30.

Russ,A.P., Wattler,S., Colledge,W.H., Aparicio,S.A., Carlton,M.B., Pearce,J.J., Barton,S.C., Surani,M.A., Ryan,K., Nehls,M.C., Wilson,V., and Evans,M.J. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature *404*, 95-99.

Sabapathy,K., Klemm,M., Jaenisch,R., and Wagner,E.F. (1997). Regulation of ES cell differentiation by functional and conformational modulation of p53. EMBO J *16*, 6217-6229.

Saitou,M., Barton,S.C., and Surani,M.A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293-300.

Sakaki-Yumoto,M., Kobayashi,C., Sato,A., Fujimura,S., Matsumoto,Y., Takasato,M., Kodama,T., Aburatani,H., Asashima,M., Yoshida,N., and Nishinakamura,R. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. Development *133*, 3005-3013.

Sambrook,J. and Russell,D. (2001). Molecular Cloning: A Laboratory Manual (3rd Edition). Cold Spring Harbor Laboratory Press).

Santos-Rosa,H., Schneider,R., Bannister,A.J., Sherriff,J., Bernstein,B.E., Emre,N.C., Schreiber,S.L., Mellor,J., and Kouzarides,T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Santos-Rosa,H., Valls,E., Kouzarides,T., and Martinez-Balbas,M. (2003). Mechanisms of P/CAF auto-acetylation. Nucleic Acids Res. *31*, 4285-4292.

Sato,N., Meijer,L., Skaltsounis,L., Greengard,P., and Brivanlou,A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat. Med. *10*, 55-63.

Scacheri,P.C., Davis,S., Odom,D.T., Crawford,G.E., Perkins,S., Halawi,M.J., Agarwal,S.K., Marx,S.J., Spiegel,A.M., Meltzer,P.S., and Collins,F.S. (2006). Genome-wide analysis of menin binding provides insights into MEN1 tumorigenesis. PLoS. Genet. *2*, e51.

Schmitt,R.M., Bruyns,E., and Snodgrass,H.R. (1991). Hematopoietic development of embryonic stem cells in vitro: cytokine and receptor gene expression. Genes Dev. *5*, 728-740.

Scholer,H.R. (1991). Octamania: the POU factors in murine development. Trends Genet. *7*, 323-329.

Scholer,H.R., Ciesiolka,T., and Gruss,P. (1991). A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. Cell *66*, 291-304.

Scholer,H.R., Hatzopoulos,A.K., Balling,R., Suzuki,N., and Gruss,P. (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. EMBO J *8*, 2543-2550.

Scholer,H.R., Ruppert,S., Suzuki,N., Chowdhury,K., and Gruss,P. (1990). New type of POU domain in germ line-specific protein Oct-4. Nature *344*, 435-439.

Schoorlemmer,J., van Puijenbroek,A., van Den,E.M., Jonk,L., Pals,C., and Kruijer,W. (1994). Characterization of a negative retinoic acid response element in the murine Oct4 promoter. Mol. Cell Biol. *14*, 1122-1136.

Sen,G.L. and Blau,H.M. (2006). A brief history of RNAi: the silence of the genes. FASEB J *20*, 1293-1299.

Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y., VanBuren,V., Falco,G., Martin,P.R., Stagg,C.A., Bassey,U.C., Wang,Y., Carter,M.G., Hamatani,T., Aiba,K., Akutsu,H., Sharova,L., Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S., Nagaraja,R., Boheler,K.R., Taub,D., Hodes,R.J., Longo,D.L., Schlessinger,D., Keller,J., Klotz,E., Kelsoe,G., Umezawa,A., Vescovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A., D'Urso,M., Kelso,J., Hide,W., and Ko,M.S. (2003). Transcriptome analysis of mouse stem cells and early embryos. PLoS. Biol. *1*, E74.

Shi,Y. and Massague,J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Silva,J., Chambers,I., Pollard,S., and Smith,A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997-1001.

Silverman,J., Takai,H., Buonomo,S.B., Eisenhaber,F., and de Lange,T. (2004). Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint. Genes Dev. *18*, 2108-2119.

Simonis,M., Klous,P., Splinter,E., Moshkin,Y., Willemsen,R., de Wit,E., van Steensel,B., and de Laat,W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348-1354.

Smith,A. (2005). The battlefield of pluripotency. Cell *123*, 757-760.

Smith,A.G. (2001). Embryo-derived stem cells: of mice and men. Annu. Rev. Cell Dev. Biol. *17*, 435-462.

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.

Smith,A.G., Nichols,J., Robertson,M., and Rathjen,P.D. (1992). Differentiation inhibiting activity (DIA/LIF) and mouse development. Dev. Biol. *151*, 339-351.

Solomon,M.J., Larsen,P.L., and Varshavsky,A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell *53*, 937-947.

Soprano,D.R., Teets,B.W., and Soprano,K.J. (2007). Role of retinoic Acid in the differentiation of embryonal carcinoma and embryonic stem cells. Vitam. Horm. *75*, 69-95.

Spivakov,M. and Fisher,A.G. (2007). Epigenetic signatures of stem-cell identity. Nat. Rev. Genet. *8*, 263-271.

Stears,R.L., Martinsky,T., and Schena,M. (2003). Trends in microarray analysis. Nat. Med. *9*, 140-145.

Stewart,C.L. and Cullinan,E.B. (1997). Preimplantation development of the mammalian embryo and its regulation by growth factors. Dev. Genet. *21*, 91-101.

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.

Sun,H., Lesche,R., Li,D.M., Liliental,J., Zhang,H., Gao,J., Gavrilova,N., Mueller,B., Liu,X., and Wu,H. (1999). PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. Proc. Natl. Acad. Sci. U. S. A *96*, 6199-6204.

Suzuki,A., Raya,A., Kawakami,Y., Morita,M., Matsui,T., Nakashima,K., Gage,F.H., Rodriguez-Esteban,C., and Izpisua Belmonte,J.C. (2006). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A *103*, 10294-10299.

Sylvester,I. and Scholer,H.R. (1994). Regulation of the Oct-4 gene by nuclear receptors. Nucleic Acids Res. *22*, 901-911.

Tada,M., Takahama,Y., Abe,K., Nakatsuji,N., and Tada,T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. *11*, 1553-1558.

Takahashi,K., Mitsui,K., and Yamanaka,S. (2003). Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. Nature *423*, 541-545.

Takahashi,K. and Yamanaka,S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Takai,H., Smogorzewska,A., and de Lange,T. (2003). DNA damage foci at dysfunctional telomeres. Curr. Biol. *13*, 1549-1556.

Takayama,K., Kaneshiro,K., Tsutsumi,S., Horie-Inoue,K., Ikeda,K., Urano,T., Ijichi,N., Ouchi,Y., Shirahige,K., Aburatani,H., and Inoue,S. (2007). Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis. Oncogene *26*, 4453-4463.

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. U. S. A *94*, 3801-3804.

Tanaka,S., Kunath,T., Hadjantonakis,A.K., Nagy,A., and Rossant,J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. Science *282*, 2072-2075.

Tanaka,T.S., Kunath,T., Kimber,W.L., Jaradat,S.A., Stagg,C.A., Usuda,M., Yokota,T., Niwa,H., Rossant,J., and Ko,M.S. (2002). Gene expression profiling of embryo-derived

stem cells reveals candidate genes associated with pluripotency and lineage specificity. Genome Res. *12*, 1921-1928.

Tanaka,Y., Patestos,N.P., Maekawa,T., and Ishii,S. (1999). B-myb is required for inner cell mass formation at an early stage of development. J Biol. Chem. *274*, 28067-28070.

Thomson,J.A., Itskovitz-Eldor,J., Shapiro,S.S., Waknitz,M.A., Swiergiel,J.J., Marshall,V.S., and Jones,J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Tjian,R. and Maniatis,T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. Cell *77*, 5-8.

Tokuzawa,Y., Kaiho,E., Maruyama,M., Takahashi,K., Mitsui,K., Maeda,M., Niwa,H., and Yamanaka,S. (2003). Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. Mol. Cell Biol. *23*, 2699-2708.

Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F., and de Laat,W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol. Cell *10*, 1453-1465.

Tolkunova,E., Cavaleri,F., Eckardt,S., Reinbold,R., Christenson,L.K., Scholer,H.R., and Tomilin,A. (2006). The caudal-related protein cdx2 promotes trophoblast differentiation of mouse embryonic stem cells. Stem Cells *24*, 139-144.

Tomioka,M., Nishimoto,M., Miyagi,S., Katayanagi,T., Fukui,N., Niwa,H., Muramatsu,M., and Okuda,A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res. *30*, 3202-3213.

Tremblay,G.B., Kunath,T., Bergeron,D., Lapointe,L., Champigny,C., Bader,J.A., Rossant,J., and Giguere,V. (2001a). Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERR beta. Genes Dev. *15*, 833-838.

Tremblay,K.D., Dunn,N.R., and Robertson,E.J. (2001b). Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development *128*, 3609-3621.

Tropepe,V., Hitoshi,S., Sirard,C., Mak,T.W., Rossant,J., and van der,K.D. (2001). Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. Neuron *30*, 65-78.

Ui-Tei,K., Naito,Y., Takahashi,F., Haraguchi,T., Ohki-Hamazaki,H., Juni,A., Ueda,R., and Saigo,K. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. Nucleic Acids Res. *32*, 936-948.

Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., Gocayne,J.D., Amanatides,P., Ballew,R.M., Huson,D.H., Wortman,J.R., Zhang,Q., Kodira,C.D., Zheng,X.H., Chen,L., Skupski,M.,

Subramanian,G., Thomas,P.D., Zhang,J., Gabor Miklos,G.L., Nelson,C., Broder,S., Clark,A.G., Nadeau,J., McKusick,V.A., Zinder,N., Levine,A.J., Roberts,R.J., Simon,M., Slayman,C., Hunkapiller,M., Bolanos,R., Delcher,A., Dew,I., Fasulo,D., Flanigan,M., Florea,L., Halpern,A., Hannenhalli,S., Kravitz,S., Levy,S., Mobarry,C., Reinert,K., Remington,K., Abu-Threideh,J., Beasley,E., Biddick,K., Bonazzi,V., Brandon,R., Cargill,M., Chandramouliswaran,I., Charlab,R., Chaturvedi,K., Deng,Z., Di,F., V, Dunn,P., Eilbeck,K., Evangelista,C., Gabrielian,A.E., Gan,W., Ge,W., Gong,F., Gu,Z., Guan,P., Heiman,T.J., Higgins,M.E., Ji,R.R., Ke,Z., Ketchum,K.A., Lai,Z., Lei,Y., Li,Z., Li,J., Liang,Y., Lin,X., Lu,F., Merkulov,G.V., Milshina,N., Moore,H.M., Naik,A.K., Narayan,V.A., Neelam,B., Nusskern,D., Rusch,D.B., Salzberg,S., Shao,W., Shue,B., Sun,J., Wang,Z., Wang,A., Wang,X., Wang,J., Wei,M., Wides,R., Xiao,C., Yan,C., Yao,A., Ye,J., Zhan,M., Zhang,W., Zhang,H., Zhao,Q., Zheng,L., Zhong,F., Zhong,W., Zhu,S., Zhao,S., Gilbert,D., Baumhueter,S., Spier,G., Carter,C., Cravchik,A., Woodage,T., Ali,F., An,H., Awe,A., Baldwin,D., Baden,H., Barnstead,M., Barrow,I., Beeson,K., Busam,D., Carver,A., Center,A., Cheng,M.L., Curry,L., Danaher,S., Davenport,L., Desilets,R., Dietz,S., Dodson,K., Doup,L., Ferriera,S., Garg,N., Gluecksmann,A., Hart,B., Haynes,J., Haynes,C., Heiner,C., Hladun,S., Hostin,D., Houck,J., Howland,T., Ibegwam,C., Johnson,J., Kalush,F., Kline,L., Koduru,S., Love,A., Mann,F., May,D., McCawley,S., McIntosh,T., McMullen,I., Moy,M., Moy,L., Murphy,B., Nelson,K., Pfannkoch,C., Pratts,E., Puri,V., Qureshi,H., Reardon,M., Rodriguez,R., Rogers,Y.H., Romblad,D., Ruhfel,B., Scott,R., Sitter,C., Smallwood,M., Stewart,E., Strong,R., Suh,E., Thomas,R., Tint,N.N., Tse,S., Vech,C., Wang,G., Wetter,J., Williams,S., Williams,M., Windsor,S., Winn-Deen,E., Wolfe,K., Zaveri,J., Zaveri,K., Abril,J.F., Guigo,R., Campbell,M.J., Sjolander,K.V., Karlak,B., Kejariwal,A., Mi,H., Lazareva,B., Hatton,T., Narechania,A., Diemer,K., Muruganujan,A., Guo,N., Sato,S., Bafna,V., Istrail,S., Lippert,R., Schwartz,R., Walenz,B., Yooseph,S., Allen,D., Basu,A., Baxendale,J., Blick,L., Caminha,M., Carnes-Stine,J., Caulk,P., Chiang,Y.H., Coyne,M., Dahlke,C., Mays,A., Dombroski,M., Donnelly,M., Ely,D., Esparham,S., Fosler,C., Gire,H., Glanowski,S., Glasser,K., Glodek,A., Gorokhov,M., Graham,K., Gropman,B., Harris,M., Heil,J., Henderson,S., Hoover,J., Jennings,D., Jordan,C., Jordan,J., Kasha,J., Kagan,L., Kraft,C., Levitsky,A., Lewis,M., Liu,X., Lopez,J., Ma,D., Majoros,W., McDaniel,J., Murphy,S., Newman,M., Nguyen,T., Nguyen,N., and Nodell,M. (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vernimmen,D., De Gobbi,M., Sloane-Stanley,J., Wood,W., and Higgs,D. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J *26*, 2041-2051.

Verrijzer,C.P., van Oosterhout,J.A., van Weperen,W.W., and van der Vliet,P.C. (1991). POU proteins bend DNA via the POU-specific domain. EMBO J *10*, 3007-3014.

Vigano,M.A. and Staudt,L.M. (1996). Transcriptional activation by Oct-3: evidence for a specific role of the POU-specific domain in mediating functional interaction with Oct-1. Nucleic Acids Res. *24*, 2112-2118.

Vittet,D., Prandini,M.H., Berthier,R., Schweitzer,A., Martin-Sisteron,H., Uzan,G., and Dejana,E. (1996). Embryonic stem cells differentiate in vitro to endothelial cells through successive maturation steps. Blood *88*, 3424-3431.

von,B., I, Silvestri,C., Erdemir,T., Lickert,H., Walls,J.R., Henkelman,R.M., Rossant,J., Harvey,R.P., Attisano,L., and Wrana,J.L. (2004). Foxh1 is essential for development of the anterior heart field. Dev. Cell *7*, 331-345.

Wang,B., Matsuoka,S., Carpenter,P.B., and Elledge,S.J. (2002). 53BP1, a mediator of the DNA damage checkpoint. Science *298*, 1435-1438.

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.

Wang,Z.X., Kueh,J.L., Teh,C.H., Rossbach,M., Lim,L., Li,P., Wong,K.Y., Lufkin,T., Robson,P., and Stanton,L.W. (2007). Zfp206 Is A Transcription Factor That Controls Pluripotency of Embryonic Stem Cells. Stem Cells.

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., and . (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.

Watanabe,S., Umehara,H., Murayama,K., Okabe,M., Kimura,T., and Nakano,T. (2006). Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. Oncogene *25*, 2697-2707.

Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M., An,P., Antonarakis,S.E., Attwood,J., Baertsch,R., Bailey,J., Barlow,K., Beck,S., Berry,E., Birren,B., Bloom,T., Bork,P., Botcherby,M., Bray,N., Brent,M.R., Brown,D.G., Brown,S.D., Bult,C., Burton,J., Butler,J., Campbell,R.D., Carninci,P., Cawley,S., Chiaromonte,F., Chinwalla,A.T., Church,D.M., Clamp,M., Clee,C., Collins,F.S., Cook,L.L., Copley,R.R., Coulson,A., Couronne,O., Cuff,J., Curwen,V., Cutts,T., Daly,M., David,R., Davies,J., Delehaunty,K.D., Deri,J., Dermitzakis,E.T., Dewey,C., Dickens,N.J., Diekhans,M., Dodge,S., Dubchak,I., Dunn,D.M., Eddy,S.R., Elnitski,L., Emes,R.D., Eswara,P., Eyras,E., Felsenfeld,A., Fewell,G.A., Flicek,P., Foley,K., Frankel,W.N., Fulton,L.A., Fulton,R.S., Furey,T.S., Gage,D., Gibbs,R.A., Glusman,G., Gnerre,S., Goldman,N., Goodstadt,L., Grafham,D., Graves,T.A., Green,E.D., Gregory,S., Guigo,R., Guyer,M., Hardison,R.C., Haussler,D., Hayashizaki,Y., Hillier,L.W., Hinrichs,A., Hlavina,W., Holzer,T., Hsu,F., Hua,A., Hubbard,T., Hunt,A., Jackson,I., Jaffe,D.B., Johnson,L.S., Jones,M., Jones,T.A., Joy,A., Kamal,M., Karlsson,E.K., Karolchik,D., Kasprzyk,A., Kawai,J., Keibler,E., Kells,C., Kent,W.J., Kirby,A., Kolbe,D.L., Korf,I., Kucherlapati,R.S., Kulbokas,E.J., Kulp,D., Landers,T., Leger,J.P., Leonard,S., Letunic,I., Levine,R., Li,J., Li,M., Lloyd,C., Lucas,S., Ma,B., Maglott,D.R., Mardis,E.R., Matthews,L., Mauceli,E., Mayer,J.H., McCarthy,M., McCombie,W.R., McLaren,S.,

McLay,K., McPherson,J.D., Meldrim,J., Meredith,B., Mesirov,J.P., Miller,W., Miner,T.L., Mongin,E., Montgomery,K.T., Morgan,M., Mott,R., Mullikin,J.C., Muzny,D.M., Nash,W.E., Nelson,J.O., Nhan,M.N., Nicol,R., Ning,Z., Nusbaum,C., O'Connor,M.J., Okazaki,Y., Oliver,K., Overton-Larty,E., Pachter,L., Parra,G., Pepin,K.H., Peterson,J., Pevzner,P., Plumb,R., Pohl,C.S., Poliakov,A., Ponce,T.C., Ponting,C.P., Potter,S., Quail,M., Reymond,A., Roe,B.A., Roskin,K.M., Rubin,E.M., Rust,A.G., Santos,R., Sapojnikov,V., Schultz,B., Schultz,J., Schwartz,M.S., Schwartz,S., Scott,C., Seaman,S., Searle,S., Sharpe,T., Sheridan,A., Shownkeen,R., Sims,S., Singer,J.B., Slater,G., Smit,A., Smith,D.R., Spencer,B., Stabenau,A., Stange-Thomann,N., Sugnet,C., Suyama,M., Tesler,G., Thompson,J., Torrents,D., Trevaskis,E., Tromp,J., Ucla,C., Ureta-Vidal,A., Vinson,J.P., Von Niederhausern,A.C., Wade,C.M., Wall,M., Weber,R.J., Weiss,R.B., Wendl,M.C., West,A.P., Wetterstrand,K., Wheeler,R., Whelan,S., Wierzbowski,J., Willey,D., Williams,S., Wilson,R.K., Winter,E., Worley,K.C., Wyman,D., Yang,S., Yang,S.P., Zdobnov,E.M., Zody,M.C., and Lander,E.S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wegner,M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. *27*, 1409-1420.

Wei,C.L., Miura,T., Robson,P., Lim,S.K., Xu,X.Q., Lee,M.Y., Gupta,S., Stanton,L., Luo,Y., Schmitt,J., Thies,S., Wang,W., Khrebtukova,I., Zhou,D., Liu,E.T., Ruan,Y.J., Rao,M., and Lim,B. (2005). Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. Stem Cells *23*, 166-185.

Wei,C.L., Wu,Q., Vega,V.B., Chiu,K.P., Ng,P., Zhang,T., Shahab,A., Yong,H.C., Fu,Y., Weng,Z., Liu,J., Zhao,X.D., Chew,J.L., Lee,Y.L., Kuznetsov,V.A., Sung,W.K., Miller,L.D., Lim,B., Liu,E.T., Yu,Q., Ng,H.H., and Ruan,Y. (2006). A global map of p53 transcription-factor binding sites in the human genome. Cell *124*, 207-219.

Weinmann,A.S., Yan,P.S., Oberley,M.J., Huang,T.H., and Farnham,P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev. *16*, 235-244.

Weissman,I.L., Anderson,D.J., and Gage,F. (2001). Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. Annu. Rev. Cell Dev. Biol. *17*, 387-403.

Wells,J., Yan,P.S., Cechvala,M., Huang,T., and Farnham,P.J. (2003). Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. Oncogene *22*, 1445-1460.

Wernig,M., Meissner,A., Foreman,R., Brambrink,T., Ku,M., Hochedlinger,K., Bernstein,B.E., and Jaenisch,R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.

Williams,D.C., Jr., Cai,M., and Clore,G.M. (2004). Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-

kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex. J Biol. Chem. *279*, 1449-1457.

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.

Wingender,E., Chen,X., Hehl,R., Karas,H., Liebich,I., Matys,V., Meinhardt,T., Pruss,M., Reuter,I., and Schacherer,F. (2000). TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. *28*, 316-319.

Winnier,G., Blessing,M., Labosky,P.A., and Hogan,B.L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev. *9*, 2105-2116.

Wolffe,A.P., Wong,J., and Pruss,D. (1997). Activators and repressors: making use of chromatin to regulate transcription. Genes Cells *2*, 291-302.

Wormald,S., Hilton,D.J., Smyth,G.K., and Speed,T.P. (2006). Proximal genomic localization of STAT1 binding and regulated transcriptional activity. BMC. Genomics *7*, 254.

Wotton,D., Lo,R.S., Lee,S., and Massague,J. (1999). A Smad transcriptional corepressor. Cell *97*, 29-39.

Wu,Q., Chen,X., Zhang,J., Loh,Y.H., Low,T.Y., Zhang,W., Zhang,W., Sze,S.K., Lim,B., and Ng,H.H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol. Chem. *281*, 24090-24094.

Wurtele,H. and Chartrand,P. (2006). Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. Chromosome. Res. *14*, 477-495.

Xu,C., Inokuma,M.S., Denham,J., Golds,K., Kundu,P., Gold,J.D., and Carpenter,M.K. (2001). Feeder-free growth of undifferentiated human embryonic stem cells. Nat. Biotechnol. *19*, 971-974.

Xu,L. and Blackburn,E.H. (2004). Human Rif1 protein binds aberrant telomeres and aligns along anaphase midzone microtubules. J Cell Biol. *167*, 819-830.

Xu,Y. (2005). A new role for p53 in maintaining genetic stability in embryonic stem cells. Cell Cycle *4*, 363-364.

Yeom,Y.I., Fuhrmann,G., Ovitt,C.E., Brehm,A., Ohbo,K., Gross,M., Hubner,K., and Scholer,H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881-894.

Yeom,Y.I., Ha,H.S., Balling,R., Scholer,H.R., and Artzt,K. (1991). Structure, expression and chromosomal location of the Oct-4 gene. Mech. Dev. *35*, 171-179.

Ying,Q.L., Nichols,J., Chambers,I., and 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.

Yoshida,K., Taga,T., Saito,M., Suematsu,S., Kumanogoh,A., Tanaka,T., Fujiwara,H., Hirata,M., Yamagami,T., Nakahata,T., Hirabayashi,T., Yoneda,Y., Tanaka,K., Wang,W.Z., Mori,C., Shiota,K., Yoshida,N., and Kishimoto,T. (1996). Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. Proc. Natl. Acad. Sci. U. S. A *93*, 407-411.

Yuan,H., Corbi,N., Basilico,C., and Dailey,L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. *9*, 2635-2645.

Zappone,M.V., Galli,R., Catena,R., Meani,N., De Biasi,S., Mattei,E., Tiveron,C., Vescovi,A.L., Lovell-Badge,R., Ottolenghi,S., and Nicolis,S.K. (2000). Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. Development *127*, 2367-2382.

Zhang,J. and Li,L. (2005). BMP signaling and stem cell regulation. Dev. Biol. *284*, 1-11.

Zhang,J.Z., Gao,W., Yang,H.B., Zhang,B., Zhu,Z.Y., and Xue,Y.F. (2006). Screening for genes essential for mouse embryonic stem cell self-renewal using a subtractive RNA interference library. Stem Cells *24*, 2661-2668.

Zhao,Z., Tavoosidana,G., Sjolinder,M., Gondor,A., Mariano,P., Wang,S., Kanduri,C., Lezcano,M., Sandhu,K.S., Singh,U., Pant,V., Tiwari,V., Kurukuti,S., and Ohlsson,R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341-1347.

Zhou,Q. and Liu,J.S. (2004). Modeling within-motif dependence for transcription factor binding site predictions. Bioinformatics. *20*, 909-916.

Appendix A: Coordinates of 1083 Oct4 binding loci and their associated genes.

This table shows a list of 1083 Oct4 bound loci, sorted by the numbers of overlapping PETs at each locus. Cluster ID is a unique ID that was generated for each Oct4 binding locus. MPSS (ES) column indicates the number of unique MPSS tag associated with each candidate Oct4 bound gene in ES cells. MPSS (EB) and MPSS (NS) columns show the quantification of MPSS tags in embryoid bodies and neurospheres respectively. The Molecular Function based on Panther classification is shown.

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr2.52030891	chr2:52030891-52032766	17	Rif1	587	5' proximal	coding	125	0	23	#DIV/0!	GeneID:51869	Molecular function unclassified
chr17.6215383	chr17:6215383-6219112	15	BB862374 EST	2500	5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr14.78241274	chr14:78241274-78242903	15	AK049993	100000	gene desert	coding	0	0	0	#DIV/0!	NA	NA
chr13.44169368	chr13:44169368-44170997	14	Jarid2	25165	5' distal	coding	97	20	0	4.85	GeneID:16468	Other transcription factor;Nucleic acid binding
chr19.4597833	chr19:4597833-4599903	13	Rbm14	6402	intragenic (intron)	coding	89	36	18	2.47	GeneID:56275	Molecular function unclassified
chr15.75387723	chr15:75387723-75389094	13	Ly6c	41929	5' distal	coding	0	0	0	#DIV/0!	GeneID:17067	Molecular function unclassified
chr3.122525943	chr3:122525943-122527756	12	Abca4	2106	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:11304	ATP-binding cassette (ABC) transporter
chr12.75308203	chr12:75308203-75309760	12	Zfp36l1	97624	3' distal	coding	1	0	93	#DIV/0!	GeneID:12192	Nuclease
chr6.87283487	chr6:87283487-87284723	11	CX569013 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr5.26608593	chr5:26608593-26610244	11	9630008K15Rik	58	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:269637	Molecular function unclassified
chr16.16510489	chr16:16510489-16514420	11	Spag6	17602	3' distal	coding	25	32	91	0.78	GeneID:50525	Molecular function unclassified
chr1.72651779	chr1:72651779-72653457	11	AK044410	8000	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr6.114408767	chr6:114408767-114410137	10	AK051287	4136	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr9.118375727	chr9:118375727-118376953	9	BF016740 EST	200	5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr8.88317164	chr8:88317164-88318110	9	Sall1	24864	3' distal	coding	0	0	0	#DIV/0!	GeneID:58198	Zinc finger transcription factor;Nucleic acid binding
chr3.53739446	chr3:53739446-53740623	9	Frem2	6793	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:242022	NA
chr3.18290400	chr3:18290400-18291857	9	AK049712		intragenic (exon)	noncoding	0	0	0	#DIV/0!	NA	NA
chr2.44484915	chr2:44484915-44486186	9	Gtdc1	38905	3' distal	coding	14	8	20	1.75	GeneID:227835	Molecular function unclassified
chr17.48606569	chr17:48606569-48607407	9	Ptcl2	81554	5' distal	coding	0	21	0	0.00	GeneID:224860	Phospholipase;Select calcium binding protein
chr17.34357035	chr17:34357035-34357787	9	Myg1	25	5' proximal	coding	30	18	59	1.67	GeneID:60315	Other hydrolase
chr17.34012244	chr17:34012244-34014108	9	Pou5f1	1988	5' proximal	coding	388	21	0	18.48	GeneID:18999	Homeobox transcription factor;Nucleic acid binding
chr14.62399173	chr14:62399173-62400634	9	Epb4.9	3477	5' proximal	coding	19	0	0	#DIV/0!	NA	NA
chr12.51489439	chr12:51489439-51490705	9	Mbip	30042	3' distal	coding	0	0	15	#DIV/0!	GeneID:217588	Kinase inhibitor
chr10.24790759	chr10:24790759-24792362	9	BB849809 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr1.53325503	chr1:53325503-53326632	9	1700019D03Rik	918	5' proximal	coding	0	0	0	#DIV/0!	GeneID:67080	Molecular function unclassified
chr9.94511327	chr9:94511327-94512909	8	Slc9a9	7941	5' proximal	coding	0	0	0	#DIV/0!	GeneID:331004	Cation transporter
chr9.64598866	chr9:64598866-64602304	8	AV340375	20554	5' distal	coding	1	0	31	#DIV/0!	GeneID:213550	Exoribonuclease;Hydrolase
chr9.58486528	chr9:58486528-58488087	8	Loxl1	8706	3' proximal	coding	0	42	0	0.00	GeneID:16949	Oxidase;Other extracellular matrix
chr6.39372789	chr6:39372789-39374414	8	CZ466057		intragenic (exon)	noncoding	0	0	38	#DIV/0!	NA	NA
chr6.34241860	chr6:34241860-34242479	8	Akr1b8	9935	5' proximal	coding	5	21	0	0.24	GeneID:14187	Oxidoreductase
chr4.134603940	chr4:134603940-134605373	8	AK031499	0	intragenic (exon)	coding	9	28	195	0.32	NA	NA
chr3.89063126	chr3:89063126-89064012	8	Ubqln4	2396	3' proximal	coding	145	171	37	0.85	GeneID:94232	Other miscellaneous function protein
chr3.5886509	chr3:5886509-5886962	8	AK045941	207196	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr3.30852371	chr3:30852371-30852851	8	Skil	13886	5' distal	coding	84	34	136	2.47	GeneID:20482	Other transcription factor
chr2.174328371	chr2:174328371-174329104	8	mir296	5000	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr2.169021975	chr2:169021975-169022981	8	Zfp64	4344	5' proximal	coding	30	20	45	1.50	GeneID:22722	KRAB box transcription factor;Nucleic acid binding
chr2.168276815	chr2:168276815-168277824	8	Dpm1	1207	intragenic (intron)	coding	53	30	43	1.77	GeneID:13480	Transferase
chr2.165872605	chr2:165872605-165873644	8	Prkcbp1	2160	intragenic (intron)	coding	0	52	32	0.00	GeneID:228880	Other signaling molecule
chr19.37572421	chr19:37572421-37573268	8	AK085625	150	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr15.80950121	chr15:80950121-80952061	8	Grap2	4106	5' proximal	coding	94	81	101	1.16	GeneID:17444	Other miscellaneous function protein
chr14.49723866	chr14:49723866-49725177	8	Lats2	967	intragenic (intron)	coding	0	0	13	#DIV/0!	GeneID:50523	Non-receptor serine/threonine protein kinase
chr13.63869658	chr13:63869658-63870418	8	AK021218	34	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr13.63830925	chr13:63830925-63833789	8	C330014B19Rik	2740	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:382770	G-protein coupled receptor
chr12.33802153	chr12:33802153-33803209	8	Etv1	799	5' proximal	coding	0	0	33	#DIV/0!	GeneID:14009	Other transcription factor;Nucleic acid binding
chr12.105883922	chr12:105883922-105888124	8	S71494	200000		target is repeat	-	-	-	#VALUE!	NA	NA
chr10.66856801	chr10:66856801-66857952	8	D10Ucla1	3555	5' proximal	coding	11	12	166	0.92	GeneID:28193	Molecular function unclassified
chrX.92487654	chrX:92487654-92488412	7	Slc7a3	923	intragenic (intron)	coding	72	0	0	#DIV/0!	GeneID:11989	Amino acid transporter
chrX.90960231	chrX:90960231-90961149	7	Pja1	86482	5' distal	coding	0	4	10	0.00	GeneID:18744	Ubiquitin-protein ligase
chr9.46993418	chr9:46993418-46995325	7	2900052N01Rik	23573	5' distal	coding	0	0	0	#DIV/0!	GeneID:73040	Molecular function unclassified
chr9.13532462	chr9:13532462-13533126	7	AK129448	2234	intragenic (intron)	coding	0	0	61	#DIV/0!	NA	NA
chr8.22365894	chr8:22365894-22366893	7	Zmat4	107535	gene desert	coding	15	0	0	#DIV/0!	NA	NA
chr8.14012514	chr8:14012514-14013547	7	2610019F03Rik	11635	5' distal	coding	0	0	0	#DIV/0!	GeneID:72148	Molecular function unclassified
chr8.115771779	chr8:115771779-115772889	7	BY735552 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr7.50942759	chr7:50942759-50944209	7	AA183224 EST	5000	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr5.73647624	chr5:73647624-73649082	7	BB863351 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr5.137353358	chr5:137353358-137354599	7	A930017N06Rik	26027	intragenic (intron)	coding	0	0	15	#DIV/0!	GeneID:243312	Cell adhesion molecule
chr4.6895606	chr4:6895606-6896563	7	Tox	21433	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:252838	HMG box transcription factor;Chromatin/chromatin-binding protein
chr4.124835924	chr4:124835924-124837357	7	Eif2c1	2195	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:236511	Translation initiation factor
chr3.34558694	chr3:34558694-34561322	7	B230215L15Rik	77131	3' distal	noncoding	10	0	0	#DIV/0!	GeneID:320478	Molecular function unclassified
chr3.142906642	chr3:142906642-142907768	7	Pdlim5	3806	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:56376	Non-motor actin binding protein
chr2.37726595	chr2:37726595-37727950	7	Strbp	64195	5' distal	coding	0	0	12	#DIV/0!	GeneID:20744	Other RNA-binding protein
chr2.20723862	chr2:20723862-20725799	7	AB093289	7440	intragenic (intron)	coding	6	8	8	0.75	NA	NA
chr2.174354527	chr2:174354527-174356162	7	Gnas	0	intragenic (intron)	coding	270	1656	2414	0.16	GeneID:14683	Large G-protein
chr2.170835199	chr2:170835199-170835772	7	Dok5	40181	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:76829	Molecular function unclassified

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular function
chr2:163094813	chr2:163094813-163096445	7	BC037708	712	intragenic (intron)	coding	0	0	9	#DIV/0!	GeneID:245866	Molecular function unclassified
chr17:9846966	chr17:9846966-9847517	7	AK005771	29511	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr17:26089869	chr17:26089869-26090997	7	Grm4	28837	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:268934	G-protein coupled receptor
chr17:13851666	chr17:13851666-13853358	7	2410011O22Rik	116	5' proximal	coding	0	0	0	#DIV/0!	GeneID:78483	Molecular function unclassified
chr16:34987756	chr16:34987756-34988332	7	Adcy5	9664	5' proximal	coding	0	4	0	0.00	GeneID:224129	Adenylate cyclase
chr16:13866711	chr16:13866711-13867782	7	AK053159	1759	intragenic (intron)	coding	9	9	12	1.00	NA	NA
chr14:70053515	chr14:70053515-70054351	7	Tnfsf11	4602	intragenic (intron)	coding	18	0	0	#DIV/0!	GeneID:21943	Cytokine
chr14:3017774	chr14:3017774-3019103	7	2610042L04Rik	47740	3' distal	coding	0	0	0	#DIV/0!	GeneID:67055	Molecular function unclassified
chr13:51594659	chr13:51594659-51595651	7	Syk	21025	5' distal	coding	0	0	43	#DIV/0!	GeneID:20963	Non-receptor tyrosine protein kinase
chr13:50872439	chr13:50872439-50873093	7	Gadd45g	611	5' proximal	coding	0	0	66	#DIV/0!	GeneID:23882	Other miscellaneous function protein
chr12:93913503	chr12:93913503-93915706	7	Ches1	8683	intragenic (intron)	coding	39	2	101	19.50	GeneID:71375	Transcription factor;Nucleic acid binding
chr12:68363923	chr12:68363923-68364981	7	Six4	13779	3' distal	coding	0	0	0	#DIV/0!	GeneID:20474	Homeobox transcription factor
chr12:13065920	chr12:13065920-13067987	7	Nmyc1	1676	5' proximal	coding	132	20	1	6.60	GeneID:18109	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr11:95651917	chr11:95651917-95652775	7	Igf2bp1	1336	intragenic (intron)	coding	48	25	0	1.92	GeneID:140486	Other RNA-binding protein
chr11:69193787	chr11:69193787-69195226	7	Trp53	1531	intragenic (intron)	coding	518	571	127	0.91	GeneID:22059	Other transcription factor;Nucleic acid binding
chr11:33422606	chr11:33422606-33424197	7	Ranbp17	13919	5' distal	coding	0	36	0	0.00	GeneID:66011	Other transfer/carrier protein
chr11:30644353	chr11:30644353-30645911	7	AK049252	1500	5' proximal	noncoding	6	0	0	#DIV/0!	NA	NA
chr11:16240105	chr11:16240105-16241802	7	BC027127	22918	intragenic (intron)	coding	375	0	180	#DIV/0!	GeneID:211739	Molecular function unclassified
chr10:59702815	chr10:59702815-59704061	7	Ddit4	6382	5' proximal	coding	190	429	162	0.44	GeneID:74747	Molecular function unclassified
chr10:56415219	chr10:56415219-56416552	7	Gja1	3072	5' proximal	coding	191	202	40	0.95	GeneID:14609	Gap junction
chr10:41365529	chr10:41365529-41366655	7	Zbtb24	1973	3' proximal	coding	0	0	0	#DIV/0!	GeneID:268294	KRAB box transcription factor
chr10:116351670	chr10:116351670-116353734	7	Cnot2	13355	intragenic (intron)	coding	82	9	55	9.11	GeneID:72068	Other transcription factor
chr1:181009997	chr1:181009997-181011343	7	Lefty1	1185	5' proximal	coding	194	4	0	48.50	GeneID:13590	TGF-beta superfamily member
chr1:171141751	chr1:171141751-171142882	7	Fcgr3	5096	5' proximal	coding	0	0	0	#DIV/0!	GeneID:14131	Immunoglobulin receptor family member;Defense/immunity protein
chr1:132096189	chr1:132096189-132097225	7	AK016010	5000	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr1:118407974	chr1:118407974-118408936	7	Tcfcp2l1	4197	intragenic (intron)	coding	81	4	0	20.25	GeneID:81879	Other transcription factor
chrX:95389628	chrX:95389628-95391106	6	Rnf12	5179	5' proximal	coding	0	7	0	0.00	GeneID:19820	Transcription cofactor
chrX:49692117	chrX:49692117-49692966	6	Zic3	3973	3' proximal	coding	1	0	1	#DIV/0!	GeneID:22773	KRAB box transcription factor;Nucleic acid binding
chr9:92093254	chr9:92093254-92094125	6	Plscr1	1737	intragenic (intron)	coding	1	0	0	#DIV/0!	GeneID:22038	Other transfer/carrier protein
chr9:73127988	chr9:73127988-73130308	6	BY752628 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr9:40264324	chr9:40264324-40265968	6	Zfp202	26778	3' distal	coding	6	0	0	#DIV/0!	GeneID:80902	KRAB box transcription factor;Nucleic acid binding
chr9:104024459	chr9:104024459-104025624	6	Ccrl1	737	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:252837	G-protein coupled receptor
chr8:81132626	chr8:81132626-81133606	6	BC064813	2800	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr8:69576916	chr8:69576916-69579061	6	Ell	374	5' proximal	coding	16	0	0	#DIV/0!	GeneID:13716	Transcription cofactor
chr8:26028823	chr8:26028823-26029538	6	Adrb3	12205	5' distal	coding	612	456	113	1.34	GeneID:11556	G-protein coupled receptor
chr7:67829035	chr7:67829035-67830648	6	Iqgap1	2343	intragenic (intron)	coding	67	99	25	0.68	GeneID:29875	Other G-protein modulator
chr7:26472023	chr7:26472023-26475087	6	C80913	10027	5' distal	coding	0	0	0	#DIV/0!	GeneID:19777	Transcription cofactor;Chaperone;Defense/immunity protein
chr7:127711594	chr7:127711594-127713929	6	Drd1ip	8887	3' proximal	coding	0	0	0	#DIV/0!	GeneID:68566	Transmembrane receptor regulatory/adaptor protein
chr7:109576433	chr7:109576433-109577869	6	Prkcb1	9553	5' proximal	coding	0	0	36	#DIV/0!	GeneID:18751	Transfer/carrier protein;Non-receptor serine/threonine protein kinase
chr6:97773995	chr6:97773995-97774977	6	Frmd4b	522	intragenic (intron)	coding	18	0	0	#DIV/0!	GeneID:232288	Molecular function unclassified
chr6:84130936	chr6:84130936-84132157	6	Zfml	23827	5' distal	coding	18	84	50	0.21	GeneID:18139	Other DNA-binding protein
chr6:83009296	chr6:83009296-83010368	6	Hk2	790	5' proximal	coding	65	6	14	10.83	GeneID:15277	Carbohydrate kinase
chr6:67350258	chr6:67350258-67351535	6	1200009K13Rik	28868	5' distal	coding	3588	1679	586	2.14	GeneID:68870	Other RNA-binding protein
chr6:67223592	chr6:67223592-67225133	6	AK039826	4500	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr6:30156516	chr6:30156516-30157779	6	UBE2H	10473	intragenic (intron)	coding	54	6	0	9.00	NA	NA
chr6:149138231	chr6:149138231-149139200	6	D030011O10Rik	21517	3' distal	coding	0	0	0	#DIV/0!	GeneID:320560	Guanyl-nucleotide exchange factor
chr6:149131991	chr6:149131991-149132906	6	AK079183	2370	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr6:140717488	chr6:140717488-140718334	6	Plekha5	31358	3' distal	coding	17	25	0	0.68	GeneID:109135	Molecular function unclassified
chr6:122972545	chr6:122972545-122974405	6	Phc1	2551	5' proximal	coding	334	118	19	2.83	GeneID:13619	Other transcription factor;Chromatin/chromatin-binding protein
chr5:75054710	chr5:75054710-75056227	6	Pdcl2	2428	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:79455	Other chaperones;G-protein modulator;Other miscellaneous function
chr5:73780151	chr5:73780151-73782355	6	Gsh2	7431	3' proximal	coding	0	0	6	#DIV/0!	GeneID:14843	Homeobox transcription factor;Nucleic acid binding
chr5:31983518	chr5:31983518-31984499	6	AB187257		intragenic (exon)	noncoding	0	0	0	#DIV/0!	NA	NA
chr5:31022968	chr5:31022968-31023627	6	Yes1	4797	intragenic (intron)	coding	15	0	15	#DIV/0!	GeneID:22612	Non-receptor tyrosine protein kinase
chr5:146768492	chr5:146768492-146769717	6	BC028471		intragenic (exon)	coding	0	0	0	#DIV/0!	NA	NA
chr5:137582437	chr5:137582437-137583821	6	Mad1l1	19551	intragenic (intron)	coding	38	14	7	2.71	GeneID:17120	Molecular function unclassified
chr5:113872476	chr5:113872476-113873019	6	1500001A10Rik	667	3' proximal	coding	7	0	0	#DIV/0!	GeneID:68955	Molecular function unclassified
chr5:105518414	chr5:105518414-105519298	6	Mtf2	3625	intragenic (intron)	coding	103	81	43	1.27	GeneID:17765	Other zinc finger transcription factor;Nucleic acid binding
chr5:104582343	chr5:104582343-104583110	6	Tgfbr3	235	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21814	TGF-beta receptor
chr4:63565396	chr4:63565396-63566989	6	Pappa	780	intragenic (intron)	coding	0	0	10	#DIV/0!	GeneID:18491	Metalloprotease
chr4:102713811	chr4:102713811-102714780	6	Dab1	1127	intragenic (intron)	coding	7	0	3	#DIV/0!	GeneID:13131	Other signaling molecule
chr3:84376964	chr3:84376964-84377884	6	D930015E06Rik	19906	intragenic (intron)	coding	19	56	99	0.34	GeneID:229473	Molecular function unclassified
chr3:138599072	chr3:138599072-138600246	6	D830013H12	6749	3' proximal	noncoding	0	0	0	#DIV/0!	GeneID:329763	Molecular function unclassified
chr2:58557621	chr2:58557621-58558834	6	Upp2	37372	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:76654	Phosphorylase
chr2:3367792	chr2:3367792-3368970	6	Meig1	1222	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:104362	Molecular function unclassified

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr2.20264786	chr2:20264786-20265747	6	AK052020	80408	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr2.180397020	chr2:180397020-180397828	6	AK129117	659	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr2.121051725	chr2:121051725-121052561	6	Trp53bp1	0	5' proximal	coding	135	63	27	2.14	GeneID:27223	Transcription cofactor;Nucleic acid binding
chr2.100043871	chr2:100043871-100044906	6	AK012553	6172	intragenic (intron)	coding	12	0	28	#DIV/0!	NA	NA
chr19.22348294	chr19:22348294-22349236	6	AK019121	58832	5' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr18.4365702	chr18:4365702-4368447	6	Lyzl1	13922	3' distal	coding	0	0	0	#DIV/0!	GeneID:67328	Hydrolase;Defense/immunity protein
chr18.40751753	chr18:40751753-40752091	6	AK220381		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr17.69256519	chr17:69256519-69257555	6	BC062120	8624	3' proximal	coding	66	222	3	0.30	NA	NA
chr17.61523417	chr17:61523417-61525188	6	BC058120	48467	intragenic (intron)	coding	19	0	0	#DIV/0!	NA	NA
chr17.46788686	chr17:46788686-46789373	6	Unc5cl	144309	gene desert	coding	0	0	0	#DIV/0!	GeneID:76589	Molecular function unclassified
chr17.35600957	chr17:35600957-35602653	6	Gabbr1	4134	3' proximal	coding	0	0	0	#DIV/0!	GeneID:54393	G-protein coupled receptor
chr17.14198472	chr17:14198472-14199351	6	Dll1	5129	5' proximal	coding	0	0	79	#DIV/0!	GeneID:13388	Receptor;Membrane-bound signaling molecule;Defense/immunity
chr17.13838536	chr17:13838536-13841782	6	Tcte3	171	5' proximal	coding	0	8	0	0.00	GeneID:21647	Molecular function unclassified
chr16.85151642	chr16:85151642-85152236	6	Jam2	5087	5' proximal	coding	0	0	71	#DIV/0!	GeneID:67374	Molecular function unclassified
chr16.46057018	chr16:46057018-46058614	6	Cd96	2697	intragenic (intron)	coding	49	3	29	16.33	GeneID:84544	Defense/immunity protein
chr16.41499402	chr16:41499402-41500268	6	Lsamp	100568	intragenic (intron)	coding	0	0	18	#DIV/0!	GeneID:268890	CAM family adhesion molecule
chr16.38264286	chr16:38264286-38265302	6	Pla1a	87	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:85031	Phospholipase
chr16.32504656	chr16:32504656-32505739	6	Tfrc	4591	3' proximal	coding	11	33	7	0.33	GeneID:22042	Other receptor
chr16.22238878	chr16:22238878-22239621	6	Etv5	4266	intragenic (intron)	coding	0	0	98	#DIV/0!	GeneID:104156	Other transcription factor;Nucleic acid binding
chr15.84805634	chr15:84805634-84806715	6	BC062953	3375	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr15.76994711	chr15:76994711-76996621	6	Foxh1	42	5' proximal	coding	48	145	0	0.33	GeneID:14106	Other transcription factor;Nucleic acid binding
chr15.63174511	chr15:63174511-63176268	6	U16672	169750	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr15.25373666	chr15:25373666-25374291	6	AK019612	300	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr14.42122924	chr14:42122924-42124643	6	Slc35f4	78084	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:75288	Molecular function unclassified
chr14.3423310	chr14:3423310-3427082	6	BC060995		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr14.3414134	chr14:3414134-3418785	6	BC055874	32949	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr14.3247642	chr14:3247642-3249264	6	2610042L04Rik	918	5' proximal	coding	17	240	0	0.07	GeneID:67055	Molecular function unclassified
chr14.3088778	chr14:3088778-3089721	6	2610042L04Rik	1097	5' proximal	coding	17	240	0	0.07	GeneID:67055	Molecular function unclassified
chr13.4611466	chr13:4611466-4614715	6	AK030184	7525	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr13.44044311	chr13:44044311-44045558	6	Jarid2	150467	gene desert	coding	97	20	0	4.85	GeneID:16468	Other transcription factor;Nucleic acid binding
chr13.109400650	chr13:109400650-109401478	6	Il6st	9998	5' proximal	coding	0	2	0	0.00	GeneID:16195	Interleukin receptor
chr12.81824700	chr12:81824700-81825598	6	Esrrb	7111	intragenic (intron)	coding	250	24	0	10.42	GeneID:26380	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr12.77495019	chr12:77495019-77496256	6	Sipa1l1	23644	intragenic (intron)	coding	2	0	10	#DIV/0!	GeneID:217692	Other G-protein modulator
chr12.71984229	chr12:71984229-71985170	6	S66283	14574	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr12.36799783	chr12:36799783-36800614	6	Immp2l	23655	intragenic (intron)	coding	0	13	0	0.00	GeneID:93757	Other proteases
chr12.12926410	chr12:12926410-12927098	6	Nmyc1	132343	gene desert	coding	132	20	1	6.60	GeneID:18109	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr11.9010476	chr11:9010476-9012696	6	Upp1	1847	5' proximal	coding	911	388	24	2.35	GeneID:22271	Phosphorylase
chr11.77506314	chr11:77506314-77507170	6	Pipox	258	5' proximal	coding	10	0	96	#DIV/0!	GeneID:19193	Oxidase
chr11.25997924	chr11:25997924-25999766	6	AK080999	106534	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr11.10992707	chr11:10992707-10993913	6	A930041G11Rik	16306	5' distal	coding	15	0	6	#DIV/0!	GeneID:319922	Molecular function unclassified
chr10.95230693	chr10:95230693-95231552	6	Socs2	40425	5' distal	coding	0	0	0	#DIV/0!	GeneID:216233	Other signaling molecule
chr10.85204764	chr10:85204764-85206468	6	Btbd11	6738	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:74007	Molecular function unclassified
chr10.83202103	chr10:83202103-83203458	6	AK038582	4261	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr10.83005589	chr10:83005589-83008237	6	Slc41a2	6572	5' proximal	coding	0	0	0	#DIV/0!	GeneID:338365	Molecular function unclassified
chr10.77488449	chr10:77488449-77489710	6	Sumo3	3037	5' proximal	coding	173	65	123	2.66	GeneID:20610	Molecular function unclassified
chr10.21710656	chr10:21710656-21713498	6	Sgk	6104	5' proximal	coding	132	85	0	1.55	GeneID:20393	Non-receptor serine/threonine protein kinase
chr10.20806206	chr10:20806206-20807532	6	BF531694 EST	100	3' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr10.117849232	chr10:117849232-117850839	6	BU937038 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr1.85399679	chr1:85399679-85400114	6	Fbxo36	8845	intragenic (intron)	coding	0	0	20	#DIV/0!	GeneID:66153	Molecular function unclassified
chr1.65134050	chr1:65134050-65135692	6	Fzd5	894	3' proximal	coding	0	0	0	#DIV/0!	GeneID:14367	G-protein coupled receptor
chr1.59825907	chr1:59825907-59826786	6	S71494	42789		target is repeat	-	-	-	#VALUE!	NA	NA
chr1.54869615	chr1:54869615-54872525	6	CN675669 EST	1500	3' proximal	coding	-	-	-	#VALUE!	NA	NA
chrX.93123814	chrX:93123814-93124640	5	Cxcr3	10913	3' distal	coding	0	0	0	#DIV/0!	GeneID:12766	G-protein coupled receptor
chrX.69124752	chrX:69124752-69125751	5	Tbl1x	65300	5' distal	coding	20	35	28	0.57	GeneID:21372	Other miscellaneous function protein
chr9.8780655	chr9:8780655-8782068	5	Pgr	118962	gene desert	coding	0	0	0	#DIV/0!	GeneID:18667	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr9.64785141	chr9:64785141-64786032	5	2410080H04Rik	1801	intragenic (intron)	coding	5	0	83	#DIV/0!	GeneID:214058	Receptor;Extracellular matrix structural protein
chr9.58662174	chr9:58662174-58662924	5	AK015493		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr9.58426992	chr9:58426992-58432276	5	Pml	198	intragenic (intron)	coding	13	0	0	#DIV/0!	GeneID:18854	RING finger transcription factor
chr9.47618147	chr9:47618147-47619751	5	Igsf4a	15724	5' distal	coding	0	0	49	#DIV/0!	GeneID:54725	Receptor
chr9.32081687	chr9:32081687-32083062	5	Grit	111189	gene desert	coding	14	0	60	#DIV/0!	GeneID:330914	Other G-protein modulator
chr9.21663847	chr9:21663847-21664756	5	Ldlr	220	intragenic (intron)	coding	61	0	25	#DIV/0!	GeneID:16835	Other receptor
chr9.18964307	chr9:18964307-18965365	5	Olfr835	61071	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:257872	G-protein coupled receptor
chr9.120685110	chr9:120685110-120686449	5	Rpl14	116864	gene desert	coding	3	0	24	#DIV/0!	GeneID:67115	Ribosomal protein
chr9.114569907	chr9:114569907-114570513	5	Cnot10	230	5' proximal	coding	0	0	0	#DIV/0!	GeneID:78893	Molecular function unclassified

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr9.107650140	chr9:107650140-107651905	5	Gnai2	4291	intragenic (intron)	coding	0	0	549	#DIV/0!	GeneID:14678	Large G-protein
chr9.100818305	chr9:100818305-100818823	5	Stag1	11663	3' distal	coding	0	19	8	0.00	GeneID:20842	Chromatin/chromatin-binding protein
chr8.90116075	chr8:90116075-90117374	5	AK034446	33763	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr8.84721401	chr8:84721401-84722507	5	Mast1		intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:56527	Non-receptor serine/threonine protein kinase
chr8.82999727	chr8:82999727-83005630	5	Cd97	232	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:26364	G-protein coupled receptor
chr8.59143252	chr8:59143252-59144397	5	U16670	24649		target is repeat	-	-	-	#VALUE!	NA	NA
chr8.32569532	chr8:32569532-32570771	5	Ppp2cb	4073	intragenic (intron)	coding	123	191	241	0.64	GeneID:19053	Protein phosphatase;Other select calcium binding proteins
chr8.20937252	chr8:20937252-20938611	5	Ckap2	91	5' proximal	coding	0	0	0	#DIV/0!	GeneID:80986	Molecular function unclassified
chr8.112948082	chr8:112948082-112951630	5	Mon1b	37798	5' distal	coding	3	0	1	#DIV/0!	GeneID:270096	Molecular function unclassified
chr8.106905844	chr8:106905844-106906847	5	Psmd7	6142	5' proximal	coding	212	132	127	1.61	GeneID:17463	Other miscellaneous function protein
chr8.103450230	chr8:103450230-103451513	5	Cdh5	9146	5' proximal	coding	0	0	0	#DIV/0!	GeneID:12562	Cadherin
chr8.103442503	chr8:103442503-103448645	5	Cdh5	11276	5' distal	coding	0	0	0	#DIV/0!	GeneID:12562	Cadherin
chr7.99961908	chr7:99961908-99962695	5	Tead1	11832	intragenic (intron)	coding	0	0	24	#DIV/0!	GeneID:21676	Other transcription factor;Nucleic acid binding
chr7.77771142	chr7:77771142-77773061	5	1110001A23Rik	63932	5' distal	coding	3	72	43	0.04	GeneID:68472	Molecular function unclassified
chr7.29012187	chr7:29012187-29012804	5	2810426N06Rik	51086	5' distal	coding	3	0	11	#DIV/0!	GeneID:67607	KRAB box transcription factor;Nucleic acid binding
chr7.23770590	chr7:23770590-23777108	5	Rhpn2	2379	5' proximal	coding	0	0	0	#DIV/0!	GeneID:52428	Other G-protein modulator
chr7.19440781	chr7:19440781-19442909	5	Nphs1	292	intragenic (intron)	coding	20	0	11	#DIV/0!	GeneID:54631	CAM family adhesion molecule
chr7.19010955	chr7:19010955-19011980	5	4732429I09Rik	5336	3' proximal	coding	0	0	15	#DIV/0!	GeneID:243906	KRAB box transcription factor;Nucleic acid binding
chr7.132598025	chr7:132598025-132598907	5	Fgf15	7694	5' proximal	coding	0	0	0	#DIV/0!	GeneID:14170	Growth factor
chr7.131811803	chr7:131811803-131812686	5	AB099695	68595	5' distal	coding	0	0	12	#DIV/0!	NA	NA
chr7.126487822	chr7:126487822-126488992	5	Ppp2r2d	3566	intragenic (intron)	coding	279	35	161	7.97	GeneID:52432	Protein phosphatase
chr7.113866045	chr7:113866045-113868157	5	Nfatc2ip	1257	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18020	Molecular function unclassified
chr6.92598080	chr6:92598080-92599431	5	Trh	2867	5' proximal	coding	757	428	0	1.77	GeneID:22044	Neuropeptide
chr6.51952914	chr6:51952914-51954176	5	Scap2	2366	intragenic (intron)	coding	3	0	14	#DIV/0!	GeneID:54353	Molecular function unclassified
chr6.108959355	chr6:108959355-108960264	5	Itpr1	19690	3' distal	coding	0	0	0	#DIV/0!	GeneID:16438	Other ligand-gated ion channel;Ion channel
chr6.100758684	chr6:100758684-100759790	5	Rybp	91655	5' distal	coding	44	20	0	2.20	GeneID:56353	Molecular function unclassified
chr5.75998267	chr5:75998267-75998629	5	Rest	3171	5' proximal	coding	150	42	10	3.57	GeneID:19712	Molecular function unclassified
chr5.71539024	chr5:71539024-71540452	5	5033405K12Rik	10069	3' distal	coding	0	0	0	#DIV/0!	GeneID:75991	Molecular function unclassified
chr5.30042541	chr5:30042541-30043140	5	Rbks	9882	intragenic (intron)	coding	23	41	57	0.56	GeneID:71336	Carbohydrate kinase
chr5.25090671	chr5:25090671-25091906	5	Dpp6	113311	gene desert	coding	0	0	0	#DIV/0!	GeneID:13483	Serine protease
chr5.20204794	chr5:20204794-20206121	5	Pres	3041	intragenic (intron)	coding	245	238	90	1.03	GeneID:80979	Other transporter
chr5.148494525	chr5:148494525-148495173	5	Stard13	11945	3' distal	coding	0	0	5	#DIV/0!	GeneID:243362	Other G-protein modulator
chr5.144629069	chr5:144629069-144632228	5	Cdx2	1211	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:12591	Homeobox transcription factor;Nucleic acid binding
chr5.139552972	chr5:139552972-139553632	5	Sdk1	1910	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:330222	CAM family adhesion molecule
chr5.129110218	chr5:129110218-129111327	5	AK086339	38536	5' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr5.124300311	chr5:124300311-124300948	5	U16672	392143	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr5.122751006	chr5:122751006-122754631	5	Scarb1	61644	3' distal	coding	0	29	47	0.00	GeneID:20778	Other receptor
chr5.119705057	chr5:119705057-119705999	5	Ppp1cc	16252	3' distal	coding	132	350	224	0.38	GeneID:19047	Protein phosphatase;Other select calcium binding proteins
chr5.117164821	chr5:117164821-117168680	5	Tbx3	315	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21386	Other transcription factor;Nucleic acid binding
chr5.112212604	chr5:112212604-112213697	5	1100001D10Rik	328	intragenic (intron)	coding	21	50	20	0.42	GeneID:68420	Molecular function unclassified
chr5.108372409	chr5:108372409-108373441	5	BB865591 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr5.107193680	chr5:107193680-107195062	5	AK019690	617	5' proximal	coding	0	90	94	0.00	NA	NA
chr4.99157067	chr4:99157067-99158005	5	B430218L07Rik	34703	intragenic (intron)	coding	0	8	14	0.00	GeneID:320508	Voltage-gated calcium channel
chr4.94013851	chr4:94013851-94016129	5	2310009E04Rik	61236	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:75578	Carbohydrate kinase
chr4.80705460	chr4:80705460-80706810	5	Nfib	47383	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18028	Other transcription factor;Nucleic acid binding
chr4.57038165	chr4:57038165-57039321	5	Akap2	12634	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:11641	Other miscellaneous function protein
chr4.45060041	chr4:45060041-45062080	5	Igfbpl1	203	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:75426	Miscellaneous function
chr4.45055384	chr4:45055384-45057323	5	Igfbpl1	0	intragenic (exon, 3'utr)	coding	0	0	0	#DIV/0!	GeneID:75426	Miscellaneous function
chr4.44448330	chr4:44448330-44449007	5	AK122409	41184	5' distal	coding	0	0	67	#DIV/0!	NA	NA
chr4.29359627	chr4:29359627-29360129	5	BY720273 EST	50000	3' distal	noncoding	-	-	-	#VALUE!	NA	NA
chr4.145407867	chr4:145407867-145411217	5	BC058613	8012	3' proximal	coding	137	4	0	34.25	NA	NA
chr4.139094855	chr4:139094855-139096987	5	2810441C07Rik	9060	intragenic (intron)	coding	0	0	36	#DIV/0!	GeneID:72754	Molecular function unclassified
chr4.138862653	chr4:138862653-138864643	5	AK030056	83522	3' distal	noncoding	0	0	36	#DIV/0!	NA	NA
chr4.138365702	chr4:138365702-138368289	5	BE850820 EST	500	5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr4.137686188	chr4:137686188-137686718	5	Capzb	6263	intragenic (intron)	coding	81	45	70	1.80	GeneID:12345	Non-motor actin binding protein
chr4.136235783	chr4:136235783-136236301	5	Akp2	18272	intragenic (intron)	coding	10	0	0	#DIV/0!	GeneID:11647	Other phosphatase
chr4.135927651	chr4:135927651-135928879	5	Hspg2	1384	intragenic (intron)	coding	19	58	5	0.33	NA	NA
chr4.134775656	chr4:134775656-134776245	5	Hnrpr	802	intragenic (intron)	coding	87	0	104	#DIV/0!	GeneID:74326	mRNA processing factor;Ribonucleoprotein
chr4.125597030	chr4:125597030-125598442	5	Dlgap3	5977	intragenic (intron)	coding	26	1	16	26.00	GeneID:242667	Molecular function unclassified
chr4.123915035	chr4:123915035-123916227	5	Ftl1	40254	3' distal	coding	5	1	12	5.00	GeneID:14325	Storage protein
chr4.114807245	chr4:114807245-114809850	5	Mast2	290	intragenic (intron)	coding	56	29	44	1.93	GeneID:17776	Non-receptor serine/threonine protein kinase
chr4.106527586	chr4:106527586-106530755	5	BC022150	426	intragenic (intron)	coding	60	0	1	#DIV/0!	GeneID:230590	Molecular function unclassified
chr3.98496161	chr3:98496161-98497230	5	Hsd3b5	45702	3' distal	coding	0	0	0	#DIV/0!	GeneID:15496	Dehydrogenase
chr3.88463565	chr3:88463565-88464540	5	Nes	331	intragenic (intron)	coding	3	20	163	0.15	GeneID:18008	Molecular function unclassified

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr3.81430898	chr3:81430898-81431571	5	Pdgfc	1494	5' proximal	coding	0	0	0	#DIV/0!	GeneID:54635	Other receptor;Growth factor
chr3.65895542	chr3:65895542-65896176	5	Tiparp	111166	gene desert	coding	56	111	50	0.50	GeneID:99929	Molecular function unclassified
chr3.34456545	chr3:34456545-34457139	5	Sox2		3' proximal	coding	188	50	377	3.76	GeneID:20674	HMG box transcription factor;Nucleic acid binding
chr3.157936014	chr3:157936014-157938438	5	Negr1		5' proximal	coding	0	0	0	#DIV/0!	GeneID:320840	CAM family adhesion molecule
chr3.14651855	chr3:14651855-14656112	5	1200008A14Rik	1535	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:71710	Molecular function unclassified
chr3.136092666	chr3:136092666-136093616	5	Manba	968	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:110173	Glycosidase
chr3.130084020	chr3:130084020-130085239	5	Elovl6	19183	intragenic (intron)	coding	1	28	26	0.04	GeneID:170439	Transferase
chr3.128024071	chr3:128024071-128027049	5	D3Wsu161e	488	5' proximal	coding	0	0	0	#DIV/0!	GeneID:28036	Molecular function unclassified
chr3.108813089	chr3:108813089-108814016	5	Sars1	390	intragenic (intron)	coding	254	77	192	3.30	GeneID:20226	Other RNA-binding protein;Aminoacyl-tRNA synthetase;Ligase
chr3.103137363	chr3:103137363-103139351	5	AK010586	9210	intragenic (intron)	noncoding	40	20	49	2.00	NA	NA
chr2.70217033	chr2:70217033-70217739	5	A430065P19Rik	10634	5' distal	coding	0	0	0	#DIV/0!	GeneID:329421	Actin binding motor protein
chr2.68986317	chr2:68986317-68987092	5	Lass6	2702	5' proximal	coding	0	0	0	#DIV/0!	GeneID:241447	Molecular function unclassified
chr2.33540335	chr2:33540335-33541106	5	Lmx1b	11529	intragenic (intron)	coding	0	0	122	#DIV/0!	GeneID:16917	Other zinc finger transcription factor
chr2.30719561	chr2:30719561-30720955	5	AV054662		intragenic (intron)	noncoding	0	0	5	#DIV/0!	NA	NA
chr2.29780262	chr2:29780262-29781081	5	Slc27a4	3130	3' proximal	coding	11	0	12	#DIV/0!	GeneID:26569	Transporter;Other ligase
chr2.29080992	chr2:29080992-29081888	5	AK044730	3824	5' proximal	coding	0	0	5	#DIV/0!	NA	NA
chr2.172561282	chr2:172561282-172562574	5	AK006503	24540	3' distal	coding	17	18	34	0.94	NA	NA
chr2.165706580	chr2:165706580-165707693	5	U71208		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr2.154490902	chr2:154490902-154491834	5	CA560714	300	5' proximal	coding	0	0	25	#DIV/0!	NA	NA
chr2.154435691	chr2:154435691-154436740	5	Cdk5rap1	132	5' proximal	coding	10	0	0	#DIV/0!	GeneID:66971	Molecular function unclassified
chr2.153434062	chr2:153434062-153434758	5	Asxl1	13735	intragenic (intron)	coding	23	5	4	4.60	NA	NA
chr2.150045794	chr2:150045794-150047135	5	BP428189 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr2.129444739	chr2:129444739-129445282	5	Pdyn	2388	intragenic (intron)	coding	0	0	2	#DIV/0!	GeneID:18610	Neuropeptide
chr19.6974624	chr19:6974624-6975463	5	Rcor2	6974	5' proximal	coding	212	91	47	2.33	GeneID:104383	Molecular function unclassified
chr19.37105064	chr19:37105064-37106127	5	BC025649	17551	3' distal	coding	1	27	4	0.04	NA	NA
chr19.37101382	chr19:37101382-37102613	5	BC025649	21451	3' distal	coding	1	27	4	0.04	NA	NA
chr19.29330611	chr19:29330611-29331307	5	Uhrf2	1339	5' proximal	coding	3	0	29	#DIV/0!	GeneID:109113	RING finger transcription factor
chr19.27912819	chr19:27912819-27913862	5	Glis3	94879	5' distal	coding	0	0	9	#DIV/0!	GeneID:226075	KRAB box transcription factor;Nucleic acid binding
chr19.20921331	chr19:20921331-20924511	5	1110059E24Rik	0	5' proximal	coding	0	0	0	#DIV/0!	GeneID:66206	Molecular function unclassified
chr19.10071681	chr19:10071681-10073167	5	Ms4a10	1721	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:69826	Molecular function unclassified
chr18.74587666	chr18:74587666-74593291	5	Mapk4	586	intragenic (intron)	coding	0	0	23	#DIV/0!	GeneID:225724	Non-receptor serine/threonine protein kinase
chr18.68024045	chr18:68024045-68025156	5	Spire1	349	intragenic (intron)	coding	10	27	61	0.37	GeneID:68166	Molecular function unclassified
chr18.42711553	chr18:42711553-42715746	5	AK018420	1521	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr17.74585043	chr17:74585043-74586097	5	2810405J04Rik	565176	gene desert	coding	0	0	0	#DIV/0!	GeneID:72722	Molecular function unclassified
chr17.6328135	chr17:6328135-6330129	5	Sytl3	484	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:83672	Membrane traffic regulatory protein
chr17.35602670	chr17:35602670-35603830	5	AK020454		5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr17.33523480	chr17:33523480-33524253	5	AK019275	12	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr16.97642295	chr16:97642295-97643412	5	Dscam	12298	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:13508	CAM family adhesion molecule
chr16.90599428	chr16:90599428-90600536	5	Tiam1	204666	gene desert	coding	0	0	0	#DIV/0!	GeneID:21844	Guanyl-nucleotide exchange factor
chr15.91451195	chr15:91451195-91451612	5	Kif21a	6444	5' proximal	coding	0	7	20	0.00	GeneID:16564	Microtubule binding motor protein
chr15.6489149	chr15:6489149-6490795	5	Fyb	827	3' proximal	coding	0	0	0	#DIV/0!	GeneID:23880	Other miscellaneous function protein
chr15.38739497	chr15:38739497-38740794	5	Atp6v1c1	170	5' proximal	coding	3	13	64	0.23	GeneID:66335	Hydrogen transporter;ATP synthase;Other hydrolase
chr14.98160614	chr14:98160614-98161350	5	Spry2	32871	3' distal	coding	0	0	8	#DIV/0!	GeneID:24064	Other signaling molecule;Other transcription factor
chr14.70504056	chr14:70504056-70505529	5	AK033642	117097	gene desert	coding	16	0	0	#DIV/0!	NA	NA
chr14.55488184	chr14:55488184-55489494	5	Tdh	119	intragenic (intron)	coding	44	0	0	#DIV/0!	GeneID:58865	Molecular function unclassified
chr14.52951825	chr14:52951825-52952755	5	Tnfrsf19	5814	3' proximal	coding	0	0	25	#DIV/0!	GeneID:29820	Tumor necrosis factor receptor
chr14.44749754	chr14:44749754-44751374	5	Zfp219	516	intragenic (intron)	coding	83	10	102	8.30	GeneID:69890	KRAB box transcription factor
chr14.44390821	chr14:44390821-44392492	5	BC049649	468	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr14.3090005	chr14:3090005-3090749	5	2610042L04Rik	2317	5' proximal	coding	17	240	0	0.07	GeneID:67055	Molecular function unclassified
chr14.3083292	chr14:3083292-3086061	5	BC055874	722	intragenic (intron)	coding	17	240	0	0.07	NA	NA
chr14.22274188	chr14:22274188-22274845	5	1110051B16Rik	7978	5' proximal	coding	8	0	0	#DIV/0!	GeneID:278672	Homeobox transcription factor;Nucleic acid binding
chr14.21744072	chr14:21744072-21744639	5	Rai17	10548	5' distal	coding	6	24	98	0.25	GeneID:328365	Other ligase
chr13.91432554	chr13:91432554-91434013	5	Tbca	82249	5' distal	coding	0	56	0	0.00	GeneID:21371	Molecular function unclassified
chr13.86319409	chr13:86319409-86320493	5	Cspg2	5358	intragenic (intron)	coding	0	50	21	0.00	GeneID:13003	Extracellular matrix glycoprotein
chr13.63809277	chr13:63809277-63810976	5	AK076976		5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr13.62313699	chr13:62313699-62315404	5	Hsd17b3	78814	3' distal	coding	0	0	0	#DIV/0!	GeneID:15487	Dehydrogenase
chr13.61899220	chr13:61899220-61900610	5	Ptch1	885	5' proximal	coding	6	0	60	#DIV/0!	GeneID:19206	Other receptor
chr13.44175638	chr13:44175638-44176581	5	Jarid2	19645	5' distal	coding	97	20	0	4.85	GeneID:16468	Other transcription factor;Nucleic acid binding
chr13.43471022	chr13:43471022-43471680	5	BY740427 EST		5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr13.17958735	chr13:17958735-17959936	5	Pou6f2	87767	5' distal	coding	2	0	77	#DIV/0!	GeneID:218030	Homeobox transcription factor;Nucleic acid binding
chr13.106418798	chr13:106418798-106422649	5	Pde4d	54587	intragenic (intron)	coding	8	0	0	#DIV/0!	GeneID:238871	Phosphodiesterase
chr12.99935460	chr12:99935460-99936513	5	Tcl1	514	5' proximal	coding	67	1	0	67.00	GeneID:21432	Kinase activator
chr12.86187961	chr12:86187961-86189009	5	Gtf2a1	3135	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:83602	Molecular function unclassified
chr12.75415014	chr12:75415014-75415762	5	Zfp36l1	3911	5' proximal	coding	1	0	93	#DIV/0!	GeneID:12192	Nuclease

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr12.75409456	chr12:75409456-75411877	5	Zfp36l1	868	intragenic (intron)	coding	1	0	93	#DIV/0!	GeneID:12192	Nuclease
chr12.56815521	chr12:56815521-56817024	5	Lrfn5	57713	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:238205	Nuclease
chr12.47613749	chr12:47613749-47614327	5	D930036F22Rik	84411	3' distal	coding	5	6	0	0.83	NA	NA
chr12.106417686	chr12:106417686-106418901	5	Ckb	558	3' proximal	coding	0	0	0	#DIV/0!	GeneID:12709	Other kinase
chr12.105446706	chr12:105446706-105449387	5	Hspca	6392	5' proximal	coding	4343	3158	432	1.38	GeneID:15519	Chaperone
chr12.104696378	chr12:104696378-104697323	5	Ppp2r5c	46094	intragenic (intron)	coding	0	68	102	0.00	GeneID:26931	Protein phosphatase
chr11.9945862	chr11:9945862-9947138	5	S57425	75693		target is repeat	-	-	-	#VALUE!	NA	NA
chr11.88136002	chr11:88136002-88136283	5	Msi2h	54445	intragenic (intron)	coding	96	17	102	5.65	GeneID:76626	Ribonucleoprotein
chr11.87686381	chr11:87686381-87686823	5	Vezf1	240	3' proximal	coding	0	0	15	#DIV/0!	GeneID:22344	Molecular function unclassified
chr11.79592354	chr11:79592354-79594028	5	Suz12	12291	5' distal	coding	120	103	4	1.17	GeneID:52615	Storage protein
chr11.77859815	chr11:77859815-77860739	5	Sdf2	1725	intragenic (intron)	coding	3	37	96	0.08	GeneID:20316	Glycosyltransferase
chr11.77538192	chr11:77538192-77539329	5	AK036145	2797	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr11.71114607	chr11:71114607-71117199	5	U16672	119864	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr11.62962282	chr11:62962282-62963278	5	S74315	117468	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr11.6075594	chr11:6075594-6076339	5	Nudcd3	12171	intragenic (intron)	coding	46	90	12	0.51	GeneID:209586	Molecular function unclassified
chr11.54478910	chr11:54478910-54479787	5	AK006294	1179	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr11.53238686	chr11:53238686-53239619	5	Kif3a	726	intragenic (intron)	coding	7	144	23	0.05	GeneID:16568	Microtubule binding motor protein
chr11.48426438	chr11:48426438-48428062	5	Gnb2-rs1	13230	5' distal	coding	0	0	0	#DIV/0!	GeneID:14694	Other enzyme regulator
chr11.28941626	chr11:28941626-28944120	5	Pnpt1	82765	5' distal	coding	0	0	0	#DIV/0!	GeneID:71701	Exoribonuclease;Nucleotidyltransferase;Esterase
chr11.25889472	chr11:25889472-25890080	5	AK080999	7516	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr11.119066370	chr11:119066370-119067294	5	AK037444	857	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr11.116752925	chr11:116752925-116755401	5	AK012687	3500	3' proximal	noncoding	86	37	74	2.32	NA	NA
chr11.112541575	chr11:112541575-112542595	5	4933434M16Rik	38474	3' distal	coding	0	0	6	#DIV/0!	GeneID:71203	Molecular function unclassified
chr10.79946544	chr10:79946544-79948160	5	Gpx4	3113	5' proximal	coding	2179	3108	595	0.70	GeneID:14779	Peroxidase
chr10.75743506	chr10:75743506-75744470	5	Mif	12971	5' distal	coding	0	0	0	#DIV/0!	GeneID:17319	Other cytokine;Isomerase
chr10.67913315	chr10:67913315-67913851	5	Arid5b	2491	intragenic (intron)	coding	1	0	12	#DIV/0!	GeneID:71371	Transcription cofactor
chr10.6094321	chr10:6094321-6096188	5	Akap12	11919	5' distal	coding	0	0	0	#DIV/0!	GeneID:83397	Kinase modulator
chr10.37241703	chr10:37241703-37242199	5	AK042044		intragenic (exon)	noncoding	0	0	0	#DIV/0!	NA	NA
chr10.120216965	chr10:120216965-120217776	5	Hmga2	49937	5' distal	coding	0	2	5	0.00	GeneID:15364	Molecular function unclassified
chr10.107966295	chr10:107966295-107967538	5	AK035230	6952	3' proximal	coding	6	26	34	0.23	NA	NA
chr10.107162353	chr10:107162353-107163271	5	Myf5	3059	5' proximal	coding	0	0	0	#DIV/0!	GeneID:17877	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr1.93133706	chr1:93133706-93136195	5	AK035048	23748	5' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr1.87131243	chr1:87131243-87132779	5	Eif4e2	353	intragenic (intron)	coding	219	0	50	#DIV/0!	GeneID:26987	Translation initiation factor
chr1.86196336	chr1:86196336-86196954	5	B3gnt7	3889	3' proximal	coding	66	2	0	33.00	GeneID:227327	Glycosyltransferase
chr1.83561973	chr1:83561973-83563045	5	Slc19a3	0	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:80721	Other transporter
chr1.67095448	chr1:67095448-67096429	5	Rpe	835	5' proximal	coding	3	3	54	1.00	GeneID:66646	Epimerase/racemase
chr1.59780637	chr1:59780637-59781750	5	S71494	2132		target is repeat	-	-	-	#VALUE!	NA	NA
chr1.58985586	chr1:58985586-58986190	5	Ndufb3	152	5' proximal	coding	1	0	20	#DIV/0!	GeneID:66495	Oxidoreductase
chr1.47002795	chr1:47002795-47004151	5	Z83810	29290	3' distal	coding	0	4	24	0.00	NA	NA
chr1.37125972	chr1:37125972-37127358	5	D1Bwg0491e	271	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:56030	Molecular function unclassified
chr1.36263026	chr1:36263026-36263954	5	Hs6st1	123747	gene desert	coding	5	0	30	#DIV/0!	GeneID:50785	Other transferase
chr1.182115189	chr1:182115189-182116005	5	Enah	10694	5' distal	coding	216	10	81	21.60	GeneID:13800	Non-motor actin binding protein
chr1.181001039	chr1:181001039-181002599	5	Lefty1	10298	5' distal	coding	194	4	0	48.50	GeneID:13590	TGF-beta superfamily member
chr1.178048676	chr1:178048676-178049433	5	Adss	119404	gene desert	coding	0	0	0	#DIV/0!	GeneID:11566	Synthetase;Other ligase
chr1.167297387	chr1:167297387-167299431	5	Uck2	13725	intragenic (intron)	coding	53	49	52	1.08	GeneID:80914	Nucleotide kinase
chr1.12935238	chr1:12935238-12936060	5	Sulf1	0	intragenic (exon)	coding	7	0	0	#DIV/0!	GeneID:240725	Esterase
chr1.110798677	chr1:110798677-110799374	5	Cdh7	654455	gene desert	coding	0	0	0	#DIV/0!	GeneID:241201	Cadherin
chrX.95385178	chrX:95385178-95386103	4	Rnf12	714	5' proximal	coding	0	7	0	0.00	GeneID:19820	Transcription cofactor
chrX.79708798	chrX:79708798-79709297	4	AK081272	42543	5' distal	coding	110	0	0	#DIV/0!	NA	NA
chrX.60565121	chrX:60565121-60566631	4	AK045941	23610		target is repeat	-	-	-	#VALUE!	NA	NA
chrX.5689011	chrX:5689011-5691335	4	2010204K13Rik	11101	3' distal	coding	147	19	7	7.74	GeneID:68355	Molecular function unclassified
chrX.22286222	chrX:22286222-22287691	4	BC068151	148364	gene desert	coding	0	0	11	#DIV/0!	NA	NA
chrX.160104061	chrX:160104061-160104310	4	Mid1	807	3' proximal	coding	0	272	0	0.00	GeneID:17318	Non-motor microtubule binding protein
chrX.156808500	chrX:156808500-156809626	4	Egfl6	110874	gene desert	coding	5	0	0	#DIV/0!	GeneID:54156	Extracellular matrix structural protein
chrX.145177773	chrX:145177773-145178697	4	Acate2	100	5' proximal	coding	15	2	0	7.50	GeneID:56360	Esterase
chrX.141011400	chrX:141011400-141011938	4	2310007F12Rik	27586	5' distal	coding	0	0	88	#DIV/0!	GeneID:69499	Molecular function unclassified
chrX.11205285	chrX:11205285-11205674	4	AK053279		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr9.97871130	chr9:97871130-97873332	4	Clstn2	289	5' proximal	coding	0	0	9	#DIV/0!	GeneID:64085	Cell adhesion molecule;Calmodulin related protein;Annexin
chr9.96958321	chr9:96958321-96959212	4	C330005L02Rik	2797	5' proximal	coding	65	7	40	9.29	GeneID:192287	Mitochondrial carrier protein
chr9.85913561	chr9:85913561-85916351	4	AK077387	201605	gene desert	coding	17	77	4	0.22	NA	NA
chr9.82680561	chr9:82680561-82682870	4	4930486G11Rik	50085	3' distal	coding	0	0	0	#DIV/0!	GeneID:75033	Molecular function unclassified
chr9.78765229	chr9:78765229-78767731	4	Dppa5	1291	5' proximal	coding	1444	1381	0	1.05	GeneID:13915	Molecular function unclassified
chr9.78411705	chr9:78411705-78412892	4	Gcm1	5193	5' proximal	coding	36	3	96	12.00	GeneID:14531	Transcription factor;Nucleic acid binding
chr9.75200667	chr9:75200667-75201812	4	Onecut1	4297	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:15379	Homeobox transcription factor;Nucleic acid binding

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr9.71813786	chr9:71813786-71814717	4	Grinl1a	8055	5' proximal	coding	197	128	102	1.54	GeneID:28015	Molecular function unclassified
chr9.70178758	chr9:70178758-70181371	4	Bnip2	56684	5' distal	coding	11	0	44	#DIV/0!	GeneID:12175	Other miscellaneous function protein
chr9.63383416	chr9:63383416-63384164	4	C230094B15Rik	60	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:207667	Other transcription factor
chr9.61593830	chr9:61593830-61595081	4	Tle3	721	5' proximal	coding	0	0	13	#DIV/0!	GeneID:21887	Transcription cofactor;Other miscellaneous function protein
chr9.60639095	chr9:60639095-60640952	4	BB638949 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr9.50261200	chr9:50261200-50262564	4	Ncam1	336296	gene desert	coding	0	0	28	#DIV/0!	GeneID:17967	CAM family adhesion molecule
chr9.47616913	chr9:47616913-47618122	4	Igsf4a	16795	5' distal	coding	0	0	49	#DIV/0!	GeneID:54725	Receptor
chr9.46995699	chr9:46995699-46996532	4	2900052N01Rik	21908	5' distal	coding	0	0	0	#DIV/0!	GeneID:73040	Molecular function unclassified
chr9.45931218	chr9:45931218-45931787	4	CF104778 EST	800	5' proximal	coding	-	-	-	#VALUE!	NA	NA
chr9.34535926	chr9:34535926-34537594	4	Kirrel3	69082	intragenic (intron)	coding	0	0	13	#DIV/0!	GeneID:67703	CAM family adhesion molecule
chr9.30572838	chr9:30572838-30574228	4	S74315	43708		target is repeat	-	-	-	#VALUE!	NA	NA
chr9.29626025	chr9:29626025-29626983	4	Hnt	246107	intragenic (intron)	coding	0	0	20	#DIV/0!	GeneID:235106	CAM family adhesion molecule
chr9.23204541	chr9:23204541-23206548	4	Bmper	5386	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:73230	Extracellular matrix glycoprotein
chr9.21741505	chr9:21741505-21742496	4	Ankrd25	338	5' proximal	coding	0	0	0	#DIV/0!	GeneID:235041	Transcription factor;Nuclease
chr9.21704099	chr9:21704099-21706060	4	Spbc24	1899	5' proximal	coding	0	0	0	#DIV/0!	GeneID:67629	Molecular function unclassified
chr9.21701622	chr9:21701622-21703881	4	Spbc24	437	5' proximal	coding	0	0	0	#DIV/0!	GeneID:67629	Molecular function unclassified
chr9.21529922	chr9:21529922-21531670	4	1810026J23Rik	0	intragenic (exon 1)	coding	0	0	0	#DIV/0!	GeneID:69773	Molecular function unclassified
chr9.122590934	chr9:122590934-122595781	4	AK048438	5000	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr9.118359240	chr9:118359240-118360706	4	Eomes	88332	5' distal	coding	0	28	0	0.00	GeneID:13813	Other transcription factor;Nucleic acid binding
chr9.117049385	chr9:117049385-117055094	4	Rbms3	21396	intragenic (intron)	coding	0	0	9	#DIV/0!	GeneID:207181	Single-stranded DNA-binding protein;Replication origin binding protein
chr9.114512279	chr9:114512279-114513529	4	Cnot10	1250	3' proximal	coding	121	35	58	3.46	GeneID:78893	Molecular function unclassified
chr9.110951310	chr9:110951310-110953014	4	Tdgf1	2256	5' proximal	coding	426	172	0	2.48	GeneID:21667	Growth factor
chr9.10463277	chr9:10463277-10464146	4	BY723902 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr9.102155855	chr9:102155855-102156381	4	Ephb1	38760	intragenic (intron)	coding	0	0	29	#DIV/0!	GeneID:270190	Tyrosine protein kinase receptor;Protein kinase
chr8.95153959	chr8:95153959-95155047	4	BC031853	27948	3' distal	coding	3	0	4	#DIV/0!	GeneID:234595	Other transporter
chr8.93561669	chr8:93561669-93562304	4	Nup93	42	intragenic (intron)	coding	145	69	100	2.10	GeneID:71805	Molecular function unclassified
chr8.92678708	chr8:92678708-92679339	4	Ces1	560	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:12623	Esterase
chr8.83999853	chr8:83999853-84005074	4	Nfix	475	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18032	Other transcription factor;Nucleic acid binding
chr8.81934556	chr8:81934556-81936427	4	BY017886 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr8.79663409	chr8:79663409-79664008	4	Gypa	321	5' proximal	coding	0	0	0	#DIV/0!	GeneID:14934	Molecular function unclassified
chr8.78203357	chr8:78203357-78204358	4	AK052736	37894	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr8.77628784	chr8:77628784-77629333	4	2410193C02Rik	688	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:76775	Transporter;Mitochondrial carrier protein
chr8.71397799	chr8:71397799-71400656	4	Eps15-rs	1927	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:13859	Other membrane traffic protein
chr8.70557950	chr8:70557950-70560762	4	2310015I10Rik	7763	3' proximal	coding	58	264	76	0.22	NA	NA
chr8.68146527	chr8:68146527-68147918	4	Atp6v1b2	8220	3' proximal	coding	111	89	99	1.25	GeneID:11966	Other ligand-gated ion channel;Anion channel;Hydrogen transporter
chr8.67519697	chr8:67519697-67520698	4	4732435N03Rik	32389	intragenic (intron)	coding	0	0	26	#DIV/0!	GeneID:234356	Glycosyltransferase
chr8.66067095	chr8:66067095-66069102	4	U16672	131609	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr8.56517173	chr8:56517173-56517744	4	AK038453	5630	intragenic (intron)	noncoding	0	640	0	0.00	NA	NA
chr8.53248699	chr8:53248699-53250274	4	AK214828	1200	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr8.46889090	chr8:46889090-46890182	4	AK035426	13376	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr8.46774810	chr8:46774810-46775638	4	Ing1l	52821	3' distal	coding	41	37	18	1.11	GeneID:69260	Acetyltransferase
chr8.44097204	chr8:44097204-44098033	4	AJ250768	17344	intragenic (intron)	coding	102	109	152	0.94	NA	NA
chr8.33692118	chr8:33692118-33693244	4	Dusp4	74437	5' distal	coding	1	11	2	0.09	GeneID:319520	Kinase inhibitor;Protein phosphatase
chr8.3327183	chr8:3327183-3329223	4	Arhgef18	12350	intragenic (intron)	coding	33	71	5	0.46	GeneID:102098	Guanyl-nucleotide exchange factor
chr8.32879323	chr8:32879323-32880551	4	Rbpms	13305	intragenic (intron)	coding	28	105	12	0.27	GeneID:19663	Nuclease
chr8.24320346	chr8:24320346-24321952	4	Fgfr1	11157	intragenic (intron)	coding	14	171	194	0.08	GeneID:14182	Tyrosine protein kinase receptor;Protein kinase
chr8.23162322	chr8:23162322-23162718	4	CB182672 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr8.22958831	chr8:22958831-22960021	4	AK042983	20731	5' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr8.18714331	chr8:18714331-18717012	4	U16670	155862	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr8.15883517	chr8:15883517-15883961	4	Csmd1	51355	3' distal	coding	0	0	0	#DIV/0!	GeneID:94109	Complement component
chr8.14041342	chr8:14041342-14041889	4	2610019F03Rik	40373	5' distal	coding	0	0	0	#DIV/0!	GeneID:72148	Molecular function unclassified
chr8.126447645	chr8:126447645-126448962	4	Pard3	455	5' proximal	coding	2	33	109	0.06	GeneID:93742	Tight junction
chr8.123003626	chr8:123003626-123004149	4	Gas8	1090	3' proximal	coding	44	60	18	0.73	GeneID:104346	Molecular function unclassified
chr8.121998684	chr8:121998684-122000244	4	BB865444 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr8.121703194	chr8:121703194-121704400	4	AK196871		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr8.119724398	chr8:119724398-119725982	4	6430548M08Rik	110037	gene desert	coding	0	0	50	#DIV/0!	GeneID:234797	Molecular function unclassified
chr8.113823688	chr8:113823688-113824474	4	Wwox	5596	intragenic (intron)	coding	0	0	6	#DIV/0!	GeneID:80707	Oxidoreductase
chr8.111401766	chr8:111401766-111402513	4	Terf2ip	41310	3' distal	coding	1	0	14	#DIV/0!	GeneID:57321	Other nucleic acid binding
chr8.109291943	chr8:109291943-109294732	4	Tat	13462	5' distal	coding	0	0	6	#DIV/0!	GeneID:234724	Synthase;Transaminase
chr8.108034927	chr8:108034927-108036199	4	Atbf1	7730	intragenic (intron)	coding	0	2	3	0.00	GeneID:11906	Other zinc finger transcription factor;Nucleic acid binding
chr8.107711579	chr8:107711579-107715079	4	U16672	59093		target is repeat	-	-	-	#VALUE!	NA	NA
chr8.105911998	chr8:105911998-105912710	4	Cdh1	247	intragenic (intron)	coding	264	416	0	0.63	GeneID:12550	Cadherin
chr7.94820379	chr7:94820379-94821448	4	Syt9	4704	3' proximal	coding	0	0	0	#DIV/0!	GeneID:60510	Membrane traffic regulatory protein
chr7.92925934	chr7:92925934-92928183	4	4632419K20Rik	100	5' proximal	coding	0	0	0	#DIV/0!	GeneID:74349	Molecular function unclassified

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr7.87256259	chr7:87256259-87256763	4	CD351614 EST	50	5' proximal	coding	-	-	-	#VALUE!	NA	NA
chr7.85146479	chr7:85146479-85151987	4	4921513O04	66	intragenic (intron)	coding	0	20	20	0.00	GeneID:233537	Molecular function unclassified
chr7.83636864	chr7:83636864-83638954	4	Odz4	18126	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:23966	Other receptor;Membrane-bound signaling molecule
chr7.83423873	chr7:83423873-83424949	4	Odz4	5309	5' proximal	coding	0	0	0	#DIV/0!	GeneID:23966	Other receptor;Membrane-bound signaling molecule
chr7.70685767	chr7:70685767-70690649	4	AK082155	348	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr7.67266900	chr7:67266900-67267387	4	AK028896	769	5' proximal	coding	0	0	7	#DIV/0!	NA	NA
chr7.64837434	chr7:64837434-64839647	4	AK045941	103389	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr7.60177562	chr7:60177562-60178626	4	AK045941	132361	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr7.57629899	chr7:57629899-57630720	4	BB640622 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr7.52904552	chr7:52904552-52905113	4	Pcsk6	11213	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18553	Serine protease
chr7.31890283	chr7:31890283-31890975	4	2310044H10Rik	82	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:69683	Molecular function unclassified
chr7.18012953	chr7:18012953-18014140	4	Ryr1	232	intragenic (intron)	coding	0	23	0	0.00	GeneID:20190	Other ligand-gated ion channel;Ion channel
chr7.17421119	chr7:17421119-17426238	4	Lrfn1	1634	5' proximal	coding	0	0	0	#DIV/0!	GeneID:80749	Nuclease
chr7.13254128	chr7:13254128-13254554	4	2410005H09Rik	1459	3' proximal	coding	78	4	82	19.50	GeneID:232969	Molecular function unclassified
chr7.131597995	chr7:131597995-131598470	4	BU936261 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr7.121646606	chr7:121646606-121647095	4	Adam12	9987	intragenic (intron)	coding	0	0	55	#DIV/0!	GeneID:11489	Metalloprotease
chr7.121571662	chr7:121571662-121572271	4	Adam12	9633	intragenic (intron)	coding	0	0	55	#DIV/0!	GeneID:11489	Metalloprotease
chr7.119636262	chr7:119636262-119637158	4	Cpxm2	27055	3' distal	coding	15	0	22	#DIV/0!	GeneID:55987	Metalloprotease
chr7.116926958	chr7:116926958-116930828	4	Wdr11	170529	gene desert	coding	6	62	24	0.10	GeneID:207425	Molecular function unclassified
chr7.115475726	chr7:115475726-115476519	4	BC067209	0	intragenic (exonic, 5' utr	coding	0	0	0	#DIV/0!	NA	NA
chr7.115281019	chr7:115281019-115284036	4	Stx1b2	4	intragenic (intron)	coding	11	65	82	0.17	GeneID:56216	SNARE protein
chr7.102395146	chr7:102395146-102396433	4	3830422K02Rik	262193	gene desert	coding	0	0	0	#DIV/0!	GeneID:233752	Molecular function unclassified
chr7.100333557	chr7:100333557-100334491	4	AK087240	16000	5' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr6.99902997	chr6:99902997-99904011	4	AK047332		intragenic (exon)	coding	0	0	0	#DIV/0!	NA	NA
chr6.98338680	chr6:98338680-98339325	4	AK052895	20828	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr6.98332671	chr6:98332671-98334183	4	AK052895	15657	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr6.98095188	chr6:98095188-98095978	4	S57425	64198		target is repeat	-	-	-	#VALUE!	NA	NA
chr6.95115634	chr6:95115634-95116786	4	Lrig1	50936	5' distal	coding	1	19	599	0.05	GeneID:16206	Other cell adhesion molecule
chr6.94175274	chr6:94175274-94176663	4	Baiap1	59779	intragenic (intron)	coding	4	0	32	#DIV/0!	GeneID:14924	Kinase
chr6.86626880	chr6:86626880-86627494	4	Tgfa	21032	3' distal	coding	0	0	0	#DIV/0!	GeneID:21802	Other cytokine
chr6.86533615	chr6:86533615-86537412	4	Tgfa	9207	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21802	Other cytokine
chr6.85512394	chr6:85512394-85517491	4	Sfxn5	1701	intragenic (intron)	coding	5	0	0	#DIV/0!	GeneID:94282	Receptor;Transporter;Other transfer/carrier protein
chr6.84304114	chr6:84304114-84306561	4	Dysf	729	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:26903	Other membrane traffic protein
chr6.83193863	chr6:83193863-83194864	4	D6Mm5e	198	intragenic (intron)	coding	72	0	14	#DIV/0!	GeneID:110958	Molecular function unclassified
chr6.77433325	chr6:77433325-77434593	4	Catna2	47990	intragenic (intron)	coding	7	0	53	#DIV/0!	GeneID:12386	Cell adhesion molecule;Non-motor actin binding protein
chr6.73014602	chr6:73014602-73015570	4	Tcf3	712	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21415	HMG box transcription factor;Nucleic acid binding
chr6.72443064	chr6:72443064-72444523	4	Atoh8	6483	intragenic (intron)	coding	71	30	98	2.37	GeneID:71093	Transcription factor
chr6.71990203	chr6:71990203-71994036	4	D6Ertd253e	4108	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:52250	Molecular function unclassified
chr6.71852818	chr6:71852818-71853834	4	AK129204	76	intragenic (intron)	coding	9	17	32	0.53	NA	NA
chr6.70973214	chr6:70973214-70974683	4	Rpia	0	intragenic (exon 1)	coding	3	2	23	1.50	GeneID:19895	Epimerase/racemase
chr6.68984807	chr6:68984807-68986144	4	AF206028	84104	5' distal	coding	3	2	19	1.50	NA	NA
chr6.67783795	chr6:67783795-67787852	4	AB045138		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr6.65148447	chr6:65148447-65149665	4	Smarcad1	385	intragenic (intron)	coding	7	5	11	1.40	GeneID:13990	DNA helicase;Hydrolase
chr6.63223384	chr6:63223384-63223948	4	AK045941	110469	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr6.5171724	chr6:5171724-5172364	4	Asb4	13777	3' distal	coding	19	8	52	2.38	GeneID:65255	Other signaling molecule
chr6.48721382	chr6:48721382-48722021	4	1810009M01Rik	33312	3' distal	coding	0	4	157	0.00	GeneID:65963	Other receptor
chr6.48717417	chr6:48717417-48718901	4	Gimap3	35515	5' distal	coding	0	0	0	#DIV/0!	GeneID:83408	Other miscellaneous function protein
chr6.39347626	chr6:39347626-39349207	4	Mkrn1	518	5' proximal	coding	386	348	3	1.11	GeneID:54484	Nucleic acid binding
chr6.38223856	chr6:38223856-38224814	4	B130055L09Rik	297	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:209032	Molecular function unclassified
chr6.37786748	chr6:37786748-37787183	4	Trim24	13063	5' distal	coding	131	32	20	4.09	GeneID:21848	Transcription cofactor;Nucleic acid binding
chr6.35136885	chr6:35136885-35137582	4	BC045524	273	intragenic (intron)	coding	24	23	75	1.04	NA	NA
chr6.34084857	chr6:34084857-34085656	4	Slc35b4	10693	5' distal	coding	3	48	54	0.06	GeneID:58246	Other transporter
chr6.32081606	chr6:32081606-32082640	4	Plxna4	683	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:243743	Tyrosine protein kinase receptor;Protein kinase
chr6.29489546	chr6:29489546-29490110	4	Tnpo3	22480	5' distal	coding	219	39	46	5.62	GeneID:320938	Other transporter
chr6.29467503	chr6:29467503-29468541	4	Tnpo3	248	5' proximal	coding	219	39	46	5.62	GeneID:320938	Other transporter
chr6.17654114	chr6:17654114-17655644	4	St7	5774	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:64213	Other miscellaneous function protein
chr6.140742834	chr6:140742834-140743752	4	Aebp2	1708	5' proximal	coding	38	3	42	12.67	GeneID:11569	KRAB box transcription factor;Nucleic acid binding
chr6.140305912	chr6:140305912-140306987	4	Plekha5	6389	5' proximal	coding	0	0	0	#DIV/0!	GeneID:109135	Molecular function unclassified
chr6.13503151	chr6:13503151-13503653	4	B630005N14Rik	70000	5' distal	coding	0	0	0	#DIV/0!	GeneID:101148	Molecular function unclassified
chr6.129209191	chr6:129209191-129215995	4	BC034132	7957	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr6.129115538	chr6:129115538-129116635	4	AK086937	11795	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr6.129100438	chr6:129100438-129101663	4	AK086937	6154	intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr6.128539157	chr6:128539157-128540590	4	Tspan11	5881	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:68498	Membrane-bound signaling molecule;Other cell adhesion molecule
chr6.128346935	chr6:128346935-128348113	4	Hrmt1l4	3217	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:381813	Methyltransferase

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr6.123344393	chr6:123344393-123346326	4	Nanog	24	5' proximal	coding	112	456	0	0.25	GeneID:71950	Homeobox transcription factor;Nucleic acid binding
chr6.123126281	chr6:123126281-123128025	4	BC052078	18640	5' distal	coding	0	0	25	#DIV/0!	NA	NA
chr6.120705087	chr6:120705087-120707043	4	AK052280	12012	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr6.119715929	chr6:119715929-119716600	4	AI841794	502	intragenic (intron)	coding	37	47	25	0.79	GeneID:211187	Molecular function unclassified
chr6.108497095	chr6:108497095-108498284	4	Sumf1	0	intragenic (exon, 3'utr)	coding	4	0	4	#DIV/0!	GeneID:58911	Molecular function unclassified
chr6.102356760	chr6:102356760-102357560	4	U16670	60188		target is repeat	-	-	-	#VALUE!	NA	NA
chr5.99401499	chr5:99401499-99402517	4	Wdfy3	13825	intragenic (intron)	coding	3	4	52	0.75	GeneID:72145	Other membrane traffic protein
chr5.98885208	chr5:98885208-98886392	4	Nkx6-1	117792	gene desert	coding	0	4	35	0.00	GeneID:18096	Homeobox transcription factor
chr5.95625391	chr5:95625391-95628989	4	BC049770	21290	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr5.91927863	chr5:91927863-91929273	4	D5Ertd606e	7463	intragenic (intron)	coding	124	58	184	2.14	GeneID:52398	Other cytoskeletal proteins
chr5.91723468	chr5:91723468-91725840	4	Shrm	21482	intragenic (intron)	coding	2	0	0	#DIV/0!	GeneID:27428	Non-motor actin binding protein
chr5.86564321	chr5:86564321-86566300	4	Csng	2272	5' proximal	coding	0	0	0	#DIV/0!	GeneID:12993	Molecular function unclassified
chr5.86142178	chr5:86142178-86143022	4	AW113456 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr5.85650060	chr5:85650060-85651028	4	BC022697	21206	3' distal	coding	35	26	20	1.35	NA	NA
chr5.76042137	chr5:76042137-76044029	4	2610024G14Rik	219	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:56412	Molecular function unclassified
chr5.71099522	chr5:71099522-71100205	4	Corin	6902	intragenic (intron)	coding	0	0	18	#DIV/0!	GeneID:53419	Receptor;Serine protease
chr5.62566719	chr5:62566719-62567658	4	AA536743	2391	intragenic (intron)	coding	36	56	94	0.64	GeneID:100532	Molecular function unclassified
chr5.52129203	chr5:52129203-52130373	4	C0429845 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr5.49783292	chr5:49783292-49784126	4	Ppargc1a	296248	gene desert	coding	0	0	8	#DIV/0!	GeneID:19017	Transcription cofactor
chr5.46091751	chr5:46091751-46092306	4	AK045941	180097	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr5.44793203	chr5:44793203-44794366	4	Mlr1	315381	gene desert	coding	58	17	28	3.41	GeneID:209707	Molecular function unclassified
chr5.42424203	chr5:42424203-42425381	4	Bst1	4336	5' proximal	coding	0	0	0	#DIV/0!	GeneID:12182	Glycosidase;Cyclase
chr5.37260080	chr5:37260080-37260810	4	Clnk	8086	3' proximal	coding	0	0	0	#DIV/0!	GeneID:27278	Other signaling molecule
chr5.33085153	chr5:33085153-33086513	4	2610033H07Rik	701	intragenic (intron)	coding	71	64	62	1.11	GeneID:75416	Molecular function unclassified
chr5.32383796	chr5:32383796-32384492	4	Whsc2	94	5' proximal	coding	0	27	21	0.00	GeneID:24116	Molecular function unclassified
chr5.26857092	chr5:26857092-26858091	4	Shh	529	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:20423	Other signaling molecule;Protease
chr5.26200307	chr5:26200307-26200849	4	Paxip1	23009	5' distal	coding	0	0	0	#DIV/0!	GeneID:55982	Molecular function unclassified
chr5.23965960	chr5:23965960-23966722	4	Gm443	39909	3' distal	coding	0	0	0	#DIV/0!	GeneID:242891	Chaperonin
chr5.17162077	chr5:17162077-17163101	4	AK045941	188347	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr5.15331322	chr5:15331322-15332328	4	AK045941	26707		target is repeat	-	-	-	#VALUE!	NA	NA
chr5.146683678	chr5:146683678-146685254	4	Alox5ap	61948	3' distal	coding	159	8	35	19.88	GeneID:11690	Other miscellaneous function protein
chr5.145086866	chr5:145086866-145087365	4	Flt1	36224	5' distal	coding	0	0	0	#DIV/0!	GeneID:14254	Tyrosine protein kinase receptor;Protein kinase
chr5.141946134	chr5:141946134-141947295	4	AK049027	32606	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr5.141512694	chr5:141512694-141513224	4	Ocm	2776	5' proximal	coding	0	0	0	#DIV/0!	GeneID:18261	Calmodulin related protein
chr5.137532455	chr5:137532455-137533387	4	Mad1l1	6552	intragenic (intron)	coding	38	14	7	2.71	GeneID:17120	Molecular function unclassified
chr5.134736286	chr5:134736286-134737671	4	Ephb4	1131	intragenic (intron)	coding	3	22	31	0.14	GeneID:13846	Tyrosine protein kinase receptor;Protein kinase
chr5.133102122	chr5:133102122-133102766	4	Usmg1	12829	5' distal	coding	0	0	16	#DIV/0!	GeneID:83676	Molecular function unclassified
chr5.133074425	chr5:133074425-133076689	4	Rhbdl7	2846	5' proximal	coding	3	30	1	0.10	GeneID:215160	Molecular function unclassified
chr5.132388815	chr5:132388815-132390633	4	Cldn4	0	intragenic (exon, 3'utr)	coding	0	0	0	#DIV/0!	GeneID:12740	Tight junction
chr5.131395114	chr5:131395114-131395700	4	AK045941	112367	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr5.130065285	chr5:130065285-130065729	4	AK052418	100989	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr5.126155472	chr5:126155472-126157985	4	Fzd10	144	5' proximal	coding	0	0	0	#DIV/0!	GeneID:93897	G-protein coupled receptor
chr5.125104638	chr5:125104638-125108339	4	AK035053	1043	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr5.120689757	chr5:120689757-120691167	4	BC035291	132	intragenic (intron)	coding	23	0	6	#DIV/0!	GeneID:208043	Zinc finger transcription factor;Nucleic acid binding
chr5.118323579	chr5:118323579-118326746	4	Oas3	6067	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:246727	Nucleic acid binding;Synthetase;Nucleotidyltransferase;Defense
chr5.117899621	chr5:117899621-117901573	4	Lhx5	22604	5' distal	coding	0	0	0	#DIV/0!	GeneID:16873	Other zinc finger transcription factor
chr5.115967040	chr5:115967040-115968139	4	AK047401	12381	5' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr5.113374076	chr5:113374076-113375248	4	Cit	49	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:12704	Non-receptor serine/threonine protein kinase
chr5.111925481	chr5:111925481-111927082	4	Mvk	53590	3' distal	coding	2	0	55	#DIV/0!	GeneID:17855	Other kinase
chr5.110689641	chr5:110689641-110690630	4	Rutbc2	22382	5' distal	coding	9	0	0	#DIV/0!	GeneID:52850	Other G-protein modulator;Membrane traffic regulatory protein
chr5.110681250	chr5:110681250-110682080	4	AK015101	3000	5' proximal	noncoding	9	0	0	#DIV/0!	NA	NA
chr5.110583354	chr5:110583354-110584342	4	BY724384 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr5.109561091	chr5:109561091-109564359	4	AK028326		intragenic (exon)	noncoding	0	0	0	#DIV/0!	NA	NA
chr5.107765850	chr5:107765850-107768331	4	BC059932		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr5.107736938	chr5:107736938-107738148	4	2410025L10Rik	20354	5' distal	coding	0	0	29	#DIV/0!	GeneID:381668	Molecular function unclassified
chr5.101097325	chr5:101097325-101098373	4	Mllt2h	18406	5' distal	coding	0	20	0	0.00	GeneID:17355	Other transcription factor
chr4.98292289	chr4:98292289-98293814	4	Pgm2	14657	intragenic (intron)	coding	24	11	126	2.18	GeneID:72157	Mutase
chr4.84516223	chr4:84516223-84516819	4	CV555842		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr4.80799843	chr4:80799843-80801632	4	AW106080 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr4.72533123	chr4:72533123-72534211	4	Jmjd2c	3410	intragenic (intron)	coding	28	0	2	#DIV/0!	GeneID:76804	Other transcription factor
chr4.70472882	chr4:70472882-70474883	4	Tle1	872	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21885	Transcription cofactor;Other miscellaneous function protein
chr4.64657774	chr4:64657774-64658689	4	Astn2	64485	intragenic (intron)	coding	0	0	20	#DIV/0!	GeneID:56079	Molecular function unclassified
chr4.60317138	chr4:60317138-60317925	4	Mup3	93331	3' distal	coding	0	0	0	#DIV/0!	GeneID:17842	Other transfer/carrier protein
chr4.57048978	chr4:57048978-57049549	4	Akap2	23398	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:11641	Other miscellaneous function protein

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr4.56429404	chr4:56429404-56434085	4	AK052790	10000	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr4.56121888	chr4:56121888-56122770	4	Catnal1	6344	intragenic (intron)	coding	2	0	0	#DIV/0!	GeneID:54366	Cell adhesion molecule;Non-motor actin binding protein
chr4.54450767	chr4:54450767-54451480	4	Zfp462	124613	gene desert	coding	283	267	168	1.06	GeneID:242466	KRAB box transcription factor
chr4.53033766	chr4:53033766-53034730	4	Tal2	2590	3' proximal	coding	0	0	0	#DIV/0!	GeneID:21350	Basic helix-loop-helix transcription factor
chr4.52441407	chr4:52441407-52442203	4	AK015122	8000	3' proximal	noncoding	0	0	19	#DIV/0!	NA	NA
chr4.48612656	chr4:48612656-48613513	4	C730036D15Rik	11605	3' distal	coding	0	0	0	#DIV/0!	GeneID:209186	Esterase
chr4.47755767	chr4:47755767-47758539	4	5730528L13Rik	29447	5' distal	coding	15	0	46	#DIV/0!	GeneID:66665	Molecular function unclassified
chr4.45063959	chr4:45063959-45064517	4	Igfbpl1	2680	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:75426	Miscellaneous function
chr4.45057327	chr4:45057327-45058744	4	Igfbpl1	5	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:75426	Miscellaneous function
chr4.44231959	chr4:44231959-44234733	4	Grhpr	101	intragenic (intron)	coding	44	176	12	0.25	GeneID:76238	Dehydrogenase
chr4.41786501	chr4:41786501-41788699	4	Cntfr	413	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:12804	Interleukin receptor
chr4.41774135	chr4:41774135-41776617	4	Cntfr	2757	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:12804	Interleukin receptor
chr4.41660585	chr4:41660585-41661242	4	Dnaic1	186	5' proximal	coding	0	0	0	#DIV/0!	GeneID:68922	Microtubule family cytoskeletal protein
chr4.36685813	chr4:36685813-36688019	4	U16670	14789		target is repeat	-	-	-	#VALUE!	NA	NA
chr4.33174704	chr4:33174704-33176168	4	Ankrd6	20698	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:140577	Transcription factor;Nucleic acid binding
chr4.18321870	chr4:18321870-18323571	4	Mmp16	276937	gene desert	coding	0	0	0	#DIV/0!	GeneID:17389	Metalloprotease;Other extracellular matrix
chr4.148563633	chr4:148563633-148564243	4	Dnb5	25846	5' distal	coding	0	18	12	0.00	GeneID:242773	Transporter
chr4.148332448	chr4:148332448-148333482	4	BQ561509 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr4.148181302	chr4:148181302-148182311	4	Eno1	47304	3' distal	coding	8423	5926	1407	1.42	GeneID:13806	Dehydratase
chr4.147555328	chr4:147555328-147558271	4	Pik3cd	2504	intragenic (intron)	coding	0	1	2	0.00	GeneID:18707	Other kinase
chr4.147481564	chr4:147481564-147483415	4	Clstn1	9293	intragenic (intron)	coding	0	0	84	#DIV/0!	GeneID:65945	Cell adhesion molecule;Calmodulin related protein;Annexin
chr4.147329935	chr4:147329935-147331644	4	Rbp7	4522	3' proximal	coding	0	0	0	#DIV/0!	GeneID:63954	Other transfer/carrier protein
chr4.143773264	chr4:143773264-143777530	4	CF164191 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr4.143606708	chr4:143606708-143607969	4	Tnfrsf8	2244	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21941	Other receptor
chr4.142647542	chr4:142647542-142649165	4	BC052827	32456	5' distal	coding	6	0	0	#DIV/0!	NA	NA
chr4.140796275	chr4:140796275-140796873	4	AK052952	0	intragenic (exon, 3'utr)	noncoding	0	0	7	#DIV/0!	NA	NA
chr4.140526893	chr4:140526893-140528449	4	BC003277	12273	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:214359	Molecular function unclassified
chr4.140021399	chr4:140021399-140022628	4	2410043F08Rik	11382	3' distal	coding	179	94	0	1.90	GeneID:74202	Other zinc finger transcription factor
chr4.139324679	chr4:139324679-139325150	4	AK015793	4500	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr4.13828052	chr4:13828052-13828958	4	Cbfa2t1h	10751	3' distal	coding	14	0	83	#DIV/0!	GeneID:12395	Other transcription factor
chr4.137154475	chr4:137154475-137155788	4	Ubxd3	17118	3' distal	coding	0	0	0	#DIV/0!	GeneID:212190	Molecular function unclassified
chr4.136869843	chr4:136869843-136870583	4	0610009K11Rik	18138	5' distal	coding	0	0	14	#DIV/0!	GeneID:68350	Molecular function unclassified
chr4.133961303	chr4:133961303-133962027	4	4930555I21Rik	2564	intragenic (intron)	coding	0	0	3	#DIV/0!	GeneID:78806	Molecular function unclassified
chr4.133184701	chr4:133184701-133185516	4	Ldlrap1	14442	3' distal	coding	4	16	0	0.25	GeneID:100017	Other signaling molecule
chr4.130829320	chr4:130829320-130830555	4	Med18	188	intragenic (intron)	coding	33	21	0	1.57	GeneID:67219	Molecular function unclassified
chr4.130519117	chr4:130519117-130519740	4	Taf12	3847	5' proximal	coding	0	27	16	0.00	GeneID:66464	Basal transcription factor;Nucleic acid binding
chr4.128888898	chr4:128888898-128889753	4	Pum1	23141	5' distal	coding	40	2	12	20.00	GeneID:80912	Other RNA-binding protein;Translation factor
chr4.128871192	chr4:128871192-128872853	4	2610200G18Rik	10882	3' distal	coding	34	29	270	1.17	GeneID:67149	Molecular function unclassified
chr4.128330221	chr4:128330221-128331394	4	AK005651	250	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr4.12589208	chr4:12589208-12590245	4	U16670	48116		target is repeat	-	-	-	#VALUE!	NA	NA
chr4.123978294	chr4:123978294-123980198	4	AF245444	50822	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr4.123834708	chr4:123834708-123835157	4	Ftl1	39495	5' distal	coding	5	1	12	5.00	GeneID:14325	Storage protein
chr4.122811350	chr4:122811350-122812844	4	CA464466 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr4.117371750	chr4:117371750-117372829	4	AA833215 EST	4000	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr4.115758745	chr4:115758745-115759528	4	4931406I20Rik	37973	intragenic (intron)	coding	67	102	36	0.66	GeneID:66743	Molecular function unclassified
chr4.115707525	chr4:115707525-115709110	4	4931406I20Rik	12259	intragenic (intron)	coding	67	102	36	0.66	GeneID:66743	Molecular function unclassified
chr4.114223297	chr4:114223297-114224694	4	Mknk1	0	intragenic (exon, 3'utr)	coding	0	14	7	0.00	GeneID:17346	Protein kinase
chr4.107372144	chr4:107372144-107372983	4	Nrd1	535	intragenic (intron)	coding	6	193	0	0.03	GeneID:230598	Other hydrolase;Metalloprotease
chr4.106596271	chr4:106596271-106599825	4	AK122545	2883	5' proximal	coding	60	0	1	#DIV/0!	NA	NA
chr4.106564293	chr4:106564293-106565276	4	AK122545	11170	3' distal	coding	60	0	1	#DIV/0!	NA	NA
chr4.106336868	chr4:106336868-106337976	4	Slc1a7	1074	intragenic (intron)	coding	1	0	141	#DIV/0!	GeneID:242607	Other transporter
chr4.104905632	chr4:104905632-104906175	4	Dhcr24	637	3' proximal	coding	10	26	20	0.38	GeneID:74754	Reductase
chr4.100234682	chr4:100234682-100235434	4	C130073F10Rik	493	3' proximal	coding	19	3	0	6.33	GeneID:242574	Molecular function unclassified
chr3.97160114	chr3:97160114-97160673	4	Acp6	134	intragenic (intron)	coding	28	16	42	1.75	GeneID:66659	Other phosphatase
chr3.96390757	chr3:96390757-96392151	4	AK195651		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr3.95696981	chr3:95696981-95697871	4	Tsrc1	14556	5' distal	coding	0	0	0	#DIV/0!	GeneID:229595	Metalloprotease
chr3.9286570	chr3:9286570-9287390	4	AK083443	34156	5' distal	coding	0	0	4	#DIV/0!	NA	NA
chr3.90721137	chr3:90721137-90722117	4	Gatad2b	941	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr3.89398540	chr3:89398540-89399734	4	AK005385	7	5' proximal	coding	4	0	43	#DIV/0!	NA	NA
chr3.88726185	chr3:88726185-88727466	4	Rhbg	5627	3' proximal	coding	0	0	0	#DIV/0!	GeneID:58176	Cation transporter
chr3.88393287	chr3:88393287-88396054	4	Hdgf	427	5' proximal	coding	1127	1255	546	0.90	GeneID:15191	Growth factor
chr3.87023531	chr3:87023531-87027525	4	Lrba	2017	5' proximal	coding	10	0	407	#DIV/0!	GeneID:80877	Other membrane traffic protein
chr3.84384024	chr3:84384024-84388393	4	D930015E06Rik	19906	intragenic (intron)	coding	19	56	99	0.34	GeneID:229473	Molecular function unclassified
chr3.81498413	chr3:81498413-81499185	4	Pdgfc	39028	intragenic (intron)	coding	0	0	13	#DIV/0!	GeneID:54635	Other receptor;Growth factor

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr3.6949070	chr3:6949070-6950507	4	AK045941	158234	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr3.69466204	chr3:69466204-69467843	4	Kpna4	154	5' proximal	coding	1	4	14	0.25	GeneID:16649	Membrane traffic regulatory protein
chr3.65904323	chr3:65904323-65905095	4	AK089841		5' proximal	noncoding				#DIV/0!	NA	NA
chr3.64553609	chr3:64553609-64554968	4	4930518C23Rik	18965	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:319210	G-protein coupled receptor
chr3.58647934	chr3:58647934-58649890	4	AK078519	0	5' proximal	coding	132	2	83	66.00	NA	NA
chr3.3672657	chr3:3672657-3673737	4	Hnf4g	9567	3' proximal	coding	0	0	0	#DIV/0!	GeneID:30942	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr3.33593904	chr3:33593904-33594817	4	Ttc14	237	5' proximal	coding	8	33	6	0.24	GeneID:67120	Glycosyltransferase
chr3.30489379	chr3:30489379-30490241	4	AK044227	21326	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr3.30485642	chr3:30485642-30487954	4	AK044227	21326	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr3.28608946	chr3:28608946-28610077	4	AK045941	25315		target is repeat	-	-	-	#VALUE!	NA	NA
chr3.28419659	chr3:28419659-28420244	4	AK088459	751	intragenic (intron)	coding	0	0	5	#DIV/0!	NA	NA
chr3.21260811	chr3:21260811-21262936	4	AK021294	500468	gene desert	noncoding	0	0	0	#DIV/0!	NA	NA
chr3.160243417	chr3:160243417-160245293	4	Rpe65	314	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:19892	Oxygenase
chr3.158549129	chr3:158549129-158550232	4	AK007475	6441	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr3.154276084	chr3:154276084-154276597	4	AK017676	113227	gene desert	noncoding	0	0	0	#DIV/0!	NA	NA
chr3.153961108	chr3:153961108-153963748	4	St6galnac3	153610	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:20447	Glycosyltransferase
chr3.15055608	chr3:15055608-15056222	4	AK044583		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr3.148538954	chr3:148538954-148539281	4	AK045941	249143	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr3.144979456	chr3:144979456-144980154	4	Hs2st1	17931	3' distal	coding	0	0	4	#DIV/0!	GeneID:23908	Other transferase
chr3.143208849	chr3:143208849-143209902	4	Gbp2	12439	3' distal	coding	0	10	3	0.00	GeneID:14469	Large G-protein
chr3.142407173	chr3:142407173-142408502	4	Bmpr1b	4022	intragenic (intron)	coding	0	0	41	#DIV/0!	GeneID:12167	TGF-beta receptor;Serine/threonine protein kinase receptor;Protein kinase
chr3.138337483	chr3:138337483-138338172	4	H2afz	84371	5' distal	coding	895	243	298	3.68	GeneID:51788	Histone
chr3.136663949	chr3:136663949-136664678	4	AY178734	12100	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr3.134078577	chr3:134078577-134081454	4	BB531403 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr3.131302275	chr3:131302275-131305814	4	AK003177	5640	3' proximal	coding	790	766	376	1.03	NA	NA
chr3.118780993	chr3:118780993-118781893	4	AK051312	3589	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr3.114948338	chr3:114948338-114949115	4	Olfm3	301322	gene desert	coding	0	0	0	#DIV/0!	GeneID:229759	Extracellular matrix glycoprotein
chr3.113340104	chr3:113340104-113341485	4	AK045941	113548	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr3.108212305	chr3:108212305-108214147	4	AK052120		intragenic (exon)	noncoding	0	0	0	#DIV/0!	NA	NA
chr3.107064234	chr3:107064234-107065117	4	DN174290 EST	50	5' proximal	coding	-	-	-	#VALUE!	NA	NA
chr3.103057254	chr3:103057254-103058334	4	AK044389		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr3.101941528	chr3:101941528-101943084	4	Nhlh2	691	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18072	Nuclease
chr3.101341290	chr3:101341290-101343028	4	Igsf3	19598	intragenic (intron)	coding	10	106	0	0.09	GeneID:78908	Other miscellaneous function protein
chr2.9855043	chr2:9855043-9855755	4	AK084574	11093	3' distal	coding	4	0	6	#DIV/0!	NA	NA
chr2.92699130	chr2:92699130-92700721	4	Syt13	109079	gene desert	coding	5	0	0	#DIV/0!	GeneID:80976	Membrane traffic regulatory protein
chr2.92663690	chr2:92663690-92666371	4	Syt13	144331	gene desert	coding	5	0	0	#DIV/0!	GeneID:80976	Membrane traffic regulatory protein
chr2.90368756	chr2:90368756-90369392	4	Ptprj	45	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:19271	Other receptor;Protein phosphatase
chr2.77400361	chr2:77400361-77401382	4	Zfp533	2812	intragenic (intron)	coding	0	0	3	#DIV/0!	GeneID:241494	Molecular function unclassified
chr2.75964304	chr2:75964304-75964977	4	AK042383		intragenic (intron)	noncoding	0	0	5	#DIV/0!	NA	NA
chr2.75662328	chr2:75662328-75663049	4	Agps	63782	5' distal	coding	2	0	29	#DIV/0!	GeneID:228061	Synthase;Transferase
chr2.71598415	chr2:71598415-71599160	4	CO813457 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr2.6542104	chr2:6542104-6543018	4	Cugbp2	5218	intragenic (intron)	coding	1	8	40	0.13	GeneID:14007	Ribonucleoprotein
chr2.58578584	chr2:58578584-58579881	4	Upp2	58575	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:76654	Phosphorylase
chr2.5662565	chr2:5662565-5662988	4	Camk1d	442	5' proximal	coding	0	0	10	#DIV/0!	GeneID:227541	Non-receptor serine/threonine protein kinase
chr2.43938832	chr2:43938832-43940207	4	Arhgap15	84459	intragenic (intron)	coding	14	8	20	1.75	GeneID:76117	Molecular function unclassified
chr2.38356496	chr2:38356496-38358410	4	Lhx2	28162	3' distal	coding	0	0	112	#DIV/0!	GeneID:16870	Other zinc finger transcription factor
chr2.34359533	chr2:34359533-34360023	4	Mapkap1	7221	5' proximal	coding	0	0	11	#DIV/0!	GeneID:227743	Molecular function unclassified
chr2.34209365	chr2:34209365-34209938	4	Pbx3	25106	intragenic (intron)	coding	1	0	14	#DIV/0!	GeneID:18516	Homeobox transcription factor
chr2.33398610	chr2:33398610-33400108	4	Zfp297b	11413	3' distal	coding	0	0	122	#DIV/0!	GeneID:71834	KRAB box transcription factor
chr2.32496494	chr2:32496494-32497372	4	C230093N12Rik	483	intragenic (intron)	coding	2	0	26	#DIV/0!	GeneID:98952	Molecular function unclassified
chr2.31531659	chr2:31531659-31532607	4	BY251886	250	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr2.31505255	chr2:31505255-31505806	4	Ass1	24779	3' distal	coding	243	157	36	1.55	GeneID:11898	Other ligase
chr2.31123481	chr2:31123481-31124170	4	AI790205	3275	intragenic (intron)	coding	0	18	17	0.00	GeneID:277463	Molecular function unclassified
chr2.30862268	chr2:30862268-30864109	4	Ptges	37	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:64292	Synthase and synthetase
chr2.29591190	chr2:29591190-29592402	4	Rapgef1	12078	intragenic (intron)	coding	47	1	20	47.00	GeneID:107746	G-protein modulator
chr2.27308773	chr2:27308773-27313474	4	Vav2	15806	intragenic (intron)	coding	17	0	0	#DIV/0!	GeneID:22325	Guanyl-nucleotide exchange factor
chr2.24869292	chr2:24869292-24872859	4	Ehmt1	7088	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:77683	Methyltransferase
chr2.22016163	chr2:22016163-22021530	4	U16670	36323		target is repeat	-	-	-	#VALUE!	NA	NA
chr2.20708678	chr2:20708678-20709430	4	AB093289	8714	intragenic (intron)	coding				#DIV/0!	NA	NA
chr2.20231746	chr2:20231746-20232950	4	AK052020	59685	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr2.19455401	chr2:19455401-19457267	4	Msrb2	5429	3' proximal	coding	0	0	0	#DIV/0!	GeneID:76467	Reductase
chr2.18741573	chr2:18741573-18743225	4	Bmi1	455	3' proximal	coding	0	0	8	#DIV/0!	GeneID:12151	RING finger transcription factor
chr2.17908344	chr2:17908344-17909107	4	CO812708 EST	4000	5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr2.176968462	chr2:176968462-176969169	4	BB187163 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr2.176867282	chr2:176867282-176869258	4	AK078446		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr2.172929979	chr2:172929979-172931678	4	Bmp7	2390	3' proximal	coding	0	0	42	#DIV/0!	GeneID:12162	TGF-beta superfamily member
chr2.167592999	chr2:167592999-167593741	4	Snai1	7960	5' proximal	coding	112	86	41	1.30	GeneID:20613	KRAB box transcription factor
chr2.166219860	chr2:166219860-166221130	4	AK034712	1681	5' proximal	coding	28	105	97	0.27	NA	NA
chr2.163427821	chr2:163427821-163431993	4	Jph2	7577	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:59091	Molecular function unclassified
chr2.160719615	chr2:160719615-160719959	4	Top1	13646	intragenic (intron)	coding	21	73	44	0.29	GeneID:21969	Nucleic acid binding;Select regulatory molecule;Isomerase
chr2.15876487	chr2:15876487-15877260	4	AK045941	9465		target is repeat	-	-	-	#VALUE!	NA	NA
chr2.157629173	chr2:157629173-157630204	4	Blcap	0	5' proximal	coding	15	0	0	#DIV/0!	GeneID:53619	Molecular function unclassified
chr2.157311311	chr2:157311311-157311962	4	AK029507		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr2.156750081	chr2:156750081-156752878	4	BY122638 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr2.156009711	chr2:156009711-156010373	4	Gdf5	1360	5' proximal	coding	0	0	0	#DIV/0!	GeneID:14563	TGF-beta superfamily member
chr2.154489240	chr2:154489240-154490350	4	CA560714		intragenic (intron)	coding				#DIV/0!	NA	NA
chr2.154411057	chr2:154411057-154412294	4	Cdk5rap1	0	intragenic (exon)	coding				#DIV/0!	GeneID:66971	Molecular function unclassified
chr2.146383628	chr2:146383628-146384320	4	AK038838	119399	gene desert	coding				#DIV/0!	NA	NA
chr2.146054663	chr2:146054663-146055328	4	BC053994	12528	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:381383	Other G-protein modulator
chr2.136863637	chr2:136863637-136865929	4	Jag1	303	5' proximal	coding	0	19	42	0.00	GeneID:16449	Receptor;Membrane-bound signaling molecule;Defense/immunity protein
chr2.131882524	chr2:131882524-131882720	4	Slc23a2	31881	intragenic (intron)	coding	6	0	0	#DIV/0!	GeneID:54338	Other transporter
chr2.127588480	chr2:127588480-127588861	4	AA880581 EST		intragenic (intron)	coding	3	2	2	1.50	NA	NA
chr2.124145197	chr2:124145197-124148915	4	Sema6d	99210	intragenic (intron)	coding	0	0	176	#DIV/0!	GeneID:214968	Membrane-bound signaling molecule
chr2.119585582	chr2:119585582-119586717	4	Tyro3	351	intragenic (intron)	coding	68	29	150	2.34	GeneID:22174	Tyrosine protein kinase receptor;Protein kinase
chr2.115249945	chr2:115249945-115251432	4	BC052040	112151	gene desert	coding				#DIV/0!	GeneID:399568	Molecular function unclassified
chr2.11521975	chr2:11521975-11522492	4	Rbm17	14217	3' distal	coding	10	0	20	#DIV/0!	GeneID:76938	mRNA splicing factor
chr2.10303015	chr2:10303015-10304172	4	AK076687		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr19.9403286	chr19:9403286-9404418	4	BY721332		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr19.8552532	chr19:8552532-8555398	4	Asrgl1	318662	gene desert	coding	8	3	67	2.67	GeneID:66514	Hydrolase
chr19.8289525	chr19:8289525-8295227	4	Asrgl1	57304	5' distal	coding	8	3	67	2.67	GeneID:66514	Hydrolase
chr19.6976740	chr19:6976740-6978929	4	Rcor2	4041	5' proximal	coding	212	91	47	2.33	GeneID:104383	Molecular function unclassified
chr19.6905540	chr19:6905540-6907864	4	D930010J01Rik	3368	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:107227	Molecular function unclassified
chr19.6660600	chr19:6660600-6663696	4	Bad	2551	intragenic (intron)	coding	0	0	32	#DIV/0!	GeneID:12015	Other miscellaneous function protein
chr19.6358367	chr19:6358367-6359845	4	AK208454		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr19.59482912	chr19:59482912-59486707	4	E330013P04Rik	53829	5' distal	coding	0	0	0	#DIV/0!	GeneID:107376	Molecular function unclassified
chr19.54963925	chr19:54963925-54964647	4	Vti1a	35063	intragenic (intron)	coding	0	0	23	#DIV/0!	GeneID:53611	SNARE protein
chr19.54937813	chr19:54937813-54939137	4	Vti1a	60633	intragenic (intron)	coding	0	0	23	#DIV/0!	GeneID:53611	SNARE protein
chr19.54599957	chr19:54599957-54601330	4	AY395631	698	intragenic (intron)	coding	1	0	4	#DIV/0!	NA	NA
chr19.49939881	chr19:49939881-49947123	4	Sorcs1	81866	intragenic (intron)	coding	0	0	9	#DIV/0!	GeneID:58178	Other receptor
chr19.48279932	chr19:48279932-48286660	4	AK045941	6619		target is repeat	-	-	-	#VALUE!	NA	NA
chr19.45585203	chr19:45585203-45585788	4	Gbf1	1234	intragenic (intron)	coding	87	3	49	29.00	GeneID:107338	Guanyl-nucleotide exchange factor
chr19.45235759	chr19:45235759-45236824	4	9130011E15Rik	4528	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:71617	Molecular function unclassified
chr19.41746171	chr19:41746171-41747075	4	Crtac1	4167	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:72832	Calmodulin related protein;Annexin
chr19.41472169	chr19:41472169-41472659	4	Avpi1	7263	5' proximal	coding	170	14	2	12.14	GeneID:69534	Molecular function unclassified
chr19.40969517	chr19:40969517-40972403	4	Slit1	98	intragenic (intron)	coding	0	0	41	#DIV/0!	GeneID:20562	Membrane-bound signaling molecule
chr19.38801706	chr19:38801706-38804621	4	Cyp2c39	6491	intragenic (intron)	coding	50	70	7	0.71	GeneID:13098	Oxygenase
chr19.36712776	chr19:36712776-36713450	4	Hhex	48351	3' distal	coding	170	68	47	2.50	GeneID:15242	Homeobox transcription factor;Nucleic acid binding
chr19.3202486	chr19:3202486-3203865	4	Gal	212	5' proximal	coding	71	2	238	35.50	GeneID:14419	Peptide hormone
chr19.30418882	chr19:30418882-30419368	4	Cstf2t	21401	5' distal	coding	10	13	143	0.77	GeneID:83410	Ribonucleoprotein
chr19.29853534	chr19:29853534-29855430	4	Dkk1	4431	5' proximal	coding	0	0	0	#DIV/0!	GeneID:13380	Other signaling molecule
chr19.28583222	chr19:28583222-28583705	4	Jak2	853	intragenic (intron)	coding	0	0	9	#DIV/0!	GeneID:16452	Non-receptor tyrosine protein kinase
chr19.28305540	chr19:28305540-28306882	4	Cdc37l1	587	intragenic (intron)	coding	1	0	3	#DIV/0!	GeneID:67072	Other chaperones;Kinase activator
chr19.23336997	chr19:23336997-23337748	4	AK006227	67	5' proximal	noncoding	63	0	33	#DIV/0!	NA	NA
chr19.15354668	chr19:15354668-15359351	4	Gnaq	52276	5' distal	coding	22	4	49	5.50	GeneID:14682	Large G-protein
chr19.13953603	chr19:13953603-13954363	4	Tle4	79122	5' distal	coding	0	0	6	#DIV/0!	GeneID:21888	Transcription cofactor;Other miscellaneous function protein
chr19.11408373	chr19:11408373-11409625	4	Olfr76	38178	5' distal	coding	0	0	54	#DIV/0!	GeneID:258677	G-protein coupled receptor
chr18.85445493	chr18:85445493-85446911	4	Fbxo15	600	5' proximal	coding	129	21	0	6.14	NA	NA
chr18.83260053	chr18:83260053-83261908	4	AK008989	22000	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr18.76265621	chr18:76265621-76268662	4	AV209748	18000	3' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr18.76117115	chr18:76117115-76118887	4	Gm672	20187	intragenic (intron)	coding				#DIV/0!	GeneID:269037	Molecular function unclassified
chr18.75874300	chr18:75874300-75876663	4	Smad7	22282	5' distal	coding	19	0	1	#DIV/0!	GeneID:17131	Other transcription factor
chr18.74763926	chr18:74763926-74770396	4	Cxxc1	16460	3' distal	coding	3	6	19	0.50	GeneID:74322	Other nucleic acid binding
chr18.6726300	chr18:6726300-6726785	4	Epc1	822	5' proximal	coding	30	2	39	15.00	GeneID:13831	Chromatin/chromatin-binding protein
chr18.66362874	chr18:66362874-66363369	4	Sec11l3	15790	3' distal	coding	0	0	0	#DIV/0!	GeneID:66286	Serine protease
chr18.57166750	chr18:57166750-57170738	4	Lmnb1	60702	5' distal	coding	28	25	11	1.12	GeneID:16906	Intermediate filament
chr18.56190444	chr18:56190444-56191747	4	AK045941	532350	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr18.42748918	chr18:42748918-42750078	4	BC054080	2435	intragenic (intron)	coding	8	0	2	#DIV/0!	NA	NA
chr18.39402426	chr18:39402426-39403351	4	4933432P15Rik	33295	5' distal	coding	0	0	5	#DIV/0!	GeneID:71302	G-protein modulator

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr18.39044266	chr18:39044266-39045403	4	Spry4	369	5' proximal	coding	60	0	6	#DIV/0!	GeneID:24066	Other signaling molecule;Other transcription factor
chr18.38248967	chr18:38248967-38249754	4	Pcdhgc3	197	intragenic (intron)	coding	0	0	591	#DIV/0!	GeneID:93706	Cadherin
chr18.31712921	chr18:31712921-31713309	4	Rit2	10235	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:19762	Small GTPase;Hydrolase
chr18.30957334	chr18:30957334-30958060	4	Pik3c3	160144	gene desert	coding	0	0	0	#DIV/0!	GeneID:225326	Other kinase
chr18.21824452	chr18:21824452-21825560	4	AK018108	134633	gene desert	coding	37	0	0	#DIV/0!	NA	NA
chr18.17652561	chr18:17652561-17653302	4	AK045941	68327		target is repeat	-	-	-	#VALUE!	NA	NA
chr18.14076911	chr18:14076911-14078419	4	Zfp521	2309	intragenic (intron)	coding	0	0	33	#DIV/0!	GeneID:225207	KRAB box transcription factor;Nucleic acid binding
chr18.10682235	chr18:10682235-10683305	4	BC008220	105498	gene desert	coding	0	0	0	#DIV/0!	NA	NA
chr17.9928127	chr17:9928127-9929988	4	AK005771	61585	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr17.9494502	chr17:9494502-9495372	4	Qk	126567	gene desert	coding	0	0	269	#DIV/0!	GeneID:19317	Other RNA-binding protein
chr17.89617611	chr17:89617611-89618160	4	Nrxn1	141	intragenic (intron)	coding	0	0	20	#DIV/0!	GeneID:18189	Other receptor;Cell adhesion molecule
chr17.8915229	chr17:8915229-8916552	4	U16671	8470		target is repeat	-	-	-	#VALUE!	NA	NA
chr17.86955368	chr17:86955368-86956341	4	Foxn2	98609	5' distal	coding	0	24	6	0.00	GeneID:14236	Transcription factor;Nucleic acid binding
chr17.86345416	chr17:86345416-86345994	4	Kcnk12	28	3' proximal	coding	0	0	0	#DIV/0!	GeneID:210741	Other ion channel
chr17.83285215	chr17:83285215-83285839	4	Lrpprc	1071	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:72416	Oxidase
chr17.79195860	chr17:79195860-79197677	4	BG261625 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr17.7819862	chr17:7819862-7820677	4	T2	1384	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21331	Molecular function unclassified
chr17.67115736	chr17:67115736-67118286	4	A730037L19Rik	89580	5' distal	coding	0	0	0	#DIV/0!	GeneID:320236	Molecular function unclassified
chr17.64670859	chr17:64670859-64671606	4	1110012J17Rik	36118	3' distal	coding	28	4	0	7.00	GeneID:68617	Molecular function unclassified
chr17.6404177	chr17:6404177-6405198	4	Sytl3	69135	5' distal	coding	480	415	719	1.16	GeneID:83672	Membrane traffic regulatory protein
chr17.6233434	chr17:6233434-6236790	4	Tctex1	26514	3' distal	coding	480	415	719	1.16	GeneID:21648	Microtubule family cytoskeletal protein
chr17.61931377	chr17:61931377-61931764	4	AK078781	105588	gene desert	noncoding	0	0	0	#DIV/0!	NA	NA
chr17.54235494	chr17:54235494-54235985	4	Ebi3	0	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:50498	Interleukin receptor
chr17.51772091	chr17:51772091-51774392	4	Pcaf	12942	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:18519	Transcription cofactor;Acetyltransferase
chr17.45928168	chr17:45928168-45930532	4	Tcfeb	21978	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21425	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr17.45900656	chr17:45900656-45901237	4	Tcfeb	1296	5' proximal	coding	0	0	0	#DIV/0!	GeneID:21425	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr17.34618969	chr17:34618969-34619470	4	H2-T10	2705	3' proximal	coding	32	0	0	#DIV/0!	GeneID:15024	Major histocompatibility complex antigen
chr17.33661448	chr17:33661448-33662078	4	Bat2	2372	5' proximal	coding	0	63	3	0.00	GeneID:53761	Transcription factor;Nuclease
chr17.33526740	chr17:33526740-33531152	4	Msh5	10	intragenic (intron)	coding	242	139	44	1.74	GeneID:17687	Damaged DNA-binding protein
chr17.32620088	chr17:32620088-32620682	4	Brd2	2783	5' proximal	coding	121	19	83	6.37	GeneID:14312	Other kinase
chr17.31372818	chr17:31372818-31374737	4	9030612M13Rik	13507	3' distal	coding	0	42	86	0.00	GeneID:208292	KRAB box transcription factor
chr17.29220940	chr17:29220940-29228612	4	Dnahc8	1316	5' proximal	coding	0	0	0	#DIV/0!	GeneID:13417	Microtubule binding motor protein;Hydrolase
chr17.29214018	chr17:29214018-29216675	4	Glo1	237	5' proximal	coding	219	21	47	10.43	GeneID:109801	Other lyase
chr17.29134926	chr17:29134926-29136111	4	Btbd9	2827	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:224671	Molecular function unclassified
chr17.28001251	chr17:28001251-28001991	4	AJ242721		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr17.26693064	chr17:26693064-26694625	4	Tcp11	25225	5' distal	coding	0	0	0	#DIV/0!	GeneID:21463	Receptor
chr17.25605068	chr17:25605068-25606655	4	Bak1	987	3' proximal	coding	0	0	0	#DIV/0!	GeneID:12018	Other miscellaneous function protein
chr17.14494885	chr17:14494885-14495613	4	Chd1	26972	5' distal	coding	64	10	3	6.40	GeneID:12648	Transcription factor;DNA helicase;Hydrolase
chr17.14065687	chr17:14065687-14066427	4	Dll1	119358	gene desert	coding	0	0	79	#DIV/0!	GeneID:13388	Receptor;Membrane-bound signaling molecule;Defense/immunity protein
chr17.13167572	chr17:13167572-13169272	4	Smoc2	26276	intragenic (intron)	coding	0	8	0	0.00	GeneID:64074	Molecular function unclassified
chr17.12356461	chr17:12356461-12357139	4	Tcte2	43879	3' distal	coding	0	0	0	#DIV/0!	GeneID:21646	Molecular function unclassified
chr17.10478314	chr17:10478314-10480073	4	Park2	1022	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:50873	Other transfer/carrier protein;Other ligase
chr16.96489618	chr16:96489618-96490843	4	BY725911 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr16.9627507	chr16:9627507-9628856	4	Grin2a	177625	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:14811	Glutamate receptor;Ion channel
chr16.8971947	chr16:8971947-8973121	4	Grin2a	292911	gene desert	coding	0	0	0	#DIV/0!	GeneID:14811	Glutamate receptor;Ion channel
chr16.8465340	chr16:8465340-8466083	4	Usp7	61880	5' distal	coding	400	163	133	2.45	GeneID:252870	Cysteine protease
chr16.81861280	chr16:81861280-81862224	4	Ncam2	9044	intragenic (intron)	coding	0	0	28	#DIV/0!	GeneID:17968	CAM family adhesion molecule
chr16.72184837	chr16:72184837-72185758	4	AF303453	43137	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr16.48380089	chr16:48380089-48380894	4	Morc	14753	intragenic (intron)	coding	9	0	0	#DIV/0!	GeneID:17450	Molecular function unclassified
chr16.48340214	chr16:48340214-48340892	4	Morc	4240	intragenic (intron)	coding	9	0	0	#DIV/0!	GeneID:17450	Molecular function unclassified
chr16.48336028	chr16:48336028-48337499	4	Morc	424	intragenic (intron)	coding	9	0	0	#DIV/0!	GeneID:17450	Molecular function unclassified
chr16.35430279	chr16:35430279-35431378	4	Sema5b	34403	intragenic (intron)	coding	4	0	29	#DIV/0!	GeneID:20357	Other receptor;Membrane-bound signaling molecule
chr16.31209056	chr16:31209056-31210363	4	Apod	30063	5' distal	coding	0	0	0	#DIV/0!	GeneID:11815	Apolipoprotein
chr16.30510347	chr16:30510347-30511416	4	BC022623	43564	3' distal	coding	17	0	7	#DIV/0!	GeneID:224093	Molecular function unclassified
chr16.29982068	chr16:29982068-29988570	4	Hes1	55425	3' distal	coding	0	0	0	#DIV/0!	GeneID:15205	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr16.26805433	chr16:26805433-26806066	4	Il1rap	245088	gene desert	coding	0	0	0	#DIV/0!	GeneID:16180	Interleukin receptor
chr16.24288646	chr16:24288646-24290726	4	Lpp	601	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:210126	Other zinc finger transcription factor
chr16.24238852	chr16:24238852-24240611	4	Lpp	3	5' proximal	coding	0	0	0	#DIV/0!	GeneID:210126	Other zinc finger transcription factor
chr16.23752685	chr16:23752685-23753609	4	Gm605	16760	3' distal	coding	0	0	0	#DIV/0!	GeneID:224055	Molecular function unclassified
chr16.22235134	chr16:22235134-22235844	4	Etv5	801	intragenic (intron)	coding	0	0	98	#DIV/0!	GeneID:104156	Other transcription factor;Nucleic acid binding
chr16.21910539	chr16:21910539-21911228	4	C330012H03Rik	36	intragenic (intron)	coding	73	257	40	0.28	GeneID:319765	Other RNA-binding protein
chr16.20465840	chr16:20465840-20466495	4	Eif4g1	546	3' proximal	coding	548	344	336	1.59	GeneID:208643	Translation initiation factor
chr16.18366516	chr16:18366516-18368889	4	Tbx1	3816	5' proximal	coding	1	4	37	0.25	GeneID:21380	Other transcription factor;Nucleic acid binding
chr16.18176953	chr16:18176953-18178951	4	Arvcf	43	intragenic (intron)	coding	1	25	25	0.04	GeneID:11877	Cell adhesion molecule;Other cell junction protein

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr16.18025268	chr16:18025268-18026433	4	Ranbp1	0	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:19385	Other G-protein modulator
chr16.17054658	chr16:17054658-17060031	4	Pik4ca	817	3' proximal	coding	15	0	69	#DIV/0!	GeneID:224020	Other kinase
chr16.17009915	chr16:17009915-17010732	4	Hic2	77	5' proximal	coding	10	0	0	#DIV/0!	GeneID:58180	KRAB box transcription factor;Nucleic acid binding
chr16.16986895	chr16:16986895-16987699	4	mKIAA1666		intragenic (exon)	coding	0	0	0	#DIV/0!	NA	NA
chr16.13720255	chr16:13720255-13720726	4	Ifitm7	5537	3' proximal	coding	0	0	5	#DIV/0!	GeneID:74482	Other miscellaneous function protein
chr15.99685538	chr15:99685538-99686429	4	Kcnh3	231	5' proximal	coding	0	0	0	#DIV/0!	GeneID:16512	Voltage-gated potassium channel
chr15.96910904	chr15:96910904-96915547	4	Sfrs2ip	23153	5' distal	coding	39	14	79	2.79	GeneID:72193	Molecular function unclassified
chr15.96608677	chr15:96608677-96609649	4	Tmem16f	204094	gene desert	coding	0	0	0	#DIV/0!	GeneID:105722	Molecular function unclassified
chr15.93946250	chr15:93946250-93946775	4	Zfp106	12598	intragenic (exon)	coding	0	0	0	#DIV/0!	GeneID:20402	KRAB box transcription factor;Nucleic acid binding
chr15.89728841	chr15:89728841-89729854	4	Aldrl6	46	intragenic (intron)	coding	12	2	106	6.00	GeneID:56727	Oxygenase
chr15.88285115	chr15:88285115-88286183	4	AW049604	139593	gene desert	coding	0	0	9	#DIV/0!	GeneID:106014	Molecular function unclassified
chr15.84977378	chr15:84977378-84977967	4	Arhgap8	41128	5' distal	coding	1	39	0	0.03	GeneID:109270	Other G-protein modulator
chr15.80547473	chr15:80547473-80547736	4	Mgat3	4186	3' proximal	coding	7	4	140	1.75	GeneID:17309	Glycosyltransferase
chr15.79932381	chr15:79932381-79933811	4	Dmc1h	240	5' proximal	coding	6	0	4	#DIV/0!	GeneID:13404	DNA strand-pairing protein;Hydrolase
chr15.77665680	chr15:77665680-77666228	4	AK050776	4800	5' proximal	noncoding	39	2	7	19.50	NA	NA
chr15.76685055	chr15:76685055-76685785	4	BE691987 EST	2500	5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr15.76316046	chr15:76316046-76318454	4	Mapk15	800	5' proximal	coding	0	0	0	#DIV/0!	GeneID:332110	Non-receptor serine/threonine protein kinase
chr15.76272872	chr15:76272872-76274585	4	Zfp623	0	intragenic (exon)	coding	3	3	47	1.00	GeneID:78834	KRAB box transcription factor;Nucleic acid binding
chr15.75954480	chr15:75954480-75956538	4	Zfp41	8678	3' proximal	coding	11	0	0	#DIV/0!	GeneID:22701	KRAB box transcription factor;Nucleic acid binding
chr15.74879656	chr15:74879656-74880405	4	Bai1	26	intragenic (intron)	coding	7	0	14	#DIV/0!	GeneID:107831	G-protein coupled receptor;Cell adhesion molecule
chr15.6944014	chr15:6944014-6944914	4	Lifr	23619	5' distal	coding	24	0	10	#DIV/0!	GeneID:16880	Cytokine receptor
chr15.6943059	chr15:6943059-6943704	4	Lifr	24490	5' distal	coding	24	0	10	#DIV/0!	GeneID:16880	Cytokine receptor
chr15.67340290	chr15:67340290-67341407	4	St3gal1	31728	3' distal	coding	0	0	0	#DIV/0!	GeneID:20442	Glycosyltransferase
chr15.53383788	chr15:53383788-53384434	4	Ext1	687	intragenic (intron)	coding	139	13	71	10.69	GeneID:14042	Glycosyltransferase
chr15.48831852	chr15:48831852-48833163	4	AK122567	>600000	gene desert	coding	0	0	0	#DIV/0!	NA	NA
chr15.41951398	chr15:41951398-41954229	4	Oxr1	117000	gene desert	coding	0	0	0	#DIV/0!	GeneID:170719	Molecular function unclassified
chr15.40878910	chr15:40878910-40889917	4	Zfpm2	42131	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:22762	Other zinc finger transcription factor;Nucleic acid binding
chr15.25810866	chr15:25810866-25812113	4	Myo10	5926	intragenic (intron)	coding	1	160	117	0.01	GeneID:17909	Actin binding motor protein
chr15.102720707	chr15:102720707-102722839	4	Rarg	1383	intragenic (intron)	coding	275	57	32	4.82	GeneID:19411	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr15.101090163	chr15:101090163-101092085	4	BC004728	5758	intragenic (intron)	coding	3	0	0	#DIV/0!	GeneID:207818	Molecular function unclassified
chr14.70548713	chr14:70548713-70551513	4	AK033642	73076	5' distal	coding	16	0	0	#DIV/0!	NA	NA
chr14.69011776	chr14:69011776-69012621	4	D230005D02Rik	17699	intragenic (intron)	coding	18	0	0	#DIV/0!	GeneID:239188	Molecular function unclassified
chr14.65009205	chr14:65009205-65009561	4	P2y5	0	5' proximal	coding	2	0	59	#DIV/0!	GeneID:67168	G-protein coupled receptor
chr14.61259962	chr14:61259962-61260970	4	Mscp	17549	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:67712	Mitochondrial carrier protein
chr14.59062441	chr14:59062441-59063125	4	BC065100	144	intragenic (intron)	coding	9	71	50	0.13	NA	NA
chr14.58107560	chr14:58107560-58108760	4	Ephx2	1603	intragenic (intron)	coding	16	0	0	#DIV/0!	GeneID:13850	Other hydrolase
chr14.52629868	chr14:52629868-52630506	4	BY216334 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr14.51702962	chr14:51702962-51703975	4	Atp8a2	19530	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:50769	Other transporter;Other hydrolase
chr14.49795730	chr14:49795730-49797305	4	Sap18	14	3' proximal	coding	0	0	0	#DIV/0!	GeneID:20220	Chromatin/chromatin-binding protein
chr14.49518595	chr14:49518595-49519134	4	Il17d	4690	intragenic (intron)	coding	4	0	4	#DIV/0!	GeneID:239114	Interleukin
chr14.47525015	chr14:47525015-47525612	4	Pck2	1574	3' proximal	coding	8	7	108	1.14	GeneID:74551	Decarboxylase
chr14.45018559	chr14:45018559-45019533	4	Rab2b	32	5' proximal	coding	0	0	41	#DIV/0!	GeneID:76338	Small GTPase
chr14.41949649	chr14:41949649-41950751	4	Slc35f4	22064	3' distal	coding	0	0	0	#DIV/0!	GeneID:75200	Molecular function unclassified
chr14.41427564	chr14:41427564-41428062	4	Otx2	86516	5' distal	coding	106	66	0	1.61	GeneID:18424	Homeobox transcription factor;Nucleic acid binding
chr14.41374148	chr14:41374148-41374861	4	Otx2	33280	5' distal	coding	106	66	0	1.61	GeneID:18424	Homeobox transcription factor;Nucleic acid binding
chr14.37056215	chr14:37056215-37058186	4	5730469M10Rik	493	intragenic (intron)	coding	0	0	11	#DIV/0!	GeneID:70564	Molecular function unclassified
chr14.3419195	chr14:3419195-3422975	4	BC055874	35746	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr14.3236083	chr14:3236083-3237008	4	2610042L04Rik	13264	5' distal	coding	17	240	0	0.07	GeneID:67055	Molecular function unclassified
chr14.30738229	chr14:30738229-30740309	4	AK129105	867	intragenic (intron)	coding	78	12	9	6.50	NA	NA
chr14.3045189	chr14:3045189-3046901	4	CN537856 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr14.26954362	chr14:26954362-26956254	4	Itih3	9	intragenic (intron)	coding	24	4	24	6.00	GeneID:16426	Serine protease inhibitor
chr14.26611310	chr14:26611310-26613060	4	Tkt	1020	3' proximal	coding	274	296	712	0.93	GeneID:21881	Transketolase
chr14.26500669	chr14:26500669-26501568	4	Mitc1	15745	5' distal	coding	0	0	0	#DIV/0!	GeneID:75901	Transcription cofactor
chr14.22914213	chr14:22914213-22917107	4	Asb14	14852	5' distal	coding	0	0	0	#DIV/0!	GeneID:142687	Transcription factor;Nucleic acid binding;Other miscellaneous function
chr14.10519724	chr14:10519724-10523330	4	1500006O09Rik	41202	3' distal	coding	0	0	22	#DIV/0!	GeneID:66231	Molecular function unclassified
chr13.98532261	chr13:98532261-98534714	4	Pik3r1	113701	gene desert	coding	0	0	37	#DIV/0!	GeneID:18708	Kinase modulator
chr13.95224533	chr13:95224533-95225561	4	Foxd1	134661	gene desert	coding	0	0	0	#DIV/0!	GeneID:15229	Other transcription factor;Nucleic acid binding
chr13.94012760	chr13:94012760-94013978	4	AK021247	200	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr13.93936031	chr13:93936031-93936832	4	Hexb	7091	5' proximal	coding	64	18	82	3.56	GeneID:15212	Glycosidase
chr13.81775202	chr13:81775202-81776395	4	Ccnh	39949	5' distal	coding	66	0	138	#DIV/0!	GeneID:66671	Kinase activator
chr13.74030561	chr13:74030561-74034094	4	AK045941	103247	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr13.69278581	chr13:69278581-69279124	4	Irx2	202767	gene desert	coding	0	0	0	#DIV/0!	GeneID:16372	Homeobox transcription factor;Nucleic acid binding
chr13.68788301	chr13:68788301-68790949	4	Irx2	282459	gene desert	coding	0	0	0	#DIV/0!	GeneID:16372	Homeobox transcription factor;Nucleic acid binding
chr13.64090844	chr13:64090844-64091741	4	MGC102251	532	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:432769	KRAB box transcription factor;Nucleic acid binding

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr13.63864815	chr13:63864815-63865829	4	AK021218	2368	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr13.63822851	chr13:63822851-63824073	4	C330014B19Rik	0	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:382770	G-protein coupled receptor
chr13.62052999	chr13:62052999-62053938	4	AK006304		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr13.61901805	chr13:61901805-61902532	4	Ptch1	3616	5' proximal	coding	6	0	60	#DIV/0!	GeneID:19206	Other receptor
chr13.5688406	chr13:5688406-5689347	4	Klf6	31993	3' distal	coding	16	72	39	0.22	GeneID:23849	KRAB box transcription factor;Nucleic acid binding
chr13.56468061	chr13:56468061-56469109	4	AK039058	60314	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr13.54281171	chr13:54281171-54282263	4	AK019465	3500	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr13.5337095	chr13:5337095-5340283	4	Klf6	311649	gene desert	coding	16	72	39	0.22	GeneID:23849	KRAB box transcription factor;Nucleic acid binding
chr13.52272165	chr13:52272165-52273271	4	Ror2	52387	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:26564	Tyrosine protein kinase receptor;Protein kinase
chr13.51893577	chr13:51893577-51894461	4	Auh	38468	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:11992	Molecular function unclassified
chr13.37518936	chr13:37518936-37520695	4	Riok1	17633	3' distal	coding	1	0	10	#DIV/0!	GeneID:71340	Protein kinase
chr13.35289861	chr13:35289861-35290950	4	Cdyl	3385	intragenic (intron)	coding	10	1	11	10.00	GeneID:12593	Molecular function unclassified
chr13.35175013	chr13:35175013-35175505	4	Cdyl	2949	intragenic (intron)	coding	10	1	11	10.00	GeneID:12593	Molecular function unclassified
chr13.34765597	chr13:34765597-34767008	4	Cdyl	330053	gene desert	coding	10	1	11	10.00	GeneID:12593	Molecular function unclassified
chr13.34173700	chr13:34173700-34174753	4	AK007247	2247	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr13.33678363	chr13:33678363-33679153	4	BC053705	22290	intragenic (intron)	coding	0	21	11	0.00	NA	NA
chr13.31107976	chr13:31107976-31110430	4	Foxc1	77507	5' distal	coding	0	0	15	#DIV/0!	GeneID:17300	Other transcription factor;Nucleic acid binding
chr13.111229642	chr13:111229642-111230406	4	Ndufs4	8311	3' proximal	coding	0	0	62	#DIV/0!	GeneID:17993	Dehydrogenase;Reductase
chr13.109926632	chr13:109926632-109929412	4	AK086507		5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr12.94026503	chr12:94026503-94028712	4	Ches1	23426	intragenic (intron)	coding	39	2	101	19.50	GeneID:71375	Transcription factor;Nucleic acid binding
chr12.82896480	chr12:82896480-82898614	4	Adck1	30000	5' distal	coding	2	0	7	#DIV/0!	GeneID:72113	ATP-binding cassette (ABC) transporter;Kinase;Transferase
chr12.82130430	chr12:82130430-82131645	4	1110030H02Rik	75501	5' distal	coding	0	4	0	0.00	GeneID:68737	Other transcription factor
chr12.81982716	chr12:81982716-81984397	4	G630009D10Rik	23766	5' distal	coding	0	0	19	#DIV/0!	GeneID:238328	Molecular function unclassified
chr12.81830643	chr12:81830643-81831413	4	Esrrb	1090	intragenic (intron)	coding	250	24	0	10.42	GeneID:26380	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr12.79961005	chr12:79961005-79965414	4	CN687776 EST	200	5' proximal	coding	-	-	-	#VALUE!	NA	NA
chr12.79938540	chr12:79938540-79940097	4	Abcd4	1923	5' proximal	coding	353	854	1650	0.41	GeneID:19300	Transporter
chr12.76140983	chr12:76140983-76142405	4	4933426M11Rik	10645	intragenic (intron)	coding	1	9	7	0.11	GeneID:217684	Molecular function unclassified
chr12.68023271	chr12:68023271-68023928	4	Ppm1a	15875	5' distal	coding	10	108	121	0.09	GeneID:19042	Protein phosphatase
chr12.68011604	chr12:68011604-68012948	4	Ppm1a	27160	5' distal	coding	10	108	121	0.09	GeneID:19042	Protein phosphatase
chr12.67924699	chr12:67924699-67925397	4	Dhrs7	403	3' proximal	coding	0	3	29	0.00	GeneID:66375	Oxidoreductase
chr12.66727206	chr12:66727206-66727781	4	AK050197		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr12.51638976	chr12:51638976-51640221	4	Titf1	85006	3' distal	coding	0	8	0	0.00	GeneID:21869	Homeobox transcription factor;Nucleic acid binding
chr12.51093137	chr12:51093137-51094530	4	Brms1l	37568	3' distal	coding	21	35	25	0.60	GeneID:52592	Molecular function unclassified
chr12.50263436	chr12:50263436-50265437	4	Srp54	908	intragenic (intron)	coding	16	17	35	0.94	GeneID:24067	Other receptor;Other RNA-binding protein;G-protein
chr12.49415669	chr12:49415669-49417378	4	Egln3	7627	intragenic (intron)	coding	4	20	71	0.20	GeneID:112407	Molecular function unclassified
chr12.49387746	chr12:49387746-49389150	4	Egln3	13934	3' distal	coding	4	20	71	0.20	GeneID:112407	Molecular function unclassified
chr12.47074447	chr12:47074447-47078774	4	D930036F22Rik	17661	3' distal	coding	5	6	0	0.83	GeneID:320487	Molecular function unclassified
chr12.4168011	chr12:4168011-4168633	4	Ncoa1	0	intragenic (exon)	coding	93	0	25	#DIV/0!	GeneID:17977	Transcription cofactor;Nucleic acid binding
chr12.36802400	chr12:36802400-36803034	4	Immp2l	26239	intragenic (intron)	coding	0	13	0	0.00	GeneID:93757	Other proteases
chr12.3651937	chr12:3651937-3653421	4	Dtnb	5597	3' proximal	coding	3	0	25	#DIV/0!	GeneID:13528	Non-motor actin binding protein
chr12.30674787	chr12:30674787-30675981	4	Ahr	109312	gene desert	coding	0	0	0	#DIV/0!	GeneID:11622	Other receptor;Basic helix-loop-helix transcription factor
chr12.28577354	chr12:28577354-28578594	4	Twistnb	117046	gene desert	coding	18	0	0	#DIV/0!	GeneID:28071	Molecular function unclassified
chr12.22235267	chr12:22235267-22240484	4	Sox11	59580	3' distal	coding	0	73	324	0.00	GeneID:20666	HMG box transcription factor;Nucleic acid binding
chr12.113557686	chr12:113557686-113558784	4	AK084355		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr12.106566765	chr12:106566765-106570546	4	Zfyve21	728	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:68520	Molecular function unclassified
chr12.105583734	chr12:105583734-105584416	4	Rage	3241	intragenic (intron)	coding	5	0	8	#DIV/0!	GeneID:26448	Non-receptor serine/threonine protein kinase
chr12.105229674	chr12:105229674-105232010	4	Ppp2r5c	1562	5' proximal	coding	0	68	102	0.00	GeneID:26931	Protein phosphatase
chr12.100895580	chr12:100895580-100897680	4	Vrk1	99678	3' distal	coding	0	0	0	#DIV/0!	GeneID:22367	Protein kinase
chr11.98630851	chr11:98630851-98633649	4	Rara	5976	intragenic (intron)	coding	28	4	4	7.00	GeneID:19401	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr11.98168423	chr11:98168423-98170467	4	AF001293	8521	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr11.98130027	chr11:98130027-98130834	4	Grb7	4102	3' proximal	coding	0	0	0	#DIV/0!	GeneID:14786	Transmembrane receptor regulatory/adaptor protein
chr11.97145676	chr11:97145676-97147876	4	Arhgap23	14282	5' distal	coding	5	8	0	0.63	GeneID:58996	Molecular function unclassified
chr11.96451262	chr11:96451262-96451871	4	Snx11	2925	5' proximal	coding	0	0	0	#DIV/0!	GeneID:74479	Molecular function unclassified
chr11.96362962	chr11:96362962-96364475	4	BB635135 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr11.95972792	chr11:95972792-95974458	4	Hoxb6	926	3' proximal	coding	0	0	0	#DIV/0!	GeneID:15414	Homeobox transcription factor;Nucleic acid binding
chr11.95810312	chr11:95810312-95811029	4	4932418K24Rik	77	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:237930	Microtubule family cytoskeletal protein;Other ligase
chr11.94902470	chr11:94902470-94902906	4	Myst2	1369	5' proximal	coding	132	16	109	8.25	GeneID:217127	Zinc finger transcription factor;Chromatin/chromatin-binding protein
chr11.94192456	chr11:94192456-94192979	4	Mycbpap	114	5' proximal	coding	0	0	0	#DIV/0!	GeneID:104601	Molecular function unclassified
chr11.93880864	chr11:93880864-93882381	4	Tob1	1272	5' proximal	coding	0	0	2	#DIV/0!	GeneID:22057	Transcription cofactor;Other miscellaneous function protein
chr11.93834126	chr11:93834126-93834878	4	CF198268 EST	2500	5' proximal	noncoding	-	-	-	#VALUE!	NA	NA
chr11.8649164	chr11:8649164-8649855	4	AK089717	90307	5' distal	coding	0	0	9	#DIV/0!	NA	NA
chr11.80747158	chr11:80747158-80748173	4	Accn1	15994	intragenic (intron)	coding	211	274	56	0.77	GeneID:11418	Other ion channel
chr11.8041569	chr11:8041569-8043431	4	AK045941	2569		target is repeat	-	-	-	#VALUE!	NA	NA
chr11.79793231	chr11:79793231-79794387	4	Centa2	2425	3' proximal	coding	0	0	0	#DIV/0!	GeneID:216991	Membrane-bound signaling molecule;Other G-protein modulator

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr11.77560189	chr11:77560189-77561113	4	Sez6	5188	intragenic (intron)	coding	0	0	7	#DIV/0!	GeneID:20370	Other receptor
chr11.77377456	chr11:77377456-77379046	4	Myo18a	11236	5' distal	coding	0	10	29	0.00	GeneID:360013	Actin binding motor protein
chr11.7428008	chr11:7428008-7429091	4	Igfbp3	319754	gene desert	coding	0	0	0	#DIV/0!	GeneID:16009	Other miscellaneous function protein
chr11.69315025	chr11:69315025-69316992	4	Tnfsf12	8173	5' proximal	coding	6	0	3	#DIV/0!	GeneID:21944	Tumor necrosis factor family member
chr11.68012145	chr11:68012145-68013972	4	BY729287 EST	10	5' proximal	coding	-	-	-	#VALUE!	NA	NA
chr11.66631717	chr11:66631717-66633721	4	A730055C05Rik	4282	5' proximal	coding	0	0	0	#DIV/0!	GeneID:338369	Molecular function unclassified
chr11.66180785	chr11:66180785-66183057	4	A530088H08Rik	343393	gene desert	coding	0	0	0	#DIV/0!	GeneID:193003	Molecular function unclassified
chr11.63067140	chr11:63067140-63069089	4	S74315	12470		target is repeat	-	-	-	#VALUE!	NA	NA
chr11.6283601	chr11:6283601-6284227	4	D11Bwg0280e	63	5' proximal	coding	69	11	15	6.27	GeneID:52915	Other ligase
chr11.62123415	chr11:62123415-62123985	4	Pigl	1123	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:327942	Glycosyltransferase;Esterase
chr11.55596971	chr11:55596971-55598153	4	Nmur2	67567	3' distal	coding	96	103	6	#REF!	GeneID:216749	G-protein coupled receptor
chr11.51925806	chr11:51925806-51927542	4	Tcf7	4227	5' proximal	coding	54	4	0	13.50	GeneID:21414	HMG box transcription factor;Nucleic acid binding
chr11.5164650	chr11:5164650-5165381	4	Kremen1	8525	5' proximal	coding	4	8	7	0.50	GeneID:84035	Serine protease
chr11.51343022	chr11:51343022-51343545	4	Sec24a	806	intragenic (intron)	coding	0	46	0	0.00	GeneID:77371	Molecular function unclassified
chr11.50262276	chr11:50262276-50264703	4	Adamts2	20417	intragenic (intron)	coding	0	18	28	0.00	GeneID:216725	Metalloprotease
chr11.50131654	chr11:50131654-50132947	4	Rufy1	61621	5' distal	coding	0	18	28	0.00	GeneID:216724	Molecular function unclassified
chr11.50096615	chr11:50096615-50098493	4	Rufy1	26715	5' distal	coding	0	18	28	0.00	GeneID:216724	Molecular function unclassified
chr11.45710567	chr11:45710567-45712711	4	Adam19	6842	intragenic (intron)	coding	5	3	34	1.67	GeneID:11492	Metalloprotease
chr11.44503508	chr11:44503508-44504381	4	Ebf1	4928	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:13591	Other transcription factor;Nucleic acid binding
chr11.44481064	chr11:44481064-44482395	4	Ebf1	27180	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:13591	Other transcription factor;Nucleic acid binding
chr11.42290064	chr11:42290064-42291382	4	C030002O17Rik	18235	3' distal	coding	0	0	0	#DIV/0!	GeneID:78533	Molecular function unclassified
chr11.4140743	chr11:4140743-4142422	4	Osm	5402	3' proximal	coding	51	9	0	5.67	GeneID:18413	Interleukin
chr11.36234073	chr11:36234073-36234994	4	Odz2	168191	intragenic (intron)	coding	58	99	14	0.59	GeneID:23964	Other receptor;Membrane-bound signaling molecule
chr11.35007030	chr11:35007030-35007876	4	AF144629	91457	intragenic (intron)	coding	0	30	0	0.00	NA	NA
chr11.3232894	chr11:3232894-3234398	4	1500004A08Rik	2879	intragenic (intron)	coding	0	27	58	0.00	GeneID:216505	Molecular function unclassified
chr11.30077927	chr11:30077927-30082794	4	Spnb2	23808	intragenic (intron)	coding	27	258	47	0.10	GeneID:20742	Non-motor actin binding protein
chr11.29492416	chr11:29492416-29493435	4	AK006591	18917	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr11.22503339	chr11:22503339-22504106	4	Tmem17	88744	3' distal	coding	0	10	0	0.00	GeneID:103765	Molecular function unclassified
chr11.20372443	chr11:20372443-20373617	4	Sertad2	65672	5' distal	coding	34	0	14	#DIV/0!	GeneID:58172	Molecular function unclassified
chr11.20295974	chr11:20295974-20296614	4	AA066037 EST		intragenic (intron)	noncoding	0	0	0	#DIV/0!	NA	NA
chr11.16486769	chr11:16486769-16487671	4	Sec61g	83304	5' distal	coding	375	0	180	#DIV/0!	GeneID:20335	Other membrane traffic protein
chr11.120497445	chr11:120497445-120498227	4	4933434G05Rik	0	intragenic (exon, 3'utr)	coding	12	0	0	#DIV/0!	GeneID:71276	Molecular function unclassified
chr11.119315970	chr11:119315970-119316651	4	4932417H02Rik	12422	intragenic (intron)	coding	49	0	0	#DIV/0!	GeneID:74370	Molecular function unclassified
chr11.116256556	chr11:116256556-116257938	4	AK129431	5174	5' proximal	coding	1	17	12	0.06	NA	NA
chr11.115951203	chr11:115951203-115953595	4	Galr2	242	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:14428	G-protein coupled receptor
chr11.115682618	chr11:115682618-115683450	4	Galk1	537	intragenic (intron)	coding	545	880	379	0.62	GeneID:14635	Carbohydrate kinase
chr11.115104997	chr11:115104997-115105744	4	Kctd2	2801	3' proximal	coding	2	1	40	2.00	GeneID:70382	Voltage-gated potassium channel
chr11.114941280	chr11:114941280-114942421	4	Fdxr	182	intragenic (intron)	coding	0	0	13	#DIV/0!	GeneID:14149	Reductase
chr11.107601082	chr11:107601082-107602741	4	Prkca	6839	3' proximal	coding	0	0	0	#DIV/0!	GeneID:18750	Transfer/carrier protein;Non-receptor serine/threonine protein kinase
chr11.107104154	chr11:107104154-107107304	4	1110020B03Rik	98133	intragenic (intron)	coding	52	4	167	13.00	mCG3303	Other transporter;Other transfer/carrier protein
chr11.107098342	chr11:107098342-107099889	4	1110020B03Rik	41759	intragenic (intron)	coding	52	4	167	13.00	mCG3303	Other transporter;Other transfer/carrier protein
chr11.106154559	chr11:106154559-106158756	4	Ern1	3514	intragenic (intron)	coding	26	70	22	0.37	GeneID:78943	Protein kinase
chr11.104164534	chr11:104164534-104166961	4	BC023187	24885	5' distal	coding	5	98	33	0.05	NA	NA
chr11.101826452	chr11:101826452-101827698	4	2610511E22Rik	76	5' proximal	coding	12	17	63	0.71	GeneID:76547	Molecular function unclassified
chr10.95129289	chr10:95129289-95130097	4	Socs2	31206	3' distal	coding	0	0	0	#DIV/0!	GeneID:216233	Other signaling molecule
chr10.94995317	chr10:94995317-94996185	4	Cradd	44848	intragenic (intron)	coding	19	0	0	#DIV/0!	GeneID:12905	Other miscellaneous function protein
chr10.90475040	chr10:90475040-90476816	4	AK049899	28496	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr10.84421887	chr10:84421887-84423636	4	Rfx4 or AK034131	762	5' proximal	coding	0	0	72	#DIV/0!	GeneID:71137	Other transcription factor
chr10.81637738	chr10:81637738-81638644	4	AK041588		intragenic (exon)	noncoding	0	0	0	#DIV/0!	NA	NA
chr10.81066428	chr10:81066428-81071722	4	Pias4	2083	intragenic (intron)	coding	42	10	83	4.20	GeneID:59004	Transcription cofactor;Other ligase
chr10.77979893	chr10:77979893-77980619	4	Icosl	17726	3' distal	coding	0	5	0	0.00	NA	NA
chr10.77081744	chr10:77081744-77084829	4	Col18a1	39667	5' distal	coding	164	112	0	1.46	GeneID:12822	Cell adhesion molecule;Extracellular matrix structural protein
chr10.75873070	chr10:75873070-75874776	4	LOC333669	4479	3' proximal	coding	57	0	0	#DIV/0!	GeneID:333669	Cation transporter;Carbohydrate transporter
chr10.71088211	chr10:71088211-71088762	4	D10Ertd214e	19670	3' distal	coding	0	0	0	#DIV/0!	GeneID:52637	Molecular function unclassified
chr10.70132437	chr10:70132437-70134874	4	Slc16a9	73754	3' distal	coding	0	0	0	#DIV/0!	GeneID:66859	Molecular function unclassified
chr10.68568369	chr10:68568369-68569261	4	Tmem26	15139	3' distal	coding	0	0	0	#DIV/0!	GeneID:327766	Molecular function unclassified
chr10.66943875	chr10:66943875-66945520	4	AK085500	1649	intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr10.63130091	chr10:63130091-63130895	4	BE985149 EST		intragenic (intron)	noncoding	-	-	-	#VALUE!	NA	NA
chr10.61361447	chr10:61361447-61362295	4	AK015804	7000	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr10.61142891	chr10:61142891-61144437	4	Pald	314	intragenic (intron)	coding	52	29	10	1.79	mCG122064	Molecular function unclassified
chr10.6058920	chr10:6058920-6060410	4	Akap12	23572	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:83397	Kinase modulator
chr10.59918076	chr10:59918076-59920184	4	Chst3	15250	3' distal	coding	0	0	9	#DIV/0!	GeneID:53374	Other transferase
chr10.59334836	chr10:59334836-59335896	4	AK080263	33651	5' distal	noncoding	0	5	69	0.00	NA	NA
chr10.58280557	chr10:58280557-58281333	4	BC042726	5911	5' proximal	coding	0	0	0	#DIV/0!	NA	NA

Cluster ID	Cluster Location	Over-lap Size	Candidate Oct4 target gene	Distance	Binding site location	coding/ noncoding	MPSS (ES)	MPSS (EB)	MPSS (NS)	ES/EB ratio	Panther_ID	Molecular.function
chr10.52314387	chr10:52314387-52315178	4	Dcbld1	9969	intragenic (intron)	coding	3	0	0	#DIV/0!	GeneID:66686	Other receptor
chr10.4401029	chr10:4401029-4402407	4	Rgs17	31746	5' distal	coding	21	0	0	#DIV/0!	GeneID:56533	Other G-protein modulator
chr10.43459853	chr10:43459853-43460990	4	1700021F05Rik	10717	5' distal	coding	4	41	37	0.10	GeneID:67851	Molecular function unclassified
chr10.30418102	chr10:30418102-30419581	4	U58494			target is repeat	-	-	-	#VALUE!	NA	NA
chr10.29383770	chr10:29383770-29384488	4	Thsd2	70000	5' distal	coding	0	0	0	#DIV/0!	GeneID:72780	Extracellular matrix glycoprotein
chr10.26110502	chr10:26110502-26112267	4	L3mbtl3	0	intragenic (intron)	coding	2	5	6	0.40	GeneID:237339	Other transcription factor;Chromatin/chromatin-binding protein
chr10.21565789	chr10:21565789-21567143	4	AK049156	1500	5' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr10.19853735	chr10:19853735-19854546	4	Mtap7	4573	intragenic (intron)	coding	19	0	11	#DIV/0!	GeneID:17761	Non-motor microtubule binding protein
chr10.129037067	chr10:129037067-129037697	4	Cd63	183	5' proximal	coding	121	471	1631	0.26	GeneID:12512	Other signaling molecule
chr10.128688361	chr10:128688361-128691623	4	Pa2g4	336	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18813	Other transcription factor;Other nucleic acid binding
chr10.12622726	chr10:12622726-12623703	4	AK045934	4500	3' proximal	noncoding	0	0	0	#DIV/0!	NA	NA
chr10.125728333	chr10:125728333-125729304	4	Lrig3	25838	5' distal	coding	0	0	0	#DIV/0!	GeneID:320398	Other cell adhesion molecule
chr10.119859217	chr10:119859217-119861145	4	Irak3	1071	intragenic (intron)	coding	0	6	2	0.00	GeneID:73914	Serine/threonine protein kinase receptor;Protein kinase
chr10.106192262	chr10:106192262-106193841	4	Ppfia2	38751	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:327814	Receptor;Nuclease
chr1.97976722	chr1:97976722-97977747	4	Pam	8929	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:18484	Oxygenase
chr1.92832839	chr1:92832839-92835465	4	Ankmy1	769	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:241158	Molecular function unclassified
chr1.89848268	chr1:89848268-89851930	4	Asb18	9915	3' proximal	coding	30	0	0	#DIV/0!	GeneID:208372	Other signaling molecule
chr1.89396571	chr1:89396571-89397468	4	Centg2	32967	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:347722	G-protein;Other G-protein modulator;Other enzyme activator
chr1.88524544	chr1:88524544-88525270	4	A630084M22Rik	85475	3' distal	coding	0	0	0	#DIV/0!	mCG49726	Small GTPase
chr1.88513900	chr1:88513900-88514779	4	AK030268	40000	3' distal	noncoding	0	0	0	#DIV/0!	NA	NA
chr1.86234394	chr1:86234394-86236317	4	Ncl	82	3' proximal	coding	116	65	74	1.78	GeneID:17975	Ribonucleoprotein
chr1.86193435	chr1:86193435-86195761	4	B3gnt7	1925	3' proximal	coding	66	2	0	33.00	GeneID:227327	Glycosyltransferase
chr1.85380430	chr1:85380430-85380776	4	Fbxo36	12085	intragenic (intron)	coding	0	0	20	#DIV/0!	GeneID:66153	Molecular function unclassified
chr1.77915587	chr1:77915587-77918725	4	Epha4	16873	intragenic (intron)	coding	0	0	89	#DIV/0!	GeneID:13838	Tyrosine protein kinase receptor;Protein kinase
chr1.65284295	chr1:65284295-65285562	4	9430067K14Rik	1983	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:241075	Molecular function unclassified
chr1.57757670	chr1:57757670-57758769	4	AK006906	8509	5' proximal	coding	0	0	0	#DIV/0!	NA	NA
chr1.54139383	chr1:54139383-54140107	4	Stk17b	4200	3' proximal	coding	0	0	0	#DIV/0!	GeneID:98267	Non-receptor serine/threonine protein kinase
chr1.51662616	chr1:51662616-51665272	4	Sdpr	1034	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:20324	Other transcription factor
chr1.40246490	chr1:40246490-40247991	4	Map4k4	21738	intragenic (intron)	coding	23	8	101	2.88	GeneID:26921	Protein kinase
chr1.39752609	chr1:39752609-39757849	4	Tbc1d8	4281	intragenic (intron)	coding	0	1	9	0.00	GeneID:54610	Other G-protein modulator
chr1.34847596	chr1:34847596-34848963	4	Tesp2	568	5' proximal	coding	0	4	0	0.00	GeneID:21756	Serine protease
chr1.194509514	chr1:194509514-194510828	4	Plxna2	3667	5' proximal	coding	0	0	23	#DIV/0!	GeneID:18845	Tyrosine protein kinase receptor;Protein kinase
chr1.180374532	chr1:180374532-180377112	4	MGC68323	26111	3' distal	coding	0	0	0	#DIV/0!	GeneID:277333	Dehydrogenase
chr1.178869653	chr1:178869653-178870786	4	BC068141		intragenic (intron)	coding	0	0	0	#DIV/0!	NA	NA
chr1.177851609	chr1:177851609-177852974	4	AK006014	10982	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr1.170971580	chr1:170971580-170972638	4	Dusp12	9197	5' proximal	coding	7	7	16	1.00	GeneID:80915	Protein phosphatase
chr1.167306094	chr1:167306094-167307085	4	Uck2	5083	intragenic (intron)	coding	53	49	52	1.08	GeneID:80914	Nucleotide kinase
chr1.166149913	chr1:166149913-166151341	4	Gpa33	2239	intragenic (intron)	coding	26	0	0	#DIV/0!	GeneID:59290	Receptor
chr1.164019990	chr1:164019990-164024272	4	2810422O20Rik	7500	3' proximal	coding	12	0	0	#DIV/0!	GeneID:69962	Molecular function unclassified
chr1.16294859	chr1:16294859-16295736	4	Rdh10	856	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:98711	Dehydrogenase
chr1.159821278	chr1:159821278-159822242	4	Tnr	36952	intragenic (intron)	coding	0	0	0	#DIV/0!	GeneID:21960	Extracellular matrix glycoprotein
chr1.156979483	chr1:156979483-156980273	4	Ralgps2	40956	5' proximal	coding	5	0	0	#DIV/0!	GeneID:78255	Guanyl-nucleotide exchange factor
chr1.153069030	chr1:153069030-153069857	4	Nmnat2	1379	intragenic (intron)	coding	9	1	68	9.00	GeneID:226518	Nucleotidyltransferase
chr1.14610916	chr1:14610916-14614442	4	Eya1	149065	gene desert	coding	1591	4424	2899	0.36	GeneID:14048	Hydrolase
chr1.140368041	chr1:140368041-140368930	4	AK054538	12500	5' distal	coding	0	0	0	#DIV/0!	NA	NA
chr1.140186303	chr1:140186303-140188164	4	Cfh	118167	gene desert	coding	0	0	0	#DIV/0!	GeneID:12628	Complement component
chr1.134422735	chr1:134422735-134425012	4	4931440L10Rik	5662	3' proximal	coding	0	0	0	#DIV/0!	GeneID:71001	Glycosyltransferase
chr1.133555268	chr1:133555268-133556435	4	AK035705	6806	5' proximal	coding	3	1	31	3.00	NA	NA
chr1.132731985	chr1:132731985-132733475	4	Lrrn2	29374	5' distal	coding	0	0	28	#DIV/0!	GeneID:16980	Other receptor
chr1.132069458	chr1:132069458-132070148	4	BC066109		intragenic (intron)	coding	0	31	0	0.00	NA	NA
chr1.131761571	chr1:131761571-131762428	4	Rab7l1	18296	3' distal	coding	0	0	0	#DIV/0!	GeneID:226422	Small GTPase
chr1.125364189	chr1:125364189-125365452	4	Slc35f5	27309	5' distal	coding	5	0	3	#DIV/0!	GeneID:74150	Molecular function unclassified
chr1.122401112	chr1:122401112-122401615	4	AK045941	526553	gene desert	target is repeat	-	-	-	#VALUE!	NA	NA
chr1.120415349	chr1:120415349-120418076	4	AK011565	19118	intragenic (intron)	noncoding	0	0	16	#DIV/0!	NA	NA
chr1.120127402	chr1:120127402-120127888	4	C1ql2	11520	3' distal	coding	0	0	16	#DIV/0!	GeneID:226359	Complement component
chr1.119262249	chr1:119262249-119264199	4	Ralb	6964	intragenic (intron)	coding	37	20	61	1.85	GeneID:64143	Small GTPase
chr1.119153875	chr1:119153875-119155033	4	BC048845	26182	3' distal	coding	0	0	0	#DIV/0!	NA	NA
chr1.11356527	chr1:11356527-11357886	4	BY719998 EST		intragenic (intron)	coding	-	-	-	#VALUE!	NA	NA
chr1.108551207	chr1:108551207-108552165	4	AK045941	25579		target is repeat	-	-	-	#VALUE!	NA	NA

Appendix B: Coordinates of 1133 Sox2 binding loci and their associated genes.

This table shows a list of 1133 Sox2 bound loci, sorted by the numbers of overlapping PETs at each locus. Cluster ID is a unique ID that was generated for each Oct4 binding locus. MPSS (ES) column indicates the number of unique MPSS tag associated with each candidate Oct4 bound gene in ESCs. MPSS (EB) and MPSS (NS) columns show the quantification of MPSS tags in embryoid bodies and neurospheres respectively. The Molecular Function based on Panther classification is shown.

Cluster ID	Overlap Location	Overlap Size	Candidate Sox2 Target Gene	Distance	Binding Site Location	MPSS (mES)	MPSS (mEp)	MPSS (mEB)	MPSS (mNS)	Gene ID (Panther)	Panther Molecular Function
chr5.26608950	chr5:26609314-26609347	26	9630008K15Rik	45	Intragenic (intron)	9	9	0	0	mCG56335	Molecular function unclassified
chr17.38290245	chr17:38292632-38292634	22	Pigt	0	Intragenic (exon)	244	256	37	948	GeneID:78928	Molecular function unclassified
chr3.122526608	chr3:122527192-122527196	21	Abca4	1605	Intragenic (intron)	0	0	0	0	mCG9528	ATP-binding cassette (ABC) transporter
chr12.30674723	chr12:30675763-30675780	17	D38417	109261	Gene desert	0	0	0	0	NA	NA
chr2.52031545	chr2:52032209-52032249	16	Rif1	592	5' end	142	22	159	25	GeneID:51869	Molecular function unclassified
chr3.103137726	chr3:103138295-103138328	16	AK010586	9113	Intragenic (intron)	0	0	0	0	NA	NA
chr2.44485050	chr2:44485507-44485516	15	Gtdc1	38891	3' end	14	4	8	20	GeneID:227835	Molecular function unclassified
chr7.128648627	chr7:128649291-128649321	15	Ifitm2	3643	5' end	80	0	0	0	mCG22584	Other miscellaneous function protein
chr2.25140065	chr2:25140684-25140735	15	Nrarp	69	5' end	0	140	0	54	mCG56531	Cytoskeletal protein
chr14.52951673	chr14:52952232-52952246	14	Tnfrsf19	5815	3' end	0	0	0	25	GeneID:29820	Tumor necrosis factor receptor
chr8.77187079	chr8:77187515-77187531	14	1700031F13Rik	138320	Gene desert	0	0	0	0	GeneID:73301	Molecular function unclassified
chr5.75997963	chr5:75998501-75998574	14	Rest	3121	5' end	100	63	11	24	mCG15860	Molecular function unclassified
chr14.78241487	chr14:78241911-78241918	14	Diap3	259060	Gene desert	2	11	3	6	mCG8502	Endoribonuclease;Non-motor actin binding protein;Esterase
chr2.106644030	chr2:106644297-106644314	14	AK012553	6163	Intragenic (intron)	0	8	0	18	NA	NA
chr10.21467771	chr10:21468375-21468400	13	LOC432436	52233	3' end	0	0	0	0	GeneID:432436	Molecular function unclassified
chr19.28305316	chr19:28305808-28305879	13	Cdc37l1	527	Intragenic (intron)	1	39	0	3	GeneID:67072	Other chaperones;Kinase activator
chr12.93914650	chr12:93915331-93915340	13	Ches1	8720	Intragenic (intron)	35	34	0	85	GeneID:71375	Transcription factor;Nucleic acid binding
chr4.34055087	chr4:34055388-34055443	13	Cnr1	119556	Gene desert	0	0	0	0	mCG12569	G-protein coupled receptor
chr16.16510497	chr16:16510711-16510743	13	Spag6	17680	3' end	0	0	0	0	mCG131165	Molecular function unclassified
chr8.88317221	chr8:88317815-88317831	13	Sall1	24834	3' end	0	0	0	11	mCG13637	Zinc finger transcription factor;Nucleic acid binding
chr12.75414852	chr12:75415631-75415657	13	Zfp36l1	4051	5' end	0	0	0	0	mCG19208	Nuclease
chr3.108813280	chr3:108813662-108813716	13	Sars1	393	Intragenic (intron)	254	289	84	192	mCG4072	Other RNA-binding protein;Aminoacyl-tRNA synthetase;Ligase
chr14.16503021	chr14:16504153-16504198	13	Kcnk5	39113	5' end	0	0	0	0	mCG5936	Other ion channel
chr17.46788660	chr17:46789017-46789046	13	AK039924	75295	3' end	0	0	0	0	NA	NA
chr7.41564851	chr7:41569521-41569551	12	Mrps33	306885	Gene desert	29	64	132	218	GeneID:14548	Molecular function unclassified
chr5.15331286	chr5:15331866-15331880	12	Hgf	351575	Gene desert (3' end)	0	0	0	0	GeneID:15234	Growth factor;Serine protease;Other select calcium binding proteins;Annexin
chr1.133374262	chr1:133374582-133374606	12	Sox13	68727	5' end	25	34	16	4	GeneID:20668	HMG box transcription factor;Nucleic acid binding
chr12.81827074	chr12:81827836-81827843	12	Esrrb	4199	Intragenic (intron)	281	211	24	0	GeneID:26380	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr1.37308324	chr1:37308676-37308739	12	D1Bwg0491e	46847	5' end	0	0	0	0	GeneID:56030	Molecular function unclassified
chr13.91432975	chr13:91433499-91433499	12	Tbca	82181	5' end	0	0	0	0	mCG122050	Molecular function unclassified
chr10.116352560	chr10:116352993-116353009	12	Cnot2	13269	Intragenic (intron)	82	14	9	55	mCG17335	Other transcription factor
chr17.6217770	chr17:6218320-6218330	12	Tulp4	27068	3' end	0	0	0	0	mCG21746	Other transcription factor
chr4.121994344	chr4:121994947-121995017	12	AF150755	59301	Intragenic (intron)	5	9	5	10	NA	NA
chr19.6973672	chr19:6975037-6975082	11	Rcor2	7028	5' end	212	157	91	47	GeneID:104383	Molecular function unclassified
chr17.36537089	chr17:36537608-36537664	11	Iap	29776	Intragenic (intron)	0	0	0	0	GeneID:15598	Nucleic acid binding;Protease;Other viral protein
chr14.3088563	chr14:3089037-3089080	11	2610042L04Rik	1011	5' end	17	76	240	0	mCG112781	Molecular function unclassified
chr5.105518255	chr5:105518955-105518955	11	Mtf2	3599	Intragenic (intron)	103	209	81	43	mCG13604	Other zinc finger transcription factor;Nucleic acid binding
chr9.120134520	chr9:120134875-120134875	11	Mobp	209	5' end	0	0	0	0	mCG14123	Actin binding motor protein;Non-motor actin binding protein;Other G-protein modulator
chr8.115771859	chr8:115772152-115772211	11	Dncl2b	117250	Gene desert	0	0	0	0	mCG4360	Other microtubule family cytoskeletal protein
chr3.97159886	chr3:97160418-97160453	11	AK089736	0	Intragenic (exon)	0	0	0	0	NA	NA
chr13.75032068	chr13:75032389-75032471	11	AK077067	108727	Gene desert	0	0	0	0	NA	NA
chr16.26804748	chr16:26805716-26805722	10	Il1rap	244988	Gene desert (3' end)	0	0	0	80	GeneID:16180	Interleukin receptor
chr17.64899009	chr17:64899879-64899912	10	Rab12	14556	3' end	7	62	3	193	GeneID:19328	Molecular function unclassified
chr13.63831171	chr13:63831584-63831658	10	C330014B19Rik	2738	Intragenic (intron)	0	0	0	0	GeneID:382770	G-protein coupled receptor
chr16.13719357	chr16:13720318-13720432	10	Ifitm7	4951	5' end	0	0	0	5	GeneID:74482	Other miscellaneous function protein
chr15.77682383	chr15:77682680-77682787	10	9130022K13Rik	32715	3' end	0	0	0	0	GeneID:75761	Transporter;Apolipoprotein
chr8.47912045	chr8:47912451-47912518	10	Odz3	77260	5' end	0	0	0	16	mCG114540	Other receptor;Membrane-bound signaling molecule
chr3.92909305	chr3:92909668-92909716	10	1110031B11Rik	11044	Intragenic (intron)	0	0	0	0	mCG119410	Molecular function unclassified
chr8.85677496	chr8:85677709-85677877	10	Abcc12	110913	Gene desert	0	0	0	0	mCG123087	ATP-binding cassette (ABC) transporter
chr6.92597400	chr6:92598254-92598288	10	Trh	1926	5' end	757	154	428	0	mCG127886	Neuropeptide
chr17.34013042	chr17:34013517-34013562	10	Pou5f1	1998	5' end	388	246	21	0	mCG19893	Homeobox transcription factor;Nucleic acid binding
chr19.22367211	chr19:22367571-22367646	10	Bteb1	46859	5' end	336	520	5	67	mCG5314	KRAB box transcription factor
chr3.102012767	chr3:102013755-102013789	10	Casq2	2491	5' end	0	0	0	0	mCG5476	Other select calcium binding proteins
chr4.89777696	chr4:89778341-89778487	10	Elavl2	48932	5' end	17	0	0	9	mCG7069	Ribonucleoprotein
chr1.86193362	chr1:86194738-86194750	10	B3gnt7	2090	3' end	107	337	104	0	mCG8506	Glycosyltransferase
chr11.25999058	chr11:25999446-25999462	10	AK080999	106602	Intragenic (intron)	0	0	0	0	NA	NA
chr13.109400491	chr13:109400990-109401039	9	Il6st	10044	5' end	0	20	2	0	GeneID:16195	Interleukin receptor
chr17.12346404	chr17:12346898-12346944	9	Tcte2	53916	3' end	0	0	0	0	GeneID:21646	Molecular function unclassified
chr11.20126578	chr11:20126800-20126877	9	BC027174	2212	3' end	1	22	35	8	GeneID:216543	Molecular function unclassified
chr11.55586046	chr11:55586566-55586566	9	Nmur2	78165	3' end	0	0	0	0	GeneID:216749	G-protein coupled receptor
chr7.60442615	chr7:60442912-60442986	9	Rgma	18832	Intragenic (intron)	0	0	0	193	GeneID:244058	Molecular function unclassified
chr5.37259772	chr5:37260192-37260207	9	Clnk	8291	3' end	0	0	0	0	GeneID:27278	Other signaling molecule

chr1.72648937	chr1:72652467-72652558	9	Wdt2	16826	5' end	49	194	251	0	GeneID:381269	Molecular function unclassified
chr1.42382372	chr1:42383057-42383061	9	2610017I09Rik	590896	Gene desert (3' end)	0	0	0	4	GeneID:66297	Molecular function unclassified
chr11.77506307	chr11:77506606-77506657	9	Pipox	242	5' end	10	22	0	96	mCG10817	Oxidase
chr16.17705704	chr16:17705983-17706046	9	Slc25a1	929	5' end	0	0	0	0	mCG131850	Mitochondrial carrier protein
chr11.75962001	chr11:75963948-75963975	9	Nxn	47293	Intragenic (intron)	56	68	91	38	mCG140720	Other oxidoreductase
chr12.82652548	chr12:82652885-82652913	9	Sptlc2	3967	Intragenic (intron)	120	46	35	35	mCG17474	Other transferase
chr1.128227822	chr1:128228168-128228180	9	Dars	1156	Intragenic (intron)	0	0	0	0	mCG20044	Other RNA-binding protein;Aminoacyl-tRNA synthetase
chr7.116935250	chr7:116935689-116935791	9	Wdr11	162149	Gene desert	6	19	62	24	mCG22163	Molecular function unclassified
chr17.14022474	chr17:14023059-14023136	9	2410011O22Rik	118347	Gene desert	0	0	0	0	mCG4507	Molecular function unclassified
chr19.4598365	chr19:4598855-4598962	9	Rbm14	732	5' end	90	167	36	19	mCG8382	Ribonucleoprotein
chr15.89096542	chr15:89096729-89096817	9	BC049267	6862	5' end	57	170	16	10	NA	NA
chr3.136663510	chr3:136664255-136664338	9	AY178734	12089	Intragenic (intron)	0	0	0	0	NA	NA
chr3.40848399	chr3:40848587-40848725	9	BC044759	16847	5' end	81	76	7	82	NA	NA
chr4.145302951	chr4:145303443-145303497	9	BC058613	93223	Intragenic (intron)	137	1	4	0	NA	NA
chr17.5000553	chr17:5000760-5000769	9	AK129314	292665	Gene desert	7	67	52	31	NA	NA
chr15.84976653	chr15:84977824-84977850	8	Arhgap8	41013	5' end	1	65	39	0	GeneID:109270	Other G-protein modulator
chr15.39620461	chr15:39620717-39620725	8	Rims2	27768	Intragenic (intron)	0	0	0	0	GeneID:116838	Other G-protein modulator
chr1.181001034	chr1:181001542-181001577	8	Lefty1	10307	5' end	194	125	4	0	GeneID:13590	TGF-beta superfamily member
chr6.122971708	chr6:122972081-122972148	8	Phc1	905	5' end	334	209	118	19	GeneID:13619	Other transcription factor;Chromatin/chromatin-binding protein
chr18.39346746	chr18:39347025-39347058	8	Fgf1	14454	Intragenic (intron)	0	0	0	8	GeneID:14164	Growth factor
chr9.58486567	chr9:58487251-58487286	8	Loxl1	8552	3' end	0	50	42	0	GeneID:16949	Oxidase;Other extracellular matrix
chr15.75388334	chr15:75388830-75388838	8	Ly6c	41749	5' end	0	38	0	0	GeneID:17067	Molecular function unclassified
chr4.80705447	chr4:80705917-80705997	8	Nfib	47451	Intragenic (intron)	0	0	0	166	GeneID:18028	Other transcription factor;Nucleic acid binding
chr6.84131245	chr6:84131589-84131656	8	Zfml	23836	5' end	0	0	0	0	GeneID:18139	Other DNA-binding protein
chr13.56467978	chr13:56468488-56468523	8	Spock1	66465	3' end	0	0	0	0	GeneID:20745	Cysteine protease inhibitor
chr18.14077235	chr18:14077527-14077559	8	Zfp521	2354	Intragenic (intron)	0	0	0	0	GeneID:225207	KRAB box transcription factor;Nucleic acid binding
chr2.165872765	chr2:165873157-165873173	8	Prkcbp1	2123	Intragenic (intron)	5	154	0	13	GeneID:228880	Other signaling molecule
chr7.19010589	chr7:19011263-19011312	8	4732429I09Rik	5268	3' end	0	0	0	15	GeneID:243906	KRAB box transcription factor;Nucleic acid binding
chr11.33422842	chr11:33423044-33423109	8	Ranbp17	13879	5' end	0	0	0	0	GeneID:66011	Other transfer/carrier protein
chr1.134424278	chr1:134424520-134424557	8	4931440L10Rik	5701	3' end	0	0	0	0	GeneID:71001	Glycosyltransferase
chr10.6015469	chr10:6015994-6016027	8	Akap12	14281	Intragenic (intron)	0	0	0	0	GeneID:83397	Kinase modulator
chr2.153231123	chr2:153231509-153231620	8	Tm9sf4	6604	Intragenic (intron)	91	56	4	103	GeneID:99237	Molecular function unclassified
chr1.73227003	chr1:73227292-73227400	8	Igfbp2	31549	5' end	0	0	0	0	mCG112979	Other miscellaneous function protein
chr8.47084848	chr8:47085091-47085111	8	D8Ertd594e	287	Intragenic (intron)	0	7	0	0	mCG115484	Molecular function unclassified
chr10.49426164	chr10:49426671-49426715	8	Grik2	2332	Intragenic (intron)	0	0	0	10	mCG120652	Glutamate receptor;Ion channel
chr1.13185209	chr1:13185862-13185972	8	Ncoa2	54104	3' end	6	6	0	11	mCG122320	Transcription cofactor
chr5.116518111	chr5:116518294-116518413	8	Thrap2	301892	Gene desert	13	58	76	0	mCG124529	Transcription cofactor
chr6.82847444	chr6:82847874-82847915	8	Pole4	34029	3' end	0	0	0	0	mCG126250	Other transcription factor;Nucleic acid binding
chr6.71156794	chr6:71157515-71157567	8	Eif2ak3	58892	3' end	0	0	5	2	mCG127277	Non-receptor serine/threonine protein kinase
chr6.100757895	chr6:100759171-100759174	8	Rybp	91524	5' end	44	110	20	0	mCG127676	Molecular function unclassified
chr6.95115217	chr6:95115673-95115690	8	Lrig1	50794	5' end	1	0	19	599	mCG127715	Other cell adhesion molecule
chr5.71539577	chr5:71540033-71540163	8	5033405K12Rik	10104	5' end	0	7	0	11	mCG128942	Molecular function unclassified
chr16.90599867	chr16:90600454-90600492	8	Tiam1	21875	5' end	0	0	0	0	mCG129085	Guanyl-nucleotide exchange factor
chr5.101811981	chr5:101812361-101812453	8	Spp1	2020	3' end	349	68	6	0	mCG14598	Other cytokine;Cell adhesion molecule;Defense/immunity protein;Other extracellular matrix
chr11.119608617	chr11:119609115-119609169	8	Baiap2	5077	5' end	12	29	0	38	mCG15032	Receptor;Kinase
chr11.117646332	chr11:117646753-117646784	8	Socs3	6477	5' end	16	70	12	6	mCG19293	Other miscellaneous function protein
chr17.35601690	chr17:35602080-35602131	8	Gabbr1	4089	3' end	3	1	0	18	mCG22614	G-protein coupled receptor
chr1.131761485	chr1:131762109-131762125	8	8430423A01Rik	20722	5' end	17	3	0	12	mCG2625	Molecular function unclassified
chr2.147045157	chr2:147045526-147045562	8	Pax1	68940	5' end	0	0	0	0	mCG4599	Homeobox transcription factor;Nucleic acid binding
chr3.30852381	chr3:30852709-30852719	8	Skil	13890	5' end	0	0	0	0	mCG4728	Other transcription factor
chr11.33432242	chr11:33433040-33433102	8	Gabrp	13279	3' end	0	0	0	0	mCG6676	GABA receptor;Ion channel
chr3.149805998	chr3:149806993-149807065	8	AK029623	1676	3' end	0	0	0	0	NA	NA
chr10.107966298	chr10:107967050-107967100	8	AK035230	7380	3' end	6	3	26	34	NA	NA
chr8.44105383	chr8:44105702-44105738	8	AJ250768	9411	Intragenic (intron)	0	0	0	39	NA	NA
chr16.22644480	chr16:22644665-22644727	8	AK039905	11670	3' end	20	0	0	49	NA	NA
chr17.82928272	chr17:82928665-82928707	8	BC052885	13370	Intragenic (intron)	0	0	0	0	NA	NA
chr5.124891398	chr5:124891851-124891953	8	AK028289	20013	Intragenic (intron)	0	7	0	0	NA	NA
chr17.79196685	chr17:79197150-79197262	8	BC005781	20902	5' end	0	10	50	21	NA	NA
chr2.20724638	chr2:20725151-20725259	8	AB093289	39949	Intragenic (intron)	6	45	3	8	NA	NA
chr5.129147573	chr5:129148255-129148309	8	BC021509	88671	5' end	0	4	0	17	NA	NA
chr2.168276963	chr2:168277324-168277411	7	Dpm1	1221	Intragenic (intron)	53	26	30	55	GeneID:13480	Transferase
chr18.34329714	chr18:34329967-34330063	7	Epb4.1l4a	772	Intragenic (intron)	0	0	0	0	GeneID:13824	Protein phosphatase
chr15.53383781	chr15:53384033-53384131	7	Ext1	729	5' end	139	144	13	71	GeneID:14042	Glycosyltransferase

chr13.44169744	chr13:44170220-44170276	7	Jarid2	25156	5' end	98	141	34	6	GeneID:16468	Other transcription factor;Nucleic acid binding
chr16.42655371	chr16:42656173-42656176	7	Ndufs5	175095	Gene desert	26	0	0	130	GeneID:170658	Oxidoreductase
chr12.13065824	chr12:13066750-13066753	7	Nmyc1	1964	5' end	197	263	20	2	GeneID:18109	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr13.54336445	chr13:54336760-54336870	7	Nsd1	0	Intragenic (exon)	78	17	16	27	GeneID:18193	
chr11.71114095	chr11:71114913-71114944	7	Nalp1	360006	Gene desert (5' end)	0	0	0	0	GeneID:195046	Other enzyme activator
chr17.80878515	chr17:80878739-80878895	7	Srst	499666	Gene desert (3' end)	0	0	0	0	GeneID:20819	Molecular function unclassified
chr14.29174001	chr14:29177784-29178027	7	Arhgap22	85527	5' end	0	0	0	0	GeneID:239027	Molecular function unclassified
chr4.56958633	chr4:56959155-56959221	7	Palm2	1545	Intragenic (intron)	0	0	0	0	GeneID:242481	Other miscellaneous function protein
chr18.75996189	chr18:75996856-75996866	7	Gm672	6855	Intragenic (intron)	0	0	0	127	GeneID:269037	Molecular function unclassified
chr7.94503259	chr7:94503675-94503701	7	LOC434225	5940	5' end	0	0	0	0	GeneID:434225	Ribonucleoprotein
chr11.79592951	chr11:79593236-79593375	7	Suz12	12355	5' end	120	0	103	4	GeneID:52615	Storage protein
chr10.20806202	chr10:20806559-20806577	7	Ahi1	21541	3' end	0	0	0	0	GeneID:52906	Molecular function unclassified
chr11.101352929	chr11:101353268-101353290	7	Arfl4	14324	3' end	0	0	0	0	GeneID:66182	Small GTPase
chr10.95230405	chr10:95231117-95231164	7	D10Ertd322e	24749	3' end	0	0	0	0	GeneID:67270	Molecular function unclassified
chr9.72083378	chr9:72084104-72084125	7	Cgnl1	8040	Intragenic (intron)	0	0	0	0	GeneID:68178	Molecular function unclassified
chr15.98044042	chr15:98044699-98044805	7	D15Ertd682e	80761	3' end	0	0	0	0	GeneID:71919	Molecular function unclassified
chr10.85204821	chr10:85204989-85205017	7	Btbd11	6814	Intragenic (intron)	0	0	0	0	GeneID:74007	Molecular function unclassified
chr10.85116766	chr10:85117183-85117188	7	Btbd11	63334	Intragenic (intron)	0	0	0	0	GeneID:74007	Molecular function unclassified
chr14.42123187	chr14:42123718-42123787	7	Slc35f4	78171	Intragenic (intron)	0	0	0	0	GeneID:75288	Molecular function unclassified
chr11.30386921	chr11:30387631-30387652	7	Acyp2	13916	3' end	0	0	0	0	GeneID:75572	Other phosphatase
chr18.38591013	chr18:38592046-38592050	7	2010005A06Rik	47802	3' end	0	49	0	0	GeneID:75599	Cadherin
chr8.86182116	chr8:86182339-86182455	7	BC004022	27323	5' end	164	74	3	15	GeneID:80750	
chr3.89062720	chr3:89063382-89063391	7	Ubqln4	2383	3' end	156	141	171	41	GeneID:94232	Other miscellaneous function protein
chr5.120689927	chr5:120690148-120690197	7	BC035291	343	Intragenic (intron)	23	14	0	6	mCG11130	Molecular function unclassified
chr2.33540563	chr2:33540795-33540839	7	Lmx1b	11544	Intragenic (intron)	0	0	0	0	mCG11256	Other zinc finger transcription factor
chr9.98993107	chr9:98993364-98993383	7	Pik3cb	13785	5' end	51	37	23	9	mCG113823	Other kinase
chr13.87540550	chr13:87541841-87541876	7	Rps23	53727	5' end	0	0	0	0	mCG113987	Ribosomal protein
chr2.159481603	chr2:159481967-159482046	7	Dhx35	560444	Gene desert (3' end)	6	7	0	30	mCG114772	RNA helicase;Hydrolase
chrX.152801576	chrX:152802080-152802166	7	4932441K18	2870	5' end	70	1	0	0	mCG116087	Molecular function unclassified
chr8.71709326	chr8:71709947-71710021	7	A230063L24Rik	2064	Intragenic (intron)	0	0	0	0	mCG118317	Molecular function unclassified
chr18.6485080	chr18:6485295-6485347	7	Kif5b	34194	5' end	0	0	0	8	mCG118617	Microtubule binding motor protein
chr7.29012246	chr7:29012460-29012491	7	2010426N06Rik	50094	5' end	3	0	0	11	mCG119772	KRAD box transcription factor;Nucleic acid binding
chr2.20551972	chr2:20552195-20552217	7	9430077C05Rik	13590	Intragenic (intron)	6	45	8	8	mCG120838	Molecular function unclassified
chr18.66469725	chr18:66470063-66470153	7	Rax	2200	5' end	0	0	0	0	mCG12626	Homeobox transcription factor;Nucleic acid binding
chr16.30510499	chr16:30510699-30510748	7	BC022623	43600	3' end	17	96	0	7	mCG126621	Molecular function unclassified
chr16.48328283	chr16:48328909-48328922	7	Morc	7191	5' end	9	33	0	0	mCG127772	Molecular function unclassified
chr16.85151669	chr16:85152005-85152034	7	Jam2	5064	5' end	0	22	0	84	mCG129050	Molecular function unclassified
chr11.61827342	chr11:61827769-61827790	7	Adora2b	33828	5' end	0	0	0	11	mCG129818	Molecular function unclassified
chr6.118269056	chr6:118270504-118270532	7	Zfp239	9099	3' end	0	24	2	24	mCG133008	KRAB box transcription factor;Nucleic acid binding
chr1.86410657	chr1:86410789-86410866	7	Ptma	105	5' end	0	0	0	0	mCG133549	Molecular function unclassified
chr11.102000553	chr11:102000811-102000903	7	Ubtf	10772	5' end	0	0	7	0	mCG140722	HMG box transcription factor;Nucleic acid binding
chr1.101655513	chr1:101656577-101656716	7	C230078M14	1277044	Gene desert (3' end)	0	0	0	0	mCG142065	Other cell adhesion molecule
chr17.29662864	chr17:29663145-29663155	7	Abcg1	368	5' end	0	0	0	0	mCG14574	Transporter
chr9.78765133	chr9:78765425-78765473	7	Dppa5	1304	5' end	1444	6569	1381	0	mCG15237	Molecular function unclassified
chr12.51489716	chr12:51490149-51490159	7	Mbip	30076	3' end	0	0	0	0	mCG15696	Kinase inhibitor
chr7.99961199	chr7:99962021-99962051	7	Tead1	11581	Intragenic (intron)	0	0	0	0	mCG15837	Other transcription factor;Nucleic acid binding
chr3.96078982	chr3:96079375-96079385	7	Vps45	26142	5' end	0	0	0	8	mCG16753	Membrane traffic regulatory protein
chr2.108734941	chr2:108735710-108735724	7	0610027B03Rik	248104	Gene desert	0	0	0	0	mCG16888	Methyltransferase
chr10.44441010	chr10:44441354-44441412	7	Prdm1	61732	5' end	0	0	0	0	mCG17867	KRAB box transcription factor;Nucleic acid binding
chr12.75307740	chr12:75309083-75309181	7	Zfp36l1	97523	3' end	0	0	0	0	mCG19208	Nuclease
chr18.5181777	chr18:5182696-5182805	7	Svil	24766	Intragenic (intron)	1	0	0	10	mCG19929	Non-motor actin binding protein
chr4.140594102	chr4:140594870-140594910	7	9030409G11Rik	20308	Intragenic (intron)	2	43	5	0	mCG19998	Molecular function unclassified
chr4.54720452	chr4:54720656-54720721	7	Klf4	51498	3' end	19	13	0	0	mCG20055	KRAB box transcription factor;Nucleic acid binding
chr17.54231274	chr17:54231546-54231586	7	Ebi3	496	5' end	0	0	0	0	mCG22978	Interleukin receptor
chr11.10992924	chr11:10993350-10993392	7	A930041G11Rik	16164	5' end	0	0	0	0	mCG3991	Molecular function unclassified
chr4.6895772	chr4:6896094-6896197	7	Tox	21390	Intragenic (intron)	0	0	0	11	mCG4373	HMG box transcription factor;Chromatin/chromatin-binding protein
chr9.17731507	chr9:17731856-17731885	7	Chordc1	408739	Gene desert (5' end)	79	48	4	22	mCG4776	Other signaling molecule
chr5.59533658	chr5:59534091-59534224	7	G6pd2	925092	Gene desert (5' end)	0	0	0	0	mCG50373	Dehydrogenase
chr10.18906638	chr10:18907410-18907510	7	Olig3	148485	Gene desert	0	42	0	0	mCG54581	Basic helix-loop-helix transcription factor
chr2.99452030	chr2:99452279-99452395	7	6430556C10Rik	1926689	Gene desert (3' end)	0	0	0	0	mCG6379	Other cell adhesion molecule
chr10.24791709	chr10:24792026-24792072	7	Arg1	101468	Gene desert	0	0	0	0	mCG9003	Other hydrolase
chr6.35136893	chr6:35137208-35137217	7	BC045524	363	5' end	24	70	23	75	NA	NA
chr15.12334446	chr15:12334834-12334894	7	AK020883	4960	Intragenic (intron)	0	0	0	0	NA	NA

chr2.18044553	chr2:18045044-18045097	7	BC061062	7103	3' end	0	0	0	0	NA	NA
chr2.93820922	chr2:93821852-93821945	7	AK014837	52118	3' end	0	0	0	0	NA	NA
chr1.57859198	chr1:57859614-57859640	7	AK048123	52693	3' end	0	0	0	4	NA	NA
chr17.79362462	chr17:79362852-79362914	7	AK082127	71985	3' end	0	0	0	0	NA	NA
chr11.30644536	chr11:30645210-30645278	6	Psme4	22385	5' end	184	114	66	48	GeneID:103554	Molecular function unclassified
chr10.61594366	chr10:61594578-61594638	6	Col13a1	9139	3' end	0	0	0	0	GeneID:12817	Extracellular matrix structural protein
chr6.122973017	chr6:122973691-122973790	6	Phc1	2531	5' end	334	209	118	19	GeneID:13619	Other transcription factor;Chromatin/chromatin-binding protein
chr13.44192850	chr13:44193186-44193225	6	Jarid2	2199	5' end	98	141	34	6	GeneID:16468	Other transcription factor;Nucleic acid binding
chr8.112038448	chr8:112038663-112038759	6	Cntnap4	31386	Intragenic (intron)	0	0	0	0	GeneID:170571	Other cell adhesion molecule
chr11.18969482	chr11:18969826-18969994	6	Meis1	55382	5' end	0	78	0	11	GeneID:17268	Homeobox transcription factor;Nucleic acid binding
chr5.101086310	chr5:101086799-101086828	6	Mllt2h	29659	5' end	41	17	20	0	GeneID:17355	Other transcription factor
chr17.51742391	chr17:51742996-51743065	6	Pcaf	5491	5' end	0	0	0	0	GeneID:18519	Transcription cofactor;Acetyltransferase
chr3.27802646	chr3:27803573-27803608	6	Pld1	1467	Intragenic (intron)	0	0	0	22	GeneID:18805	Phospholipase
chr6.144605196	chr6:144605935-144605985	6	Sox5	245651	Gene desert	0	0	0	90	GeneID:20678	HMG box transcription factor;Nucleic acid binding
chr1.108551299	chr1:108551668-108551875	6	Serpinb8	937641	Gene desert (3' end)	0	0	0	0	GeneID:20725	Serine protease inhibitor
chr6.73013984	chr6:73014948-73014989	6	Tcf3	595	5' end	0	0	0	0	GeneID:21415	HMG box transcription factor;Nucleic acid binding
chr9.45451040	chr9:45451459-45451569	6	Tmprss13	1809	3' end	0	0	0	0	GeneID:214531	Serine protease
chr15.22109036	chr15:22109399-22109468	6	Cdh12	483705	Gene desert	0	0	0	0	GeneID:215654	Cadherin
chr19.13953388	chr19:13953964-13954025	6	Tle4	79035	5' end	0	0	0	6	GeneID:21888	Transcription cofactor;Other miscellaneous function protein
chr14.30739248	chr14:30739450-30739535	6	BC037674	867	Intragenic (intron)	78	59	66	9	GeneID:218914	Molecular function unclassified
chr8.110984406	chr8:110984806-110984847	6	Zfp1	5328	Intragenic (intron)	0	0	0	0	GeneID:22640	KRAB box transcription factor;Nucleic acid binding
chr1.179746063	chr1:179746460-179746535	6	Elys	21648	3' end	29	122	69	12	GeneID:226747	Molecular function unclassified
chr4.138458206	chr4:138458513-138458566	6	BC055811	24614	3' end	0	0	0	0	GeneID:230868	Other receptor;Other defense and immunity protein
chr4.146624249	chr4:146624573-146624637	6	Tardbp	111667	Gene desert	6	15	7	18	GeneID:230908	Other transcription factor;Nucleic acid binding
chr5.85199770	chr5:85200073-85200089	6	Tmprss11d	34054	5' end	0	0	0	0	GeneID:231382	Serine protease
chr6.89396996	chr6:89397637-89397822	6	4933427D06Rik	36827	5' end	0	0	0	0	GeneID:232217	Molecular function unclassified
chr7.13254161	chr7:13254336-13254465	6	2410005H09Rik	1531	3' end	78	87	4	82	GeneID:232969	Molecular function unclassified
chr9.21711346	chr9:21711799-21711909	6	Ankrd25	0	Intragenic (exon)	0	0	0	0	GeneID:235041	Transcription factor;Nuclease
chr9.32671391	chr9:32671692-32671790	6	Ets1	10072	5' end	0	0	0	21	GeneID:23871	Other transcription factor;Nucleic acid binding
chr6.4724516	chr6:4724819-4724943	6	Ppp1r9a	26425	Intragenic (intron)	0	0	0	0	GeneID:243725	Non-motor actin binding protein
chr8.89138992	chr8:89139391-89139434	6	Tnrc9	423927	Gene desert (3' end)	0	4	0	30	GeneID:244579	HMG box transcription factor;Chromatin/chromatin-binding protein
chr7.72241404	chr7:72242209-72242200	6	Olfr290	107031	Gene desert	0	0	0	0	GeneID:250411	G-protein coupled receptor
chr9.102098498	chr9:102099565-102099604	6	Ephb1	10285	Intragenic (intron)	0	0	0	29	GeneID:270190	Tyrosine protein kinase receptor;Protein kinase
chr6.149277771	chr6:149278132-149278231	6	4833442J19Rik	10381	5' end	0	0	0	0	GeneID:320204	Molecular function unclassified
chr12.61849597	chr12:61849903-61849903	6	Mamdc1	24909	Intragenic (intron)	0	0	0	0	GeneID:320772	CAM family adhesion molecule
chr14.49722334	chr14:49724571-49724571	6	Lats2	1038	Intragenic (intron)	0	0	0	13	GeneID:50523	Non-receptor serine/threonine protein kinase
chr17.11372227	chr17:11372949-11373014	6	Park2	40111	Intragenic (intron)	0	0	0	0	GeneID:50873	Other transfer/carrier protein;Other ligase
chr6.149063859	chr6:149064152-149064191	6	Tera	2054	3' end	14	119	0	0	GeneID:56306	Molecular function unclassified
chr9.43799297	chr9:43799540-43799636	6	Pvrl1	46400	5' end	3	0	0	24	GeneID:58235	Other receptor;Cell adhesion molecule;Other defense and immunity protein
chr11.60052094	chr11:60052573-60052691	6	4933439F18Rik	1513	3' end	17	102	10	51	GeneID:66771	Molecular function unclassified
chr10.130524542	chr10:130525264-130525287	6	5830405N20Rik	58711	3' end	0	0	0	0	GeneID:67596	Molecular function unclassified
chr16.8771674	chr16:8771805-8771935	6	AK007485	229809	Gene desert	0	0	0	0	GeneID:69053	Molecular function unclassified
chr10.79942135	chr10:79942433-79942487	6	6330406L22Rik	14541	3' end	3	53	0	0	GeneID:70719	Other G-protein modulator
chr5.31983872	chr5:31984151-31984186	6	2410018C17Rik	64631	3' end	0	0	0	0	GeneID:74504	Molecular function unclassified
chr3.30489381	chr3:30489932-30489990	6	Samd7	21327	5' end	0	0	0	0	GeneID:75953	Molecular function unclassified
chr14.41427405	chr14:41427570-41427607	6	A930006J02Rik	23673	5' end	0	0	0	0	GeneID:77790	Molecular function unclassified
chr16.16458506	chr16:16458738-16458915	6	Olfr19	10729	5' end	0	0	0	0	mCG1038100	G-protein coupled receptor
chr4.57187704	chr4:57188315-57188391	6	D630039A03Rik	3957	5' end	0	0	0	0	mCG1041274	Molecular function unclassified
chr10.86071927	chr10:86072147-86072156	6	Timp3	162	Intragenic (intron)	0	0	0	75	mCG11195	Metalloprotease inhibitor
chr1.54871488	chr1:54872400-54872422	6	D230012E17Rik	69438	3' end	0	0	0	0	mCG113061	Molecular function unclassified
chr10.62854712	chr10:62855392-62855395	6	Atoh7	10503	5' end	0	0	0	0	mCG11319	Transcription factor;Nuclease
chr15.91450919	chr15:91451326-91451392	6	Kif21a	6361	5' end	0	21	7	20	mCG114919	Microtubule binding motor protein
chr13.99400188	chr13:99400442-99400468	6	Ly78	38218	5' end	0	0	0	0	mCG115540	Other receptor
chr17.27788373	chr17:27788729-27788916	6	Cpne5	956	Intragenic (intron)	0	0	0	0	mCG117104	Other miscellaneous function protein;Other membrane traffic protein
chr17.39332192	chr17:39333492-39333532	6	Mut	223667	Gene desert (3' end)	10	18	0	47	mCG11778	Mutase
chr12.106354902	chr12:106355387-106355458	6	Mark3	3561	Intragenic (intron)	0	49	0	88	mCG119299	Non-receptor serine/threonine protein kinase
chr4.18322308	chr4:18322570-18322595	6	Mmp16	278851	Gene desert (3' end)	10	0	0	8	mCG121174	Metalloprotease;Other extracellular matrix
chr1.12843746	chr1:12844623-12844733	6	Sulf1	16439	Intragenic (intron)	7	0	0	0	mCG123544	Esterase
chr7.118392130	chr7:118392648-118392758	6	Ate1	37070	Intragenic (intron)	17	0	25	3	mCG126337	Molecular function unclassified
chr6.100770760	chr6:100771295-100771359	6	Rybp	103679	Gene desert	44	110	20	0	mCG127676	Molecular function unclassified
chr6.100803584	chr6:100804316-100804396	6	Rybp	136708	Gene desert	44	110	20	0	mCG127676	Molecular function unclassified
chr16.78617694	chr16:78618743-78618814	6	Cxadr	38028	5' end	0	0	0	0	mCG128445	Other receptor
chr16.81861270	chr16:81861635-81861798	6	Ncam2	9009	Intragenic (intron)	0	0	0	0	mCG130138	CAM family adhesion molecule

chr1.137268569	chr1:137268994-137269074	6	Nr5a2	417642	Gene desert (5' end)	44	19	0	0	mCG130962	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr1.135396415	chr1:135396676-135396719	6	Nav1	2177	Intragenic (intron)	0	3	0	43	mCG131193	Molecular function unclassified
chr11.96451099	chr11:96451532-96451589	6	Snx11	2906	5' end	0	0	0	0	mCG13233	Molecular function unclassified
chr2.70217006	chr2:70217467-70217659	6	A430065P19Rik	10413	5' end	0	0	0	0	mCG142009	Actin binding motor protein
chr9.78556109	chr9:78556490-78556573	6	Gsta4	1077	5' end	310	613	53	45	mCG14386	Other transferase
chr6.60504065	chr6:60504317-60504384	6	A530053G22Rik	5566	5' end	0	0	0	0	mCG144863	Molecular function unclassified
chr10.99134641	chr10:99134772-99134932	6	B530045E10Rik	49187	Intragenic (intron)	0	0	0	0	mCG145343	Molecular function unclassified
chr2.116900868	chr2:116901548-116901582	6	Spred1	1459	5' end	5	5	0	45	mCG14558	Non-motor actin binding protein
chr2.38628316	chr2:38628898-38629007	6	Nr5a1	23664	3' end	0	0	0	0	mCG1494	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr12.60848431	chr12:60848808-60848824	6	Wdr20	361837	Gene desert (3' end)	0	0	0	0	mCG15070	Molecular function unclassified
chr12.112245175	chr12:112245603-112245623	6	Ptprn2	5570	Intragenic (intron)	0	0	0	0	mCG15292	Other receptor;Protein phosphatase
chr9.72199476	chr9:72200299-72200363	6	Tcf12	3078	Intragenic (intron)	0	0	0	0	mCG16396	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr10.44638327	chr10:44638643-44638646	6	Prdm1	258993	Gene desert	0	0	0	0	mCG17867	KRAB box transcription factor;Nucleic acid binding
chr9.106154713	chr9:106155128-106155184	6	Ptk9l	274	3' end	243	418	113	59	mCG19506	Non-motor actin binding protein
chr7.8044995	chr7:8045434-8045552	6	Uble1a	1402	Intragenic (intron)	671	638	316	227	mCG2058	Other ligase
chr2.169022386	chr2:169022672-169022752	6	Zfp64	4325	5' end	30	59	20	47	mCG20647	KRAB box transcription factor;Nucleic acid binding
chr1.80916009	chr1:80916277-80916332	6	Cul3	80028	5' end	0	4	3	56	mCG21668	Other miscellaneous function protein
chr13.111356367	chr13:111356706-111356863	6	Ndufs4	18611	5' end	0	0	0	0	mCG21737	Dehydrogenase;Reductase
chr9.9603967	chr9:9604258-9604297	6	Pgr	638979	Gene desert (3' end)	0	0	0	0	mCG21848	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr9.11122189	chr9:11122418-11122481	6	Pgr	2157151	Gene desert (3' end)	0	0	0	0	mCG21848	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr4.57037941	chr4:57038419-57038517	6	Akap2	12686	Intragenic (intron)	0	9	0	0	mCG2708	Other miscellaneous function protein
chr5.42471754	chr5:42472432-42472525	6	Cd38	8024	5' end	0	0	0	0	mCG4157	Glycosidase;Cyclase
chr1.182115200	chr1:182115699-182115743	6	Enah	10787	5' end	204	213	10	81	mCG4719	Non-motor actin binding protein
chr13.12776563	chr13:12776796-12776832	6	Nid1	12625	Intragenic (intron)	0	0	0	0	mCG5304	Extracellular matrix linker protein
chrX.85215258	chrX:85215557-85215567	6	Pcyt1b	8818	Intragenic (intron)	18	1	0	2	mCG5463	Nucleotidyltransferase
chr3.102122113	chr3:102122747-102122774	6	Vangl1	1367	Intragenic (intron)	0	0	0	34	mCG5474	Molecular function unclassified
chr19.30430372	chr19:30430818-30430836	6	Cstf2t	32953	3' end	10	166	21	159	mCG55983	Ribonucleoprotein
chr8.84255209	chr8:84255552-84255611	6	Junb	1338	5' end	24	683	15	63	mCG5902	Other transcription factor;Nucleic acid binding
chr6.34242054	chr6:34242240-34242292	6	Akr1b8	9866	5' end	0	0	0	0	mCG6069	Oxidoreductase
chr3.18290781	chr3:18291479-18291504	6	Cyp7b1	396242	Gene desert (5' end)	0	0	0	0	mCG6269	Oxygenase
chr11.34390360	chr11:34390558-34390568	6	Dock2	6176	Intragenic (intron)	0	0	0	0	mCG6670	Other G-protein modulator
chr15.9040560	chr15:9040947-9040967	6	1700020N11Rik	5029	Intragenic (intron)	0	0	0	0	mCG7031	Molecular function unclassified
chr12.72190697	chr12:72191105-72191178	6	Fntb	4338	Intragenic (intron)	73	60	3	67	mCG7924	Acyltransferase
chr6.134015120	chr6:134015376-134015454	6	Etv6	72533	5' end	0	0	0	4	mCG7964	Other transcription factor;Nucleic acid binding
chr3.6948788	chr3:6949737-6949761	6	Pkia	440459	Gene desert (5' end)	0	0	0	40	mCG8935	Kinase inhibitor
chr1.139294333	chr1:139294885-139295013	6	9830132G07Rik	9813	5' end	0	0	0	24	mCG9171	Molecular function unclassified
chr8.106183214	chr8:106183619-106183669	6	Has3	1266	5' end	0	0	0	0	mCG9338	Synthase;Glycosyltransferase
chr1.164033723	chr1:164034274-164034336	6	2810422O20Rik	21293	3' end	12	4	0	45	mCG9378	Molecular function unclassified
chr9.7610144	chr9:7610545-7610617	6	Mmp20	1030	Intragenic (intron)	0	0	0	0	mCG9883	Metalloprotease;Other extracellular matrix
chr5.107185577	chr5:107185712-107185856	6	AK019690	4691	Intragenic (intron)	0	0	0	0	NA	NA
chr7.60577912	chr7:60579813-60579867	6	BC062275	5516	Intragenic (intron)	0	0	0	0	NA	NA
chr11.8463357	chr11:8464325-8464335	6	AK052680	9544	Intragenic (intron)	0	8	0	4	NA	NA
chr5.108499048	chr5:108499347-108499413	6	AK087089	9898	3' end	0	0	0	0	NA	NA
chr16.13492702	chr16:13492940-13493003	6	BC006659	18478	Intragenic (intron)	0	0	0	0	NA	NA
chr19.37101564	chr19:37102105-37102139	6	BC025649	21254	3' end	1	4	27	4	NA	NA
chr15.6942524	chr15:6942793-6942900	6	Lifr	24000	5' end	24	0	0	0	NA	NA
chr19.37572566	chr19:37573094-37573109	6	BC057932	27361	5' end	0	0	0	0	NA	NA
chr9.30997511	chr9:30998411-30998474	6	AK046026	32144	Intragenic (intron)	0	0	0	0	NA	NA
chr5.129109877	chr5:129110710-129110748	6	AK086339	38618	5' end	0	0	0	0	NA	NA
chr18.50202412	chr18:50202719-50202740	6	AK004450	42248	5' end	0	0	0	0	NA	NA
chr17.61523477	chr17:61523678-61523761	6	BC058120	48424	Intragenic (intron)	0	0	0	0	NA	NA
chr5.52131475	chr5:52132005-52132125	6	AY512934	49656	5' end	0	0	0	0	NA	NA
chr10.27782957	chr10:27783482-27783619	6	AK078614	79860	5' end	0	0	0	0	NA	NA
chr6.68984788	chr6:68985427-68985551	6	AF206028	84087	5' end	0	0	0	0	NA	NA
chr3.6796366	chr3:6797053-6797169	6	AK005786	147696	Gene desert	0	0	0	0	NA	NA
chr7.121384639	chr7:121384855-121384969	5	Dhx32	269	Intragenic (intron)	0	0	0	0	GeneID:101437	RNA helicase
chr15.4955528	chr15:4955709-4955752	5	Prkaa1	8426	5' end	0	0	0	0	GeneID:105787	Non-receptor serine/threonine protein kinase
chr16.36555641	chr16:36556018-36556166	5	Ildr1	310	Intragenic (intron)	0	0	0	0	GeneID:106347	Other receptor
chr5.11917172	chr5:11917291-11917336	5	Sema3d	284231	Gene desert	0	0	0	0	GeneID:108151	Membrane-bound signaling molecule
chr15.28109772	chr15:28110334-28110410	5	Dnahc5	131913	Gene desert	0	0	0	0	GeneID:110082	Microtubule binding motor protein;Hydrolase
chr3.136092808	chr3:136093140-136093298	5	Manba	1002	Intragenic (intron)	0	0	0	0	GeneID:110173	Glycosidase
chr13.77338663	chr13:77339138-77339245	5	Mass1	369008	Gene desert (3' end)	0	0	110	0	GeneID:110789	Other transporter
chr15.11370193	chr15:11370434-11370570	5	Tars	61783	5' end	0	0	0	0	GeneID:110960	Other RNA-binding protein;Aminoacyl-tRNA synthetase;Ligase

chr8.65293288	chr8:65293783-65293900	5	Mar1	46436	Intragenic (intron)	0	0	0	0	GeneID:111383	
chr4.52583974	chr4:52584859-52584988	5	Abca1	179986	Gene desert (5' end)	0	9	0	24	GeneID:11303	ATP-binding cassette (ABC) transporter
chr14.17377186	chr14:17377574-17377592	5	Adk	3430	Intragenic (intron)	0	0	0	0	GeneID:11534	Nucleotide kinase
chr2.127588121	chr2:127588668-127588691	5	Bub1	7803	5' end	0	0	0	0	GeneID:12235	Non-receptor serine/threonine protein kinase
chr8.117941707	chr8:117942457-117942566	5	Cdh13	18445	Intragenic (intron)	0	0	0	0	GeneID:12554	Cadherin
chr11.12900088	chr11:12900395-12900486	5	Cobl	540847	Gene desert (5' end)	22	11	0	0	GeneID:12808	Molecular function unclassified
chr11.14076592	chr11:14076970-14077089	5	Cobl	1717436	Gene desert (5' end)	22	0	0	0	GeneID:12808	Molecular function unclassified
chr12.46740491	chr12:46740854-46740881	5	Coch	71451	5' end	0	0	0	0	GeneID:12810	Defense/immunity protein
chr3.18364591	chr3:18364646-18364872	5	Cyp7b1	469510	Gene desert (5' end)	0	0	0	0	GeneID:13123	Oxygenase
chr4.102714386	chr4:102714506-102714525	5	Dab1	1157	5' end	7	0	0	3	GeneID:13131	Other signaling molecule
chr18.20475482	chr18:20475837-20475913	5	Dsc1	3607	3' end	0	0	0	0	GeneID:13505	Cadherin
chr1.181010244	chr1:181010597-181010768	5	Lefty1	1184	5' end	194	125	4	0	GeneID:13590	TGF-beta superfamily member
chr8.26028778	chr8:26029144-26029149	5	Eif4ebp1	19897	5' end	612	0	456	113	GeneID:13685	Translation factor
chr9.118375976	chr9:118376354-118376417	5	Eomes	72109	5' end	0	0	28	0	GeneID:13813	Other transcription factor;Nucleic acid binding
chr9.118359570	chr9:118360200-118360216	5	Eomes	88286	5' end	0	0	28	0	GeneID:13813	Other transcription factor;Nucleic acid binding
chr18.34471046	chr18:34471645-34471820	5	Epb4.1l4a	25449	5' end	0	0	0	0	GeneID:13824	Protein phosphatase
chr8.24232760	chr8:24232988-24233166	5	Fgfr1	63043	5' end	0	0	0	0	GeneID:14182	Tyrosine protein kinase receptor;Protein kinase
chr13.67947850	chr13:67948278-67948366	5	Irx1	449857	Gene desert (5' end)	0	0	0	0	GeneID:16371	Homeobox transcription factor;Nucleic acid binding
chr13.69389424	chr13:69390091-69390187	5	Irx2	313972	Gene desert	0	0	0	0	GeneID:16372	Homeobox transcription factor;Nucleic acid binding
chr10.38486853	chr10:38487488-38487532	5	Lama4	345615	Gene desert (5' end)	0	0	13	0	GeneID:16775	Extracellular matrix linker protein
chr3.9642031	chr3:9642266-9642355	5	Gig1	17309	Intragenic (intron)	0	0	0	0	GeneID:170753	Transcription cofactor
chr1.9770616	chr1:9770874-9770939	5	Mybl1	2037	5' end	0	0	0	0	GeneID:17864	Other transcription factor;Nucleic acid binding
chr15.25795099	chr15:25795323-25795480	5	Myo10	812	Intragenic (intron)	2	167	160	126	GeneID:17909	Actin binding motor protein
chr3.88463586	chr3:88464036-88464213	5	Nes	332	5' end	37	33	20	490	GeneID:18008	Molecular function unclassified
chr11.4700964	chr11:4701200-4701227	5	Nf2	89	Intragenic (intron)	6	0	0	12	GeneID:18016	Other cytoskeletal proteins
chr4.80675997	chr4:80676378-80676558	5	Nfib	17962	Intragenic (intron)	0	0	0	0	GeneID:18028	Other transcription factor;Nucleic acid binding
chr9.113474054	chr9:113474786-113474892	5	Pdcd6ip	194735	Gene desert	0	20	54	63	GeneID:18571	Transmembrane receptor regulatory/adaptor protein
chr2.92037832	chr2:92038030-92038051	5	Phf21a	40227	5' end	3	114	0	42	GeneID:192285	Molecular function unclassified
chr3.6011987	chr3:6012504-6012586	5	Pxmp3	409177	Gene desert (5' end)	0	0	0	0	GeneID:19302	Other miscellaneous function protein
chr11.118826807	chr11:118827378-118827540	5	Tbc1d16	162	Intragenic (intron)	0	0	0	0	GeneID:207592	Molecular function unclassified
chr6.39373680	chr6:39373814-39373999	5	B930096L08Rik	16375	3' end	0	1	0	38	GeneID:209773	Molecular function unclassified
chr15.21402301	chr15:21402597-21402012	5	Cdh12	7491	Intragenic (intron)	0	0	0	0	GeneID:215654	Cadherin
chr2.24871856	chr2:24872500-24872541	5	Arrdc1	12768	3' end	4	6	15	0	GeneID:215705	Molecular function unclassified
chr11.20667947	chr11:20668316-20668433	5	9130023F12Rik	31310	5' end	15	37	17	21	GeneID:216549	Molecular function unclassified
chr9.89961861	chr9:89962233-89962348	5	Morf4l1	18981	5' end	149	102	0	71	GeneID:21761	Other transcription factor;Chromatin/chromatin-binding protein
chr17.44851866	chr17:44852187-44852250	5	Ppp2r5d	5118	Intragenic (intron)	67	56	65	111	GeneID:21770	Protein phosphatase
chr6.37696724	chr6:37697097-37697192	5	Trim24	102927	Gene desert	131	34	32	20	GeneID:21848	Transcription cofactor;Nucleic acid binding
chr11.69316467	chr11:69316853-69316902	5	Tnfsf12	8567	5' end	65	16	0	13	GeneID:21944	Tumor necrosis factor family member
chr18.11088040	chr18:11088289-11088481	5	Mib1	54176	3' end	0	0	0	70	GeneID:225164	Molecular function unclassified
chr17.22195646	chr17:22196212-22196388	5	Zfp13	1913	5' end	0	10	0	14	GeneID:22654	KRAB box transcription factor;Nucleic acid binding
chr2.145814248	chr2:145814665-145814676	5	BC024760	14568	Intragenic (intron)	0	0	0	0	GeneID:228726	
chr15.92280364	chr15:92280823-92280981	5	Muc19	6054	Intragenic (intron)	0	0	0	0	GeneID:239611	Extracellular matrix glycoprotein
chr2.21448054	chr2:21448752-21448823	5	Gpr158	6283	Intragenic (intron)	0	0	0	0	GeneID:241263	Other transporter
chr4.97491352	chr4:97491837-97491840	5	Apg4c	32369	5' end	1	77	10	9	GeneID:242557	Molecular function unclassified
chr8.19086700	chr8:19086925-19087108	5	Mcph1	115241	Gene desert	11	16	6	0	GeneID:244329	Molecular function unclassified
chr4.117056915	chr4:117057573-117057680	5	Olfr1341	1651	3' end	0	0	0	0	GeneID:258852	G-protein coupled receptor
chr12.81824263	chr12:81824836-81824931	5	Esrrb	7155	Intragenic (intron)	281	211	24	0	GeneID:26380	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr2.29780552	chr2:29780699-29780749	5	Slc27a4	3180	3' end	11	2	0	12	GeneID:26569	Transporter;Other ligase
chr10.25076447	chr10:25076761-25076883	5	Akap7	15255	5' end	0	0	0	7	GeneID:268287	Kinase activator
chr11.59124039	chr11:59124287-59124359	5	Zfp496	5575	5' end	14	44	46	101	GeneID:268417	KRAB box transcription factor
chr12.104696390	chr12:104696702-104696811	5	Ppp2r5c	46191	Intragenic (intron)	0	0	0	0	GeneID:26931	Protein phosphatase
chr1.7143986	chr1:7144501-7144557	5	A030012M09Rik	22315	5' end	67	10	32	41	GeneID:319263	Molecular function unclassified
chr12.113802352	chr12:113802891-113802891	5	Sp8	80959	5' end	0	0	0	0	GeneID:320145	KRAB box transcription factor
chr15.66017587	chr15:66017813-66017863	5	D030063F01Rik	37269	5' end	0	0	0	0	GeneID:320269	Molecular function unclassified
chr10.109705206	chr10:109705518-109705581	5	9630020C08Rik	121187	Intragenic (intron)	0	0	0	0	GeneID:327819	
chr11.62123228	chr11:62123554-62123667	5	Pigl	1074	3' end	0	0	0	0	GeneID:327942	Glycosyltransferase;Esterase
chr3.73212474	chr3:73212641-73212689	5	Slitrk3	177927	Gene desert	0	0	0	0	GeneID:386750	Antibacterial response protein
chr7.57630042	chr7:57630417-57630428	5	LOC434198	179756	Gene desert	0	0	0	0	GeneID:434198	Molecular function unclassified
chrX.6496113	chrX:6496390-6496546	5	Porcn	988	5' end	7	8	63	59	GeneID:53627	Membrane traffic regulatory protein
chr6.136533980	chr6:136534170-136534285	5	Atf7ip	48358	5' end	0	0	0	47	GeneID:54343	Transcription cofactor;Hydrolase
chr5.90938068	chr5:90938797-90938844	5	Vdp	25043	5' end	7	23	55	46	GeneID:56041	Other membrane traffic protein
chr13.6702825	chr13:6703290-6703397	5	Pfkp	244995	Gene desert (5' end)	121	98	26	47	GeneID:56421	Carbohydrate kinase
chr8.84721562	chr8:84721702-84721726	5	Sast	77364	Intragenic (intron)	0	0	0	0	GeneID:56527	Non-receptor serine/threonine protein kinase

chr6.34084705	chr6:34084991-34085071	5	Slc35b4	10535	5' end	3	0	48	54	GeneID:58246	Other transporter
chr10.75938403	chr10:75938650-75938674	5	Slc5a4b	3762	Intragenic (intron)	0	0	0	0	GeneID:64454	Cation transporter;Carbohydrate transporter
chr2.181605790	chr2:181606390-181606425	5	Polr3k	48127	3' end	0	43	2	42	GeneID:67005	DNA-directed RNA polymerase;Nucleotidyltransferase
chr18.4017642	chr18:4018185-4018224	5	Lyzl1	318121	Gene desert (5' end)	0	0	0	0	GeneID:67328	Hydrolase;Defense/immunity protein
chr18.52412598	chr18:52412925-52412982	5	Ftmt	399047	Gene desert (5' end)	0	0	0	0	GeneID:67634	Storage protein
chr14.67265293	chr14:67265443-67265590	5	Nurit	77472	3' end	0	0	0	0	GeneID:67926	Molecular function unclassified
chr16.8465471	chr16:8465636-8465799	5	AK007485	50244	5' end	0	0	0	0	GeneID:69053	Molecular function unclassified
chr2.80532746	chr2:80533266-80533363	5	Nup35	323	5' end	0	0	0	0	GeneID:69482	Molecular function unclassified
chr7.31879405	chr7:31880019-31880133	5	2310044H10Rik	4577	3' end	20	25	8	229	GeneID:69683	Molecular function unclassified
chr12.18387220	chr12:18387595-18387649	5	2410018L13Rik	176006	Gene desert	0	0	0	0	GeneID:69732	KRAB box transcription factor
chr10.8566107	chr10:8566624-8566681	5	Sash1	3875	5' end	16	12	0	98	GeneID:70097	Molecular function unclassified
chr1.44962315	chr1:44963316-44963378	5	Gulp1	68553	Intragenic (intron)	0	2	0	21	GeneID:70676	Other signaling molecule
chr8.74090677	chr8:74092607-74092685	5	Hmgb2l1	1795	Intragenic (intron)	0	7	0	0	GeneID:70823	HMG box transcription factor;Chromatin/chromatin-binding protein
chr1.8629653	chr1:8630451-8630612	5	Sntg1	1193	Intragenic (intron)	0	0	0	0	GeneID:71096	Other cytoskeletal proteins
chr10.84525655	chr10:84527016-84527151	5	Rfx4	788	Intragenic (intron)	0	0	0	0	GeneID:71137	Other transcription factor
chr5.107557499	chr5:107557713-107557880	5	D5Ertd585e	3420	5' end	0	24	0	40	GeneID:71782	Molecular function unclassified
chr11.107023189	chr11:107023931-107024111	5	Pitpnc1	15336	Intragenic (intron)	52	8	4	167	GeneID:71795	Other transporter;Other transfer/carrier protein
chr10.39091829	chr10:39092246-39092271	5	Tube1	74502	3' end	0	0	0	0	GeneID:71924	Tubulin
chr3.27508730	chr3:27509388-27509472	5	Fndc3b	33666	5' end	0	0	0	27	GeneID:72007	Other actin family cytoskeletal protein
chr8.63921747	chr8:63922369-63922394	5	3110005G23Rik	4084	5' end	0	0	0	0	GeneID:73067	Molecular function unclassified
chr16.49712890	chr16:49713196-49713207	5	Esrrbl1	39053	3' end	0	0	0	0	GeneID:73916	Molecular function unclassified
chr17.21233085	chr17:21233409-21233481	5	1300003B13Rik	63415	3' end	0	0	0	0	GeneID:74149	KRAB box transcription factor;Nucleic acid binding
chr3.11269153	chr3:11269535-11269542	5	Snx16	770485	Gene desert (5' end)	0	0	0	0	GeneID:74718	Molecular function unclassified
chr4.153555813	chr4:153556274-153556358	5	Zzank1	1560	5' end	16	11	1	6	GeneID:76580	Molecular function unclassified
chr2.58579141	chr2:58579865-58579886	5	Upp2	58979	Intragenic (intron)	0	0	0	0	GeneID:76654	Phosphorylase
chr11.120334792	chr11:120334884-120335030	5	5730593N15Rik	5638	5' end	21	0	0	0	GeneID:77583	Molecular function unclassified
chr9.114512367	chr9:114512504-114512616	5	Cnot10	1239	3' end	0	0	0	0	GeneID:78893	Molecular function unclassified
chr9.40264932	chr9:40265238-40265387	5	Zfp202	26738	3' end	0	0	0	0	GeneID:80902	KRAB box transcription factor;Nucleic acid binding
chr3.21355698	chr3:21356573-21356700	5	Tbl1xr1	406694	Gene desert	0	8	0	20	GeneID:81004	Other miscellaneous function protein
chr3.60265933	chr3:60267328-60267372	5	Sucnr1	19210	3' end	0	0	0	0	GeneID:84112	G-protein coupled receptor
chr3.89066472	chr3:89066793-89066909	5	Ubqln4	5848	3' end	156	141	171	51	GeneID:94232	Other miscellaneous function protein
chr1.54130042	chr1:54130570-54130692	5	Stk17b	5200	3' end	0	0	0	0	GeneID:90267	Non-receptor serine/threonine protein kinase
chr19.39180176	chr19:39180371-39180472	5	Cyp2c40	77609	5' end	0	0	0	0	mCG10001	Oxygenase
chr13.69544172	chr13:69544610-69544758	5	Irx4	160995	Gene desert	0	0	0	0	mCG10022	Homeobox transcription factor;Nucleic acid binding
chr3.131684805	chr3:131684985-131685089	5	Lef1	8346	5' end	0	0	0	0	mCG10208	HMG box transcription factor;Nucleic acid binding
chr6.53752076	chr6:53752389-53752454	5	1200009O22Rik	13951	3' end	1	0	0	99	mCG1036418	Receptor
chr1.34752059	chr1:34752324-34752437	5	Ptpn18	383	Intragenic (intron)	3	2	98	0	mCG10409	Protein phosphatase
chr10.56415670	chr10:56415886-56415947	5	Gja1	2927	5' end	224	294	238	56	mCG10836	Gap junction
chr4.125699307	chr4:125699557-125699597	5	Gjb3	1239	5' end	23	152	0	0	mCG10907	Gap junction
chr4.20897861	chr4:20898377-20898416	5	E130310K16Rik	193613	Gene desert (5' end)	0	0	0	16	mCG10984	Molecular function unclassified
chr4.20898925	chr4:20899632-20899813	5	E130310K16Rik	194939	Gene desert (5' end)	0	0	0	16	mCG10984	Molecular function unclassified
chr13.40311337	chr13:40312331-40312524	5	Gcnt2	8107	Intragenic (intron)	51	10	0	0	mCG10994	Glycosyltransferase
chr12.4431448	chr12:4432642-4432685	5	Sh3d1B	55866	5' end	0	0	0	0	mCG11010	Other membrane traffic protein
chr8.22365888	chr8:22366183-22366227	5	Zmat4	107582	Gene desert	15	0	0	0	mCG113560	Molecular function unclassified
chr2.26463899	chr2:26464172-26464289	5	Notch1	571	5' end	32	23	0	461	mCG11364	Receptor;Membrane-bound signaling molecule;Defense/immunity protein
chr14.91735778	chr14:91736547-91736584	5	Klf5	199417	Gene desert	0	0	0	0	mCG113841	KRAB box transcription factor;Nucleic acid binding
chr14.86960574	chr14:86960816-86960937	5	4921530L21Rik	1089702	Gene desert (5' end)	0	0	0	0	mCG113900	Molecular function unclassified
chr1.176773933	chr1:176774199-176774206	5	4933426L22Rik	89245	3' end	0	0	0	0	mCG114397	Molecular function unclassified
chr13.93798210	chr13:93798496-93798509	5	1700029F12Rik	26861	5' end	0	0	0	0	mCG114606	Molecular function unclassified
chr10.84577764	chr10:84577922-84578053	5	MGC58970	7325	5' end	8	0	0	12	mCG115197	Molecular function unclassified
chr1.53325798	chr1:53326054-53326105	5	1700019D03Rik	838	5' end	0	27	0	0	mCG115472	Molecular function unclassified
chr13.86904191	chr13:86904357-86904386	5	0610041E09Rik	138693	Gene desert	1	6	14	7	mCG116561	Molecular function unclassified
chrX.34691716	chrX:34692023-34692099	5	Stag2	67588	3' end	0	0	0	0	mCG116901	Chromatin/chromatin-binding protein
chr14.23385507	chr14:23385773-23385827	5	Arhgef3	11225	Intragenic (intron)	29	0	0	22	mCG117573	Guanyl-nucleotide exchange factor
chr19.20195787	chr19:20196053-20196286	5	Tmc1	10451	Intragenic (intron)	0	0	0	0	mCG119958	Molecular function unclassified
chr3.93592914	chr3:93593068-93593175	5	Rptn	15451	5' end	0	0	0	0	mCG120170	Calmodulin related protein;Annexin
chr2.20265008	chr2:20265193-20265292	5	9430077C05Rik	80401	Intragenic (intron)	6	45	8	8	mCG120838	Molecular function unclassified
chr14.110736823	chr14:110737130-110737255	5	Sox21	179477	Gene desert	0	23	0	43	mCG120858	Molecular function unclassified
chr3.155152296	chr3:155152740-155152967	5	5230400J09Rik	29112	3' end	0	0	0	0	mCG120907	Molecular function unclassified
chr2.170834059	chr2:170835241-170835377	5	Dok5	28657	Intragenic (intron)	0	0	0	0	mCG120918	Molecular function unclassified
chr14.17762936	chr14:17763418-17763530	5	Myst4	25043	5' end	3	38	5	34	mCG123147	Other zinc finger transcription factor;Chromatin/chromatin-binding protein;Acetyltransferase
chr3.19697736	chr3:19698127-19698273	5	Hps3	1138	Intragenic (intron)	30	0	0	0	mCG123402	Molecular function unclassified
chr7.50942475	chr7:50942973-50943059	5	Klf13	18100	5' end	0	0	0	0	mCG123556	KRAB box transcription factor

chr16.12705646	chr16:12705939-12705985	5	Ercc4	152125	Gene desert	0	0	0	0	mCG123887	Endodeoxyribonuclease;Hydrolase
chr3.134071216	chr3:134071485-134071536	5	E130014J05Rik	49876	5' end	2	28	8	11	mCG123956	Molecular function unclassified
chr6.58599879	chr6:58600635-58600820	5	Abcg2	59915	5' end	0	0	0	0	mCG124607	Transporter
chr9.46686504	chr9:46686914-46686989	5	D030060M11Rik	268634	Gene desert	11	15	25	5	mCG126323	Molecular function unclassified
chr9.40423463	chr9:40423822-40423837	5	A930008A22Rik	5860	Intragenic (intron)	6	0	0	0	mCG127192	Molecular function unclassified
chr16.86560990	chr16:86561697-86561794	5	Adamts5	286923	Gene desert	0	0	0	15	mCG129418	Signaling molecule;Metalloprotease
chr6.126048801	chr6:126049153-126049204	5	Cd9	26865	3' end	33	390	140	33	mCG129426	Other cell adhesion molecule
chr6.123402644	chr6:123402918-123402952	5	Slc2a3	22859	5' end	1209	2814	896	5	mCG132202	Carbohydrate transporter
chr6.123261883	chr6:123262305-123262308	5	Dppa3	1697	5' end	0	7	0	0	mCG132218	Molecular function unclassified
chr10.51257441	chr10:51257955-51258010	5	Sim1	248554	Gene desert (3' end)	0	0	0	0	mCG13325	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr3.89650377	chr3:89650616-89650708	5	Thbs3	201	3' end	2	0	0	18	mCG133395	Cell adhesion molecule;Extracellular matrix glycoprotein
chr10.79947436	chr10:79947529-79947650	5	Gpx4	3094	5' end	0	0	0	0	mCG13421	Peroxidase
chr4.128871782	chr4:128872620-128872680	5	2610200G18Rik	11186	3' end	0	0	0	7	mCG13543	Molecular function unclassified
chr3.53740009	chr3:53740263-53740310	5	Stoml3	143055	Gene desert (3' end)	0	0	0	0	mCG13685	Other cytoskeletal proteins
chr7.55044882	chr7:55045178-55045202	5	Igf1r	10338	Intragenic (intron)	0	31	230	114	mCG13842	Tyrosine protein kinase receptor;Protein kinase
chr13.115235768	chr13:115236083-115236193	5	Mrps30	115654	Gene desert	211	62	43	17	mCG1400	Molecular function unclassified
chr2.180397021	chr2:180397400-180397547	5	Dido1	7166	5' end	1	7	0	20	mCG140082	Other zinc finger transcription factor;Nucleic acid binding
chr8.90921676	chr8:90922019-90922213	5	Fto	66689	Intragenic (intron)	0	0	0	30	mCG141364	Molecular function unclassified
chr8.106273787	chr8:106274395-106274395	5	Sntb2	17735	Intragenic (intron)	0	0	0	0	mCG141366	Other cytoskeletal proteins
chr9.122000216	chr9:122000377-122000460	5	C85492	5992	Intragenic (intron)	0	9	0	1	mCG141435	Molecular function unclassified
chr4.45060268	chr4:45060720-45060808	5	Igfbpl1	83	Intragenic (intron)	0	0	0	0	mCG14155	Miscellaneous function
chr16.22477672	chr16:22478395-22478418	5	Dgkg	15076	Intragenic (intron)	0	0	0	0	mCG141782	Transfer/carrier protein;Other kinase;Calmodulin related protein;Annexin
chr1.15952957	chr1:15953538-15953593	5	Terf1	13862	5' end	0	0	0	0	mCG14187	Other nucleic acid binding
chr2.40870646	chr2:40871054-40871121	5	Lrp1b	7617	Intragenic (intron)	0	0	0	0	mCG141878	Other receptor
chr16.41888032	chr16:41888242-41888347	5	Lsamp	67431	Intragenic (intron)	0	0	0	0	mCG141940	CAM family adhesion molecule
chr12.36798919	chr12:36800150-36800249	5	Immp2l	23788	Intragenic (intron)	0	0	0	0	mCG142422	Other proteases
chr17.45878321	chr17:45878538-45878559	5	Frs3	9137	3' end	0	0	0	0	mCG144020	Molecular function unclassified
chr5.146385712	chr5:146385952-146386120	5	Hmgb1	14	5' end	0	0	0	0	mCG144566	HMG box transcription factor;Chromatin/chromatin-binding protein
chr6.60505186	chr6:60505709-60505732	5	A530053G22Rik	6936	5' end	0	0	0	0	mCG144863	Molecular function unclassified
chr6.67224097	chr6:67224472-67224510	5	A430010J10Rik	53205	5' end	0	0	0	0	mCG144876	Molecular function unclassified
chr7.36891394	chr7:36891546-36891633	5	5330421F07Rik	3519	Intragenic (intron)	0	0	0	0	mCG145358	Non-motor actin binding protein
chr7.76517370	chr7:76517706-76517012	5	2010439K00Rik	3410	5' end	0	74	0	0	mCG15145	Molecular function unclassified
chr17.83527963	chr17:83528537-83528650	5	Ppm1b	893	5' end	2	156	20	13	mCG15599	Protein phosphatase
chr14.55488847	chr14:55489116-55489187	5	Tdh	251	Intragenic (intron)	44	0	0	0	mCG15867	Molecular function unclassified
chr1.105412478	chr1:105412808-105412985	5	Pign	29709	3' end	0	0	0	0	mCG16111	Molecular function unclassified
chr13.82998012	chr13:82998714-82998915	5	Cox7c	310965	Gene desert (5' end)	1172	756	307	1129	mCG1735	Oxidase
chr15.102880309	chr15:102881303-102881305	5	Sp1	6649	5' end	0	0	0	0	mCG17604	KRAB box transcription factor
chr1.36379851	chr1:36380333-36380416	5	Hs6st1	6994	5' end	15	89	1	33	mCG17739	Other transferase
chr14.69281006	chr14:69281201-69281289	5	D230005D02Rik	11654	Intragenic (intron)	18	0	0	0	mCG17768	Molecular function unclassified
chr5.73780463	chr5:73780807-73780891	5	Gsh2	7486	3' end	0	0	0	0	mCG18262	Molecular function unclassified
chr5.112880736	chr5:112881714-112881849	5	Pla2g1b	1285	5' end	0	0	0	0	mCG19247	Phospholipase
chr2.71136952	chr2:71137432-71137550	5	Dncic2	3362	Intragenic (intron)	38	0	0	53	mCG19348	Microtubule family cytoskeletal protein
chr13.43470907	chr13:43471212-43471380	5	Cd83	215776	Gene desert	0	0	0	0	mCG19429	Immunoglobulin
chr5.14785623	chr5:14786217-14786264	5	Cacna2d1	55532	3' end	0	0	0	10	mCG19433	Voltage-gated calcium channel
chr3.55084374	chr3:55084693-55084839	5	Ccna1	55305	3' end	0	0	0	0	mCG19561	Kinase activator
chr6.37617777	chr6:37618776-37618831	5	Akr1d1	120963	Gene desert	0	0	0	0	mCG19697	Oxidoreductase
chr4.54708194	chr4:54709465-54709499	5	Klf4	62704	3' end	19	13	0	0	mCG20055	KRAB box transcription factor;Nucleic acid binding
chr1.92360827	chr1:92361281-92361332	5	Ndufa10	6624	3' end	96	338	48	320	mCG20119	Oxidoreductase
chr3.35514349	chr3:35515317-35515391	5	Atp11b	30603	5' end	34	0	0	122	mCG20276	Other transporter;Other hydrolase
chr13.46139597	chr13:46139962-46140085	5	Cap2	1010	3' end	0	0	0	0	mCG20458	Molecular function unclassified
chr11.69176716	chr11:69176973-69177157	5	BC021790	562	3' end	0	0	0	0	mCG20905	Molecular function unclassified
chr2.35982998	chr2:35983535-35983568	5	Ndufa8	12767	3' end	0	10	2	45	mCG22271	Hydrogen transporter;Mitochondrial carrier protein;Oxidoreductase
chr7.94606082	chr7:94606342-94606398	5	Syt9	31026	5' end	0	0	0	0	mCG2602	Membrane traffic regulatory protein
chr4.15170057	chr4:15170749-15171026	5	9630015D15Rik	22042	5' end	0	2	3	5	mCG2670	Molecular function unclassified
chr4.57048724	chr4:57049122-57049184	5	Akap2	23371	Intragenic (intron)	0	9	0	0	mCG2708	Other miscellaneous function protein
chr7.82702008	chr7:82704939-82705057	5	Odz4	500000	Gene desert (5' end)	15	0	0	0	mCG2837	Other receptor;Membrane-bound signaling molecule
chrX.90858903	chrX:90859240-90859408	5	Pja1	9989	3' end	0	9	8	26	mCG3733	Ubiquitin-protein ligase
chrX.90960244	chrX:90960554-90960626	5	Pja1	86369	5' end	0	9	8	30	mCG3733	Ubiquitin-protein ligase
chr8.115649450	chr8:115649619-115649720	5	Dncl2b	239762	Gene desert	0	0	0	0	mCG4360	Other microtubule family cytoskeletal protein
chr8.116064440	chr8:116064814-116064867	5	Cdyl2	49436	Intragenic (intron)	57	20	10	0	mCG4362	Molecular function unclassified
chr17.14495175	chr17:14495269-14495383	5	Chd1	26893	5' end	49	60	0	3	mCG4508	Transcription factor;DNA helicase;Hydrolase
chr10.78931732	chr10:78932541-78932636	5	Olfr57	6972	3' end	0	0	0	0	mCG53077	G-protein coupled receptor
chr2.163526633	chr2:163526960-163527014	5	D930001I22Rik	2960	3' end	0	0	20	4	mCG5451	Molecular function unclassified

chr5.137583019	chr5:137583286-137583391	5	Mad1l1	19607	Intragenic (intron)	38	0	14	7	mCG5835	Molecular function unclassified
chr9.116172146	chr9:116172449-116172501	5	Tgfbr2	38104	5' end	0	0	0	0	mCG6467	TGF-beta receptor;Serine/threonine protein kinase receptor;Protein kinase
chr11.72032665	chr11:72033141-72033205	5	D130058I21Rik	8917	5' end	2	20	4	0	mCG6664	Other cytoskeletal proteins
chr1.133222061	chr1:133222347-133222498	5	Ren2	10075	5' end	0	0	0	0	mCG6933	Aspartic protease
chr2.45695550	chr2:45696394-45696508	5	Zfhx1b	623236	Gene desert (5' end)	0	0	30	37	mCG8151	Homeobox transcription factor;Zinc finger transcription factor;Nucleic acid binding
chr6.30156541	chr6:30156733-30156777	5	Ube2h	10320	Intragenic (intron)	0	0	6	54	mCG8299	Other ligase
chr1.86233831	chr1:86234251-86234289	5	Ncl	712	3' end	116	239	65	74	mCG8509	Ribonucleoprotein
chr2.150045768	chr2:150046526-150046616	5	Zfp120	132956	Gene desert	11	0	0	6	mCG8721	KRAB box transcription factor;Nucleic acid binding
chr15.14021391	chr15:14021774-14021860	5	Cdh6	1025327	Gene desert (5' end)	0	0	0	0	mCG8950	Cadherin
chr15.16039868	chr15:16040296-16040380	5	Cdh6	3043848	Gene desert (5' end)	0	0	0	0	mCG8950	Cadherin
chr6.38659139	chr6:38659614-38659727	5	Hipk2	312	Intragenic (intron)	0	0	0	0	mCG9268	Non-receptor serine/threonine protein kinase
chr1.164035662	chr1:164035891-164035962	5	2810422O20Rik	22914	3' end	12	4	0	45	mCG9378	Molecular function unclassified
chr15.27170252	chr15:27170535-27170695	5	ank	335141	Gene desert (5' end)	0	6	0	4	mCG9599	Transporter;Mitochondrial carrier protein
chr2.103990509	chr2:103991055-103991061	5	Cd59a	3167	Intragenic (intron)	0	0	0	0	mCG9649	Membrane-bound signaling molecule
chr4.38462291	chr4:38462514-38462742	5	Aco1	1772296	Gene desert (5' end)	26	8	0	67	mCG9710	Dehydratase;Hydratase
chr4.134604276	chr4:134604913-134605019	5	AK031499	0	Intragenic (exon)	9	204	31	200	NA	NA
chr5.42480499	chr5:42480836-42480981	5	AK043442	0	Intragenic (exon)	0	0	0	0	NA	NA
chr14.3085358	chr14:3085974-3085988	5	BC055874	442	5' end	17	76	240	0	NA	NA
chr17.44970748	chr17:44971069-44971228	5	AB093219	551	Intragenic (intron)	0	0	0	0	NA	NA
chr18.64385190	chr18:64386054-64386116	5	AK122304	840	Intragenic (intron)	0	0	0	0	NA	NA
chr3.28419825	chr3:28420035-28420048	5	AK088459	933	Intragenic (intron)	0	0	0	0	NA	NA
chr1.127835445	chr1:127835845-127835944	5	AK050039	1439	Intragenic (intron)	0	0	0	0	NA	NA
chr1.106694469	chr1:106695188-106695234	5	BC023820	1633	3' end	34	6	0	9	NA	NA
chr17.34357093	chr17:34357462-34357511	5	AK129074 / Mdc1	2156	Intragenic (intron)	9	8	19	36	NA	NA
chr8.81132358	chr8:81132763-81132782	5	BC064813	2573	Intragenic (intron)	0	0	0	0	NA	NA
chr18.66116581	chr18:66117063-66117117	5	BC067032	4358	Intragenic (intron)	1	0	5	24	NA	NA
chr13.13641094	chr13:13641356-13641430	5	AB083710	4550	Intragenic (intron)	0	0	0	0	NA	NA
chr3.157935841	chr3:157936424-157936467	5	AK122576	4786	3' end	0	0	0	0	NA	NA
chr6.87283922	chr6:87284106-87284177	5	AK129272	5244	5' end	0	0	0	0	NA	NA
chr7.77198435	chr7:77199188-77199241	5	AK036708	5472	5' end	0	0	0	0	NA	NA
chr11.77292390	chr11:77292786-77292830	5	AK122494	5901	5' end	34	62	25	4	NA	NA
chr7.77192414	chr7:77197424-77197529	5	AK036700	7210	5' end	0	0	0	0	NA	NA
chrX.142539953	chrX:142540302-142540366	5	AK049850	8404	5' end	0	0	0	0	NA	NA
chr15.8704423	chr15:8705396-8705435	5	S57425	9492	5' end	0	0	0	0	NA	NA
chr1.31512817	chr1:31513130-31513200	5	AK015858	9618	3' end	0	0	0	0	NA	NA
chr19.11448227	chr19:11448984-11449094	5	NM_146414	10030	5' end	0	0	0	0	NA	NA
chr19.11320095	chr19:11320299-11320372	5	NM_146681	10231	5' end	0	0	0	0	NA	NA
chr5.107766896	chr5:107767262-107767366	5	AK038433	10240	5' end	0	0	0	0	NA	NA
chr3.17438597	chr3:17439069-17439148	5	BC030462	11129	5' end	0	0	0	0	NA	NA
chr11.77285925	chr11:77286227-77286289	5	AK122494	12451	5' end	34	62	25	4	NA	NA
chr12.13105258	chr12:13106220-13106362	5	AK017761	14010	5' end	0	0	0	0	NA	NA
chr2.119638063	chr2:119639133-119639311	5	AK018123	14324	3' end	0	0	0	0	NA	NA
chr18.76111138	chr18:76111813-76111870	5	BC058104	14495	Intragenic (intron)	0	0	0	138	NA	NA
chr4.84457697	chr4:84458271-84458377	5	BC062814	15400	Intragenic (intron)	0	0	0	0	NA	NA
chr11.29491940	chr11:29492870-29492924	5	AK006591	18980	3' end	0	0	0	0	NA	NA
chr11.6669535	chr11:6673409-6673708	5	BC057932	21305	5' end	0	0	0	0	NA	NA
chr16.90744108	chr16:90744620-90744748	5	BC053096	25681	5' end	0	41	0	2	NA	NA
chr17.83713774	chr17:83714212-83714351	5	AK007130	28959	3' end	0	0	0	21	NA	NA
chr10.14968704	chr10:14969245-14969383	5	AK042566	36877	5' end	0	0	0	0	NA	NA
chr13.94310376	chr13:94310722-94310855	5	AK077132	37085	5' end	0	0	0	0	NA	NA
chr2.31028354	chr2:31028710-31028903	5	AK122308	37144	Intragenic (intron)	0	0	0	15	NA	NA
chr5.31906719	chr5:31907042-31907119	5	BC061483	39046	3' end	0	0	0	7	NA	NA
chr11.58680819	chr11:58681193-58681312	5	BC044882	39216	5' end	0	0	0	0	NA	NA
chr8.12520111	chr8:12520526-12520654	5	AK085316	50010	3' end	0	0	0	0	NA	NA
chr18.40751764	chr18:40751859-40751965	5	AK013700	50489	3' end	0	0	0	0	NA	NA
chr17.61880373	chr17:61881020-61881035	5	AK078781	55086	5' end	0	0	0	4	NA	NA
chr5.115897049	chr5:115897637-115897654	5	AK047401	55805	3' end	0	0	0	0	NA	NA
chr7.60647235	chr7:60647542-60647568	5	BC062275	61714	5' end	0	0	0	0	NA	NA
chr18.10682576	chr18:10683020-10683213	5	BC008220	105477	Gene desert	0	0	0	0	NA	NA
chr18.30570428	chr18:30570575-30570640	5	M73818	105896	Gene desert	0	0	0	0	NA	NA
chr13.47350069	chr13:47350414-47350518	5	AK016490	105899	Gene desert	0	0	0	0	NA	NA
chr17.46852969	chr17:46853340-46853369	5	AK039924	139618	Gene desert	0	0	0	0	NA	NA
chr4.81603914	chr4:81604264-81604420	4	Psip1	127887	Gene desert	52	2	25	98	GeneID:101739	Transcription cofactor

chr17:27603929	chr17:27604329-27604574	4	Stk38	525	5' end	40	83	15	65	GeneID:106504	Non-receptor serine/threonine protein kinase
chr2:29689210	chr2:29689592-29689659	4	Rapgef1	923	Intragenic (intron)	47	0	1	22	GeneID:107746	G-protein modulator
chr6:140717592	chr6:140717874-140717982	4	Plekha5	6294	3' end	17	0	25	0	GeneID:109135	Molecular function unclassified
chr1:139377047	chr1:139377207-139377319	4	Calmbp1	794	3' end	0	0	0	0	GeneID:12316	Molecular function unclassified
chr12:103068969	chr12:103069095-103069402	4	Ccnk	117331	Gene desert	107	406	153	128	GeneID:12454	Kinase activator
chr8:118597407	chr8:118597572-118597689	4	Cdh13	28400	Intragenic (intron)	0	0	0	0	GeneID:12554	Cadherin
chr11:12330545	chr11:12330929-12330983	4	Cobl	24023	Intragenic (intron)	22	11	0	0	GeneID:12808	Molecular function unclassified
chr7:120738162	chr7:120738431-120738482	4	Ctbp2	8090	Intragenic (intron)	636	526	248	167	GeneID:13017	Transcription cofactor;Dehydrogenase
chr17:29482430	chr17:29482592-29482664	4	Dnahc8	3670	3' end	0	0	0	0	GeneID:13417	Microtubule binding motor protein;Hydrolase
chr5:25138136	chr5:25138801-25138964	4	Dpp6	65902	5' end	0	0	0	0	GeneID:13483	Serine protease
chr11:44503656	chr11:44503888-44503958	4	Ebf1	4902	Intragenic (intron)	0	21	0	0	GeneID:13591	Other transcription factor;Nucleic acid binding
chr11:44481229	chr11:44481550-44481702	4	Ebf1	27199	Intragenic (intron)	0	21	0	0	GeneID:13591	Other transcription factor;Nucleic acid binding
chr18:34456911	chr18:34457183-34457256	4	Epb4.1l4a	10936	5' end	0	0	0	0	GeneID:13624	Protein phosphatase
chr6:65126818	chr6:65127106-65127128	4	Smarcad1	20583	5' end	7	51	5	11	GeneID:13990	DNA helicase;Hydrolase
chr12:33800838	chr12:33801336-33801442	4	Etv1	1780	5' end	0	0	0	33	GeneID:14009	Other transcription factor;Nucleic acid binding
chr4:114253940	chr4:114254330-114254464	4	Otx3	6364	3' end	0	0	0	0	GeneID:140477	Homeobox transcription factor;Nucleic acid binding
chr5:94406844	chr5:94407193-94407226	4	Bmp2k	0	Intragenic (exon)	0	0	0	0	GeneID:140780	Non-receptor serine/threonine protein kinase
chr17:62166299	chr17:62166821-62166980	4	Fert2	123112	Gene desert	1	36	0	7	GeneID:14158	Non-receptor tyrosine protein kinase
chr19:23497371	chr19:23498028-23498116	4	Fxn	36821	3' end	0	0	0	0	GeneID:14297	Other miscellaneous function protein
chr1:89842704	chr1:89843210-89843383	4	Gbx2	3634	5' end	0	0	0	0	GeneID:14472	Homeobox transcription factor;Nucleic acid binding
chr12:20061899	chr12:20062261-20062300	4	Id2	2856	5' end	30	61	20	250	GeneID:15902	Other transcription factor
chr13:47748373	chr13:47748586-47748715	4	Id4	8702	5' end	0	7	7	6	GeneID:15904	Other transcription factor
chr13:47655814	chr13:47656885-47657123	4	Id4	100348	Gene desert	0	7	7	6	GeneID:15904	Other transcription factor
chr5:106115709	chr5:106116208-106116335	4	Idua	622	5' end	0	0	0	32	GeneID:15932	Glycosidase
chr10:87675097	chr10:87675224-87675467	4	Igf1	3338	3' end	0	0	0	0	GeneID:16000	Growth factor
chr3:7602899	chr3:7603230-7603384	4	Il7	2513	Intragenic (intron)	0	0	0	0	GeneID:16196	Interleukin
chr13:68788098	chr13:68788567-68788632	4	Irx2	282460	Gene desert	0	0	0	0	GeneID:16372	Homeobox transcription factor;Nucleic acid binding
chr8:127959588	chr8:127959777-127959999	4	Itgb1	145168	Gene desert	211	268	216	497	GeneID:16412	Other receptor;Cell adhesion molecule
chr11:110452636	chr11:110452920-110453081	4	Kcnj16	186197	Gene desert	0	0	0	28	GeneID:16517	Voltage-gated potassium channel
chr6:145438845	chr6:145438964-145439011	4	Kras	28787	5' end	13	23	22	13	GeneID:16653	Small GTPase
chr13:22367999	chr13:22372858-22373014	4	V1rh6	18435	5' end	0	0	0	0	GeneID:171249	G-protein coupled receptor
chr10:3143505	chr10:3143794-3143994	4	V1rg10	116935	Gene desert (3' end)	0	0	0	0	GeneID:171266	G-protein coupled receptor
chr18:75876147	chr18:75876280-75876324	4	Smad7	22449	5' end	19	239	0	1	GeneID:17131	Other transcription factor
chr10:41305200	chr10:41305398-41305482	4	Nical	50000	5' end	0	3	4	107	GeneID:171580	Molecular function unclassified
chr9:109893948	chr9:109894394-109894574	4	Mtap4	20767	5' end	95	218	347	206	GeneID:17758	Non-motor microtubule binding protein
chr9:50261130	chr9:50261499-50261616	4	Ncam1	336176	Gene desert	0	3	84	58	GeneID:17967	CAM family adhesion molecule
chr15:67340337	chr15:67340825-67340845	4	Ndrg1	101546	Gene desert	39	35	0	1	GeneID:17988	Molecular function unclassified
chr13:40903918	chr13:40904086-40904178	4	Nedd9	28504	Intragenic (intron)	0	0	0	38	GeneID:18003	Other cytoskeletal proteins
chr3:102535143	chr3:102535693-102535948	4	Ngfb	84563	3' end	20	0	0	0	GeneID:18049	Neurotrophic factor
chr3:102647764	chr3:102648246-102648391	4	Ngfb	197061	Gene desert	20	0	0	0	GeneID:18049	Neurotrophic factor
chr12:12951214	chr12:12951455-12951542	4	Nmyc1	107673	Gene desert	97	263	20	2	GeneID:18109	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr12:12926537	chr12:12926810-12926958	4	Nmyc1	132287	Gene desert	97	263	20	2	GeneID:18109	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr13:58029335	chr13:58029750-58029769	4	Ntrk2	26810	Intragenic (intron)	3	0	0	399	GeneID:18212	Tyrosine protein kinase receptor;Protein kinase
chr11:53779863	chr11:53780191-53780257	4	P4ha2	8820	3' end	0	0	0	0	GeneID:18452	Hydroxylase
chr9:113204445	chr9:113204685-113204845	4	Pdcd6ip	464809	Gene desert	0	20	54	63	GeneID:18571	Transmembrane receptor regulatory/adaptor protein
chr5:136395146	chr5:136395281-136395434	4	Pdgfa	1348	5' end	0	0	0	0	GeneID:18590	Growth factor
chr11:106395139	chr11:106395678-106395730	4	Pecam1	9314	Intragenic (intron)	10	0	43	0	GeneID:18613	Immunoglobulin receptor family member;Other cell adhesion molecule;Defense/immunity protein
chr3:148502552	chr3:148502847-148503011	4	Prkacb	1115592	Gene desert (5' end)	1	2	33	29	GeneID:18749	Protein kinase
chr2:134709228	chr2:134709606-134709889	4	Plcb1	259990	Gene desert	0	0	0	0	GeneID:18795	Phospholipase;Select calcium binding protein
chr2:135363555	chr2:135363850-135363959	4	Plcb4	102002	Gene desert	4	3	12	3	GeneID:18798	Phospholipase;Select calcium binding protein
chr9:112121642	chr9:112122180-112122345	4	Arpp21	1278	Intragenic (intron)	0	0	0	0	GeneID:19050	Protein phosphatase;Esterase
chr15:10007697	chr15:10008091-10008150	4	Prlr	72051	5' end	0	6	0	0	GeneID:19116	Other receptor
chr5:19343572	chr5:19343890-19343924	4	Ptpn12	21981	3' end	0	0	0	9	GeneID:19248	Protein phosphatase
chr10:21946624	chr10:21947117-21947129	4	Raet1a	28527	3' end	0	0	0	0	GeneID:19368	Other receptor;Other defense and immunity protein
chr8:77627150	chr8:77628971-77629123	4	Rbmxrt	6000	5' end	199	0	260	69	GeneID:19656	Ribonucleoprotein
chr7:26472104	chr7:26472385-26472454	4	C80913	8940	5' end	22	53	0	16	GeneID:19777	Transcription cofactor;Chaperone;Defense/immunity protein
chr10:89229470	chr10:89229890-89229989	4	Nr1h4	2733	Intragenic (intron)	0	0	0	0	GeneID:20186	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr3:97874071	chr3:97874647-97874754	4	Sec22l1	26315	5' end	3	9	0	29	GeneID:20333	SNARE protein
chr1:133268331	chr1:133268462-133268524	4	Sox13	104	3' end	25	34	16	4	GeneID:20668	HMG box transcription factor;Nucleic acid binding
chr6:144512231	chr6:144512638-144512796	4	Sox5	152408	Gene desert	0	0	0	90	GeneID:20678	HMG box transcription factor;Nucleic acid binding
chr13:33352597	chr13:33352791-33352834	4	Serpinb6a	1782	3' end	0	0	0	0	GeneID:20719	Serine protease inhibitor
chr3:28604134	chr3:28604352-28604435	4	Eif5a2	38794	3' end	19	6	0	2	GeneID:208691	Translation initiation factor
chr1:93132373	chr1:93134283-93134492	4	Sned1	38225	5' end	34	3	4	30	GeneID:208777	Receptor;Membrane-bound signaling molecule;Defense/immunity protein

chr2.32820798	chr2:32820891-32821167	4	Stxbp1	13874	5' end	7	6	7	80	GeneID:20910	Membrane traffic regulatory protein
chr14.62128297	chr14:62128479-62128551	4	Slc39a14	693	5' end	10	16	0	0	GeneID:213053	Transporter
chr17.7819911	chr17:7820010-7820241	4	T2	1325	Intragenic (intron)	0	0	0	0	GeneID:21331	Molecular function unclassified
chr17.82924669	chr17:82924902-82924936	4	Plekhh2	157882	Gene desert	0	7	4	0	GeneID:213556	Transfer/carrier protein
chrX.69124724	chrX:69125095-69125144	4	Tbl1x	65421	5' end	23	97	43	42	GeneID:21372	Other miscellaneous function protein
chr5.116770995	chr5:116771110-116771181	4	Tbx3	384768	Gene desert	3	290	8	0	GeneID:21386	Other transcription factor;Nucleic acid binding
chr1.4867040	chr1:4867595-4867686	4	Tcea1	2022	5' end	41	6	0	70	GeneID:21399	Basal transcription factor;Nucleic acid binding
chr10.61262636	chr10:61262844-61262943	4	Lrrc20	19414	Intragenic (intron)	0	0	0	0	GeneID:216011	Molecular function unclassified
chr11.115836225	chr11:115837684-115837894	4	Fbf1	736	5' end	1	51	111	84	GeneID:217335	Molecular function unclassified
chr11.117256676	chr11:117256820-117256862	4	Tmc6	180275	Gene desert	0	8	9	8	GeneID:217353	Molecular function unclassified
chr12.15937882	chr12:15938117-15938232	4	Trib2	535	5' end	1	0	0	1152	GeneID:217410	Protein kinase;Other miscellaneous function protein
chr12.95478181	chr12:95478807-95478836	4	9030617O03Rik	2428	Intragenic (intron)	0	0	0	45	GeneID:217830	Molecular function unclassified
chr13.17958773	chr13:17958974-17959087	4	Pou6f2	87763	5' end	0	0	0	0	GeneID:218030	Homeobox transcription factor;Nucleic acid binding
chr13.105071301	chr13:105071869-105072073	4	Depdc1b	14259	5' end	0	6	0	0	GeneID:218581	Molecular function unclassified
chr14.13081479	chr14:13082174-13082175	4	Rarb	62	3' end	0	0	0	0	GeneID:218772	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr19.14063305	chr19:14063931-14063950	4	Tle4	188981	Gene desert	0	0	0	6	GeneID:21888	Transcription cofactor;Other miscellaneous function protein
chr19.14111976	chr19:14112696-14112782	4	Tle4	237780	Gene desert	0	0	0	6	GeneID:21888	Transcription cofactor;Other miscellaneous function protein
chr15.93822131	chr15:93822877-93823000	4	Pphln1	1000	5' end	31	0	6	38	GeneID:223828	Molecular function unclassified
chr16.16343366	chr16:16343701-16343773	4	Fgd4	6940	Intragenic (intron)	0	0	0	0	GeneID:224014	Guanyl-nucleotide exchange factor
chr17.48606667	chr17:48606815-48607043	4	Plcl2	81620	5' end	0	0	21	0	GeneID:224860	Phospholipase;Select calcium binding protein
chr17.69201037	chr17:69201282-69201363	4	Dlgap1	6597	Intragenic (intron)	0	0	0	0	GeneID:224997	Molecular function unclassified
chr18.30823264	chr18:30823475-30823629	4	Pik3c3	26116	3' end	0	0	0	0	GeneID:225326	Other kinase
chr5.31023132	chr5:31023385-31023411	4	Yes1	4891	5' end	15	35	0	15	GeneID:22612	Non-receptor tyrosine protein kinase
chr1.119399598	chr1:119400117-119400174	4	Epb4.1l5	1390	Intragenic (intron)	0	2	11	33	GeneID:226352	Molecular function unclassified
chr1.136615590	chr1:136615932-136616005	4	Zfp281	99781	5' end	38	10	6	13	GeneID:226442	KRAB box transcription factor;Nucleic acid binding
chr1.153068936	chr1:153069689-153069776	4	Nmnat2	1443	Intragenic (intron)	9	0	1	68	GeneID:226518	Nucleotidyltransferase
chr17.22199943	chr17:22200285-22200346	4	Zfp13	5928	5' end	0	10	0	14	GeneID:22654	KRAB box transcription factor;Nucleic acid binding
chr2.26402657	chr2:26402882-26403074	4	AU024582	12805	5' end	28	0	1	25	GeneID:227648	Molecular function unclassified
chr2.32122561	chr2:32122717-32122960	4	5830434P21Rik	11524	Intragenic (intron)	86	65	154	207	GeneID:227723	Transcription factor;Nuclease
chr2.34359456	chr2:34359616-34359808	4	Mapkap1	7229	5' end	0	0	0	11	GeneID:227743	Molecular function unclassified
chr2.44498046	chr2:44498373-44498397	4	Gtdc1	26017	3' end	14	4	8	20	GeneID:227835	Molecular function unclassified
chr2.121574056	chr2:121575300-121575499	4	A930004K21Rik	19360	5' end	17	1	113	7	GeneID:220564	Molecular function unclassified
chr2.169922323	chr2:169922761-169922846	4	Sdccag33l	23839	Intragenic (intron)	0	0	0	22	GeneID:228911	Nuclease
chr4.117938095	chr4:117938223-117938309	4	Foxj3	6801	Intragenic (intron)	0	0	0	0	GeneID:230700	Other transcription factor;Nucleic acid binding
chr5.111089064	chr5:111089362-111089656	4	D5Ertd40e	37272	5' end	0	4	0	8	GeneID:231630	Molecular function unclassified
chr6.130025281	chr6:130025412-130025564	4	Clec12a	0	Intragenic (exon)	0	0	0	0	GeneID:232413	Other receptor
chr7.17660755	chr7:17661022-17661190	4	Fbxo27	185	5' end	0	0	0	0	GeneID:233040	Other receptor;Other defense and immunity protein
chr9.20414864	chr9:20415249-20415405	4	Zfp426	194	5' end	0	0	0	55	GeneID:235028	KRAB box transcription factor
chr9.49170125	chr9:49170475-49170588	4	Usp28	11968	3' end	25	40	33	29	GeneID:235323	Molecular function unclassified
chrX.42093342	chrX:42093676-42093769	4	Olfr1321	17505	5' end	0	0	0	0	GeneID:236785	G-protein coupled receptor
chr10.7282614	chr10:7282892-7283132	4	Lrp11	3661	Intragenic (intron)	0	28	6	0	GeneID:237253	Molecular function unclassified
chr11.29778520	chr11:29778787-29778927	4	C230094A16Rik	1147	Intragenic (intron)	0	0	0	0	GeneID:237711	Molecular function unclassified
chr9.59681388	chr9:59681626-59681720	4	Arih1	12619	Intragenic (intron)	44	60	8	29	GeneID:23806	Ubiquitin-protein ligase
chr12.57924719	chr12:57925259-57925383	4	Lrfn5	866510	Gene desert (3' end)	0	0	0	0	GeneID:238205	Nuclease
chr14.109197317	chr14:109197790-109197797	4	Gpc6	55377	5' end	0	2	0	55	GeneID:23888	Cell adhesion molecule;Extracellular matrix glycoprotein
chr14.109616157	chr14:109616655-109616859	4	Gpc6	109444	Intragenic (intron)	0	2	0	55	GeneID:23888	Cell adhesion molecule;Extracellular matrix glycoprotein
chr14.23713318	chr14:23713704-23713807	4	D14Ertd171e	20713	Intragenic (intron)	0	0	0	0	GeneID:238988	Other enzyme regulator
chr5.120663990	chr5:120664176-120664311	4	Rhof	2653	5' end	0	0	0	0	GeneID:23912	Small GTPase
chr15.92316489	chr15:92316916-92317091	4	Muc19	29	Intragenic (intron)	0	0	0	0	GeneID:239611	Extracellular matrix glycoprotein
chr17.31678944	chr17:31679262-31679425	4	D10628	41639	5' end	0	0	0	22	GeneID:240068	KRAB box transcription factor;Nucleic acid binding
chr1.133582534	chr1:133582847-133582997	4	Lax1	4763	5' end	0	0	0	0	GeneID:240754	Molecular function unclassified
chr3.92189729	chr3:92189969-92190047	4	Pglyrp3	49898	5' end	0	0	0	0	GeneID:242100	Other receptor
chr4.82936545	chr4:82936956-82937021	4	Bnc2	12467	Intragenic (intron)	0	0	0	0	GeneID:242509	Other zinc finger transcription factor
chr4.82802899	chr4:82803016-82803143	4	Bnc2	17606	Intragenic (intron)	0	0	0	0	GeneID:242509	Other zinc finger transcription factor
chr4.97177530	chr4:97178037-97178037	4	BC060737	30225	5' end	0	0	0	0	GeneID:242553	Transcription factor;Nuclease
chr5.23965941	chr5:23966369-23966404	4	Gm443	39823	3' end	0	0	0	0	GeneID:242891	Chaperonin
chr5.125997513	chr5:125997893-125997927	4	C630028F04Rik	10579	5' end	0	0	0	0	GeneID:243274	Molecular function unclassified
chr6.4824376	chr6:4824686-4824781	4	Ppp1r9a	2831	Intragenic (intron)	0	0	0	0	GeneID:243725	Non-motor actin binding protein
chr6.4738440	chr6:4738773-4738903	4	Ppp1r9a	12468	Intragenic (intron)	0	0	0	0	GeneID:243725	Non-motor actin binding protein
chr7.25077506	chr7:25077745-25077891	4	Zfp537	53569	5' end	0	0	0	0	GeneID:243931	Nuclease
chr7.25993382	chr7:25993506-25993767	4	Zfp536	14062	Intragenic (intron)	0	0	0	40	GeneID:243937	Molecular function unclassified
chr7.32095379	chr7:32095825-32096012	4	D030014N22Rik	26637	3' end	16	0	0	0	GeneID:243963	KRAB box transcription factor
chr8.14374452	chr8:14374810-14374977	4	Dlgap2	90263	Intragenic (intron)	0	0	0	0	GeneID:244310	Molecular function unclassified

chr8.25665360	chr8:25666090-25666134	4	Spfh2	143826	Gene desert	2	0	0	10	GeneID:244373	Molecular function unclassified
chr9.55802295	chr9:55802572-55802581	4	Zfp291	4125	Intragenic (intron)	0	0	0	0	GeneID:244891	
chr2.163094720	chr2:163095020-163095166	4	BC037708	474	Intragenic (intron)	0	0	0	9	GeneID:245866	Molecular function unclassified
chr12.35298179	chr12:35298453-35298537	4	Zfp277	35657	3' end	0	0	0	0	GeneID:246196	Nucleic acid binding
chr7.95464967	chr7:95465387-95465482	4	Olfr487	15182	3' end	0	0	0	0	GeneID:258042	G-protein coupled receptor
chr7.72689714	chr7:72689827-72689959	4	Olfr290	635480	Gene desert (3' end)	0	0	0	0	GeneID:258411	G-protein coupled receptor
chr7.94013008	chr7:94013352-94013443	4	Olfr697	4073	5' end	0	0	0	0	GeneID:258592	G-protein coupled receptor
chr9.38890399	chr9:38890585-38890731	4	Olfr926	2112	5' end	0	0	0	0	GeneID:258811	G-protein coupled receptor
chr9.38423336	chr9:38423573-38423628	4	Olfr147	3793	3' end	0	0	0	0	GeneID:258869	G-protein coupled receptor
chr7.6817395	chr7:6817872-6818022	4	Ehd2	140269	Gene desert (3' end)	0	0	0	0	GeneID:259300	Calmodulin related protein;Membrane traffic regulatory protein
chr7.6811931	chr7:6812760-6812883	4	Ehd2	145395	Gene desert (5' end)	0	0	0	0	GeneID:259300	Calmodulin related protein;Membrane traffic regulatory protein
chr7.10283993	chr7:10284258-10284402	4	Psg19	29351	3' end	0	0	0	0	GeneID:26439	CAM family adhesion molecule
chr11.116670617	chr11:116671376-116671528	4	GnT-IX	13383	3' end	0	19	0	0	GeneID:268510	Glycosyltransferase
chr17.26089977	chr17:26090324-26090413	4	Grm4	726	5' end	0	0	0	0	GeneID:268934	G-protein coupled receptor
chr1.87119635	chr1:87119861-87120022	4	Eif4e2	623	5' end	0	0	0	0	GeneID:26987	Translation initiation factor
chr8.70116251	chr8:70116811-70116974	4	Bpy2ip1	160201	Gene desert	49	7	39	59	GeneID:270058	Microtubule family cytoskeletal protein
chr8.70123104	chr8:70123547-70123565	4	Bpy2ip1	166865	Gene desert	49	7	39	59	GeneID:270058	Microtubule family cytoskeletal protein
chr8.112949849	chr8:112950172-112950269	4	5033413H12Rik	37755	5' end	3	41	0	1	GeneID:270096	Molecular function unclassified
chr6.31827662	chr6:31827778-31827893	4	Podxl	387465	Gene desert	0	0	226	0	GeneID:27205	Extracellular matrix glycoprotein
chr14.22273994	chr14:22274362-22274382	4	1110051B16Rik	8027	5' end	8	0	0	0	GeneID:278672	Homeobox transcription factor;Nucleic acid binding
chr9.110321168	chr9:110321458-110321527	4	Cspg5	61124	5' end	0	0	0	168	GeneID:29873	Extracellular matrix glycoprotein
chr8.35509696	chr8:35510587-35510618	4	6430573F11Rik	26364	3' end	0	0	0	0	GeneID:319582	Molecular function unclassified
chr13.81383060	chr13:81383610-81383644	4	Smo	270740	Gene desert	0	0	0	0	GeneID:319757	G-protein coupled receptor
chr11.52841935	chr11:52842369-52842433	4	B230374F23Rik	15315	3' end	1	0	0	0	GeneID:320027	CAM family adhesion molecule
chr1.180964213	chr1:180964591-180964737	4	Lefty2	4231	5' end	194	125	4	0	GeneID:320202	TGF-beta superfamily member
chr10.125585460	chr10:125585949-125586100	4	Lrig3	168488	Gene desert	15	0	0	10	GeneID:320398	Other cell adhesion molecule
chr12.47077491	chr12:47078032-47078091	4	D930036F22Rik	17701	3' end	5	16	6	0	GeneID:320487	Molecular function unclassified
chr6.125778272	chr6:125778602-125778753	4	A930037G23Rik	0	Intragenic (exon)	2	0	0	5	GeneID:320678	Molecular function unclassified
chr16.40479192	chr16:40482263-40482333	4	D930030D11Rik	590217	Gene desert (5' end)	0	0	0	0	GeneID:320874	Molecular function unclassified
chr13.64419566	chr13:64420081-64420468	4	Zfp459	2141	Intragenic (intron)	0	0	0	0	GeneID:328274	KRAB box transcription factor
chr14.21610386	chr14:21610588-21610685	4	Rai17	144374	Gene desert	0	0	0	39	GeneID:328365	Other ligase
chr10.03336717	chr10:03337345-03337360	4	C330029D10Rik	99424	5' end	47	0	0	7	GeneID:329003	Molecular function unclassified
chr2.120025427	chr2:120025734-120025840	4	2310026J01Rik	473	5' end	0	0	0	0	GeneID:329502	Phospholipase
chr6.5050224	chr6:5050454-5050525	4	Pon2	26873	5' end	0	0	0	114	GeneID:330260	Peroxidase;Esterase
chr6.120687841	chr6:120687952-120688177	4	AB114826	4748	5' end	0	0	0	0	GeneID:330406	Molecular function unclassified
chr10.8034448	chr10:8035100-8035239	4	Ust	35141	5' end	0	0	0	0	GeneID:338362	Other transferase
chr10.7786682	chr10:7786969-7787000	4	Ust	90746	3' end	0	0	0	0	GeneID:338362	Other transferase
chr15.98397964	chr15:98398819-98398978	4	D15Ertd405e	25167	5' end	0	0	0	31	GeneID:380967	Molecular function unclassified
chr4.60295397	chr4:60295813-60295888	4	MGC107671	81888	3' end	0	0	0	0	GeneID:381530	Other transfer/carrier protein
chr8.119871215	chr8:119871923-119871967	4	Gse1	66280	5' end	14	42	6	36	GeneID:382034	Molecular function unclassified
chr15.57812230	chr15:57812798-57812965	4	Zhx2	77078	5' end	0	0	0	0	GeneID:387609	Homeobox transcription factor;Other zinc finger transcription factor
chr19.54645140	chr19:54645257-54645363	4	Acsl5	1127	Intragenic (intron)	0	0	0	0	GeneID:433256	Other ligase
chr1.3484950	chr1:3485951-3486042	4	AY534250	61238	Intragenic (intron)	0	0	0	0	GeneID:497097	Molecular function unclassified
chr10.79391124	chr10:79391515-79391753	4	Ppap2c	30316	3' end	0	0	0	0	GeneID:50784	Other phosphatase
chr17.10478174	chr17:10478636-10478800	4	Park2	1045	Intragenic (intron)	0	0	0	0	GeneID:50873	Other transfer/carrier protein;Other ligase
chr4.53317631	chr4:53318085-53318137	4	Tmem38b	212477	Gene desert (3' end)	0	8	0	0	GeneID:52076	Nuclease
chr7.126487742	chr7:126487864-126488130	4	Ppp2r2d	3503	5' end	279	422	35	161	GeneID:52432	Protein phosphatase
chr13.113876818	chr13:113877361-113877466	4	Parp8	5758	Intragenic (intron)	0	0	0	5	GeneID:52552	Molecular function unclassified
chr10.72819481	chr10:72819767-72819853	4	Zwint	366067	Gene desert	25	11	201	243	GeneID:52696	Molecular function unclassified
chr5.110689830	chr5:110690115-110690323	4	Rutbc2	22485	5' end	9	30	0	0	GeneID:52850	Other G-protein modulator;Membrane traffic regulatory protein
chr10.59918058	chr10:59918231-59918389	4	Chst3	15335	3' end	0	0	0	9	GeneID:53374	Other transferase
chr4.6042933	chr4:6043309-6043431	4	Sdcbp	174880	Gene desert	24	102	144	175	GeneID:53378	Membrane traffic regulatory protein
chr5.71099109	chr5:71099754-71099800	4	Corin	6804	Intragenic (intron)	0	0	0	0	GeneID:53419	Receptor;Serine protease
chr19.54917672	chr19:54917885-54918005	4	Vti1a	44972	Intragenic (intron)	0	6	0	23	GeneID:53611	SNARE protein
chr6.136741828	chr6:136742165-136742232	4	Atf7ip	50000	3' end	3	56	6	7	GeneID:54343	Transcription cofactor;Hydrolase
chr6.149101864	chr6:149102201-149102351	4	Tera	10000	5' end	14	119	0	0	GeneID:56306	Molecular function unclassified
chr3.52498446	chr3:52499245-52499421	4	Foxo1	64303	3' end	0	13	0	7	GeneID:56458	Other transcription factor;Nucleic acid binding
chr8.84752715	chr8:84752984-84753142	4	Sast	0	Intragenic (intron)	0	0	0	0	GeneID:56527	Non-receptor serine/threonine protein kinase
chr6.123984089	chr6:123984388-123984495	4	Clec4e	42780	5' end	0	0	0	0	GeneID:56619	Other receptor
chr10.33622130	chr10:33623587-33623755	4	Sult3a1	60851	5' end	0	0	0	0	GeneID:57430	Other transferase
chr19.49787492	chr19:49787624-49787837	4	Sorcs1	35352	Intragenic (intron)	0	0	0	48	GeneID:58178	Other receptor
chr6.141724817	chr6:141725257-141725320	4	Slco1c1	26659	3' end	0	0	0	0	GeneID:58807	Other transporter
chr10.75873094	chr10:75873340-75873473	4	Suhw2	30084	5' end	0	0	0	12	GeneID:64453	Nuclease

chr6.88790214	chr6:88790741-88790817	4	Eefsec	8210	5' end	30	76	64	12	GeneID:65967	Translation elongation factor
chr15.100333389	chr15:100333931-100334012	4	D15Ertd366e	2381	5' end	3	25	0	36	GeneID:65970	Other cytoskeletal proteins
chr14.3853734	chr14:3853842-3854014	4	Abhd6	459	3' end	2	1	0	21	GeneID:66082	Lipase
chr7.128714136	chr7:128714413-128714438	4	Ifitm6	7953	5' end	0	0	0	0	GeneID:66141	Other miscellaneous function protein
chr3.10420535	chr3:10420790-10420880	4	Zfand1	10730	5' end	26	13	3	37	GeneID:66361	Molecular function unclassified
chr4.44533058	chr4:44533190-44533395	4	Exosc3	30000	3' end	11	0	0	9	GeneID:66362	Exoribonuclease;Esterase
chr8.106445483	chr8:106445888-106445988	4	1810044O22Rik	14409	5' end	10	0	6	59	GeneID:66427	Oxidase
chr1.67095543	chr1:67095858-67095989	4	Rpe	882	5' end	3	0	3	105	GeneID:66646	Epimerase/racemase
chr4.115719069	chr4:115719305-115719449	4	4931406I20Rik	1576	Intragenic (intron)	67	151	102	36	GeneID:66743	Molecular function unclassified
chr2.59565987	chr2:59566290-59566554	4	1200003E16Rik	15590	Intragenic (intron)	8	9	0	2	GeneID:66860	Molecular function unclassified
chr9.121342197	chr9:121342796-121342905	4	2310001H13Rik	19513	Intragenic (intron)	13	29	0	12	GeneID:67095	Molecular function unclassified
chr9.121336008	chr9:121336258-121336393	4	2310001H13Rik	26038	Intragenic (intron)	13	29	0	12	GeneID:67095	Molecular function unclassified
chr18.86017576	chr18:86017743-86017872	4	1700034H14Rik	554296	Gene desert (5' end)	3	0	0	22	GeneID:67105	Molecular function unclassified
chr9.120674070	chr9:120674183-120674248	4	Rpl14	105028	Gene desert	3	71	0	24	GeneID:67115	Ribosomal protein
chr11.72220275	chr11:72220947-72221027	4	UBE2G1	797	5' end	32	0	0	0	GeneID:67128	Other ligase
chr19.25884009	chr19:25884421-25884481	4	Smarca2	3426	5' end	37	2	0	42	GeneID:67155	Transcription cofactor;DNA helicase;Hydrolase
chrX.134228998	chrX:134229112-134229228	4	Glt28d1	12859	3' end	0	0	0	0	GeneID:67574	Molecular function unclassified
chr9.3906264	chr9:3906742-3907004	4	Aasdhppt	311713	Gene desert (3' end)	5	4	0	0	GeneID:67618	Molecular function unclassified
chr2.26015261	chr2:26015553-26015671	4	Btbd14a	3416	3' end	0	0	0	0	GeneID:67991	Molecular function unclassified
chr5.71859404	chr5:71859671-71859741	4	Ociad1	73744	5' end	34	146	26	65	GeneID:68095	Molecular function unclassified
chr4.12589188	chr4:12589693-12589861	4	6720467C03Rik	491457	Gene desert (5' end)	14	0	0	83	GeneID:68099	Molecular function unclassified
chr18.68023710	chr18:68024615-68024709	4	Spire1	658	Intragenic (intron)	10	56	27	81	GeneID:68166	Molecular function unclassified
chr11.79748149	chr11:79748819-79748954	4	1110002N22Rik	348	3' end	10	0	22	6	GeneID:68550	Molecular function unclassified
chr11.79744167	chr11:79744713-79744745	4	1110002N22Rik	4505	3' end	10	0	22	6	GeneID:68550	Molecular function unclassified
chr12.80708091	chr12:80708411-80708515	4	1110014C03Rik	11635	5' end	80	258	120	162	GeneID:68581	Transfer/carrier protein;Vesicle coat protein
chr11.29532341	chr11:29532460-29532642	4	Rtn4	55961	5' end	328	355	209	331	GeneID:68585	Transmembrane receptor regulatory/adaptor protein
chr5.19089731	chr5:19090136-19090143	4	Phtf2	48174	3' end	0	0	0	0	GeneID:68770	Homeobox transcription factor;Nucleic acid binding
chr16.8524582	chr16:8524852-8525006	4	AK007485	3901	Intragenic (intron)	0	0	0	0	GeneID:69053	Molecular function unclassified
chr16.8780160	chr16:8780560-8780777	4	AK007485	238607	Gene desert	0	0	0	0	GeneID:69053	Molecular function unclassified
chr16.8819496	chr16:8820165-8820213	4	AK007485	278128	Gene desert	0	0	0	0	GeneID:69053	Molecular function unclassified
chr19.38412497	chr19:38413362-38413381	4	Cyp2c66	32059	3' end	0	0	0	0	GeneID:69888	Oxygenase
chr6.40204117	chr6:40204304-40204452	4	Mulk	903	Intragenic (intron)	0	0	0	0	GeneID:69920	Kinase
chr6.54046883	chr6:54047330-54047363	4	Chn2	168503	Gene desert	0	46	0	8	GeneID:69993	Transfer/carrier protein;G-protein modulator
chr9.106286245	chr9:106286440-106286544	4	Wdr51a	9298	Intragenic (intron)	0	16	0	0	GeneID:70235	Molecular function unclassified
chr2.30094884	chr2:30095325-30095514	4	Tbc1d13	526	5' end	25	50	0	20	GeneID:70296	Molecular function unclassified
chr10.75743666	chr10:75743948-75743959	4	Derl3	20221	5' end	0	0	0	7	GeneID:70377	Molecular function unclassified
chr16.34987908	chr16:34988016-34988188	4	Ptplb	40179	3' end	3	114	147	108	GeneID:70757	Protein phosphatase
chr6.66043212	chr6:66043387-66043448	4	Prdm5	2592	3' end	39	7	134	16	GeneID:70779	Zinc finger transcription factor;Nucleic acid binding
chr10.70482339	chr10:70482454-70482574	4	Phyhipl	109628	Gene desert	0	0	0	52	GeneID:70911	Molecular function unclassified
chr6.61092223	chr6:61092612-61092717	4	Mmrn1	4937	3' end	0	0	0	0	GeneID:70945	Molecular function unclassified
chr10.95397563	chr10:95397797-95397891	4	Nudt4	58883	5' end	0	33	78	124	GeneID:71207	Other phosphatase
chr10.67912647	chr10:67913482-67913731	4	Arid5b	2438	Intragenic (intron)	1	17	39	12	GeneID:71371	Transcription cofactor
chr4.10940310	chr4:10940492-10940639	4	Plekhf2	6682	5' end	8	44	100	0	GeneID:71801	Molecular function unclassified
chr4.10858726	chr4:10858764-10858907	4	Plekhf2	56149	3' end	8	44	100	0	GeneID:71801	Molecular function unclassified
chr8.109007542	chr8:109007698-109007739	4	2400003C14Rik	503	5' end	0	11	4	16	GeneID:71955	Molecular function unclassified
chr12.28201815	chr12:28202324-28202356	4	Atxn7l4	33461	5' end	0	0	0	0	GeneID:72174	Molecular function unclassified
chr5.142772361	chr5:142772523-142772637	4	Zfp655	54675	3' end	0	5	0	4	GeneID:72611	KRAB box transcription factor
chr17.74585490	chr17:74585793-74585813	4	2810405J04Rik	565418	Gene desert	0	0	0	0	GeneID:72722	Molecular function unclassified
chr7.18962317	chr7:18962658-18962723	4	Zfp74	40803	5' end	0	0	0	37	GeneID:72723	KRAB box transcription factor;Nucleic acid binding
chr1.120901802	chr1:120901969-120901993	4	Insig2	184486	Gene desert	0	0	38	23	GeneID:72999	Molecular function unclassified
chr8.4116579	chr8:4116816-4117081	4	LOC73072	50000	3' end	0	3	0	31	GeneID:73072	Molecular function unclassified
chr3.138105023	chr3:138105712-138105928	4	Ddit4l	52379	5' end	4	9	0	25	GeneID:73284	Molecular function unclassified
chr8.77131342	chr8:77132053-77132155	4	1700031F13Rik	193784	Gene desert (5' end)	0	0	0	0	GeneID:73301	Molecular function unclassified
chr10.119858982	chr10:119859374-119859567	4	Irak3	924	Intragenic (intron)	0	0	6	2	GeneID:73914	Serine/threonine protein kinase receptor;Protein kinase
chr10.85171410	chr10:85173157-85173280	4	Btbd11	38599	Intragenic (intron)	0	0	0	0	GeneID:74007	Molecular function unclassified
chr3.18455441	chr3:18455655-18456047	4	Arcp	345148	Gene desert (3' end)	46	9	0	10	GeneID:74252	Molecular function unclassified
chr14.41949196	chr14:41949781-41950079	4	Slc35f4	23004	3' end	0	0	0	0	GeneID:75288	Molecular function unclassified
chr8.103607203	chr8:103607374-103607482	4	Cklf	52	5' end	0	1	18	35	GeneID:75458	Chemokine
chr9.14071180	chr9:14071374-14071470	4	Sesn3	42398	5' end	0	14	15	205	GeneID:75747	Molecular function unclassified
chr13.89716930	chr13:89717113-89717205	4	Cmya5	47632	3' end	0	0	0	0	GeneID:76469	Molecular function unclassified
chr11.88361150	chr11:88361389-88361394	4	Msi2h	19006	Intragenic (intron)	96	66	17	371	GeneID:76626	Ribonucleoprotein
chr11.88135807	chr11:88136072-88136114	4	Msi2h	54372	Intragenic (intron)	96	66	17	371	GeneID:76626	Ribonucleoprotein
chr14.91262026	chr14:91262111-91262254	4	2410129H14Rik	1060	Intragenic (intron)	2	0	0	16	GeneID:76789	Molecular function unclassified

chr2.170308728	chr2:170308939-170309150	4	Bcas1	101063	Gene desert	3	0	0	44	GeneID:76960	Molecular function unclassified
chr3.84580595	chr3:84580735-84580870	4	Trim2	3640	Intragenic (intron)	0	0	0	0	GeneID:80890	RING finger transcription factor;Actin binding motor protein
chr4.128888933	chr4:128889189-128889215	4	Pum1	23162	5' end	63	29	2	44	GeneID:80912	Other RNA-binding protein;Translation factor
chr1.167296845	chr1:167297397-167297421	4	Uck2	14247	Intragenic (intron)	0	0	0	0	GeneID:80914	Nucleotide kinase
chr15.10881481	chr15:10881826-10881884	4	C1qtnf3	1180	3' end	0	0	0	0	GeneID:81799	Molecular function unclassified
chr10.6059310	chr10:6059685-6059824	4	Akap12	23503	Intragenic (intron)	0	0	0	0	GeneID:83397	Kinase modulator
chr19.21087224	chr19:21087404-21087644	4	Tmem2	248	Intragenic (intron)	0	29	0	6	GeneID:83921	Molecular function unclassified
chr19.21163926	chr19:21164464-21164515	4	Tmem2	36428	3' end	0	29	0	6	GeneID:83921	Molecular function unclassified
chr1.64403351	chr1:64403885-64403950	4	Klf7	24222	3' end	0	8	0	16	GeneID:93691	KRAB box transcription factor;Nucleic acid binding
chr18.38098555	chr18:38098982-38099000	4	Pcdhga1	6109	5' end	0	0	0	591	GeneID:93709	Cadherin
chr8.87013008	chr8:87013428-87013477	4	Zfp423	28689	Intragenic (intron)	0	0	0	0	GeneID:94187	Molecular function unclassified
chr13.7884034	chr13:7885325-7885706	4	Adarb2	171548	Gene desert	0	0	0	0	GeneID:94191	Nucleic acid binding;Deaminase
chr2.168827400	chr2:168827772-168827896	4	Sall4	2194	5' end	6	197	86	0	GeneID:99377	Zinc finger transcription factor;Nucleic acid binding
chr19.43602192	chr19:43602866-43602871	4	Scd2	27876	5' end	37	47	15	353	mCG10078	Other oxidoreductase
chr5.63226033	chr5:63226348-63226371	4	Klf3	207606	Gene desert	2	0	0	36	mCG10252	Transcription cofactor;Nucleic acid binding
chr10.91193456	chr10:91193571-91193653	4	Tmpo	251435	Gene desert	57	89	24	91	mCG10339	Peptide hormone
chr3.15055237	chr3:15055756-15055786	4	Car2	99278	3' end	9	0	24	27	mCG10347	Dehydratase
chr3.15111810	chr3:15112363-15112564	4	Car2	155970	Gene desert (3' end)	9	0	24	27	mCG10347	Dehydratase
chr9.46993933	chr9:46994251-46994335	4	2900052N01Rik	23617	5' end	0	0	0	0	mCG1035159	Molecular function unclassified
chr8.125078998	chr8:125079138-125079228	4	4933403G14Rik	25898	3' end	0	4	0	0	mCG1035826	Molecular function unclassified
chr6.80811833	chr6:80812176-80812269	4	Lrrtm4	0	Intragenic (intron)	0	0	0	0	mCG1036448	Molecular function unclassified
chr14.30406343	chr14:30406891-30407056	4	Sncg	10165	3' end	0	0	0	0	mCG10518	Other miscellaneous function protein
chr5.65742005	chr5:65742146-65742151	4	Phox2b	40210	5' end	0	0	0	0	mCG10559	Homeobox transcription factor;Nucleic acid binding
chr6.100409073	chr6:100409520-100409575	4	Tpt1	149516	Gene desert	2654	1111	1499	1996	mCG10592	Non-motor microtubule binding protein
chr14.68029940	chr14:68030226-68030345	4	Nufip1	128910	Gene desert	0	0	0	0	mCG10596	Nucleic acid binding
chrX.30914872	chrX:30915098-30915105	4	Cul4b	1710	Intragenic (intron)	46	70	37	0	mCG10889	Other miscellaneous function protein
chr4.125715265	chr4:125715547-125715625	4	Gjb3	16109	5' end	23	152	0	0	mCG10907	Gap junction
chr14.3098114	chr14:3100900-3101032	4	2610042L04Rik	12919	5' end	17	76	240	0	mCG112781	Molecular function unclassified
chr14.3046810	chr14:3047713-3047827	4	2610042L04Rik	18772	3' end	17	76	240	0	mCG112781	Molecular function unclassified
chr14.3037136	chr14:3040245-3040313	4	2610042L04Rik	26263	3' end	17	76	240	0	mCG112781	Molecular function unclassified
chr14.3030592	chr14:3030653-3030775	4	2610042L04Rik	35828	3' end	17	76	240	0	mCG112781	Molecular function unclassified
chr15.64050172	chr15:64050540-64050503	4	Adcy0	109110	Gene desert	0	0	0	15	mCG112052	Adenylate cyclase
chr10.59895965	chr10:59896210-59896361	4	Spock2	7205	3' end	0	0	0	18	mCG11290	Cysteine protease inhibitor;Select calcium binding protein
chr8.22957898	chr8:22959098-22959175	4	Zmat4	92029	3' end	0	0	0	15	mCG113560	Molecular function unclassified
chr14.113844530	chr14:113844700-113844740	4	Slc15a1	1154	5' end	0	0	0	0	mCG114131	Other transporter
chr12.99799877	chr12:99800073-99800137	4	2900070E19Rik	45357	3' end	7	3	41	43	mCG1145	Reductase
chr13.94012587	chr13:94012793-94012851	4	Enc1	29258	3' end	136	178	108	186	mCG114601	Protease
chr4.97971563	chr4:97972301-97972309	4	Foxd3	14267	5' end	25	31	40	0	mCG114630	Molecular function unclassified
chr2.158864028	chr2:158864339-158864368	4	Dhx35	369	Intragenic (intron)	6	17	0	30	mCG114772	RNA helicase;Hydrolase
chr1.63573255	chr1:63573383-63573507	4	Eef1b2	923	3' end	0	0	0	0	mCG114996	Translation elongation factor
chr1.63530429	chr1:63530780-63530872	4	Ndufs1	5322	3' end	0	0	0	0	mCG114998	Dehydrogenase;Reductase
chr12.73269536	chr12:73269799-73269841	4	Gphn	254349	Gene desert	0	0	0	0	mCG115252	Other miscellaneous function protein
chr5.3339016	chr5:3339286-3339288	4	Cdk6	11029	5' end	0	0	0	0	mCG115258	Non-receptor serine/threonine protein kinase
chr1.53582609	chr1:53582757-53582802	4	Pms1	2291	5' end	0	0	0	0	mCG115459	Nucleic acid binding
chrX.152474996	chrX:152475173-152475334	4	Reps2	1493	Intragenic (intron)	0	0	0	0	mCG116086	Molecular function unclassified
chr9.100889551	chr9:100890008-100890225	4	Stag1	75000	3' end	17	1	19	40	mCG116377	Chromatin/chromatin-binding protein
chr5.101294148	chr5:101294278-101294454	4	Klhl8	20992	5' end	0	0	0	11	mCG1165	Molecular function unclassified
chr13.28159952	chr13:28160360-28160445	4	Sox4	168322	Gene desert	108	206	228	1384	mCG11673	HMG box transcription factor;Nucleic acid binding
chr1.59850743	chr1:59851721-59851814	4	Fzd7	33417	5' end	0	0	0	0	mCG117844	G-protein coupled receptor
chr15.55496776	chr15:55497274-55497354	4	Col14a1	10057	Intragenic (intron)	0	0	0	0	mCG11867	Extracellular matrix structural protein
chr6.51490650	chr6:51490924-51491051	4	Snx10	0	Intragenic (intron)	62	16	0	85	mCG119117	Molecular function unclassified
chr1.21077136	chr1:21077553-21077581	4	Mcm3	47247	5' end	585	666	423	315	mCG119126	Replication origin binding protein
chr10.52314487	chr10:52314743-52314868	4	Dcbld1	9801	Intragenic (intron)	3	12	0	4	mCG119930	Other receptor
chrX.140928417	chrX:140928869-140928981	4	Fgd1	6095	5' end	0	90	0	37	mCG120274	Guanyl-nucleotide exchange factor
chr10.94950184	chr10:94950331-94950451	4	Cradd	0	Intragenic (exon)	19	136	0	0	mCG120384	Other miscellaneous function protein
chr10.94995839	chr10:94996033-94996049	4	Cradd	45142	Intragenic (intron)	19	136	0	0	mCG120384	Other miscellaneous function protein
chr14.41328084	chr14:41328347-41328430	4	Otx2	3457	3' end	106	10	66	0	mCG12117	Homeobox transcription factor;Nucleic acid binding
chr14.98222564	chr14:98222742-98222778	4	Spry2	24574	5' end	0	0	0	8	mCG121239	Other signaling molecule;Other transcription factor
chr4.19064125	chr4:19064692-19064724	4	Cngb3	143240	Gene desert	0	0	0	0	mCG121247	Cyclic nucleotide-gated ion channel;Ion channel
chr17.63227918	chr17:63228167-63228282	4	Man2a1	119088	Gene desert	0	0	0	28	mCG121684	Glycosidase
chr13.19456972	chr13:19457959-19458011	4	Elmo1	87733	5' end	0	0	0	50	mCG121959	Other signaling molecule
chr18.36697030	chr18:36697172-36697261	4	Pura	27007	5' end	8	39	3	69	mCG121964	Other transcription factor;Other DNA-binding protein
chr18.36858381	chr18:36858979-36859071	4	Pfdn1	11365	Intragenic (intron)	0	0	0	0	mCG121970	Molecular function unclassified

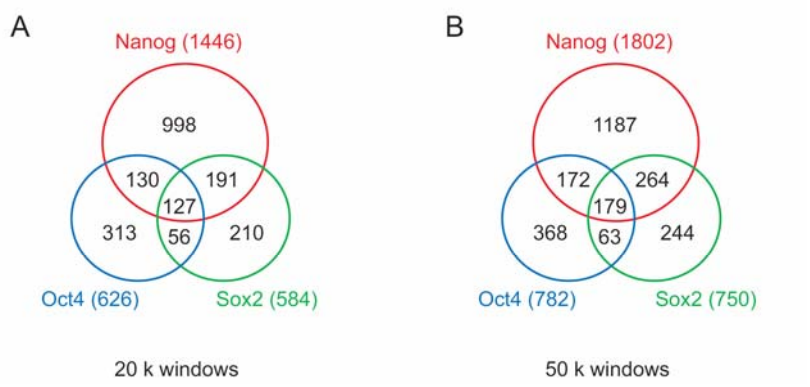
chr10.61111333	chr10:61111717-61111871	4	Pald	291	Intragenic (intron)	52	15	29	10	mCG122064	Molecular function unclassified
chr10.61007932	chr10:61009087-61009245	4	Pald	67000	3' end	52	15	29	10	mCG122064	Molecular function unclassified
chr10.61000700	chr10:61001084-61001185	4	Pald	75000	3' end	52	15	29	10	mCG122064	Molecular function unclassified
chr17.34451029	chr17:34451106-34451298	4	H2-T23	721	Intragenic (intron)	0	0	0	0	mCG12223	Major histocompatibility complex antigen
chr11.20040809	chr11:20041019-20041175	4	Actr2	32614	5' end	117	318	427	49	mCG122263	Actin and actin related protein
chr5.132002621	chr5:132002985-132003227	4	Gtf2i	1683	Intragenic (intron)	0	0	0	0	mCG122431	Basal transcription factor;Nucleic acid binding
chr10.111995763	chr10:111996175-111996271	4	Caps2	12897	3' end	0	0	0	0	mCG122580	Molecular function unclassified
chr11.97333288	chr11:97333478-97333568	4	Mllt6	1012	5' end	0	0	0	5	mCG122606	Zinc finger transcription factor;Nucleic acid binding
chr4.118956373	chr4:118956482-118956615	4	Ctps	40808	5' end	112	106	39	16	mCG122871	Synthase;Ligase
chr8.82636779	chr8:82636932-82637217	4	4933434I20Rik	3789	3' end	0	0	0	0	mCG12297	Glycosyltransferase
chr14.114593801	chr14:114594083-114594112	4	Clybl	239	Intragenic (intron)	0	0	0	28	mCG123228	Other lyase
chr14.114677484	chr14:114677735-114677745	4	Clybl	28966	Intragenic (intron)	0	0	0	28	mCG123228	Other lyase
chr5.126715392	chr5:126715753-126715983	4	Ran	131217	Gene desert	55	78	43	28	mCG123446	Small GTPase
chr18.35577828	chr18:35578131-35578216	4	Catna1	14426	Intragenic (intron)	334	503	304	273	mCG123912	Cell adhesion molecule;Non-motor actin binding protein
chr4.57810681	chr4:57810872-57811003	4	Edg2	10294	Intragenic (intron)	0	65	3	0	mCG12400	G-protein coupled receptor
chr12.49325516	chr12:49325711-49325949	4	Npas3	33667	3' end	0	0	0	21	mCG124421	Basic helix-loop-helix transcription factor;Nucleic acid binding
chr5.115888055	chr5:115888596-115888666	4	Thrap2	155883	Gene desert	13	58	76	0	mCG124529	Transcription cofactor
chr11.77358981	chr11:77359353-77359517	4	Myo18a	30384	5' end	0	0	10	29	mCG124805	Actin binding motor protein
chr17.69295024	chr17:69295602-69295714	4	Tgif	15101	5' end	0	0	0	0	mCG12491	Homeobox transcription factor;Nucleic acid binding
chr17.69256621	chr17:69257440-69257557	4	Tgif	15535	3' end	66	204	222	3	mCG12491	Homeobox transcription factor;Nucleic acid binding
chr2.101559239	chr2:101559404-101559553	4	Traf6	13091	5' end	1	0	0	4	mCG12557	Other signaling molecule;Other miscellaneous function protein
chr6.73712859	chr6:73712951-73712987	4	4931417E11Rik	16659	5' end	0	0	0	0	mCG125726	Molecular function unclassified
chr10.39460507	chr10:39460702-39460880	4	Fyn	27425	3' end	0	2	0	62	mCG125800	Non-receptor tyrosine protein kinase
chr3.64553724	chr3:64554024-64554271	4	4930518C23Rik	19123	Intragenic (intron)	0	0	0	0	mCG126038	G-protein coupled receptor
chr3.64496753	chr3:64497869-64497949	4	4930518C23Rik	75361	Intragenic (intron)	0	0	0	0	mCG126038	G-protein coupled receptor
chr7.117818218	chr7:117818842-117818961	4	Fgfr2	907	5' end	0	11	2	0	mCG126336	Tyrosine protein kinase receptor;Protein kinase
chr16.18546348	chr16:18546672-18546733	4	Cldn5	7223	5' end	0	62	2	0	mCG126582	Tight junction
chr6.97420712	chr6:97420810-97420893	4	C130034I18Rik	11229	Intragenic (intron)	0	0	0	0	mCG126764	Molecular function unclassified
chr14.49525713	chr14:49525857-49525939	4	Il17d	4937	Intragenic (intron)	4	0	0	4	mCG127048	Molecular function unclassified
chr18.32025353	chr18:32025570-32025677	4	1200007B05Rik	11245	3' end	0	0	0	9	mCG127142	Molecular function unclassified
chr16.48207108	chr16:48207391-48207415	4	Dppa4	9452	3' end	8	1	0	0	mCG127779	Molecular function unclassified
chr17.22730096	chr17:22730325-22730561	4	Pdpk1	7300	Intragenic (intron)	44	54	11	124	mCG12032	Protein kinase
chr16.78628448	chr16:78628512-78628669	4	Cxadr	28216	5' end	0	0	0	0	mCG128445	Other receptor
chr16.78846005	chr16:78846414-78846417	4	D16Ertd472e	53225	3' end	0	0	1	19	mCG128448	Molecular function unclassified
chr16.20162345	chr16:20162493-20162563	4	Abcc5	918	Intragenic (intron)	33	89	2	28	mCG128453	ATP-binding cassette (ABC) transporter
chr19.10549285	chr19:10549431-10549618	4	Ms4a5	15338	5' end	0	0	0	0	mCG128843	Molecular function unclassified
chr16.38328662	chr16:38329023-38329061	4	Cd80	1889	Intragenic (intron)	0	0	0	0	mCG129101	Immunoglobulin receptor family member;Membrane-bound signaling molecule;Defense/immunity protein
chr16.38263841	chr16:38264591-38264681	4	Pla1a	53	Intragenic (intron)	0	27	0	0	mCG129102	Phospholipase
chr16.36461781	chr16:36462037-36462052	4	Cd86	1344	Intragenic (intron)	0	0	0	0	mCG130169	Immunoglobulin receptor family member;Membrane-bound signaling molecule;Defense/immunity protein
chr16.36497462	chr16:36497547-36497739	4	Cd86	16457	Intragenic (intron)	0	0	0	0	mCG130169	Immunoglobulin receptor family member;Membrane-bound signaling molecule;Defense/immunity protein
chr2.75527754	chr2:75528285-75528306	4	Nfe2l2	45136	3' end	17	13	39	81	mCG1302	Other transcription factor;Nucleic acid binding
chr2.75767917	chr2:75768262-75768392	4	Agps	3308	Intragenic (intron)	2	0	0	29	mCG1304	Synthase;Transferase
chr5.18979686	chr5:18980140-18980231	4	Acvrinp1	1705	Intragenic (intron)	0	0	0	0	mCG131945	Other cell junction protein
chr7.16487681	chr7:16488105-16488134	4	Pld3	3008	5' end	0	0	0	0	mCG133360	Phospholipase
chr7.114696108	chr7:114696294-114696373	4	Sept1	6799	5' end	160	52	15	4	mCG134099	Other cytoskeletal proteins;Small GTPase;Other hydrolase
chr6.125447254	chr6:125447653-125447816	4	Usp5	192	5' end	154	17	10	194	mCG134304	Cysteine protease
chr11.66624091	chr11:66624415-66624609	4	2310004I24Rik	23188	3' end	8	8	0	20	mCG13456	Molecular function unclassified
chr7.105890273	chr7:105890621-105890692	4	Tmc7	9200	5' end	0	0	0	15	mCG1355	Molecular function unclassified
chrX.11537919	chrX:11538608-11538640	4	Usp9x	34478	3' end	7	0	0	0	mCG140089	Cysteine protease
chr12.4230681	chr12:4231480-4231554	4	Ncoa1	4283	Intragenic (intron)	93	3	0	25	mCG140096	Transcription cofactor;Nucleic acid binding
chr11.16503036	chr11:16503470-16503753	4	Egfr	144160	Gene desert	0	0	0	62	mCG140518	Tyrosine protein kinase receptor;Protein kinase
chr11.102002586	chr11:102002891-102002934	4	Ubtf	12827	5' end	0	0	7	0	mCG140722	HMG box transcription factor;Nucleic acid binding
chr6.20497555	chr6:20498387-20498518	4	Kcnd2	529395	Gene desert (5' end)	0	0	0	6	mCG140786	Voltage-gated potassium channel
chr6.35778771	chr6:35779044-35779184	4	Mtpn	337107	Gene desert	10	64	0	125	mCG140789	Molecular function unclassified
chr2.157903985	chr2:157904080-157904331	4	Ctnnbl1	4165	Intragenic (intron)	0	0	0	12	mCG140973	Molecular function unclassified
chr15.3577916	chr15:3578358-3578633	4	Ghr	166495	Gene desert	0	0	0	0	mCG141048	Other receptor
chr11.80501902	chr11:80502028-80502139	4	Accn1	204	Intragenic (intron)	0	0	0	0	mCG141259	Other ion channel
chr18.17788790	chr18:17789306-17789401	4	Cdh2	610356	Gene desert (5' end)	0	0	0	0	mCG141325	Cadherin
chr18.17935487	chr18:17935716-17935793	4	Cdh2	756757	Gene desert (5' end)	0	0	0	0	mCG141325	Cadherin
chr8.72138759	chr8:72139034-72139122	4	Large	5379	Intragenic (intron)	0	8	0	12	mCG141327	Glycosyltransferase
chr4.45430875	chr4:45431086-45431141	4	Xpa	324	Intragenic (intron)	0	0	0	0	mCG14151	Damaged DNA-binding protein
chr2.164763413	chr2:164763783-164763957	4	9230105I15Rik	35	Intragenic (intron)	0	0	0	0	mCG142013	Serine protease inhibitor
chr13.22404312	chr13:22405027-22405170	4	V1rh5	9362	5' end	0	0	0	0	mCG142202	G-protein coupled receptor

chr12.51712806	chr12:51713245-51713331	4	Titf1	11400	3' end	0	0	0	0	mCG142469	Homeobox transcription factor;Nucleic acid binding
chr5.118392489	chr5:118392647-118392719	4	Oas1e	0	Intragenic (intron)	0	0	0	0	mCG142506	Nucleic acid binding;Synthetase;Nucleotidyltransferase;Defense/immunity protein
chr3.34563528	chr3:34564187-34564200	4	B230215L15Rik	80470	3' end	10	0	0	0	mCG142516	Molecular function unclassified
chr19.55897047	chr19:55897291-55897329	4	Dclre1a	18910	3' end	0	0	0	0	mCG14267	Molecular function unclassified
chr6.67281018	chr6:67281498-67281521	4	A430010J10Rik	625	3' end	0	0	0	0	mCG144876	Molecular function unclassified
chr16.27747094	chr16:27747236-27747387	4	Fgf12	253955	Gene desert	0	0	0	0	mCG145691	Growth factor
chr17.29717356	chr17:29720112-29720191	4	Abcg1	491	3' end	0	0	0	0	mCG14574	Transporter
chr8.59635734	chr8:59636292-59636412	4	Clcn3	175764	Gene desert (3' end)	4	15	0	20	mCG14814	Anion channel
chr11.62883719	chr11:62883802-62883906	4	Pmp22	107000	Gene desert	0	4	21	31	mCG14822	Myelin protein;Other miscellaneous function protein
chr3.97131108	chr3:97131430-97131447	4	Bcl9	70000	3' end	41	7	0	25	mCG14866	Molecular function unclassified
chr12.105583122	chr12:105583871-105583941	4	Rage	3250	5' end	5	0	0	8	mCG14938	Non-receptor serine/threonine protein kinase
chr4.19949181	chr4:19949475-19949595	4	Ttpa	1050	Intragenic (intron)	2	0	0	24	mCG1498	Other transfer/carrier protein
chr9.78771616	chr9:78771763-78771831	4	2410146L05Rik	2210	3' end	12	2	0	0	mCG15227	Molecular function unclassified
chr9.78756618	chr9:78756706-78757038	4	Dppa5	7031	3' end	6559	1381	0	1444	mCG15237	Molecular function unclassified
chr3.34456660	chr3:34456808-34457065	4	Sox2	1605	3' end	105	541	47	184	mCG15247	HMG box transcription factor;Nucleic acid binding
chr2.160718996	chr2:160719757-160719971	4	Top1	9883	Intragenic (intron)	21	28	73	44	mCG15357	Nucleic acid binding;Select regulatory molecule;Isomerase
chr17.83285151	chr17:83285386-83285453	4	Lrpprc	1052	Intragenic (intron)	0	0	0	0	mCG15600	Oxidase
chr10.61176519	chr10:61176854-61177023	4	Nodal	1910	5' end	98	68	27	0	mCG15753	TGF-beta superfamily member
chrX.141011454	chrX:141011550-141011713	4	2310007F12Rik	27662	5' end	71	168	2	170	mCG15785	Molecular function unclassified
chrX.141054586	chrX:141056055-141056216	4	2310007F12Rik	72166	5' end	71	168	2	170	mCG15785	Molecular function unclassified
chr17.86575539	chr17:86575828-86575943	4	Msh6	1150	5' end	37	15	0	16	mCG15886	Damaged DNA-binding protein;Hydrolase
chr2.109804462	chr2:109805039-109805066	4	Lin7c	12595	3' end	4	46	2	98	mCG15974	Cell adhesion molecule;Other cell junction protein
chr2.109834689	chr2:109834824-109835157	4	Lin7c	42533	3' end	4	46	2	98	mCG15974	Cell adhesion molecule;Other cell junction protein
chr11.7946181	chr11:7946332-7946592	4	Igfbp3	837838	Gene desert (5' end)	25	43	72	143	mCG16620	Other miscellaneous function protein
chr5.96240844	chr5:96241206-96241416	4	Prkg2	10160	3' end	0	0	0	0	mCG16631	Non-receptor serine/threonine protein kinase
chr3.96334169	chr3:96334416-96334683	4	Fcgr1	34891	5' end	0	0	0	0	mCG16734	Immunoglobulin receptor family member;Defense/immunity protein
chr2.162991230	chr2:162991841-162991875	4	Sfrs6	3034	5' end	236	95	58	119	mCG1675	mRNA splicing factor
chr4.136235842	chr4:136236022-136236171	4	Akp2	18159	Intragenic (intron)	10	234	72	0	mCG17177	Other phosphatase
chr4.83482588	chr4:83483059-83483167	4	Sh3gl2	3042	5' end	0	2	0	113	mCG1718	Acyltransferase;Membrane traffic regulatory protein
chr4.122811969	chr4:122812268-122812397	4	Pou3f1	215634	Gene desert	6	159	0	0	mCG17254	Homeobox transcription factor;Nucleic acid binding
chr1.36332601	chr1:36332921-36332939	4	Hs6st1	54438	5' end	15	89	1	33	mCG17739	Other transferase
chr10.44643424	chr10:44643621-44643713	4	Prep	347041	Gene desert	0	52	0	10	mCG17069	Serine protease
chr14.83595785	chr14:83596174-83596202	4	Pcdh20	3262406	Gene desert (5' end)	0	0	0	0	mCG17884	Cadherin
chr3.84376882	chr3:84377104-84377181	4	D930015E06Rik	11730	Intragenic (intron)	0	0	0	0	mCG18092	Molecular function unclassified
chr19.56053693	chr19:56054078-56054122	4	Adrb1	56930	5' end	0	0	0	14	mCG18124	G-protein coupled receptor
chr5.73647852	chr5:73647982-73648087	4	Chic2	52527	3' end	0	0	0	0	mCG18257	Molecular function unclassified
chr2.29795659	chr2:29795768-29795896	4	2900073H19Rik	3048	Intragenic (intron)	18	34	0	33	mCG18290	Molecular function unclassified
chr8.46889333	chr8:46889628-46889828	4	Ing1l	55000	5' end	41	17	37	18	mCG1858	Acetyltransferase
chr12.97409416	chr12:97409790-97409895	4	Moap1	11688	3' end	6	2	0	7	mCG18585	Other defense and immunity protein
chr10.58280907	chr10:58281036-58281161	4	Ranbp2	38595	3' end	5	0	2	7	mCG19012	Other receptor;Membrane traffic regulatory protein
chr14.70053495	chr14:70053979-70053992	4	Tnfsf11	4706	Intragenic (intron)	0	0	0	0	mCG1914	Cytokine
chr10.33884304	chr10:33884424-33884549	4	Rwdd1	37363	5' end	78	30	54	106	mCG19188	Molecular function unclassified
chr5.112213007	chr5:112213383-112213450	4	1100001D10Rik	266	3' end	21	27	50	20	mCG19253	Molecular function unclassified
chr4.42579215	chr4:42579976-42580146	4	4930417M19Rik	627	Intragenic (intron)	0	0	0	0	mCG19279	Other transporter;Other hydrolase
chr13.41747918	chr13:41748341-41748448	4	Edn1	327	5' end	0	0	0	0	mCG19645	Peptide hormone
chr2.38502341	chr2:38502489-38502762	4	Nek6	7917	Intragenic (intron)	0	0	0	98	mCG19677	Non-receptor serine/threonine protein kinase
chr3.18978999	chr3:18979357-18979400	4	Pde7a	338	5' end	0	2	0	4	mCG19780	Phosphodiesterase
chr10.99067372	chr10:99067712-99067809	4	Dusp6	19142	3' end	57	52	12	270	mCG19867	Kinase inhibitor;Protein phosphatase
chr4.140691103	chr4:140691423-140691454	4	9030409G11Rik	24928	5' end	2	43	5	0	mCG19998	Molecular function unclassified
chr4.140021091	chr4:140021725-140021806	4	2410043F08Rik	11464	3' end	179	236	94	0	mCG20002	Other zinc finger transcription factor
chr15.100503015	chr15:100503073-100503314	4	AI317237	112643	Gene desert	0	10	0	27	mCG20145	Molecular function unclassified
chr2.87375123	chr2:87375407-87375454	4	Pramel7	7714	3' end	0	0	0	0	mCG20681	Molecular function unclassified
chr11.49647700	chr11:49648103-49648104	4	Mapk9	122270	Gene desert	19	16	2	11	mCG20796	Non-receptor serine/threonine protein kinase
chrX.147461757	chrX:147462371-147462452	4	Sms	512	5' end	0	0	0	0	mCG21158	Synthase
chr5.130441981	chr5:130442455-130442540	4	Gats	1382528	Gene desert (5' end)	0	6	0	18	mCG2174	Molecular function unclassified
chr8.46479486	chr8:46479803-46479875	4	4933409N07Rik	49852	5' end	0	0	1	34	mCG2179	Molecular function unclassified
chr9.61608694	chr9:61608996-61609090	4	Tle3	6710	Intragenic (intron)	0	45	0	13	mCG21830	Transcription cofactor;Other miscellaneous function protein
chr9.61792990	chr9:61793426-61793499	4	Tle3	152423	Gene desert	0	45	0	0	mCG21830	Transcription cofactor;Other miscellaneous function protein
chr5.17163643	chr5:17164335-17164477	4	Cd36	965481	Gene desert (5' end)	0	0	0	0	mCG21846	Other receptor;Cell adhesion molecule
chr9.8780758	chr9:8781237-8781393	4	Pgr	119002	Gene desert	0	0	0	0	mCG21848	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr9.11121023	chr9:11121311-11121434	4	Pgr	2156074	Gene desert (3' end)	0	0	0	0	mCG21848	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr9.11394339	chr9:11394991-11395068	4	Pgr	2429731	Gene desert (3' end)	0	0	0	0	mCG21848	Nuclear hormone receptor;Transcription factor;Nucleic acid binding
chr9.11457359	chr9:11458086-11458122	4	Pgr	2492806	Gene desert (3' end)	0	0	0	0	mCG21848	Nuclear hormone receptor;Transcription factor;Nucleic acid binding

chr9:8450648	chr9:8450832-8450841	4	Trpc6	72676	5' end	0	0	0	0	mCG21849	Other ion channel
chr7:128652114	chr7:128652558-128652723	4	Ifitm1	5436	5' end	0	0	0	0	mCG22584	Other miscellaneous function protein
chr7:128628729	chr7:128628782-128628876	4	BC023151	2452	5' end	3	0	3	6	mCG22585	Molecular function unclassified
chr4:123683399	chr4:123683626-123683731	4	Ftl1	190717	Gene desert	5	0	1	12	mCG23169	Storage protein
chr4:153486768	chr4:153487305-153487331	4	Nadk	10387	3' end	0	15	0	0	mCG23367	Kinase;Transferase
chr7:32636484	chr7:32636648-32636750	4	Cd37	2186	5' end	0	0	0	0	mCG23449	Membrane-bound signaling molecule;Other cell adhesion molecule
chr5:136606037	chr5:136606331-136606399	4	Unc84a	532	5' end	38	178	17	40	mCG23660	Molecular function unclassified
chr4:103515519	chr4:103516430-103516467	4	Ppap2b	12844	Intragenic (intron)	6	0	0	238	mCG2422	Other phosphatase
chr8:31625466	chr8:31625580-31625796	4	Wrn	572236	Gene desert (3' end)	0	0	0	0	mCG2443	DNA helicase;Hydrolase
chr8:32562707	chr8:32562806-32563035	4	Ppp2cb	153	5' end	0	0	0	0	mCG2445	Protein phosphatase;Other select calcium binding proteins
chr2:171941851	chr2:171942433-171942525	4	Cbln4	156973	Gene desert	0	0	0	0	mCG2580	Neuropeptide
chr2:171637354	chr2:171637599-171637608	4	Cbln4	461849	Gene desert	0	0	0	0	mCG2580	Neuropeptide
chr4:57033912	chr4:57034215-57034219	4	Akap2	8435	Intragenic (intron)	0	9	0	0	mCG2708	Other miscellaneous function protein
chr10:19966131	chr10:19966270-19966581	4	Mtap7	1905	Intragenic (intron)	19	9	0	11	mCG2820	Non-motor microtubule binding protein
chr10:21547677	chr10:21547899-21548067	4	Sgk	170583	Gene desert	138	310	85	52	mCG2829	Non-receptor serine/threonine protein kinase
chr10:21423790	chr10:21423847-21423977	4	Sgk	294654	Gene desert	138	310	85	52	mCG2829	Non-receptor serine/threonine protein kinase
chr10:117849325	chr10:117849955-117850082	4	Rap1b	223961	Gene desert	2	0	0	26	mCG3219	Small GTPase
chr13:50873229	chr13:50873260-50873363	4	Gadd45g	123	5' end	0	47	0	66	mCG3413	Other miscellaneous function protein
chr13:51222560	chr13:51222715-51222803	4	Gadd45g	347529	Gene desert (5' end)	0	47	0	66	mCG3413	Other miscellaneous function protein
chr5:106561459	chr5:106561643-106561818	4	V2r16	106842	Gene desert (5' end)	0	0	0	0	mCG3469	G-protein coupled receptor
chr5:111263602	chr5:111264179-111264215	4	Coro1c	7437	Intragenic (intron)	153	56	110	234	mCG3535	Non-motor actin binding protein
chr17:13167962	chr17:13168119-13168453	4	Smoc2	23367	Intragenic (intron)	0	0	0	0	mCG3583	Molecular function unclassified
chr11:51925737	chr11:51926346-51926372	4	Tcf7	4049	5' end	0	7	0	0	mCG3635	HMG box transcription factor;Nucleic acid binding
chr9:37499633	chr9:37500315-37500518	4	BC024479	3818	5' end	78	11	28	16	mCG4139	Molecular function unclassified
chr17:14198466	chr17:14198940-14199052	4	Dll1	5259	5' end	0	0	0	79	mCG4506	Receptor;Membrane-bound signaling molecule;Defense/immunity protein
chr2:146575841	chr2:146576176-146576250	4	Xrn2	189406	Gene desert	17	10	30	27	mCG4604	Exoribonuclease;Hydrolase
chr8:55047061	chr8:55047659-55047968	4	Hpgd	46695	5' end	0	0	4	0	mCG4609	Dehydrogenase
chr1:182266807	chr1:182267175-182267381	4	Srp9	49521	3' end	0	70	0	31	mCG4718	Other receptor;Other RNA-binding protein
chr10:37077183	chr10:37077450-37077520	4	Marcks	71174	5' end	9	6	172	129	mCG50380	Non-motor actin binding protein
chr13:21166667	chr13:21166874-21166950	4	Hist1h2an	36	Intragenic (exon)	0	0	0	0	mCG50406	Histone
chr15:5756450	chr15:5756790-5757025	4	Dab2	366319	Gene desert	33	0	0	0	mCG5086	Other signaling molecule
chrX:44505029	chrX:44505029-44505305	4	Gpc4	24170	5' end	0	0	57	0	mCG51034	Cell adhesion molecule;Extracellular matrix glycoprotein
chr12:82151805	chr12:82152044-82152408	4	6430527G18Rik	56345	3' end	82	84	12	110	mCG52190	Molecular function unclassified
chr13:35289952	chr13:35290189-35290297	4	Cdyl	3358	Intragenic (intron)	10	68	1	11	mCG53594	Molecular function unclassified
chr8:20664278	chr8:20664689-20664801	4	Defb11	5652	5' end	0		0	0	mCG56321	Antibacterial response protein
chr5:26630343	chr5:26630976-26631011	4	9630008K15Rik	5409	Intragenic (intron)	9	9	0	0	mCG56335	Molecular function unclassified
chr9:53846554	chr9:53846771-53846862	4	Cul5	25000	5' end	70	66	54	24	mCG5637	Other miscellaneous function protein
chr5:22608956	chr5:22609306-22609400	4	Nupl2	34531	3' end	0	19	12	9	mCG5716	Molecular function unclassified
chr12:66565353	chr12:66565778-66565811	4	Dact1	12965	5' end	1	0	25	54	mCG5763	Molecular function unclassified
chr7:106289190	chr7:106289399-106289526	4	Gprc5b	2403	5' end	8	13	17	53	mCG5793	G-protein coupled receptor
chr10:78275232	chr10:78275530-78275551	4	Nnp1	9129	3' end	707	389	266	689	mCG5971	Molecular function unclassified
chr15:46043954	chr15:46044050-46044183	4	Kcnv1	905556	Gene desert (5' end)	0	0	0	4	mCG6239	Voltage-gated potassium channel
chr4:117371912	chr4:117372246-117372339	4	Slc2a1	82115	5' end	335	440	112	3	mCG6287	Carbohydrate transporter
chr4:85394270	chr4:85394409-85394645	4	Slc24a2	43920	Intragenic (intron)	0	0	0	0	mCG6544	Transporter
chr14:33560645	chr14:33560858-33560907	4	Ghitm	382499	Gene desert	96	28	202	18	mCG6698	Molecular function unclassified
chr10:24352308	chr10:24352928-24353035	4	Ctgf	4700	5' end	0	0	0	0	mCG6745	Growth factor;Cell adhesion molecule
chr11:49840840	chr11:49841513-49841550	4	Sqstm1	605	3' end	115	71	94	135	mCG67971	Molecular function unclassified
chr9:90663656	chr9:90664295-90664526	4	Zic1	528308	Gene desert	0	0	0	88	mCG68156	KRAB box transcription factor;Nucleic acid binding
chr14:40241579	chr14:40242336-40242477	4	D14Ertd436e	4293	5' end	31	0	0	4	mCG6911	Molecular function unclassified
chr3:44162641	chr3:44162802-44162906	4	Pcdh10	1087770	Gene desert (5' end)	0	0	0	18	mCG7131	Cadherin
chr14:53128250	chr14:53128646-53128871	4	Sacs	53000	5' end	0	0	0	12	mCG7227	Chaperone
chr13:57587991	chr13:57590343-57590443	4	4932432N11Rik	61983	3' end	20	1	0	0	mCG7240	Molecular function unclassified
chr10:17315998	chr10:17316278-17316327	4	Cited2	100201	Gene desert	15	0	9	26	mCG7522	Transcription cofactor
chr4:134038349	chr4:134039718-134039822	4	4930555I21Rik	43952	3' end	0	61	4	0	mCG7590	Molecular function unclassified
chr7:81870829	chr7:81871369-81871561	4	A830059I20Rik	1561540	Gene desert (3' end)	0	0	0	143	mCG7876	Molecular function unclassified
chrX:152725995	chrX:152726028-152726173	4	Rbbp7	4194	5' end	112	0	96	47	mCG7886	Other miscellaneous function protein
chr5:114749800	chr5:114750111-114750258	4	Pbp	16367	5' end	3766	2278	2543	2237	mCG7941	Kinase inhibitor
chr5:114760858	chr5:114761153-114761237	4	Pbp	27378	5' end	3766	2278	2543	2237	mCG7941	Kinase inhibitor
chr1:188521262	chr1:188521599-188521659	4	Kctd3	122149	Gene desert	85	120	24	13	mCG8558	Molecular function unclassified
chr11:22992187	chr11:22992488-22992585	4	Xpo1	159099	Gene desert	114	40	47	28	mCG8765	Other receptor;Other miscellaneous function protein
chr15:81999983	chr15:82000253-82000503	4	Aco2	0	Intragenic (intron)	4	2	0	62	mCG8788	Dehydratase;Hydratase
chr3:88770687	chr3:88771518-88771684	4	Cct3	15320	5' end	905	496	339	192	mCG8828	Chaperonin
chr14:71050957	chr14:71051329-71051369	4	1190002H23Rik	10595	3' end	0	0	0	52	mCG9012	Kinase activator

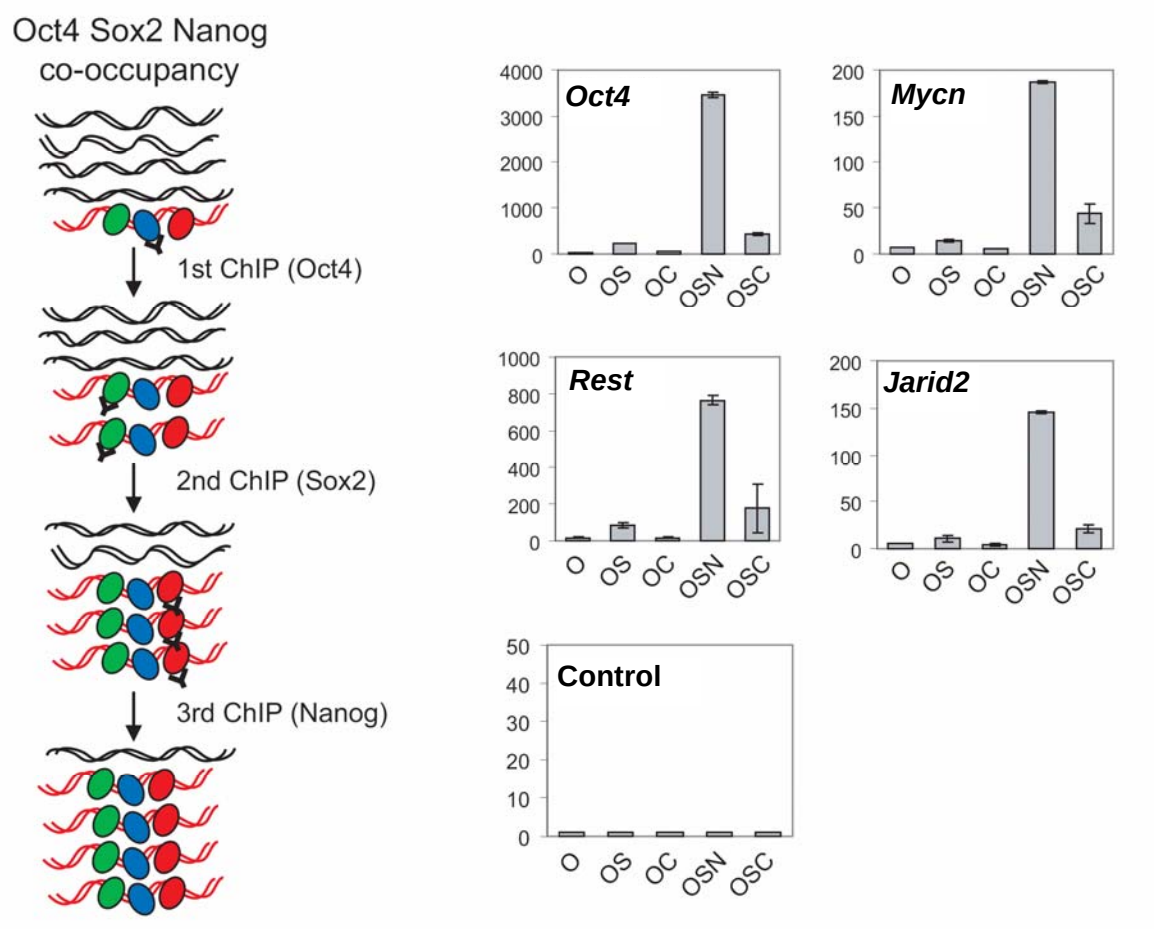
chr8.106906122	chr8:106906513-106906596	4	Psmd7	6317	5' end	212	704	132	127	mCG9335	Other miscellaneous function protein
chr15.27460730	chr15:27460995-27461068	4	Ank	44725	5' end	0	6	0	4	mCG9599	Transporter;Mitochondrial carrier protein
chr13.66142609	chr13:66143082-66143103	4	BC058543	0	Intragenic (exon)	0	0	0	0	NA	NA
chr7.100309570	chr7:100310076-100310140	4	AY053456	0	Intragenic (intron)	0	0	0	0	NA	NA
chr5.133547044	chr5:133547289-133547308	4	BC057460	43	Intragenic (intron)	0	0	0	0	NA	NA
chr11.116424571	chr11:116426148-116426223	4	AK088390	235	Intragenic (intron)	0	0	0	0	NA	NA
chr7.129385655	chr7:129385938-129386065	4	AF016695	311	Intragenic (intron)	0	0	0	0	NA	NA
chr16.25366952	chr16:25367158-25367202	4	S57425	523	5' end	0	0	0	0	NA	NA
chr18.85488928	chr18:85489916-85490260	4	AF176530	610	3' end	129	169	21	0	NA	NA
chr11.43225224	chr11:43225461-43225558	4	AK016814	1264	5' end	0	0	0	0	NA	NA
chr7.67974540	chr7:67974886-67975154	4	AK034740	1325	3' end	0	0	0	0	NA	NA
chr4.81531955	chr4:81532034-81532122	4	AK049108	1484	Intragenic (intron)	0	0	0	0	NA	NA
chr1.87135971	chr1:87136333-87136394	4	BC049077	1862	3' end	0	0	0	0	NA	NA
chr2.132437723	chr2:132438257-132438305	4	AK004090	2277	5' end	0	0	0	0	NA	NA
chr6.22752108	chr6:22752567-22752576	4	AJ428208	2459	Intragenic (intron)	0	0	0	81	NA	NA
chr14.62398602	chr14:62399409-62399606	4	Epb4.9	3497	Intragenic (intron)	19	0	0	0	NA	NA
chr7.16693084	chr7:16693429-16693579	4	BC024554	3745	Intragenic (intron)	0	0	0	0	NA	NA
chr2.29080846	chr2:29081122-29081298	4	AK044730	3802	5' end	0	0	0	0	NA	NA
chr7.120238967	chr7:120239490-120239545	4	AK009207	5739	Intragenic (intron)	0	0	0	0	NA	NA
chr6.129120924	chr6:129121159-129121291	4	AK086937	6458	Intragenic (intron)	0	0	0	0	NA	NA
chr3.21842439	chr3:21843248-21843284	4	AK033347	7606	Intragenic (intron)	0	8	0	20	NA	NA
chr2.84412443	chr2:84413009-84413016	4	AK088963	8667	Intragenic (intron)	0	0	0	0	NA	NA
chr12.22114731	chr12:22115130-22115148	4	AK044503	8845	Intragenic (intron)	0	0	0	0	NA	NA
chr1.151836264	chr1:151836687-151836724	4	BC022177	10288	5' end	0	0	0	0	NA	NA
chr2.89749568	chr2:89749909-89750075	4	AF102522	11136	5' end	0	0	0	0	NA	NA
chr4.106564324	chr4:106564667-106564721	4	AK122545	11386	3' end	64	1	14	6	NA	NA
chr9.67453697	chr9:67453881-67454049	4	AB093231	12021	3' end	21	0	0	4	NA	NA
chr8.60634485	chr8:60634824-60634905	4	AF205079	12672	5' end	0	0	0	0	NA	NA
chr3.28891402	chr3:28891561-28891655	4	AK018073	12741	Intragenic (intron)	0	0	0	0	NA	NA
chr10.23572753	chr10:23572984-23573116	4	AK003902	13271	5' end	0	0	54	4	NA	NA
chr14.44406049	chr14:44407864-44408216	4	AY358079	17153	Intragenic (intron)	0	0	0	0	NA	NA
chr6.03659652	chr6:03659953-03660145	4	AK044750	17100	Intragenic (intron)	0	0	0	0	NA	NA
chr19.37105426	chr19:37105892-37106060	4	BC025649	17400	3' end	1	4	27	4	NA	NA
chr8.66922349	chr8:66922935-66923061	4	AK028684	18435	Intragenic (intron)	0	0	0	0	NA	NA
chr3.79824271	chr3:79824592-79824787	4	AK122512	22294	3' end	6	0	0	45	NA	NA
chr15.23525567	chr15:23526584-23526726	4	AF068258	24568	3' end	0	0	0	0	NA	NA
chr6.143291554	chr6:143291705-143291895	4	AK044502	24744	5' end	0	0	0	8	NA	NA
chr13.43037388	chr13:43037733-43037803	4	AK018089	25308	5' end	0	0	0	59	NA	NA
chr4.142647670	chr4:142648372-142648412	4	BC052827	32641	5' end	0	0	0	0	NA	NA
chr6.73326510	chr6:73326763-73327181	4	AK036732	34335	5' end	0	0	0	0	NA	NA
chr2.166257275	chr2:166257631-166257732	4	AK034712	38905	5' end	6	0	98	95	NA	NA
chr17.83734090	chr17:83734220-83734276	4	AK007130	48926	3' end	0	0	0	21	NA	NA
chr2.170228581	chr2:170228992-170229092	4	BC046393	51333	5' end	0	0	0	0	NA	NA
chr2.21552670	chr2:21552793-21552961	4	AK086256	56171	Intragenic (intron)	0	0	0	0	NA	NA
chr15.6909248	chr15:6909330-6909370	4	Lifr	57000	5' end	24	0	0	0	NA	NA
chr3.28776004	chr3:28776588-28776803	4	AK018073	58968	5' end	0	0	0	0	NA	NA
chr13.94284832	chr13:94285759-94285814	4	BC054101	62000	3' end	26	37	9	0	NA	NA
chr2.17988791	chr2:17989143-17989172	4	BC061062	63016	5' end	0	0	0	0	NA	NA
chr15.73960747	chr15:73961239-73961406	4	BC058722	70841	5' end	0	20	0	0	NA	NA
chr10.70132909	chr10:70133789-70133960	4	AK077025	71906	5' end	0	0	0	0	NA	NA
chr16.8463525	chr16:8463870-8463881	4	AK075830	82403	5' end	400	266	163	133	NA	NA
chr14.3247657	chr14:3248634-3248768	4	AK007159	87920	Intragenic (intron)	0	0	0	0	NA	NA
chr11.8649094	chr11:8649380-8649437	4	AK089717	90087	5' end	0	0	0	9	NA	NA
chr4.5442026	chr4:5442774-5442793	4	X16670	93529	Intragenic (intron)	0	0	0	0	NA	NA
chr15.38150627	chr15:38150984-38150989	4	AK122398	100000	Gene desert	0	0	21	10	NA	NA
chr14.70504362	chr14:70505030-70505241	4	AK033642	116969	Gene desert	16	0	0	0	NA	NA
chr1.11356909	chr1:11357011-11357210	4	AK042743	117140	Gene desert	0	0	0	0	NA	NA
chr17.12603337	chr17:12603669-12603796	4	AF172447	126910	Intragenic (intron)	0	0	0	0	NA	NA
chr11.104615450	chr11:104615999-104616026	4	BC050801	147332	Gene desert	0	0	0	0	NA	NA
chrX.22286353	chrX:22286563-22286642	4	BC068151	148106	Gene desert	0	0	0	0	NA	NA
chr2.146425148	chr2:146425461-146425485	4	AK038838	161001	Gene desert	0	0	0	0	NA	NA
chr5.16927375	chr5:16928277-16928349	4	U38501	293125	Gene desert	0	20	0	0	NA	NA
chr15.60676893	chr15:60677317-60677351	4	X16670	330327	Intragenic (intron)	0	0	0	0	NA	NA

APPENDIX C: Overlapping Sox2, Oct4 and Nanog associated genes (triple overlaps) viewed in a (A) 20K window and a (B) 50K window.



Venn diagrams indicate the extent of overlap between genes associated with Sox2, Oct4 and Nanog binding in mESCs. A core set of genes, 127 and 179 using the 20 and 50 kb cut-offs, respectively, which were bound by all three factors were identified. These included *Rest*, *Rcor2*, *Phc1*, *Id3*, *Nmyc1*, *Tcf3*, *Dppa5*, *Esrrb*, *Otx2*, *Ilst6*, *Lifr*, *Pum1*, *Lefty1*, *Rpe*, *Uck2*, *Rif1*, as well as *Sox2*, *Pou5f1*, and *Nanog*, themselves. All three transcription factors also bound to genes independent of one another; in the case of Sox2, this represented 244 genes.

APPENDIX D: Oct4, Sox2, Nanog triple sequential ChIP



Schematic diagram of triple sequential ChIP (seqChIP). The eluate for Oct4 ChIP was used for Sox2 ChIP, followed by Nanog ChIP. Triple sequential ChIP for selected triple bound genes *Pou5f1*, *Nmyc1*, *Rest*, *Nmyc1*, *Jarid2*, and control. O: Oct4, OS: Oct4-Sox2, OC: Oct4-control, OSN: Oct4-Sox2-Nanog, OSC: Oct4-Sox2-control seqChIPs .

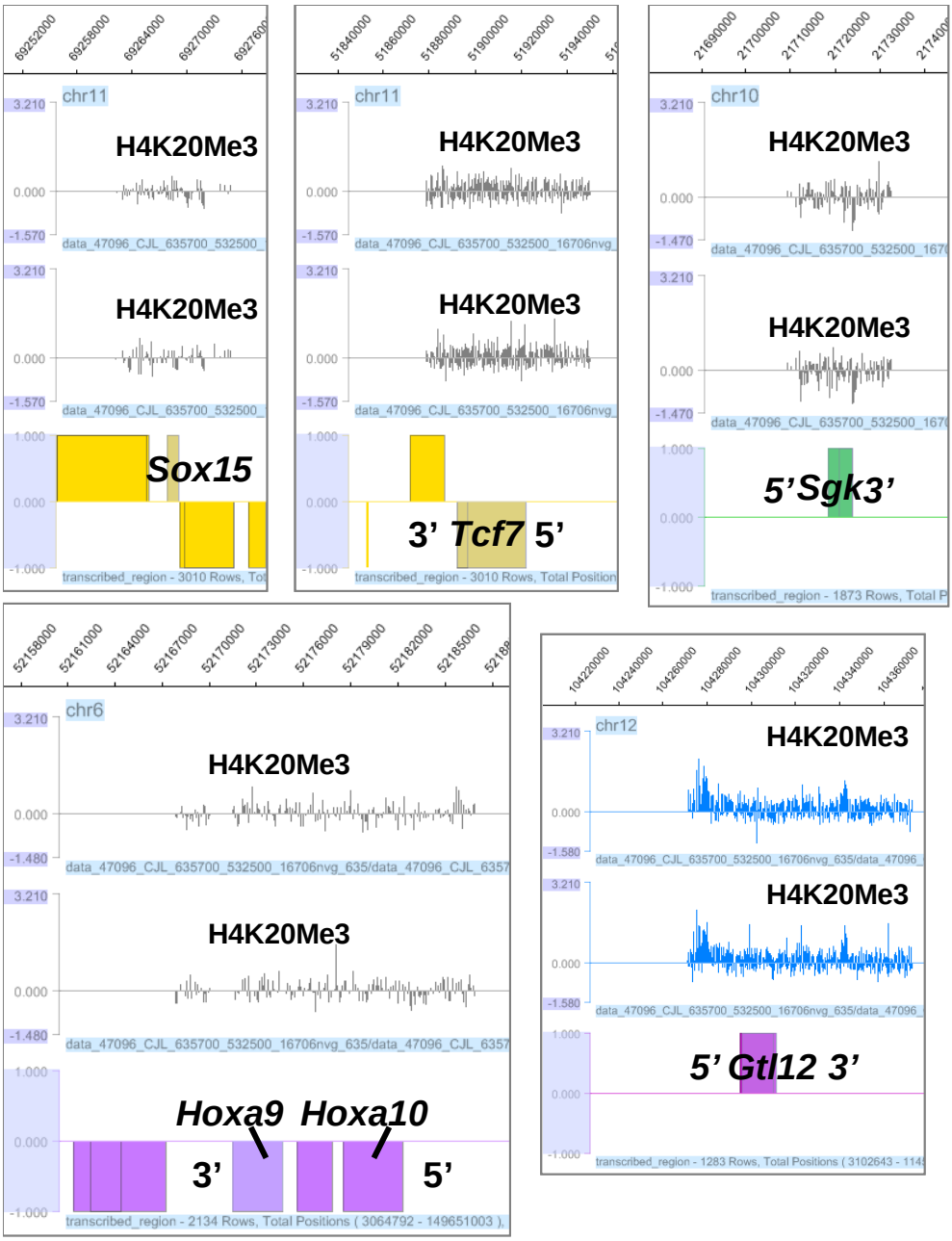
There are a number of examples where the PET clusters from all three transcription factors overlap, indicating their respective binding sites do lie in close (100 bp) proximity with one another. This apparent clustering of Sox2, Oct4, and Nanog binding sites at a number of genomic loci does not necessarily mean these transcription factors are all simultaneously binding the same DNA (ie. co-occupancy) as all binding data was generated from a pool of cells within which there was no way to calculate the percentage with which these protein-DNA interactions were occurring. Chapter 5 reported a quantitative measure of co-occupancy for Sox2 and Oct4 on a number of target genes in mESCs. This protocol was extended to include a third round of ChIP with the Nanog antibody and thereby demonstrating that a triple sequential ChIP experiment can show three proteins simultaneously co-occupying the same stretch of DNA *in vivo*. As *Pou5f1*, *Nmyc*, *Rest*, *and Jarid2* were all identified as targets of Sox2, Oct4, and Nanog by ChIP-PET and the PET clusters of each overlapped with the other, these loci were tested for triple co-occupancy. After Oct4-Sox2-Nanog sequential ChIP there was a drastic increase in fold enrichment for each of the gene loci tested, demonstrating that the three factors indeed co-occupy these loci. Control sequential ChIPs using Oct4-Ena1 and Oct4-Sox2-Ena1 showed no significant (<2-fold) increase in fold enrichment compared to single Oct4 ChIP and Oct4-Sox2 ChIP, respectively.

Appendix E: Oligo probe sequences for reporter assay

3x Ebf1:
cgcgtCATTGTTATGCAAATtagtCATTGTTATGCAAATtcacCATTGTTATGCAAATa
3x Tcf7:
cgcgtTATTGTTATGCAAATtagtTATTGTTATGCAAATtcacTATTGTTATGCAAATa
3x Rest:
cgcgtTATTGTGATGCAAATtagtTATTGTGATGCAAATtcacTATTGTGATGCAAATa
3x Rif1:
cgcgtCTTTGTTATGCACGCtagtCTTTGTTATGCACGCtcacCTTTGTTATGCACGCa

3x Ebf1 swap
cgcgtATGCAAATCATTGTTtagtATGCAAATCATTGTTtcacATGCAAATCATTGTTa
3x Tcf7 swap
cgcgtATGCAAATTATTGTTtagtATGCAAATTATTGTTtcacATGCAAATTATTGTTa
3x REST swap
cgcgtATGCAAATTATTGTGtagtATGCAAATTATTGTGtcacATGCAAATTATTGTGa
3x Rif1 swap
cgcgtATGCACGCCTTTGTTtagtATGCACGCCTTTGTTtcacATGCACGCCTTTGTTa

Appendix F: Control ChIP-on-chip (H4K20Me3) for ChIP-on-chip experiments in Figure 6.3 and Figure 7.5



ChIP-on-chip SignalMap diagram show that H4K20Me3 binding peaks were not present in all test regions except *Gtl12* which serves as a positive control for H4K20Me3 ChIP. Binding peaks are represented as bell-shape peaks. Single points showing enrichments are regarded as background.

Appendix G: Coordinates for ChIP-qPCR amplicons representing peak enrichments in Figure 7.6 and Figure 7.7

Gene names	Coordinates
Atbf1	chr8:108035509-108035712
Ctnnbl1	chr2:157904101-157904326
Dido1	chr2:180397230-180397445
Dppa3	chr6:123262084-123262235
Eif2c1	chr4:124836154-124836306
Ell	chr8:69578358-69578557
Esrrb	chr12:81827868-81828044
Etv1	chr12:33802312-33802517
Id3	chr4:134604765-134604961
Jarid2	chr13:44216298-44216511
mir296	chr2:174328760-174328971
Myc	chr15:62193112-62193306
Mycn	chr12:13066248-13066464
Nanog	chr6:123344925-123345105
Phc1	chr6:122972472-122972634
Rest	chr5:75998395-75998588
Rif1	chr2:52032076-52032272
Sgk	chr10:21712474-21712659
Tbx3	chr5:117164992-117165213
Tcf3	chr6:73014329-73014513
Trp53	chr11:69194615-69194807
Tulp4	chr17:6218243-6218402
Zfp57	chr17:35520539-35520704
Zfp64	chr2:169022611-169022830
Zic3	chrX:49692413-49692250
p21	chr17:27691796-27692040
Test region 1	chr7:132581000-132581249
Test region 2	chr3:24457307-34457477
Test region 3	chr6:123352993-123353158
Test region 4	chr17:34024263-34024412
Test region 5	chr4:93382802-93382977