REGULATION OF GENE EXPRESSION BY ESRRB IN EMBRYONIC STEM CELLS

ZHANG WEIWEI

NATIONAL UNIVERSITY OF SINGAPORE

2008

REGULATION OF GENE EXPRESSION BY ESRRB IN EMBRYONIC STEM CELLS

ZHANG WEIWEI

(*B.Med., PEKING UNIVERSITY*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2008

Acknowledgements

Sincerest thanks to my mentor, Dr Huck Hui Ng who has introduced me to the amazing world of stem cell research and has given me the opportunities to work on this project. During the entire four years of my Ph.D. study, I am impressed by Dr Ng's talents, insights, and perseverance in Science. I have learnt and matured under his inspirational mentorship.

I would like to express my heartfelt appreciation to Dr Yuin Han Loh for being a wonderful co-worker. His encouragement and selfless help were indispensable for mine completion of the project.

I would also like to thank Hwee Goon Tay, Ching Aeng Lim, Xuejing Liu, Katty Kuay, Qiu Li Tan, Kelvin Tan, Kee Yew Wong, Linda Lim, Wai Leong Tam, Boon Seng Soh and other members of the Genome Institute of Singapore. Their great help and friendship are most invaluable.

Special thanks to my collaborators for the ChIP-sequencing experiments, Chai Lin Wei (Cloning and Sequencing group of Genome Institute of Singapore), Eleanor Wong (Cloning and Sequencing group of Genome Institute of Singapore), Han Xu (Information & Mathematical Sciences Group of Genome Institute of Singapore), Vinsensius B Vega (Information & Mathematical Sciences Group of Genome Institute of Singapore), Xi Chen and Fang Fang have provided great assistance and technical support.

I am grateful to Dr Huck Hui Ng, Dr Yuin Han Loh, Dr Kian Liong Lee, Dr Andrew M. Thomson, Dr Rory Johnson, Dr Vardy Leah, Dr Max Fun, Jia Hui Ng, Clara Cheong and Yu Chun Lee for their critical comments on this thesis.

I thank the Department of Biological Sciences, National University of Singapore for their generous scholarship and full support.

Lastly, I am greatly indebted to my father, mother and sister. Their love and understanding are great motivation for mine completion of the four-year graduate study. I shall always remember my mother's advice to me "Do not be a lamster" (Do not give up).

Table of Contents

Summary

Mouse embryonic stem (ES) cells are derived from the preimplantation embryo. ES cells can be cultured indefinitely *in vitro* while retaining the capacity to give rise to any cell type of an organism. To maintain the self-renewal and pluripotency of ES cells, transcription factors play critical roles via the activation of the ES cell specific gene expression program. Nanog is a homeodomain-containing protein that has been identified to be important both for the early development of the blastocyst and the maintenance of undifferentiated ES cells. However, the mechanisms underlying the function of Nanog remain unclear. This project aims to identify the downstream effectors responsible for implementing the decision of Nanog to maintain the self-renewal state of ES cells. Through the manipulation of *Nanog* level by RNAi knockdown and overexpression, putative target genes positively regulated by Nanog were identified. Among the Nanog target genes is a gene encoding for nuclear receptor protein Esrrb (Estrogen-related receptor, beta). Interestingly, *Esrrb* is also positively regulated by another key factor of ES cells, Oct4. Chromatin immunoprecitation (ChIP) and electrophoretic mobility shift assay (EMSA) further demonstrated the specific and direct interaction of Nanog and Oct4 with the *Esrrb* gene. Thus I have identified *Esrrb* as a *bona-fide* target regulated by both Nanog and Oct4. Strikingly, short hairpin RNA (shRNA)-mediated *Esrrb* knockdown resulted in a loss of ES cell morphology, accompanied by a significant reduction of pluripotency markers and induction of differentiation genes. Hence, the project uncovered the novel role of Esrrb in maintaining the undifferentiated state of mouse ES cells. To further characterize the function of Esrrb, the transcriptional regulatory network

of Esrrb was constructed using genome-wide ChIP-sequencing technology and microarray profiling. Both ES cell-associated genes and differentiation-related genes were found to be bound and regulated by Esrrb. Thus Esrrb maintains pluripotency by promoting the expression of downstream self-renewal genes while simultaneously repressing the activity of differentiation-promoting genes. Furthermore, *Nanog* overexpression can rescue the differentiation phenotype induced by *Esrrb* depletion. Thus, *Nanog* is a key downstream target of Esrrb in maintaining pluripotency. In addition, Esrrb is involved in the regulation of genes encoding for chromatin modifiers, such as *Jmjd3*. This suggests a role for Esrrb in governing the unique chromatin structure of ES cells. Together, the findings in this thesis provide new insights into the mechanisms that underlie the critical roles of Nanog and Esrrb in maintaining the self-renewal and pluripotency of mouse ES cells.

List of Tables

List of Figures

List of Publications

1) Loh, Y.H.*, **Zhang, W**.*, Chen, X., George, J., Ng, H.H. (2007). Jmjd1a and Jmjd2c histone H3 lysine 9 demethylases regulate self-renewal in embryonic stem cells. **Genes and Development** *21*, 2545-2557.
** Zhang and Loh are co-first authors*

2) 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. Molecular Biology of the Cell *18(4)*, 1348-58.

3) 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. Nature Genetics *38*, 431–440.

4) 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. Journal of Biological Chemistry *281(34)*, 24090-24094.

List of Abbreviations

aa	amino acid
APC	axin/adenomatous polyposis coli
ARs	androgen receptors
bFGF	basic fibroblast growth factor
BIO	6-bromoindirubin-3'-oxime
BMPs	bone morphogenetic proteins
ChIP	chromatin immunoprecipitation assay
ChIP-chip	**c**hromatin immunoprecipitation coupled with DNA microarray
ChIP-PET	chromatin immunoprecipitation coupled with paired-end ditag technology
CLC	cardiotrophin-like cytokine
CNTF	ciliary neutrophic factor
COUP-TFs	chicken ovalbumin upstream promoter transcription factors
CR	conserved regions
CT-1	cardiotrophin-1
DE	distal enhancer
EC cells	embryonic carcinoma cells.
EG cells	embryonic germ cells
EMSA	electrophoretic mobility shift assays
ERE	estrogen responsive element
ERRE	estrogen related receptor element
ERRs	estrogen related receptor proteins
ES cells	embryonic stem cells
ERs	estrogen receptors
Esrrb	estrogen-related receptor beta
Fbx15	F-box containing protein 15
FDR	false discovery rate analysis
Fox	forkhead box proteins,
GCNF	germ cell nuclear factor
gp130	glycoprotein 130
GRs	glucocorticoid receptors
GSK3	glycogen-synthase kinase-3
HNF-4	hepatocyte nuclear factor 4
Hox	homeobox containing protein family,
ICM	inner cell mass
Id2	inhibitor of differentiation 2
Igf2r	insulin-like growth factor 2 receptor
Jmjd	JmjC domain-containing proteins.
LIF	leukaemia inhibitory factor
LRH-1/NR5A2	liver receptor homologue 1
MPSS	signature sequencing
MRs	mineralocorticoid receptors
Myo	myogenic basic domain proteins,

Nkx	NK transcription factor related proteins,
NR0B1	nuclear receptor subfamily 0, group B, member 1
Oct4	octamer-binding transcription factor-4
OE	over-expression
OSM	oncostatin M
Pax	paired Box and paired-like proteins ,
PcG	polycomb group
PE	proximal enhancer
PGCs	primordial germ cells
POUs	POU-specific domain
POUh	POU homeo-domain
PPARγ	peroxisome proliferators-activated receptor γ
PRs	progestin receptors
Py large T	polyoma early region encoding the large tumor
Q-PCR	quantitative PCR
RA	retinoid receptors
RAR	trans-retinoic acid receptors
RE	responsive elements
RXR	cis retinoic acid receptors
SAGE	serial analysis of gene expression
SCF	stem cell factor
SF1/NR5A1	steroidogenic factor
shRNAi	short-hairpin RNA interfering
Sox2	SRY-related HMG Box 2
SSEA1	stage-specific embryonic antigen 1
Tbx	T-box proteins
Tcf/Lef	T-cell factor/lymphoid enhancer factors
TE	trophectoderm
TR	thyroid hormone
Utf1	undifferentiated cell transcription factor 1
VDR	dihydroxyvitamin D3 receptors
Zfp42	zinc finger protein 42

CHAPTER ONE

INTRODUCTION

Chapter I. Introduction

Stem cells are defined by their unique capacity to self-renew and undergo multi-lineage differentiation. Stem cells can be found in both adult and embryonic tissues where they are important for the processes of cell regeneration, growth and embryo development. Based on their capacity in differentiation, stem cells in mammals can be grouped into three different types, including totipotent stem cells, pluripotent stem cells and multipotent stem cells. Totipotent stem cells can generate all cell types that comprise an entire organism including the placenta. This developmental potential is best exemplified by the fertilized zygote and the cells of the blastomeres up to the 8-cell stage. Pluripotent stem cells can differentiate into all cell types of the three germ layers of an organism. Examples of pluripotent stem cells include the embryonic stem (ES) cells, embryonic germ (EG) cells and embryonal carcinoma (EC) cells. Multipotent stem cells can give rise to cells of certain specialized lineages. Many adult stem cells are multipotent, including hematopoietic stem cells, mesenchymal stem cells, and other adult progenitor cells.

Pluripotent mouse ES cells are derived from the inner cell mass (ICM) of the E3.5 (embryonic day 3.5) preimplantation embryo (Evans and Kaufman, 1981; Martin, 1981). Due to their similarity to the ICM cells, mouse ES cells are a good model for the study of embryogenesis and other developmental processes. In addition, the capability of directed differentiation in culture has made mouse ES cells a potential source for cell replacement therapy (Fujikura *et al*., 2002; Kyba *et al*., 2002; Li, *et al*., 1998). On the other hand, their amenability to genetic perturbation approaches allows mouse ES cells to be a powerful

platform for the study of gene function (Nichols *et al*., 1998; Avilion *et al*., 2003; Mitsui *et al.,* 2003). However, despite the derivation of mouse ES cells more than 20 years ago, little is understood on how ES cells maintain their unique properties. Hence, insights into the molecular mechanisms underlying the self-renewal and pluripotency of ES cells are necessary to realize their clinical and scientific potentials.

1.1. Sources and properties of pluripotent stem cells

Pluripotent stem cells can be isolated from various embryonic sources. For instance, EC cells can be derived from teratocarcinomas, while ES cells and EG cells can be isolated from the ICM and the primordial germ cells (PGCs) respectively. Despite the varied sources of isolation and derivation, these pluripotent cells share the unique properties of self-renewal and broad differentiation capacities.

1.1.1. Mouse embryonal carcinoma cells

Mouse embryonal carcinoma (EC) cells are derived from the teratocarcinomas. Teratocarcinomas are malignant tumors commonly found in the gonads. Histologically, these tumors comprise of various somatic tissues, such as bone, hair and teeth, and they maintain the ability of rapid growth during repeated transplantation (Solter *et al.*, 1970; Stevens, 1970; Solter, 2006). In 1964, Kleinsmith and Pierce showed that single cells from teratocarcinomas retain the capability of tumourigenesis and differentiation into multiple lineages when injected into mice. This finding suggests that unique stem cells reside in teratocarcinomas. Furthermore, transplantation of preimplantation mouse embryos or embryonic tissues to extra-uterine sites resulted in teratocarcinoma formation

(Solter *et al.*, 1970; Stevens, 1970). This finding indicates the embryonic origin of teratocarcinoma stem cells. In 1974, the stem cells in teratocarcinomas were successfully isolated and defined as embryonal carcinoma (EC) cells (Martin and Evans, 1974).

EC cells grow in tight colonies, and are able to proliferate indefinitely (Martin and Evans, 1974) (Figure 1.1). The pluripotency of EC cells has been demonstrated by several experiments. Firstly, subcutaneous injection of EC cells resulted in teratocarcinoma formation (Martin and Evans, 1974). Brinster further found that when reintroduced into the embryo, EC cells could participate in the processes of embryogenesis and contribute to chimera generation. These findings suggest that EC cells are similar to the resident epiblast cells *in vivo* and are receptive to cues in the microenvironment of the embryo (Brinster, 1974). In addition, Martin and Evans showed that EC cells can be used for embryoid body (EB) formation and produce derivatives of all three primary germ layers, including the endoderm, the mesoderm and the ectoderm (Martin and Evans, 1975a and b).

Mouse EC cells are characterized by the expression of unique markers, which include *alkaline phosphatase*, *SSEA1* (*stage-specific embryonic antigen 1*), *TRA-1-60 antigen* and *TRA-1-81 antigen* (Solter and Knowles, 1978; Kannagi, 1983). On differentiation, the expression of these markers is alerted. For instance, *SSEA1* expression is lost during differentiation.

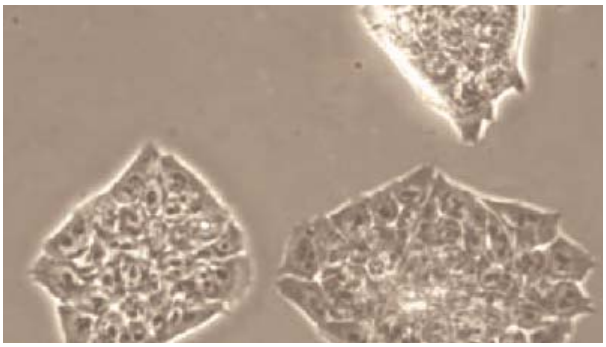


Figure 1.1 Mouse embryonal carcinoma cell line (Adapted from Solter, 2006).

However, it was soon apparent that EC cells suffer from many inherent limitations. Most EC cell lines have poor differentiation capacity and low efficiency in chimera generation. In addition, EC cells give rise to high incidences of tumor formation, thus limiting its application in generating live animals (Mintz and Illmensee, 1975; Papaioannou *et al.,* 1975; Illmensee and Mintz, 1976; Papaioannou *et al.,* 1978; Stewart and Mintz, 1981; Stewart and Mintz, 1982; Rossant and McBurney, 1982). Moreover, EC cells are always aneuploid which prevents cells from proceeding through meiosis and producing mature gametes (Smith, 2001). Therefore, it was necessary to establish "true" stem cells that are isolated from the embryo and retain full developmental potential.

1.1.2. Mouse embryonic stem cells

Mouse embryonic stem (ES) cells were first derived from the inner cell mass and cultured on division-incompetent mouse fibroblasts in the presence of serum (Evans and Kaufman, 1981; Martin, 1981). These cells have similar morphology to EC cells, but they grow in more compact colonies with a higher nucleus-cytoplasm ratio (Figure 1.2). The cytoplasmic organelles associated with non-apoptosis, such as autophagosomes, are also prevalent in mouse ES cells (Ginis *et al.,* 2004).

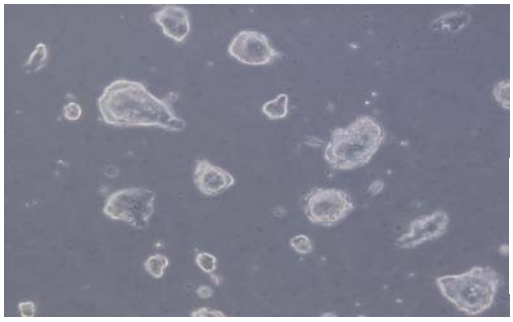


Figure 1.2 E14 mouse embryonic stem cell line.

The two most important properties of ES cells are self-renewal and pluripotency. The pluripotency of mouse ES cells has been demonstrated by extensive studies. *In vitro*, mouse ES cells can be induced to differentiate into various cell lineages (Doetschman *et al*., 1985; Nakano *et al*., 1994; Nishikawa *et al*., 1998). Upon injection into immunoincompetent mice, mouse ES cells can form teratomas consisting of all three germ layers (Evans and Kaufman, 1983). Furthermore, when ES cells are reintroduced into the preimplantation embryo, they can colonize all of the embryonic lineages and contribute to chimeras that give rise to viable offsprings (Bradley *et al.*, 1984; Beddington and Robertson, 1989; Smith, 2001). Because of their superior differentiation capability, consistent chimera generation and normal diploid karyotype, ES cells represent a better model than EC cells for the study of embryogenesis and directed differentiation.

At the molecular level, ES cells express many of the specific markers of EC cells. They include *alkaline phosphatase*, *SSEA1* and *TRA-1-60/81 antigen*. Transcription factors, such as *octamer-binding transcription factor-4* (*Oct4*), *SRY-related HMG Box 2* (*Sox2*) and *Nanog,* are also highly expressed in ES cells (Table 1.1).

Table 1.1 The characterized markers of ES cells (Adapted from Boiani and Schöler, 2005).

Undifferentiated state marker	Mouse	Human
Cell-surface and nuclear antigens		
SSEA1‡	+	–
SSEA3/4‡	–	+
TRA1–60/81§	–	+
TRA2-54	–	+
GCTM-2§	–	+
TG343§	?	+
TG30‖	?	+
CD9‖	+	+
CD133/prominin	+	+
OCT4	+	+
NANOG	+	+
SOX2	+	+
Enzymatic activities		
AP	+	+
Telomerase	+	+
In vitro *culture requirements*		
Feeder-cell dependent	+	+
LIF dependent	+	–
FGF4	+	–
Growth characteristics		
Ability to form trophoblast	–	+
Teratoma formation *in vivo*	+	+
Growth characteristics *in vitro*	Form tight, rounded, multi-layer clumps; can form EBs	Form flat, loose aggregates; can form EBs
Ability to form germ cells *in vitro*	+	NR

1.1.3. Mouse embryonic germ cells

Compared with EC and ES cells, embryonic germ (EG) cells are the least-studied pluripotent stem cells. EG cells are derived from the primordial germ cells (PGCs) of the proximal epiblast during the E8.5 to E11.5 stages of the embryo (Matsui *et al*., 1992; Resnick *et al*., 1992; Durcova-Hills *et al*., 2006). During the initial stage of PGC culture to isolate EG cells, basic fibroblast growth factor (bFGF), leukemia inhibitory factor (LIF) and feeder layers secreting the transmembrane form of stem cell factor (SCF) are required. After the isolation process, however, EG cells can be cultured routinely under the same conditions with ES cells (Matsui *et al*., 1992; Dolci *et al*., 1991).

EG cells are highly similar to ES cells. For instance, EG cells express alkaline phosphatase and SSEA1 antigen. They are immuno-reactive to TRA-1-60 and TRA-1-81. In culture, EG cells have similar morphology to ES cells and can self-renew while maintaining a normal karyotype. EG cells are also capable of giving rise to chimeras (Matsui *et al*., 1992; Labosky *et al*., 1994; Stewart *et al*., 1994). Furthermore, EG cells are capable of germ-line transmission (Labosky *et al*., 1994; Stewart *et al*., 1994). However, unlike ES cells, EG cells retain the erased imprinting pattern that was acquired during the process of germ cell development (Tada *et al*., 1997). For example, the *insulin-like growth factor 2 receptor* (*Igf2r*) gene shows different methylation status in EG cells from ES cells (Labosky *et al*., 1994). This erased imprinting pattern may compromise the developmental potential of EG cells (Kato *et al*., 1999; Tada *et al*., 1998). Kato *et al*. showed that transplantation of the EG cell nuclei into the enucleated oocyte resulted in formation of an abnormal placenta (Kato *et al*., 1999). In addition, 25-50% EG

cell-contributed chimeras were abnormal in weight and gross skeletal structure (Tada *et al*., 1998).

1.1.4. Pluripotent stem cells derived from other species

Pluripotent stem cells have been isolated from animal species other than the mouse. For examples, ES cells from human, horse, pig, dog and cat have been isolated (Thomson *et al*., 1998; Saito *et al*., 2002; Li *et al*., 2006; Hatoya *et al*., 2006; Serrano *et al*., 2006). Due to the scarcity of embryos and the lack of established cell recovery or culturing system, studies on pluripotent cells from other species have lagged significantly behind their mouse and human counterparts.

Human EC cells were isolated from human teratocarcinomas (Hogan *et al.,* 1977). Similar to mouse EC cells, human EC cells are capable of self-renewal and differentiation (Andrews *et al*., 1984). The chromosomes of human EC cells are also abnormal, which greatly limits their application in human cell-based therapy (Blelloch *et al.,* 2004). At the molecular level, human EC cells and mouse EC cells have overlapping but distinct gene expression profiles. For instance, similar with mouse EC cells, human EC cells are positive for alkaline phosphatase staining and immuno-reactive for TRA-1-60 and TRA-1-81. However, unlike mouse EC cells, human EC cells express *SSEA3* and *SSEA4* instead of *SSEA1*. On differentiation, the expression of *SSEA3* and *SSEA4* is lost while the expression of *SSEA1* turns on (Fenderson *et al*., 1987).

Human EG cells are derived from the gonadal tissue of the human embryo during the 5th-11th week post-fertilization stages (Shamblott *et al*., 1998). During *in vitro* culture, human EG cells grow as tightly compacted colonies and are relatively resistant to enzymatic disaggregation. As a result, the efficiency of human EG cell culture and expansion is low. In addition, although they are grown on a feeder cell layer, human EG cells require FGF stimulation. Human EG cells have normal karyotype and high expression of pluripotent marker genes, such as *SSEA-1*, *SSEA-4* and *OCT4*.

Human ES cells were isolated almost 17 years after the first derivation of mouse ES cells (Evans and Kaufman, 1981; Martin, 1981; Thomson *et al*., 1998). Similar to mouse ES cells, human ES cells are isolated from the blastocysts and can grow in colony morphology (Thomson *et al*., 1998). However, the doubling time of human ES cells (35-40 hours) is much longer than that of mouse ES cells (12 hours). Human ES cells are pluripotent based on their capacity for *in vitro* differentiation and teratoma formation. Notably, human ES cells can differentiate into the trophectoderm lineage while mouse ES cells typically do not (Gerami-Naini *et al*., 2003; Odorico *et al*., 2001; Thomson *et al*., 1995, Xu *et al*., 2002, Rossant and Papaioannou, 1984). By comparing the expression patterns of five human ES cell lines and 69 different human somatic cell lines or tissue samples, Sperger *et al.* identified the highly expressed genes in human ES cells. These genes include *OCT4*, *FOXD3*, *SOX2*, *DNMT3B*, *Frizzled 7/8*, and *TCF3* (Sperger *et al*., 2003). In addition, the genes encoding fibroblast growth factor (FGF) receptors are also highly expressed. This suggests a potential role for the FGF signaling pathway in maintaining the undifferentiated state of human ES cells (Sperger *et al*., 2003). Other

studies using different strategies were also performed in an attempt to identify the genes that are important for human ES cell maintenance (Brandenberger *et al*., 2004; Richards *et al*., 2004; Abeyta *et al*., 2004; Bhattacharya *et al*., 2004; Bhattacharya *et al*., 2005). Based on these studies, the list of genes enriched in human ES cells includes *OCT4*, *SOX2, NANOG*, *REX1*, *DNMT3B*, *LIN28*, *TDGF1* and *GDF3*.

Human and mouse ES cells have distinct but overlapping gene expression profile. The regulatory mechanisms underlying cell growth and proliferation also show some differences (Sato *et al*., 2003; Wei *et al*., 2005) (Table 1.1). Like mouse ES cells, human ES cells express *alkaline phosphatase, TRA-1-60 antigen* and *TRA-1-81 antigen* (Thomson *et al*., 1995; Thomson and Marshall, 1998). The expression of mouse ES cell-associated genes such as *Oct4, Sox2, Lefty, Utf-1* and *Tdgf* is also conserved in human ES cells. However, the similarity in gene expression profile between human ES cells and mouse ES cells are limited (Wei *et al*., 2005). For instance, human ES cells express *SSEA3* and *SSEA4* but not *SSEA1*, while mouse ES cells do not express *SSEA3* or *SSEA4* but express *SSEA1* instead. Human ES cells do not express *LIF receptor* (*LIFR*), *STAT3*, or *JAK*, while expressing high levels of FGF receptors. This is consistent with the requirement of the FGF signaling pathway in human ES cell culture (Brandenberger *et al*., 2004; Wei *et al*., 2005). Moreover, many genes that are involved in the Wnt, TGFβ/BMP or other signaling pathways are differentially expressed between the human and mouse ES cells (Brandenberger *et al*., 2004; Ginis *et al*., 2004; Rao, 2004; Wei *et al*., 2005).

1.1.5. Mouse ES cells as a cell model to study the ES cell biology

In this project, mouse ES cells were used as a cell model to understand the mechanisms underlying the maintenance of self-renewal and pluripotency. Compared with other pluripotent stem cells, mouse ES cells have outstanding advantages for the study of ES cell biology.

First of all, mouse ES cells provide a universal cell source for the study of embryogenesis, gene function and directed differentiation. Because mouse ES cells are derived from the ICM and maintain the capability of integration into the developmental processes of the embryo (Beddington and Robertson, 1989), they are valuable as a surrogate for deciphering the pluripotency of the ICM cells and the subsequent steps of cell commitment during early development. On the other hand, when combined with gene-targeting technology, mouse ES cells serve as an excellent "vector" to study the functions of genes in embryogenesis and other biological processes (Thomas and Capecchi, 1986; Kuehn *et al*., 1987; Doetschman *et al*., 1987; Thomas and Capecchi, 1987; Thompson *et al*., 1989; Smithies, 2005; Capecchi, 2005). Mice with thousands of specific targeted mutations have been created to decipher the importance of these genes. Recently, attempts have been made to drive directed differentiation of mouse ES cells using a combination of genetic manipulation, chemical induction and addition of growth factors and extracellular matrices (Wichterle *et al*., 2002; Kim *et al*., 2002; Li *et al*., 2003; Schmitt *et al*., 2004; Solter, 2006). This may contribute to generation of cells and tissues for replacement therapy. Hence, a better understanding of the biology of mouse ES cells is crucial for using stem cells in clinical applications.

Secondly, the lesson learnt from the study of mouse ES cells may be applied to improve the understanding of human ES cell biology. Due to ethical consideration and limited sources of human embryos, the study of human ES cells is severely restricted. However, the understanding of the roles of extracellular and intracellular signals in the maintenance of mouse ES cells holds promises to generate improved methodologies for maintenance and proliferation of human ES cells *in vitro*.

Thirdly, compared to the ES cells from other mammalian species, mouse ES cells have the most established isolation method and *in vitro* culturing protocol. Moreover, unlike mouse ES cells, the properties of ES cells from other species have not been well defined. For example, because of ethical reasons, the pluripotency of human ES cells has not been demonstrated by the chimera generation and the tetraploid complementation experiment.

1.2. Factors required for the maintenance of mouse ES cells

The capacity of self-renewal and pluripotency distinguishes ES cells from other cell types. These unique properties are maintained by extra-cellular signals and intra-cellular regulators.

1.2.1. Signaling pathways in mouse ES cells

Through binding to cell-membrane receptors, extracellular factors can induce nucleus-directed signaling pathways to modulate gene expression. In mouse ES cells, extracellular signals are required for cell growth and maintenance. Several signaling

pathways, such as the leukaemia inhibitory factor (LIF) signaling pathway, the bone morphogenetic protein (BMP) signaling pathway and the wingless-related MMTV integration site (Wnt) signaling pathway have been reported to be involved in the regulation of stem cell properties.

1.2.1.1. The leukaemia inhibitory factor (LIF) signaling pathway

When mouse ES cells were first isolated from the embryo, fibroblast feeder cells were used to support their growth in culture. Without the feeder cells, ES cells cannot be maintained in an undifferentiated state (Evans and Kaufman, 1981; Martin, 1981). This suggests that the feeder cells produce factors that prevent the differentiation of ES cells. Using fractionation strategies, the active component of the feeder cell conditioned medium was identified to be the leukaemia inhibitory factor (LIF). Furthermore, *LIF*-deficient fibroblasts are not capable of supporting ES cells (Hooper *et al.,* 1987; Smith *et al.,* 1988; Williams *et al.,* 1988; Thompson *et al.,* 1989).

LIF belongs to the interleukin-6 (IL-6) cytokine family that comprises IL-6, IL-11, oncostatin M (OSM), ciliary neutrophic factor (CNTF), cardiotrophin-1 (CT-1) and cardiotrophin-like cytokine (CLC). In many cell types, these cytokine family members share overlapping effects on gene expression. They also regulate apoptosis, proliferation and differentiation (Heinrich *et al.,* 2003). In mouse ES cells, LIF functions through the direct binding to a transmembrane heterodimeric receptor which consists of LIF-receptor β (LIFRβ) and signal transducer glycoprotein 130 (gp130). Upon the formation of the trimeric complex, LIF-LIFR-gp130, tyrosine kinase Jak (Janus kinase) is activated

through phosphorylation and serves as a docking site for Stat3 (signal transducer and activator of transcription 3). This in turn results in phosphorylation and dimerization of Stat3. The phosphorylated Stat3 will then translocate into the nucleus and activate its downstream target genes to support the self-renewal of ES cells (Niwa *et al*., 1998). The role of Stat3 as the downstream effector of the LIF signaling cascade was further demonstrated by Matsuda *et al.* who created a chimeric protein of Stat3 fused with the ligand-binding domain of estrogen receptor. When ES cells were treated with 4-hydroxy tamoxifen (4-HT), an agonist of estrogen receptor, the chimeric Stat3 became dimerized. This dimerized chimeric Stat3 can maintain ES cells in the self-renewing condition independent of LIF (Matsuda *et al*., 1999). In addition, *Myc*, a downstream target of Stat3, can confer LIF-independence on ES cells when ectopically expressed (Cartwright *et al.,* 2005).

Although the LIF-gp130-Stat3 signaling cascade is sufficient for the self-renewal of mouse ES cells (Yoshida *et al*., 1994; Niwa *et al*., 1998), there is no direct evidence to support its importance in early embryonic development. It was reported that although LIF and LIFR are expressed in the early embryo, both *LIF*-null and *LIFR*-null embryos can form normal ICM and further develop beyond the egg cylinder stage (Stewart *et al*., 1992; Li *et al*., 1995). In addition, *Stat3*-null and *gp130*-null mouse embryos can also survive beyond the blastocyst stage (Yoshida *et al*., 1996; Takeda *et al*., 1997). Further investigation has uncovered the *in vivo* role of the LIF signaling pathway during embryonic diapause. The embryo lacking gp130 fails to recover after diapause, owing to

the inability to maintain the epiblast (Nichols *et al*., 2001). This result suggests that the signaling through gp130 is essential to prolong the epiblast development.

1.2.1.2. The bone morphogenetic protein (BMP) signaling pathway

Bone morphogenetic proteins (BMPs) belong to the transforming growth factor beta (TGFβ) family that is widely involved in cell proliferation, differentiation and apoptosis in the embryo and many cultured cell types (Massague, 1998). There are over 20 BMPs, and these proteins function through receptor-mediated intracellular signaling pathways (Mishina, 2003). There are two types of BMP receptors: type I (Alk2, Alk3, and Alk6) and type II (BmprII). Interaction between different receptors determines the specificity and outcome of the BMP functions (Mishina, 2003). Generally, the interaction between BMPs and their receptors results in phosphorylation of the downstream effectors, R-Smads (receptor regulated Smad proteins) (such as Smad1, Smad5, or Smad8). Two of these phosphorylated R-Smads form a heterotrimer with a common Smad protein, Smad4. This heterotrimer translocates into the nucleus and regulates the expression of downstream target genes (Derynck and Zhang, 2003; Massague, 2000; Miyazono *et al.,* 2000; Moustakas *et al.,* 2001; Shi and Massague, 2003). In parallel, BMP-receptor signaling can also activate the mitogen activated protein kinase (MAPK) pathway mediated by TGFh1 activated tyrosine kinase 1 (TAK1), a MAPKKK tyrosine kinase. Even though the mechanism of BMP-mediated TAK1 activation remains unclear, some studies have suggested that TAK1 binding proteins (TAB1, TAB2 and TAB3) and X-linked inhibitor of apoptosis (XIAP) may be involved in this process (Derynck and Zhang, 2003; Massague et al., 2000; Yamaguchi *et al*., 1999; Behrens, 2000; Ishitani *et al*., 2003;

Ishitani *et al.,* 1999; Smit *et al.,* 2004; Massague, 2003). In addition, the BMP signaling pathway was also reported to participate in cross-talk with other growth factor/cytokine-mediated signaling to inhibit the Smad function through an unknown mechanism (Aubin *et al.,* 2004; Kretzschmar *et al*., 1997a, b).

The function of the BMP pathway in mouse ES cells was uncovered by the observation that even in the presence of LIF, mouse ES cells could not be propagated in a serum-free condition, and the cells underwent differentiation towards neuronal lineages (Ying *et al*., 2003). Further study identified BMP as the important component in the serum that sustain the undifferentiated state of ES cells through cooperation with the LIF pathway (Figure 1.3) (Ying *et al*., 2003). The LIF/gp130/Stat3 signaling cascade blocks mesoderm and endoderm differentiations but favors neuronal differentiation, while the BMP/Smad signaling pathway prevents the differentiation into neuroectoderm lineages through targeting the *inhibitor of differentiation 2* (*Id2*) gene (Ying *et al*., 2003). This was further confirmed by the overexpression of *Id2* gene which led to the inhibition of neuronal differentiation of ES cells independent of the BMP signaling pathway (Ying *et al*., 2003).

In the embryo, BMP is widely expressed during germ cell specification and body patterning (Chen *et al*., 2004; Nohe *et al*., 2004). However, *BMP4*-null or *Smad4*-null mouse embryos can still develop beyond the blastocyst stage with normal formation of the ICM (Fujiwara *et al*., 2001).

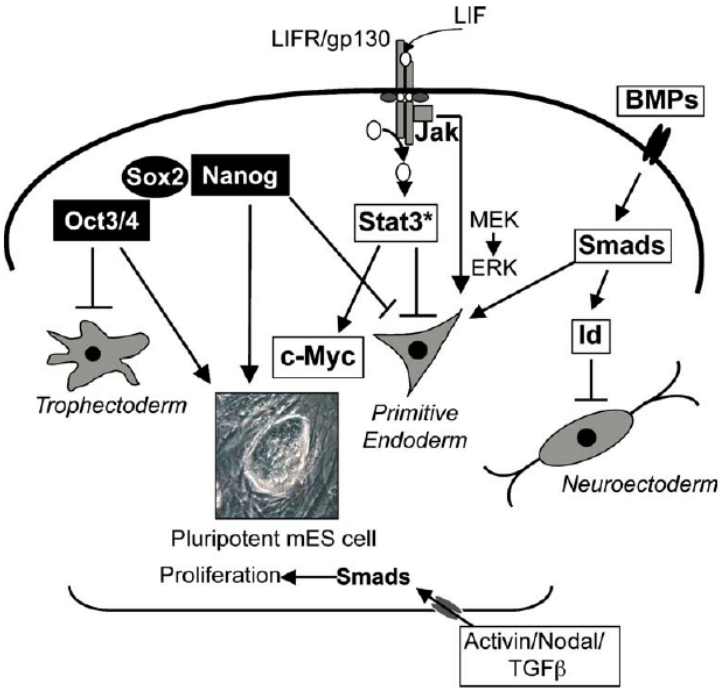


Figure 1.3 Major signaling pathways and transcription factors in maintaining the undifferentiated state of mouse ES cells. LIF and BMP cooperate to maintain the self-renewal and pluripotency of ES cells. LIF activates STAT3, which blocks non-neuronal differentiation. BMP inhibits neuronal differentiation by the induction of Smad1, 5, 8 and Id proteins. Activin/Nodal pathway promotes the proliferation of ES cells through the activation of Smad2/3. Oct4, Sox2 and Nanog are three key transcription factors in ES cells. They cooperate to block differentiation and promote self-renewal of ES cells (Adapted from Yamanaka *et al*., 2008).

1.2.1.3. The wingless-related MMTV integration site (Wnt) signaling pathway

The Wnt signaling pathway is critical for animal development (Cardigan and Nusse, 1997). Recent work has reported the palmitoylation of Wnt3a at the conserved cysteine, Cys77. This modification seems to be crucial for its signal activation (Willert *et al*., 2003). The Wnt signaling pathway is activated upon the binding of the Wnt protein to the Frizzled receptor on the cell surface. This binding leads to the inactivation of the downstream glycogen-synthase kinase-3 (GSK3) protein via the replacement of GSK3 from the axin/adenomatous polyposis coli (APC) complex. The replacement of GSK3 in

turn prevents the degradation of β-catenin, which promotes its translocation into the nucleus where it associates with the T-cell factor/lymphoid enhancer factors (Tcf/Lef) to regulate gene expression (Moon, *et al.,* 2002; Van *et al.,* 2003).

The involvement of the Wnt pathway in stem cell maintenance has been investigated by several studies. Sato *et al.* demonstrated that activation of the Wnt pathway is sufficient to support the self-renewal of both human and mouse ES cells (Sato *et al.,* 2004). In their study, ES cells were treated with the GSK3 inhibitor, 6-bromoindirubin-3'-oxime (BIO), to activate the Wnt pathway, and the cells became resistant to the differentiation. Moreover, the BIO-treated mouse ES cells sustained the expression of pluripotent genes, such as *Rex1*, in the absence of LIF (Sato *et al.,* 2004). However, BIO-induced inhibition of GSK3 mimics the effects of PI3K activation in the LIF signaling pathway (Paling *et al.,* 2004). In addition, the suppressed GSK can result in up-regulation of *Myc*, a key downstream target of the LIF pathway (Cartwright *et al.,* 2005). Hence, the Wnt pathway may maintain the self-renewal of mouse ES cells through a synergistic effect with the LIF/Stat3 signals. More studies have reported the involvement of the Wnt pathway in ES cell biology. Ogawa *et al.* demonstrated that the Wnt pathway is important in maintaining the pluripoteny of mouse ES cells (Ogawa *et al.,* 2006). Inactivating the APC complex or elevating β-catenin level results in the inhibition of neuronal differentiation (Haegele *et al.,* 2003). In addition, activating the downstream target genes of the Wnt pathway, for example *cyclin-D1*, *Myc* and *Bmp*, can partially prevent mouse ES cells from neuronal differentiation induced by the BMP4 antagonist Noggin (He, *et al.,* 1998; Shtutman *et al.,* 1999; Haegele *et al.,* 2003).

1.2.1.4. Other signaling pathways

Besides the LIF, BMP and Wnt signaling pathway, recent studies have uncovered other signaling pathways involved in ES cell maintenance. These pathways include the Activin/Nodal pathway and the Notch signaling pathway (Amit, *et al*., 2004; Ogawa *et al*., 2007; Louvi and Artavanis-Tsakonas, 2006). In ES cells, the Activin/Nodal pathway is autonomously activated to promote self-renewal (Figure 1.3). Inhibition of this pathway can dramatically decrease the proliferation of ES cells. On the other hand, serum-free culture supplemented with recombinant Activin or Nodal enhances the propagation of ES cells (Ogawa *et al*., 2007). The Notch signaling pathway is involved in cell-fate commitment through mediating cell-cell interactions in many tissues (Lai, 2004). In ES cells, interference of the Notch pathway through pharmacologic or genetic manipulation suppresses differentiation towards the neuronal lineages. Activation of the Notch pathway, in contrast, promotes cell differentiation towards the neuronal lineage (Lowell *et al*., 2006). Although these signaling pathways play potential roles in mouse ES cell maintenance, the mechanisms underlying their function remain relatively unclear.

1.2.2. Transcription factors in ES cell maintenance

In ES cells, signaling pathways play important roles in maintaining self-renewal and pluripotency. In most cases, these signaling pathways function through regulating gene expression by transcription factors. Extensive studies have shown that transcription factors, Oct4, Sox2 and Nanog, play critical roles in maintaining the undifferentiated state of mouse ES cells through regulating gene expression.

1.2.2.1 Transcription factor Oct4

Oct4, also known as Oct3/4 or Pou5f1, belongs to the POU (Pit-Oct-Unc) transcription factor family. Oct4 regulates gene expression through its binding to the octamer motif, AGTCAAAT. The DNA binding domain of Oct4 is divided into two regions (Figure 1.4), the POU-specific domain (POUs) and POU homeo-domain (POUh). The POUs and POUh domains are separated by the flexible linker residues that enable both domains to interact with DNA independently. In addition, these two domains may serve as interaction sites for cell type-specific co-regulators (Brehm *et al*., 1997). Subsequent studies have revealed that the POU-domains are central to the interaction of Oct4 and Sox2 via the HMG domain of Sox2 (described in greater detail in later section). Two other regions of Oct4, the N-terminal domain and the C-terminal domain are not responsible for the DNA-binding property of Oct4. Instead, these two domains are involved in the process of transcription regulation through their transactivation activities (Vigano and Staudt, 1996; Brehm *et al*., 1997).

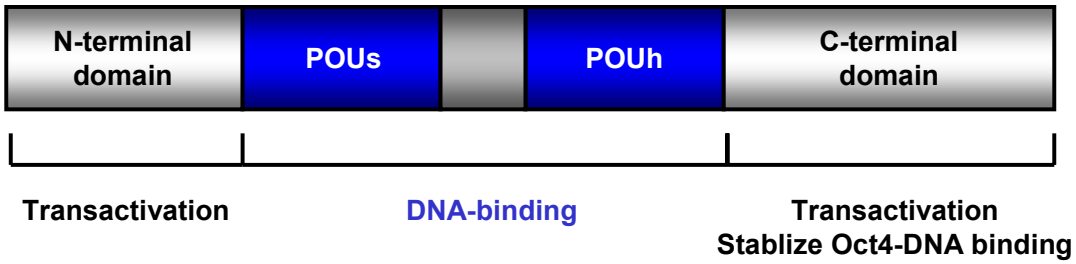


Figure 1.4 Schematic diagram illustrating the protein structure of Oct4 (Adapted from Pan *et al*., 2002).

Oct4 is expressed throughout the early embryo. Its expression is initially detected in the unfertilized egg, the totipotent zygote and all blastomeres at the cleavage stage (Schöler, 1991). At the blastocyst stage, *Oct4* expression starts to be restricted to the ICM, embryonic ectoderm and primordial germ cells (PGCs) (Palmieri *et al.,* 1994). Other lineage cells including trophectoderm (TE), primitive endoderm and extraembryonic tissues have limited expression of *Oct4* (Palmieri *et al.,* 1994). *In vitro*, *Oct4* is highly expressed in pluripotent cell lines such as ES cells, EG cells and EC cells (Rosner *et al.,* 1990; Pesce *et al.,* 1998; Nichols *et al.,* 1998).

Genetic studies uncovered the critical roles of Oct4 in pluripotent cell lines and early development of the embryo. *Oct4*-null embryo dies as early as 3.5 days post coitum (d.p.c.) and showed blastocyst-like structures that are composed of trophectoderm lineage cells without the ICM. This indicates the importance of Oct4 for ICM formation while suggesting its inhibitory role towards trophectoderm differentiation (Nichols *et al.,* 1998). Niwa *et al*. demonstrated that Oct4 controls pluripotency in a dose-dependent manner (Niwa *et al.,* 2000). A two-fold induction of *Oct4* led to ES cell differentiation into primitive endoderm and mesoderm. Loss of *Oct4*, on the other hand, triggered the differentiation into trophectoderm lineages (Figure 1.5). These observations indicate that the appropriate level of Oct4 is critical for maintenance of ES cells.

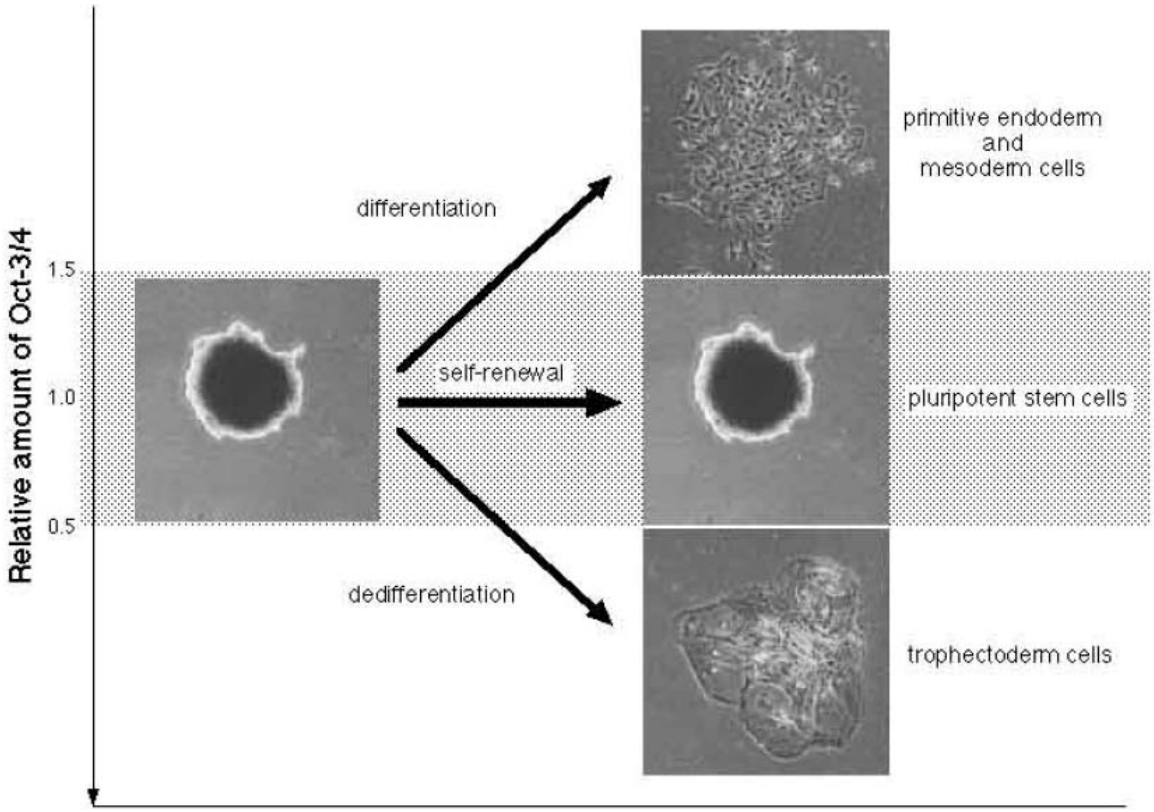


Figure 1.5 Oct4 functions in a dose-dependent manner to control pluripotency in ES cells. Increasing or reducing Oct4 level in ES cells lead to the loss of pluripotency and differentiation into primitive endoderm and trophectoderm lineages respectively. These observations indicate that appropriate Oct4 expression level is necessary for the maintenance of pluripotency (Adapted from Niwa, 2001).

Given its critical role in sustaining the developmental potential of ES cells and the early embryo, the activity of *Oct4* gene is tightly regulated to ensure proper differentiation and continuity of the germline. Expression of *Oct4* is controlled through its proximal promoter and two enhancer regions (the distal enhancer [DE] and proximal enhancer [PE] regions) (Yeom *et al*., 1996). Comparative study between species has identified four highly conserved regions (CR) 1-4 at the regulatory region of *Oct4* (Nordhoff *et al*., 2001). CR1 is located in the immediate upstream region of exon 1, while CR2, CR3 and

CR4 overlap with DE and PE (Figure 1.6). Multiple transcription factors have been identified to regulate the expression of *Oct4* by binding to the conserved regions. Germ cell nuclear factor (Gcnf) has been reported to be a repressor of *Oct4* through targeting the PE region (Hummelke and Cooney, 2001; Fuhrmann *et al*., 2001). In the embryo, loss of *Gcnf* resulted in the mis-repression of *Oct4* in germ layer cell lines after gastrulation (Fuhrmann 2001). Cdx2 is another repressor of *Oct4*. Cdx2 directly binds to the *Oct4* CR4 region and inhibits *Oct4* in a reciprocal fashion (Niwa *et al*., 2005). Several transcription factors have also been reported to activate the expression of *Oct4*. They include Oct4, Sox2 and Sall4 (Chew *et al*., 2005; Zhang *et al*., 2006). *Oct4* expression is also regulated by epigenetic modifications. It has been reported that extinction of *Oct4* activity is correlated with the methylation of the PE and DE elements at the *Oct4* promoter/enhancer regions (Ben-Shushan *et al*., 1993). This is consistent with the observation that *de novo* methylation during embryonic development accompanies the loss of *Oct4* expression in somatic cells (Jaenisch, 1997). More specifically, Feldman *et al.* found that G9a, a histone H3 lysine 9 methyltransferase represses *Oct4* expression by histone H3 lysine 9 methylation at its regulatory region (Feldman *et al*., 2006).

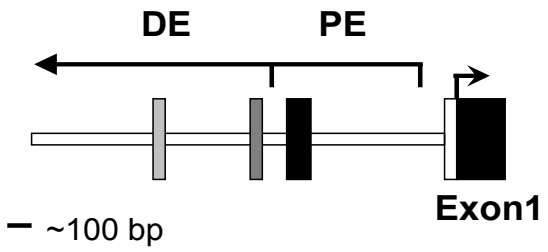


Figure 1.6 A schematic diagram illustrating the regulatory elements of the *Oct4* promoter and enhancer regions. Distal enhancer (DE) and proximal enhancer (PE) are two enhancer regions important for *Oct4* expression. Four highly conserved regions (CR) (CR1-4) are found located in the proximal promoter, DE and PE regions (Adapted from Pan *et al*., 2002).

As a transcription factor, Oct4 has been reported to bind to several target genes to regulate their expression. These genes include *Oct4*, *Sox2*, *Fbx15* (F-box containing protein 15), *Rex1/Zfp42* (*Zinc finger protein 42*), *Utf1* (*undifferentiated cell transcription factor 1*), *Opn* and *Fgf4* (*fibroblast growth factor 4*) (Chew *et al*., 2005; Tokuzawa *et al*., 2003; Catena *et al*., 2004; Nishimoto *et al*., 1999; Botquin *et al*., 1998; Ambrosetti *et al*., 2000; Dailey *et al*., 1994; Yuan *et al*., 1995).

1.2.2.2 Transcription factor Sox2

Sox2 (*SRY-related HMG box*) belongs to the *Sox* gene family that encode for HMG-domain containing transcription factors. In the early embryo, *Sox2* can be detected in the zygote and its expression is maintained throughout the cleavage, blastocyst and epiblast stages (Avilion *et al.*, 2003). However, unlike *Oct4*, *Sox2* expression can also be detected in the extraembryonic ectoderm and neuroectoderm (Avilion *et al.*, 2003). The relatively wider expression profile of *Sox2* indicates that its function may not be restricted to

pluripotent stem cells. *Sox2*-null embryo dies at 5.5 d.p.c due to the lack of epiblast (Avilion *et al.*, 2003). In ES cells, depletion of *Sox2* results in polyploidy and differentiation into trophectoderm lineage (Chew *et al*., 2005; Li *et al.*, 2007). These results implicate the essential roles of Sox2 in embryogenesis and ES cell maintenance.

The overlapping expression profiles between *Oct4* and *Sox2* suggest a close functional relationship. Sox2 is capable of heterodimerizing with Oct4 to form a ternary complex on the *Fgf4* gene (Yuan *et al*., 1995). Other known targets of Oct4, such as *Fbx15, Rex1, Utf1, Opn* and *Nanog*, are also bound by Sox2 (Tokuzawa *et al*., 2003; Nishimoto *et al*., 1999; Botquin *et al*., 1998; Kuroda *et al*., 2005; Rodda *et al*., 2005). Interestingly, Oct4 and Sox2 function as a binary complex to regulate the expression of their own encoding genes (Catena *et al*., 2004; Okumura-Nakanishi *et al*., 2005; Chew *et al*., 2005).

1.2.2.3. Transcription factor Nanog

Nanog was first identified as "ENK" (early embryo specific NK) based on the homolog of its homeodomain to NK protein family (Wang *et al*., 2003). The two studies by Chambers *et al.* and Mitsui *et al*. uncovered its critical roles in ES cell maintenance and embryonic development (Chambers *et al*., 2003; Mitsui *et al*., 2003).

The Nanog protein comprises three domains, N-terminal domain, homeodomain, and C-terminal domain. The N-terminal domain has 96 amino acids and is serine-rich with transactivation activity (Pan and Pei, 2003) (Figure 1.7). The C-terminal domain comprises a CD2 region and 10 WR repeats which are highly conserved between human

and mouse. Compared with the N-terminal domain, the C-terminal domain has a much higher transactivation activity (Pan and Pei, 2003) (Figure 1.7). The homeobox domain is a DNA binding domain with low similarity with the other members of the NK family.

Nanog has a highly restricted expression profile. Its expression is elevated in undifferentiated ES cells, EG cells and EC cells but is not detected in differentiated cells (Chamber *et al*., 2003; Mitsui *et al*., 2003). In the embryo, *Nanog* is first detected at the morula stage, and its expression persists throughout the blastocyst stage after which it starts to decline. At the post-implantation stage, *Nanog* expression is maintained in the primordial germ cells and in the subsets of the epiblast cells. During primitive streak formation, the expression of *Nanog* is down-regulated (Chambers *et al*., 2003; Mitsui *et al*., 2003; Hart *et al*., 2004).

Nanog-null embryo fails in the formation of primitive ectoderm, and dies at 4.5 d.p.c,. Hence, Nanog is required for the ICM formation and primitive ectoderm development (Mitsui *et al*., 2003). ES cells derived from *Nanog* -/- ICM differentiate into endoderm lineage cells (Chambers *et al*., 2003; Mitsui *et al*., 2003; Hatano *et al*., 2005). Interestingly, overexpression of *Nanog* allows ES cells to bypass its dependence on the LIF and BMP signaling pathway (Chambers *et al*., 2003; Mitsui *et al*., 2003). These results demonstrate the indispensable roles of Nanog in early embryonic development and ES cell maintenance.

Figure 1.7 A schematic diagram showing the protein structure of Nanog (Adapted from Pan and Pei, 2003).

Several regulators have been implicated in the regulation of *Nanog* expression (Figure 1.8). By the Chromatin Immunoprecipitation (ChIP) assay and the electrophoretic mobility shift assay, the Oct4/Sox2 complex was shown to bind to the sox-oct element in the proximal promoter of the *Nanog* gene (Rodda *et al*., 2005; Kuroda *et al*., 2005). Recently, FoxD3 was shown to bind to the promoter region of *Nanog* and activate its expression (Pan *et al*., 2006) (Figure 1.8). Besides these factors which activate its expression, *Nanog* was also found to be directly regulated by repressors. The tumour suppressor p53 binds to the promoter of *Nanog*, thereby enabling p53-dependent suppression of *Nanog* expression during differentiation (Lin *et al*., 2005). Tcf3, a downstream effector of the Wnt pathway, is another transcription regulator found to negatively regulate *Nanog* expression (Pereira and Merrill, 2006). Ablation of *Tcf3* in ES cells leads to elevated *Nanog* expression that delays cell differentiation (Pereira *et al.,* 2006).

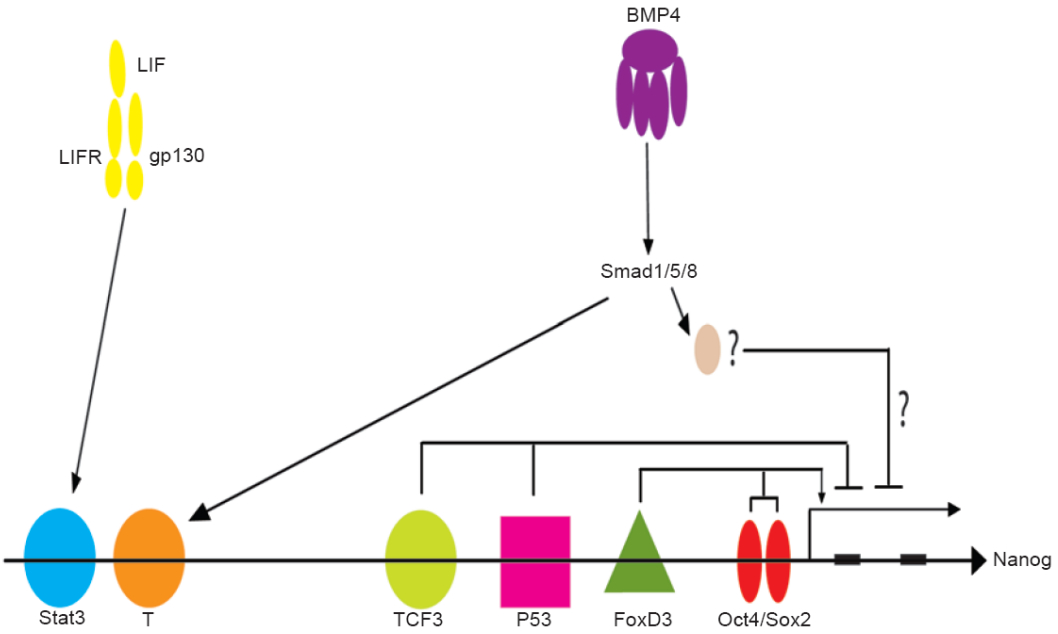


Figure 1.8 Genetic regulation of *Nanog* transcription. Oct4/Sox2 complex and FoxD3 activate Nanog expression by binding to the proximal promoter. TCF3 and p53 are involved in repression of *Nanog* expression. STAT3 from the LIF pathway and T from BMP pathway may also be involved in *Nanog* regulation (Adapted from Pan and Thomson, 2007).

1.2.2.4. Nuclear receptor proteins

Nuclear receptor proteins are transcription factors that respond to extracellular signals (Evans, 1988; Green and Chambon, 1988; Beato, 1989). Generally, they contain distinct domains A to F that exhibit different functions (Laudet *et al.,* 1992; Glass, 1994; Bourguet *et al.,* 1995). For example, domain C is a DNA-binding domain, while domain E has multiple functions including ligand binding, transactivation and dimerization (Glass, 1994; Bourguet *et al.,* 1995). Nuclear receptor proteins are divided into four classes according to their interaction with heat-shock proteins, their tendency to form dimers and DNA-binding specificity, class I nuclear receptor proteins consist of the steroid receptor proteins with larger A and B domains. They include estrogen receptors

(ERs), glucocorticoid receptors (GRs), androgen receptors (ARs), progestin receptors (PRs) and mineralocorticoid receptors (MRs). These proteins are sequestered in 9S heterocomplexes containing heat-shock proteins, immunophilins and other chaperons (Pratt and Toft, 1997; Knoblauch and Garabedian, 1999). Upon binding of ligands to these receptors, the protease-resistant conformation of the complexes is altered. This results in the dissociation of the chaperons along with the dimerization of the receptor proteins (Pratt and Toft, 1997; Knoblauch and Garabedian, 1999; Dechering *et al.,* 2000). The class II receptors consist of thyroid hormone (TR), dihydroxyvitamin D3 receptors (VDR), all trans-retinoic acid receptors (RAR), 9 cis retinoic acid receptors (RXR), and most of the orphan receptors. Unlike the class I receptors, members of the class II receptors have shorter A and B domains that do not interact with heat-shock proteins. In the absence of ligands, class II receptor proteins form heterodimers with RXR. Upon activation by the ligands, the receptor proteins in turn form homodimers or occasionally monomers to regulate gene expression (Brtko, 1994; Parker, 1995). Class III nuclear receptors mainly regulate transcription as monomers. Steroidogenic factors belong to this class of proteins. The class IV receptors share the properties of both class I and II receptor proteins and bind DNA exclusively as monomers (Jiang *et al.,* 1997). Some orphan nuclear receptor proteins such as hepatocyte nuclear factor 4 (HNF-4) belong to this class of nuclear receptors (Jiang *et al.,* 1997). Similar to other receptor proteins, nuclear receptors can respond rapidly to extra-cellular signals and function in maintaining or altering the physiological state of cells. Generally, the receptors are modified by the binding of their ligands, which increases the receptor binding affinity for their specific DNA recognition sequences, termed responsive elements (RE), resulting in the

modulation of gene transcription (Laudet, 1997; Zhang *et al.,* 2004; Tsai and O'Malley, 1994; Mangelsdorf *et al.,* 1995).

Many nuclear receptors have been shown to be involved in embryonic development and maintenance of ES cells. They include Hepatocyte nuclear factor 4α (HNF4a), Germ cell nuclear factor (Gcnf), Nuclear receptor subfamily 0, group B, member 1 (Nr0b1) and Liver receptor homologue 1 (LRH-1/Nr5a2) (Chen *et al*., 1994; Duncan *et al*., 1997; Chung *et al*., 2001; Fuhrmann *et al.,* 2001; Gu *et al*., 2005; Mitsui *et al*., 2003; Clipsham *et al.,* 2004; Palmqvist *et al.,* 2005; Niakan *et al.,* 2006; Pare *et al.,* 2004). HNF4a is the earliest nuclear protein to be expressed in vertebrate embryos and is a marker of primitive endoderm (Duncan *et al*., 1997). *HNF4a*-null mouse embryo dies at E6.5 because of apoptosis in the embryonic ectoderm (Chen *et al*., 1994). *HNF4a*-/- mouse ES cells can only give rise to the endoderm lineages, suggesting its repressive effect on endoderm differentiation (Duncan *et al*., 1997). The repressive role of HNF4a in the endoderm was further confirmed by the finding that Gata6, an endoderm marker, regulates *HNF4a* expression in the visceral endoderm (Morrisey *et al*., 1998). Gcnf is another nuclear receptor with critical roles in embryonic development. The *Gcnf*-null mice show embryonic lethality due to cardiovascular complications, neural tube closing and posterior truncation defects (Chung *et al*., 2001). Gcnf was reported to be a repressor of *Oct4* (Fuhrmann *et al.,* 2001; Gu *et al*., 2005). Nr0b1 belongs to the orphan nuclear receptor proteins, and has a recognized role in the establishment and maintenance of steroid producing tissues and steroidogenesis (Morohashi, 1997; Keegan and Hammer, 2002). In mouse ES cells, it is among the top 20 most highly expressed genes (Mitsui *et*

al., 2003). Its expression is, however, significantly down-regulated with the differentiation of ES cells (Mitsui *et al*., 2003; Clipsham *et al*., 2004; Palmqvist *et al*., 2005). siRNA-mediated knockdown or conditional knockout of *Nr0b1* led to ES cell differentiation into endoderm-like lineages (Niakan *et al*., 2006). LRH-1 is also known as fetoprotein transcription factor (FTF). It is an orphan nuclear receptor, and is highly expressed in the pluripotent cells (Gu *et al*., 2005; Pare *et al*., 2004). It can also be detected in the endoderm-derived tissues such as the liver, pancreas, and intestine (Annicotte *et al*., 2003; Nitta *et al*., 1999; Rausa *et al*., 1999). *LRH-1* null embryo dies at E6.5 to E7.5 due to the mis-regulation of endodermal genes (Pare *et al*., 2004; Gu *et al*., 2005). *LRH*-/- ES cells have similar morphology to wild type ES cells but their *Oct4* levels are more rapidly down-regulated upon LIF withdrawal-induced differentiation (Gu *et al*., 2005). LRH-1 was reported to play a role in regulating the expression of *Oct4* in ES cells and in the embryo (Gu *et al*., 2005). Disruption of *LRH1* resulted in loss of *Oct4* expression both *in vivo* and *in vitro* (Gu *et al*., 2005), indicating the essential roles of LRH1 in maintaining *Oct4* expression. Its role in activation of *Oct4* expression was confirmed by the direct binding of LRH1 to the *Oct4* proximal enhancer and promoter (Gu *et al*., 2005). However, because *LRH-1*-null embryo dies at a later stage than *Oct4*-null embryo, LRH-1 may not be required for the *Oct4* expression in the ICM. Rather, LRH-1 is required for *Oct4* expression in the epiblast. On the other hand, LRH-1 may function through the Wnt signaling cascade to maintain self-renewal, given to the similar phenotypes of the β catenin- and LRH-1-inactivated embryos (Huelsken *et al*., 2000).

Besides HNF4a, Gcnf, Nr0b1 and LRH-1, other nuclear receptor proteins, such as peroxisome proliferators-activated receptor γ (PPARγ), Retinoid acid receptors (RA), Chicken ovalbumin upstream promoter transcription factors (COUP-TFs) and Steroidogenic factor (SF1/Nr5a1), are also important for embryonic development (Barak *et al*., 1999; Lohnes *et al*., 1994; Mendelsohn *et al*., 1994, Kastner *et al*., 1994; Durston *et al*., 1989; Saga *et al*., 1999; Kastner *et al*., 1994; Sucov *et al*., 1994; Massaro *et al*., 2003; Matt *et al.,* 2003; Romand *et al*., 2002; Wendling *et al*., 2001; Qiu *et al.,* 1997; Pereira *et al.,* 1999; Luo *et al*., 1994; Luo *et al*., 1995; Sadovsky *et al.,* 1995).

1.3. Genetic perturbation and genomic approaches to understand ES cell biology

To achieve a detailed understanding of the transcription regulatory network in mouse ES cells, we need to identify the regulatory targets of transcription factors, uncover the *cis*-regulatory element in the genome and investigate the combinatorial control of gene expression by multiple transcription factors. In order to address these challenges, genetic perturbation combined with genomic approaches have been developed and extensively applied in the study of ES cell biology.

1.3.1. Alteration of gene expression by genetic perturbation

There are two main strategies of genetic perturbation used for the study of ES cell biology. They are the ‘gain-of-function’ approach and the ‘loss-of-function’ approach. Most gain-of-function studies involve introducing the cDNA of a gene into cells, ideally resulting in the hyper-activation or hyper-repression of the downstream effectors regulated by the gene product. These studies can be used to study the function of a gene

and dissect the regulatory network in the ES cell biology. For instance, overexpression of *Nanog* in mouse ES cells renders the cells independent of the LIF and BMP signals (Mitsui *et al.,* 2003; Chambers *et al.,* 2003; Ying *et al*., 2003). Through the analysis of expression arrays performed on these cells, the modulators regulated by Nanog can be identified and the interconnection of Nanog with the LIF and BMP signals can be dissected. One of the potential applications of the gain-of-function approaches is to identify a gene whose expression is specific for a certain cell lineage. For example, elevating the expression of *Gata4* or *Gata6* drives ES cells to the primitive endoderm lineage, suggesting a role of these two genes in endoderm differentiation (Fujikura *et al.,* 2002; Capo-Chichi *et al.,* 2005). On the other hand, the gain-of-function approaches have been used for directed cell differentiation through overexpressing lineage-specific genes (Kyba *et al.,* 2002; Li *et al.,* 1998). However, one caveat with these approaches is that they rely on the assumption that over-expression of a gene is sufficient to induce its downstream activity. This assumption may not be universally true because the protein may not carry the appropriate post-translational modifications or have the correct subcellular localization and as a result may be inactive despite its overexpression.

'Loss of function' approaches are classic means of elucidating gene function through suppressing the expression of specific genes. These approaches include homologous recombination, RNA interference (RNAi) technology and other gene-perturbing strategies such as random insertional mutagenesis and ribozymes and antisense RNA methodology. Homologous recombination is the most direct way to assess gene function through the deletion of the gene at the DNA level in the genome. This approach is highly

effective, but is slow and labour-intensive. RNAi technology is a reliable tool used to generate loss-of-function phenotypes transiently or stably in mammalian cells with precision and ease. RNAi-mediated gene suppression can be induced by introducing chemically synthesized small interfering RNAs (siRNAs) or plasmids expressing short hairpin RNAs (shRNAs) into cells. The siRNAs introduced or produced from shRNAs by Dicer are incorporated into the RNA-induced silencing complex (RISC) to direct sequence-specific degradation or translational suppression of the target mRNA (Bernstein *et al*., 2001; Dykxhoorn *et al*., 2003). Moreover, the recent development of inducible RNAi systems has allowed knockdown of a gene in a temporally and spatially regulated manner (Dickins *et al.,* 2005; Ngo *et al.,* 2006). Despite the efficiency of RNAi technology in the study of gene function, it remains difficult to consistently predict the capability of RNAi constructs in gene knockdown. In addition, RNAi constructs introduced may induce 'off-target' effect which suppresses genes other than the intended target. To avoid this, at least two independent RNAi constructs and RNAi-immune cDNA rescue experiments are recommended. Both recombination-mediated gene deletion and siRNA-mediated gene knockdown have been widely used to identify the genes critical for ES cell maintenance and embryonic development, such as *Oct4*, *Sox2*, *Nanog*, *Mbd3* (Nichols *et al.,* 1998; Avilion *et al.*, 2003; Chambers *et al*., 2003; Mitsui *et al*., 2003; Kaji *et al.,* 2006). Random insertional mutagenesis and ribozymes and antisense methodology are also useful tools to elucidate gene function but are less amenable to automation or high-throughput applications, limiting their application in the study of ES cell biology.

1.3.2. Genomic approaches to study gene expression in ES cells

To have a comprehensive understanding of the molecular events in the regulatory networks in ES cells, systematic approaches have been developed to analyze the expression and regulation of genes on a genome-wide scale.

Expression profiling approaches include expressed sequence tags (ESTs), differential display, serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS), large-scale *in situ* hybridization and microarray analysis (Adams *et al.,* 1991; Liang *et al.,* 1992; Schena *et al.,* 2000; Brenner *et al.,* 2000; Komiya *et al.,* 1997; Hughes and Shoemaker, 2001; Lipshutz *et al*., 1999; Schena *et al*., 1995). The EST data mining techniques are used to identify novel genes through digital differential display (DDD) analysis or *in silico* subtraction analysis. Several ES cell specific genes (such as *Nanog*, *Eras* and *Sox15*) have been identified by using the EST data mining techniques (Mitsui *et al*., 2003; Takahashi *et al*., 2003; Maruyama *et al.,* 2005). Differential display techniques, including polymerase chain reaction (PCR)-based differential display and suppression subtractive hybridization, are used for novel gene identification (Liang and Pardee, 1992; Diatchenko *et al*., 1996). For instance, Hllnagel *et al.* showed the inducd expression of the *Id* genes (*Id1*, *Id2* and *Id3*) by BMP2/4 in mouse ES cells by using PCR-based differential display analysis, while Zeng *et al.* identified 50 oocyte-specific genes (Such as *Tcl1*, *Zp1* and *Bmp15*) through suppression subtractive hybridization analysis (Hllnagel *et al.,* 1999; Zeng *et al*., 2003). Both SAGE and MPSS are sequence-based expression analyses and promise a qualitative and quantitative characterization of the transcriptome in cells (Velculescu *et al.,* 1995; Brenner *et al.,*

2000). A series of genes have been identified by the SAGE or MPSS analysis for their potential roles in ES cell maintenance (Anisimov *et al*., 2002; Richards *et al*., 2004; Wei *et al*., 2005; Miura *et al*., 2004). For instance, 192 genes, including *OCT3*/4, *NANOG*, *SOX2*, *REX1* and *GDF3*, were identified to be up-regulated in human ES cells by SAGE analysis (Richards *et al*., 2004). Large-scale *in situ* hybridization is a useful technique to detect the expression patterns of genes in embryos or ES cells (Neidhardt *et al*., 2000; Gitton *et al*., 2002; Reymond *et al*., 2002; Yoshikawa *et al*., 2006). Microarray analysis can be used to monitor the mRNA levels of genes in a parallel fashion for many cell samples at the same time (Schena *et al.,* 1995; Schena *et al.,* 1996; Lockhart *et al.,* 1996; DeRisi *et al.,* 1997). In the study of ES cells, microarray analysis has become a routine strategy to detect gene expression profiles. For example, using Affymetrix microarrays, Ramalho-Santos *et al.* compared the expression profiles between ES cells, hematopoietic stem cells and neural stem cells.

Gene expression profiling analyses provide information about the product of the regulatory events impinging on the expression of genes. It is of interest to understand how regulatory proteins control gene activity. This can be addressed by chromatin immunoprecipitation (ChIP) assay. The ChIP assay was first described by Solomon *et al*. and has been used extensively for the study of protein/DNA interactions in living cells (Solomon *et al.,* 1988). The assay comprises three main steps which include cross-linking of the proteins to DNA by formaldehyde, co-immunoprecipitating the protein-DNA fragments using a specific antibody against the protein of interest, and recovery of the immunoprecipitated DNA for sequence determination (Figure 1.9). ChIP can be

combined with paired-end ditag (PET) technology (ChIP-PET) or DNA microarray (ChIP-chip) to achieve genome-wide mapping of the ChIP-enriched DNA fragments. These strategies allow identification of binding sites of transcription regulators or chromatin-associated modifications on a global scale. In the ChIP-PET assay, the enriched DNA generated by ChIP is labeled by a pair of signature tags at the 5' and 3' ends. The paired-end ditags (PETs) are sequenced and mapped to the genome to demarcate the transcription boundaries of every gene (Ng *et al*., 2005; Wei *et al*., 2006). In the ChIP-chip approach, ChIP-enriched DNA and non-enriched control DNA are labeled with Cy5 and Cy3 dyes respectively before being hybridized to a DNA microarray. The regions enriched with Cy5 signals relative to Cy3 signals inferred to be bound by the transcription factor of interest (Buck and Lieb, 2004; Mockler *et al.,* 2005). Both the ChIP-PET assay and the ChIP-chip assay can be used for genome-wide location mapping. Moreover, combined with computational algorithms, these approaches can also be applied to decipher and characterize the DNA-binding motif of a transcription factor. Computational algorithms have been designed to identify DNA-binding motif pairs of different transcription factors which may account for the co-regulation of specific genes by these transcription factors (Pilpel *et al*., 2001). However, in spite of the successful application of these two approaches, the ChIP-PET and ChIP-chip assay have limitations. For examples, these two approaches require large amounts of DNA and they are expensive. Recently, a new genomic mapping technology, the ChIP-sequencing method (ChIP-Seq), has been developed (Barski *et al.,* 2007; Johnson *et al.,* 2007; Mikkelsen *et al.,* 2007; Robertson *et al.,* 2007). This technique combines ChIP assay with massively parallel DNA sequencing of the ChIP-enriched DNA. Compared with ChIP-PET and

ChIP-chip, ChIP-sequencing technology is more sensitive and less expensive. By using the Solexa sequencing technology, several groups have successfully performed genome-wide mapping studies to detect the binding sites of transcription factors and the distribution of epigenetic modifications throughout the genome (Barski *et al.,* 2007; Johnson *et al.,* 2007; Mikkelsen *et al.,* 2007; Robertson *et al.,* 2007).

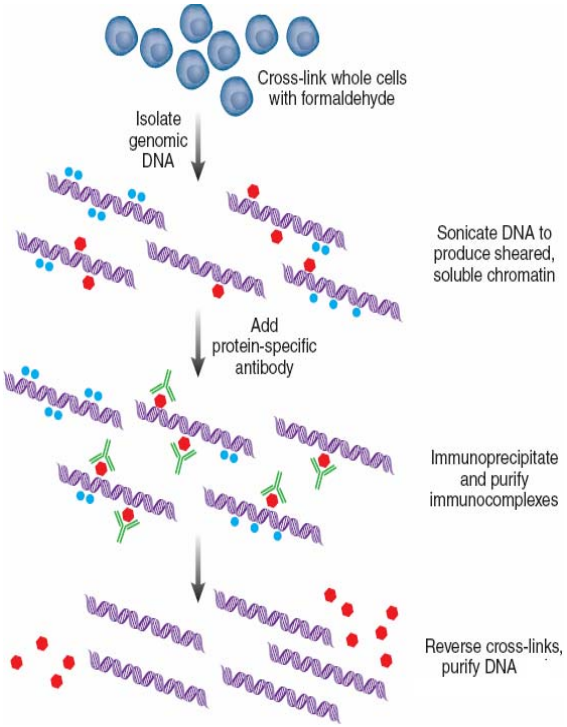


Figure 1.9 Schematic diagram depicting the process of ChIP. Cells are treated with formaldehyde to cross-link the proteins with the DNA, which are further sheared by sonication to produce soluble chromatin. Using antibody against a protein of interest, the protein-bound DNA fragments are selectively coimmunoprecipitated. After purification, the interactions between the protein-DNA complexes are reversed and the recovered DNA is further processed for sequence determination (Adapted from Mardis, 2007).

1.4. Objective and value of this project

ES cells offer great potential for regenerative medicine. As a result, there are tremendous interests in better understanding of the regulatory mechanisms that govern ES cell

biology. Besides the key transcription regulators including Oct4, Sox2 and Nanog, other novel regulators that may be critical for ES cell maintenance have not been extensively studied.

Mouse ES cells depend on signaling pathways and transcription factors to maintain their undifferentiated state. The downstream effectors of these signaling pathways and transcription factors may be as candidate regulators of the self-renewal and pluripotency of ES cells. Transcription factor Nanog plays critical roles in undifferentiated ES cells with unknown mechanisms. This project aimed to decipher the function of Nanog in mouse ES cells through identifying its downstream effectors that are involved in maintaining the self-renewing and pluripotent state of ES cells. The finding from this thesis may contribute to better understanding of ES cell biology.

CHAPTER TWO

MATERIALS AND METHODS

Chapter II. Materials and Methods

2.1. Cell culture

Mouse E14 ES cells were cultured under a feeder-free condition at 37°C with 5% CO_2. The cells were maintained on gelatin-coated dishes in Dulbecco's modified Eagle medium (DMEM; GIBCO), supplemented with 15% heat-inactivated fetal bovine serum (FBS; GIBCO), 0.055 mM β-mercaptoethanol (GIBCO), 2 mM L-glutamine, 0.1 mM MEM nonessential amino acid, 5,000 units/ml penicillin/streptomycin and 1,000 units/ml of LIF (Chemicon). 293T cells were cultured in Dulbecco's modified Eagle's medium (GIBCO) supplemented with 10% heat-inactivated fetal bovine serum (GIBCO) and maintained at 37°C with 5% CO2.

2.2. Knockdown and overexpression plasmids and cell transfection

19 base pair (bp) gene-specific oligonucleotides for RNA interference (RNAi) were designed with the criteria defined by the work of Reynolds *et al*. (2004) and Ui-Tei *et al*. (2004). The RNAi oligonucleotides were cloned into pSuperpuro (Oligoengine) with BglII and HindIII sites which carries a puromycin resistant gene driven by a PGK promoter, and 19-bp short hairpin RNAs (shRNAs) with a 9-bp loop were expressed by the pSuperpuro plasmids. To ensure the specificity of the oligonucleotides for RNAi, all sequences were analyzed by BLAST search to remove any cross effect with other genes. For rescue experiments, RNAi constructs were co-transfected with a construct expressing mutated ORF of the targeting gene through which the target site of shRNA sequence was destroyed. The sequences for shRNAs of different genes are listed here.

Gfp shRNA	5-GATCCCCGAACGGCATCAAGGTGAACTTCAAGAGAGTTCACCTTGATGCCGTTCTTTTTA-3
	5-AGCTTAAAAAGAACGGCATCAAGGTGAACTCTCTTGAAGTTCACCTTGATGCCGTTCGGG-3
Oct4 shRNA	5-GATCCCCGAAGGATGTGGTTCGAGTATTCAAGAGATACTCGAACCACATCCTTCTTTTTA-3
	5-AGCTTAAAAAGAAGGATGTGGTTCGAGTATCTCTTGAATACTCGAACCACATCCTTCGGG-3;
Nanog shRNA	5- GATCCCCGAACTATTCTTGCTTACAATTCAAGAGATTGTAAGCAAGAATAGTTCTTTTTA-3
	5- AGCTTAAAAAGAACTATTCTTGCTTACAATCTCTTGAATTGTAAGCAAGAATAGTTCGGG-3;
Esrrb shRNA1	5- GATCCCCGATCGTCTCGAATCTACTATTCAAGAGATAGTAGATTCGAGACGATCTTTTTA-3
	5- AGCTTAAAAAGATCGTCTCGAATCTACTATCTCTTGAATAGTAGATTCGAGACGATCGGG-3;
Esrrb shRNA2	5- GATCCCCGATTCGATGTACATTGAGATTCAAGAGATCTCAATGTACATCGAATCTTTTTA-3
	5- AGCTTAAAAAGATTCGATGTACATTGAGATCTCTTGAATCTCAATGTACATCGAATCGGG -3.

Transfection of shRNA and overexpression plasmids was performed using Lipofectamine 2000 (Invitrogen). For RNA and protein extraction, 2 μg of shRNA or overexpression plasmids were transfected into ES cells on 60mm plates. For RNAi ChIP assay, 18 μg of shRNA plasmids were transfected into ES cells on 15cm plates. Puromycin (Sigma) selection at 1.0 μg/ml was introduced 24 hours after transfection, and maintained for 2-6 days prior to harvesting. Alkaline phosphatase staining was carried out using a commercial ES Cell Characterization Kit from Chemicon. For the overexpression plasmid, *Nanog* ORF was cloned into pCAGpuro or pCAGhygro tagged by 3XFlag. To establish the Nanog overexpression cell lines, 2 μg of pCAGpuro or pCAGhygro inserted with *Nanog* ORF were transfected into ES cells on 60mm plates. After one-week selection with puromycin (1ug/ml) or hygromycin (300ug/ml), 10 colonies were picked. By using western blot, the expression level of *Nanog* was checked which was detected by the antibodies against Flag and Nanog. Empty pCAGpuro or pCAGhygro was used as a control.

2.3. Luciferase reporter assay

For *Nanog-luc*, a 5.5kb fragment of the mouse *Nanog* promoter was cloned into pGL3-Basic plasmid (BglII and NcoI sites) upstream of the firefly luciferase gene (Promega).

To study the effects of knockdown on the *Nanog* promoter, 100 ng of mouse *Nanog-luc* plasmid, 100 ng of shRNA plasmid (pSuperpuro; Oligoengine), and 5 ng of plasmid containing *Renilla* luciferase (pRL-SV40; Promega) were co-transfected into E14 cells. The pRL-SV40 plasmid was used as a control for normalizing the transfection efficiency. Firefly and *Renilla* luciferase activities were measured 48 hours after transfection by using the Dual Luciferase System (Promega) and a Centro LB960 96-well luminometer (Berthold Technologies).

2.4. RNA isolation, reverse transcription and real-time PCR analysis

The total RNA of cells was extracted by using Trizol reagent (Invitrogen) and further purified with the RNeasy minikit (Qiagen). 500ng of the total RNA purified were used for cDNA synthesis by the SuperScript II kit (Invitrogen). 2 μl of the reverse transcription reaction product was mixed with 5μl SYBR Green mix reagent (Applied Biosystems), 50 nM forward primers and 50 nM reverse primers in a 10 ul reaction system. And the mixture was further loaded to the ABI Prism 7900HT machine (Applied Biosystems) for the measurement of the endogenous mRNA levels. Each sample was performed in duplicate. Results were normalized with β-actin. For all the primers used, each gave a single product of the right size.

2.5. Protein extraction and western blotting

Total protein extracts were extracted by lysing cells with the whole cell extraction buffer (Tris, 50mM; Nacl, 150mM; NP40, 1%; Glycerol, 10%; EDTA, 1mM; PMSF, 1mM). 30μg of the total protein were separated by SDS–PAGE and transferred to PVDF

membrane. The membrane was blocked with 5% milk and probed with specific primary antibodies and secondary antibodies. The blots were developed with ECL Advance Western Blotting Detection Kit (Amersham). Rabbit anti-mouse Nanog antibody and anti-mouse Esrrb antibody were produced by our lab; F-3165 mouse anti-Flag antibody and mouse monoclonal anti-ß actin antibody were from Sigma.

2.6. Microarray

For *Nanog*-depleted ES cells, *Nanog*-overexpressed ES cells and RA-treated *Nanog* overexpressers, Affymetrix Mouse Genome 430 2.0 Array arrays were used to profile the gene expression. Three biological repeats were performed for each of these experimental groups. The probes for hybridization were prepared by GeneChip One-Cycle Target Labeling kit (Affymetrix). The microarray data files are deposited at GEO database. RMA (Robust Multichip Averaging) method was employed to normalize the array data. After the normalization I used Significance Analysis of Microarray (SAM) statistics to select differentially expressed genes from the two groups. SAM (http://www-stat.stanford.edu/~tibs/SAM) uses permutations of the repeated measurements to estimate the percentage of genes identified by chance. The selected genes have a median FDR (false discovery rate) of less than 0.001.

For the *Esrrb*-depleted ES cells, Illumina microarray platform (Sentrix Mouse-6 Expression BeadChip v 1.0) was used. The mRNAs derived from *Esrrb* shRNA 1 and *Gfp* shRNA treated ES cells were reverse transcribed, labeled and analyzed using. Arrays were processed following the manufacturer's instructions. Three biological replicates of the profiles were performed for *Esrrb*-depleted ES cells and *Gfp* knockdown control.

Rank Invariant normalization was use to normalize the microarrays. Significance analysis of microarrays (SAM) was used to select differentially expressed genes. The differentially expressed genes were selected based on the following three criteria (Fold change (FC)>1.5 for up-regulated, FC<0.6 for down-regulated; q value<2%; and detection probability greater than 0.99 in all samples.

2.7. ChIP assay

ES cells were cross-linked with 1% (w/v) formaldehyde for 10 min at room temperature and formaldehyde was then inactivated by the addition of 125 mM glycine. Chromatin extracts containing DNA fragments with an average size of 500 bp were immunoprecipitated using anti-Oct4 (sc-8628, Santa Cruz), anti-Nanog (Cosmo Bio), anti-Esrrb (amino-acids 1 to 200 of mouse Esrrb raised in rabbit) or anti-GFP (sc-9996, Santa Cruz) antibodies. Quantitative PCR analyses were performed in real-time using the ABI PRISM 7900 sequence detection system and SYBR green master mix. Threshold cycles (Ct) were determined for both immunoprecipitated DNA and known amount of DNA from input sample for different primer pairs. Relative occupancy values (also known as fold enrichments) were calculated by determining the IP 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 downstream of *Nanog* gene (chr6:123352993-123353158) (mm5 genome build) and defined as 1.0. For all the primers used, each gave a single product of the right size, as confirmed by agarose gel electrophoresis and dissociation curve analysis.

2.8. Electrophoretic mobility shift assay (EMSA)

The EMSA double-stranded probes (Proligo) were prepared through annealing the sense oligonucleotide strands labeled with biotin at the 5_ termini with the reverse strands in the annealing buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM EDTA). The annealed probes were purified by using an agarose gel DNA extraction kit (QIAGEN). 1ng of purified probes were mixed with 1ug of polydG/dC (Amersham), 2ul of 5x reaction buffer (10mM HEPES, pH7.5, 10mM KCl, 10mM MgCl2, 1mM DTT, 1 mM EDTA, 10% glycerol), and 120ng of Nanog homeodomain, full-length Oct4 or full-length Esrrb protein which were purified by using GSH-sepharose beads (Amersham) and dialyzed at 4°C overnight with the dialysis buffer (20mM HEPES, pH7.9, 20% glycerol, 100mM KCl, 0.83mM EDTA, 1.66mM DTT). The concentration of the protein was measured with a Bradford assay kit (Bio-Rad). The reaction mixtures were incubated at room temperature for 20 minutes. The different complexes in the reaction mixtures were separated by electrophoresis by using 10% DNA PAGE gels which had pre-run for 1 hour at 4°C. The gels were transferred to Biodyne B nylon membranes (Pierce Biotechnologies) and detected using the LightShift Chemiluminescent EMSA kit (Pierce Biotechnologies).

2.9. Esrrb ChIP sequencing library construction and data processing

10 ng of Esrrb ChIP DNA and GFP ChIP control DNA fragments prepared were processed for cluster generation and sequencing on the Illunima cluster station and 1G analyzer by using the methods described in the previous study by Robertson *et al.* (2007). The sequencing reads were extracted and mapped to the mouse genome (mm8) (Chromosome1-22 and X) through the Eland application (Illunima). No more than 2

mismatches per 26bp read sequence were permitted and only uniquely mapped read sequences were retained for further analysis. The mapped reads were then transformed into the profiles of overlapped numbers at each reference genome position, which was performed according to the method in the previous study (Robertson *et al.*, 2007). The putative binding peaks chosen located where the maximal numbers of ChIP fragments resided within a ±500 bps interval were found. To determine a peak height, FDR (false discovery rate) was calculated which is the ratio of the number of non-specific peaks which should occur randomly indicated by a background model to the number of real peaks observed. By using Monte Carlo simulation, the smallest peak height of (FDR<0.05) was chosen to be as a cutoff threshold. Through comparing the peak heights of Esrrb ChIP-sequencing library with the peak heights of the negative control GFP ChIP-sequencing library at the same locations, the non-specific peaks were further filtered out and the peaks with a height ≥ 5 fold were kept and defined as transcription binding sites (Johnson *et al.,* 2007).

CHAPTER THREE

RESULTS

Chapter III. Results

3.1. The roles of Nanog in mouse ES cells

Nanog has been regarded as a key transcription factor involved in the maintenance of ES cells and the processes of embryogenesis (Chambers *et al*., 2003; Mitsui *et al*., 2003). The instrumental roles of Nanog in ES cells have attracted great interests to uncover the mechanisms underlying its functions. This project sought to identify the downstream targets of Nanog which could be important for maintaining the undifferentiated state of ES cells. To achieve that, I had undertaken a combinatorial approach of gene expression profiling upon the genetic manipulations of *Nanog* with chromatin immunoprecipitation (ChIP) assay.

3.1.1. *Nanog* knockdown led to mouse ES cell differentiation

I speculated that if a gene is functionally targeted by Nanog, its expression should be alerted along with the perturbation of *Nanog* expression. To knockdown *Nanog* in ES cells, short-hairpin RNA interference (shRNAi) technology was employed. The *Nanog* shRNA construct was effective in reducing the mRNA and protein levels of Nanog (Figure 3.1A). Consistent with previous reports, *Nanog* depletion resulted in flattened fibroblast-like morphology and the differentiated cells lost the red staining signal of alkaline phosphatase (Figure 3.1B) (Chambers *et al.,* 2003; Mitsui *et al.,* 2003). As would be expected with the distinct differentiation morphology, the expression levels of pluripotent markers, *Oct4, Sox2 and Rex1*, were significantly decreased, while multiple lineage markers were induced (Figure 3.1C and D). Notably, endoderm markers, *Sox17* and *Gata4* were up-regulated in the *Nanog*-knockdown ES cells. This is supported by the

phenotype of *Nanog*-null embryos which are deficient in the formation of primitive ectoderm (Mitsui *et al*., 2003). *Nanog* knockdown also induced the expression of trophectoderm marker genes (*Hand1* and *Cdx2*), mesendoderm marker (*Msx1*) and mesoderm gene (*Brachyury*). Hence, the resulting cells were likely to be composed of multiple differentiated cell types. Taking into consideration of the cellular differentiation morphology, the expression reduction in pluripotency markers and induction of lineage markers, the results have shown that Nanog plays an important role in maintaining the undifferentiated state of ES cells.

To substantiate the specificity of the *Nanog* shRNA construct, a rescue experiment was performed by using the RNAi-immune *Nanog* cDNA expression plasmid. This plasmid contains two synonymous amino-acid substitutions at the shRNA targeting site (Figure 3.2A). The *Nanog* shRNA construct was co-transfected along with the RNAi-immune *Nanog* cDNA expression plasmid. As compared to the vector control, the RNAi-immune *Nanog* rescued the differentiation of ES cells that was induced by knockdown of endogenous *Nanog* (Figure 3.2A). In addition, endogenous *Nanog* was effectively depleted in the cells with co-tranfection of *Nanog* shRNA and control overexpression vector, while exogenous *Nanog* was expressed from the RNAi-immune construct in the cells co-transfected with either the *Gfp* shRNA or the *Nanog* shRNA (Figure 3.2B). To further demonstrate the rescue effect, I measured the expression levels of various ES cell-associated genes. *Oct4* mRNA level which was down-regulated to 50% by *Nanog* knockdown was restored to 80% in the *Nanog* rescue cells (Figure 3.2C). The induction of *Gata6* expression was reversed to almost the same level with the *Gfp* shRNA-

transfected control cells (Figure 3.2C). Together, these results verified the specificity of the *Nanog* shRNA construct.

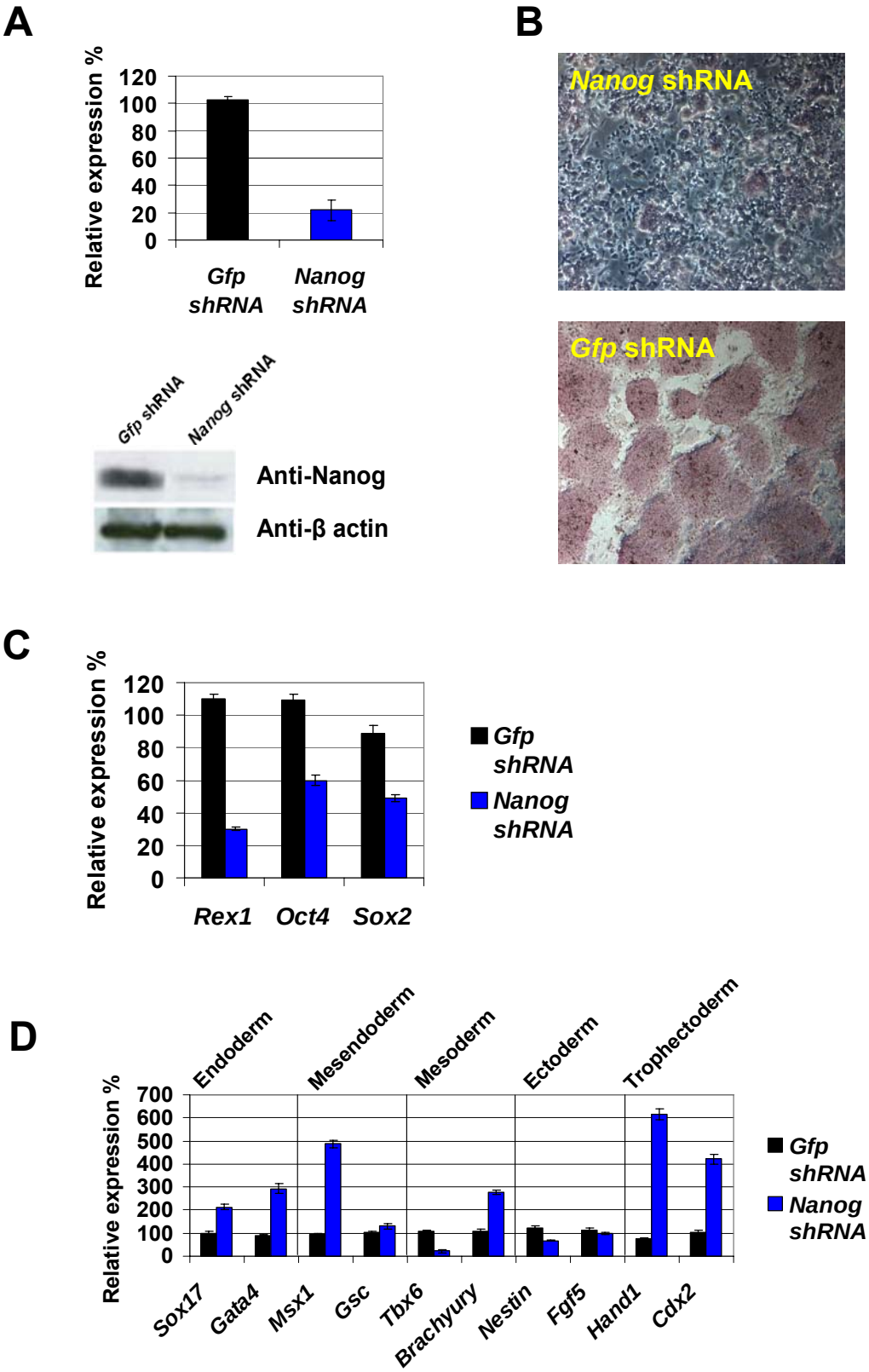
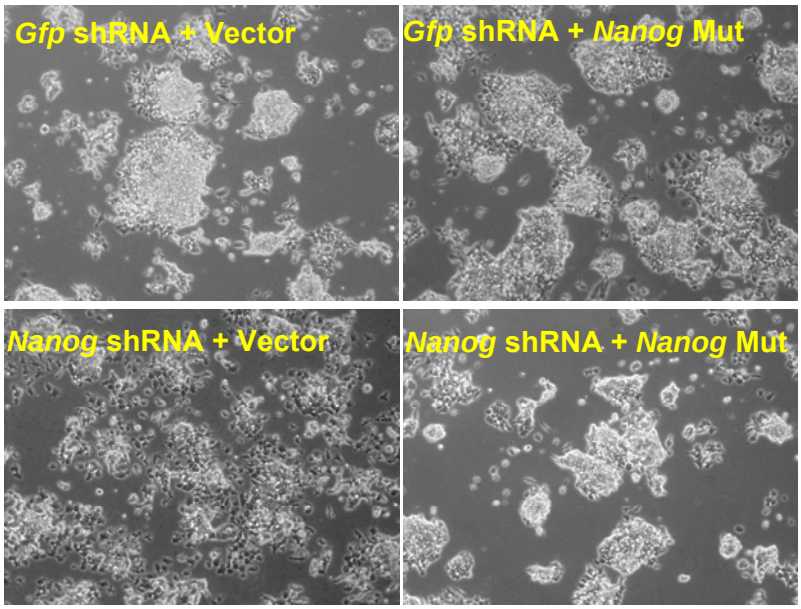

A
Relative expression %
120
100
80
60
40
20
0
Gfp shRNA
Nanog shRNA
Gfp shRNA
Nanog shRNA
Anti-Nanog
Anti-β actin
B
Nanog shRNA
Gfp shRNA
C
Relative expression %
120
100
80
60
40
20
0
Rex1
Oct4
Sox2
Gfp shRNA
Nanog shRNA
D
Endoderm
Mesendoderm
Mesoderm
Ectoderm
Trophectoderm
Relative expression %
700
600
500
400
300
200
100
0
Sox17
Gata4
Msx1
Gsc
Tbx6
Brachyury
Nestin
Fgf5
Hand1
Cdx2
Gfp shRNA
Nanog shRNA

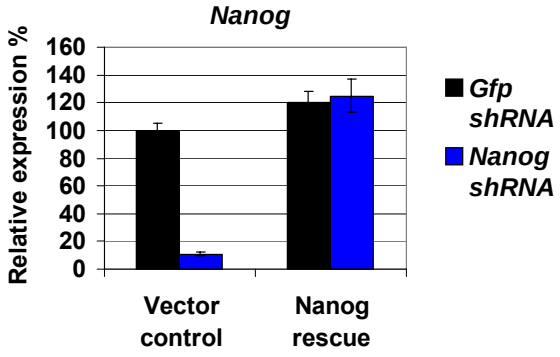
Figure 3.1 shRNA-mediated *Nanog* knockdown led to ES cell differentiation. **A**. The *Nanog* shRNA construct was effective in reducing Nanog at both mRNA (upper panel) and protein level (lower panel). *Nanog* mRNA level relative to the *Gfp* shRNA-transfected control cells was determined by reverse transcription followed by real-time PCR. Data are presented as the mean ± SEM. Nanog protein level was detected by western blotting and the lysates were probed using anti-Nanog or anti-ß actin antibodies. Both RNA and cell lysates were harvested after 4 days of puromycin selection. **B**. Flattened fibroblast-like cells were formed after *Nanog* knockdown and the cells lost the red staining signal of alkaline phosphatase, while alkaline phosphatase-positive ES cell colonies were maintained in the *Gfp* shRNA-transfected control cells. The cells were stained after 4 days of puromycin selection. **C**. Expression profiles of pluripotent markers after *Nanog* knockdown. The mRNA levels relative to the *Gfp* shRNA-transfected control cells were analyzed after 4 days of puromycin selection. Data are presented as the mean ± SEM. **D**. Real-time PCR showing the expression level changes of various lineage-specific markers in *Nanog*-depleted ES cells. The mRNA levels were analyzed after 4 days of puromycin selection and normalized against control the *Gfp* shRNA-transfected cells. Data are presented as the mean ± SEM.

A



Nanog shRNA target	GAG	AAC	TAT	TCT	TGC	TTA	CAA
Nanog Mut	GAG	AAT	TAT	AGC	TGC	CTA	CAG
Amino Acid	E	N	Y	S	C	L	Q

B



C

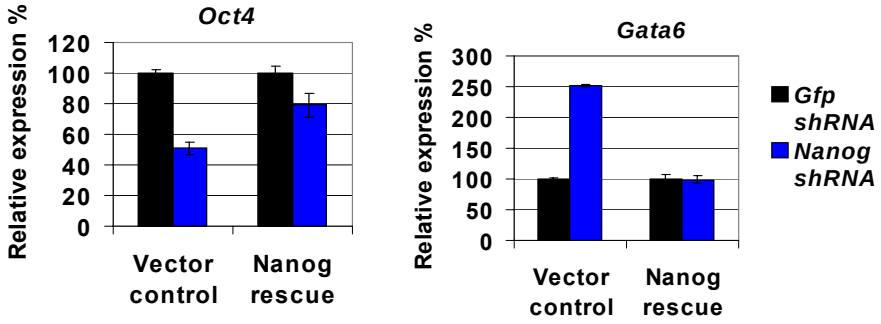


Figure 3.2 The rescue experiment demonstrating the specificity of the *Nanog* shRNA construct. **A**. The expression of RNAi-immune *Nanog* cDNA rescued *Nanog* shRNA-induced cell differentiation. Silent mutations were introduced into the *Nanog* open reading frame (ORF). The mutated region was no longer fully complementary to the shRNA, which rendered the transcript resistant to RNAi. This RNAi-immune *Nanog* expression plasmid was co-transfected along with the *Nanog* shRNA construct. *Gfp* shRNA construct and the rescue vector without *Nanog* cDNA insertion were used as controls. The pictures were taken 3 days after transfection. The lower panel shows the *Nanog* shRNA target sequence (in blue highlight) and the residues with silent mutations (in red). **B**. Endogenous *Nanog* was effectively depleted while exogenous *Nanog* was equally overexpressed in the *Gfp* shRNA- and the *Nanog* shRNA-transfected cells. The expression levels of *Nanog* were detected by reverse transcription followed by real-time PCR. Data are presented as the mean ± SEM. **C**. The expression of RNAi-immune *Nanog* cDNA rescued the effect of *Nanog* RNAi on the expression of both pluripotent gene *Oct4* (left panel) and lineage marker *Gata6* (right panel). Data are presented as the mean ± SEM.

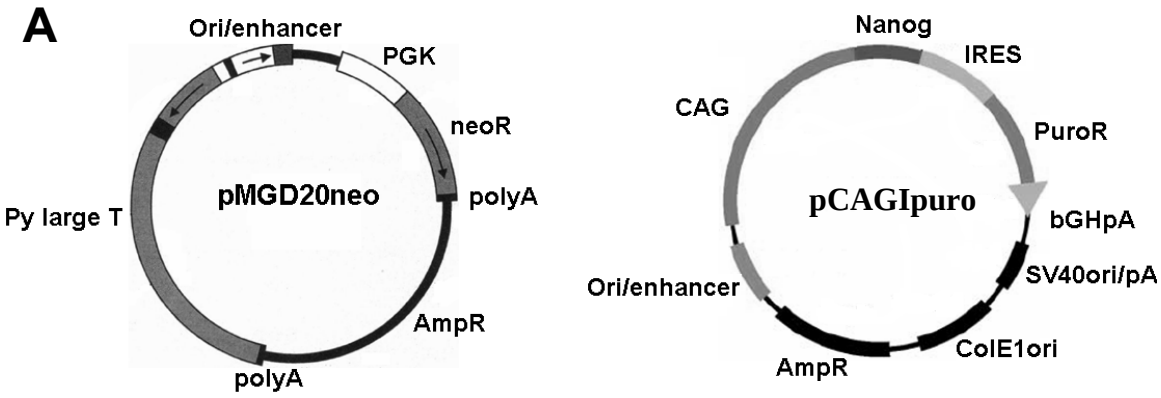
3.1.2. Establishment of *Nanog* overexpression ES cell line

To further characterize the roles of Nanog in ES cells, its expression level was elevated by using the episomal expression system (Gassmann *et al.,* 1995; Chambers *et al.,* 2003; Mitsui *et al*., 2003). In this system, two constructs, pMGD20neo and pCAGIpuro were introduced into mammalian cells (Figure 3.3A) (Gassmann *et al.,* 1995; Chambers *et al.,* 2003). The pMGD20neo construct contains a modified polyoma early region encoding the large tumor (Py large T) antigen, which can activate the *polyoma origin* (*ori*) contained in the pCAGIpuro construct. This serves to support the episomal replication of pCAGIpuro in the mammalian cells and prevents its integration into the chromosomes. Because of its high transfection efficiency (>90%) and stability, this episomal expression system is also known as the “super transfection system”.

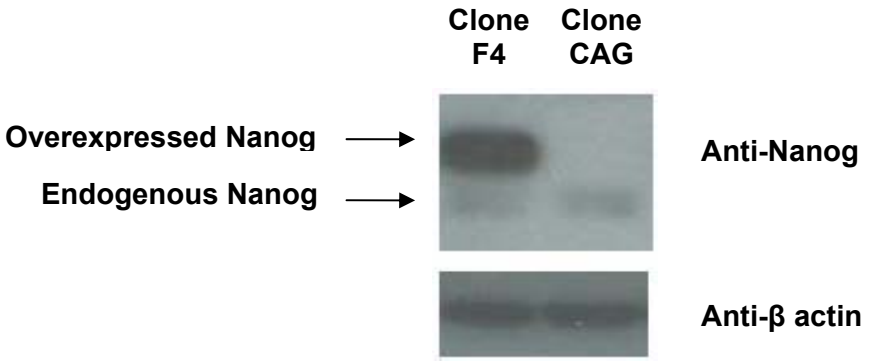
To establish the *Nanog* overexpression stable cell line, I performed two rounds of transfection. Firstly, the pMGDneo construct was transfected into ES cells. After five

days of neomycin selection (350ng/ul), 20 clones were picked and screened for their expression of *large T* gene by qPCR detection (data not shown). The clone with the highest *large T* expression was selected and subjected to a second round of transfection with the pCAGIpuro construct that expresses the *Nanog* cDNA. After 4 days of selection with both neomycin (350ng/ul) and puromycin (1ng/ul), 10 clones were picked and further checked for Nanog expression via western blotting (data not shown). Based on the criterion of high *Nanog* expression level, Clone F4 was selected and used for subsequent experiments. Transfection with mock pCAGIpuro vector was also performed to generate clone CAG to serve as a negative control. Western blot analysis showed that the expression level of *Nanog* in clone F4 was approximately eight times higher than the control cell line (Figure 3.3B). I next tested the ability of the *Nanog*-overexpressed cell line F4 in resisting differentiation. LIF was withdrawn from the ES culturing medium, which led to complete differentiation of the control CAG cells. However, the *Nanog*-overexpressed cell line F4 retained the colony morphology of ES cells (Figure 3.3C). Consistent with previous findings of Chambers *et al*. (2003), this result indicates that *Nanog* overexpression conferred the LIF independence of mouse ES cells. To check if the *Nanog*-overexpressed cells can also resist RA-induced differentiation, the cells were treated with 0.3μM of RA for 48 hours. While undifferentiated cells could hardly be found in the control CAG cells, the *Nanog*-overexpressed cell line F4 exhibited limited differentiation and retained the ES cell phenotype (Figure 3.3D). These experiments validated that the F4 cell line expressed a high level of exogenous *Nanog* that is functionally active. I subsequently used the F4 clone for further studies to dissect the roles of Nanog through uncovering its downstream target genes.

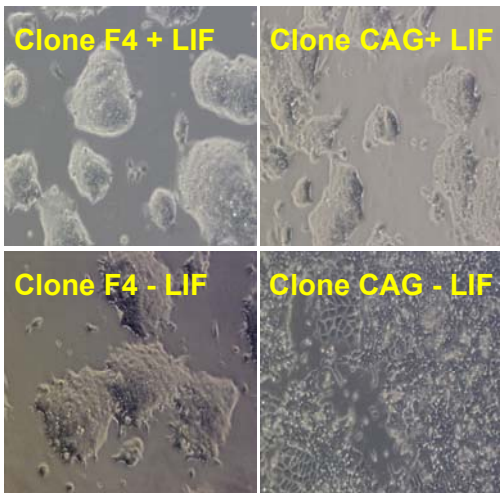
A


Ori/enhancer
PGK
neoR
polyA
pMGD20neo
Py large T
AmpR
polyA
Nanog
IRES
CAG
PuroR
pCAGIpuro
bGHpA
SV40ori/pA
Ori/enhancer
ColE1ori
AmpR

B


Clone F4
Clone CAG
Overexpressed Nanog
Endogenous Nanog
Anti-Nanog
Anti-β actin

C


Clone F4 + LIF
Clone CAG+ LIF
Clone F4 - LIF
Clone CAG - LIF

D

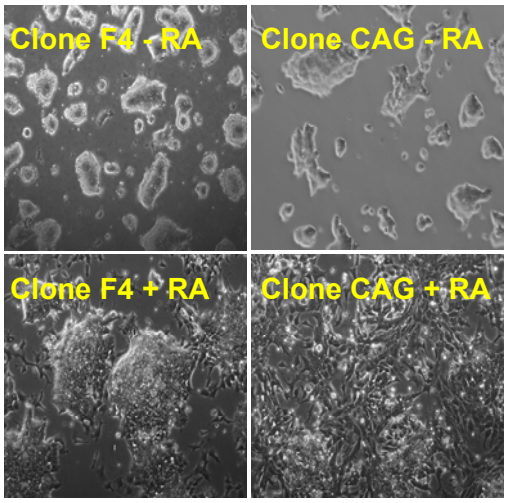

Clone F4 - RA
Clone CAG - RA
Clone F4 + RA
Clone CAG + RA

Figure 3.3 Establishment and characterization of *Nanog*-overexpressed cell line F4. **A**. The constructs used for establishing the *Nanog* overexpression ES cell line. **B**. *Nanog* was substantially overexpressed in the clone F4 cells. Cell lysates of the clone F4 cells and clone CAG control cells were analyzed by western blot using anti-Nanog and anti-ß actin antibodies. The higher molecular weight of overexpressed Nanog was due to the presence of an epitope-tag. The band representing endogenous Nanog was indicated. **C**. *Nanog* overexpression conferred ES cells resistance to LIF withdrawal. *Nanog*-overexpressed ES cells (Clone F4) and control ES cells (Clone CAG) were grown in the presence of LIF (+LIF) or absence of LIF (-LIF) for 60 hours. **D**. *Nanog* overexpression rendered ES cells resistant to RA-induced differentiation. *Nanog*-overexpressed ES cells (Clone F4) and control ES cells (Clone CAG) were grown in the presence 0.3 μM RA (+RA) or no RA (-RA) for 48 hours.

3.2. *Estrogen related receptor beta* (*Esrrb*) is a novel target of Nanog

To identify novel targets acting downstream of Nanog, affymetrix microarrays were used for expression profiling of mouse ES cells that were subjected to *Nanog* knockdown or *Nanog* overexpression (Figure 3.4). As compared to the *Gfp* shRNA knockdown control, 855 genes were down-regulated upon *Nanog* depletion (Figure 3.5A). On the other hand, the profiling of the clone F4 cells showed that 215 genes were up-regulated (Figure 3.5A). Among the responsive genes, 32 genes were positively regulated by Nanog as their expression levels were elevated in the clone F4 cells and reduced in *Nanog*-depleted ES cells (Figure 3.5A) (Appendix 1). These 32 genes are enriched for transcription factors that could potentially have critical roles in regulating the expression program specific to the ES cell state. I next asked if Nanog could maintain the expression levels of its prospective targets under differentiation-induced condition. To this end, I performed genome-wide expression profiling of the RA-treated F4 cell line. Following the RA treatment, the expression levels of 513 genes were sustained under the condition of *Nanog*-overexpression (Figure 3.5B). 15 genes were identified to be positively regulated

by Nanog and sustained in the RA-treated F4 cells (Figure 3.5B and C). I noticed that *Oct4* and *Sox2* were not upregulated upon *Nanog* overexpression even though overexpressed *Nanog* maintained Oct4 under the RA-treated condition (data not shown). Figure 3.4 provides an illustrative model by which Nanog regulates these 15 genes. In self-renewing ES cells, the activity of Nanog target genes is positively regulated by Nanog. Overexpression of *Nanog* elevates the expression of these genes, while knockdown of *Nanog* results in suppression of their expression. Under a differentiation-induced condition, the activating effect of Nanog on these genes can still maintain their expression.

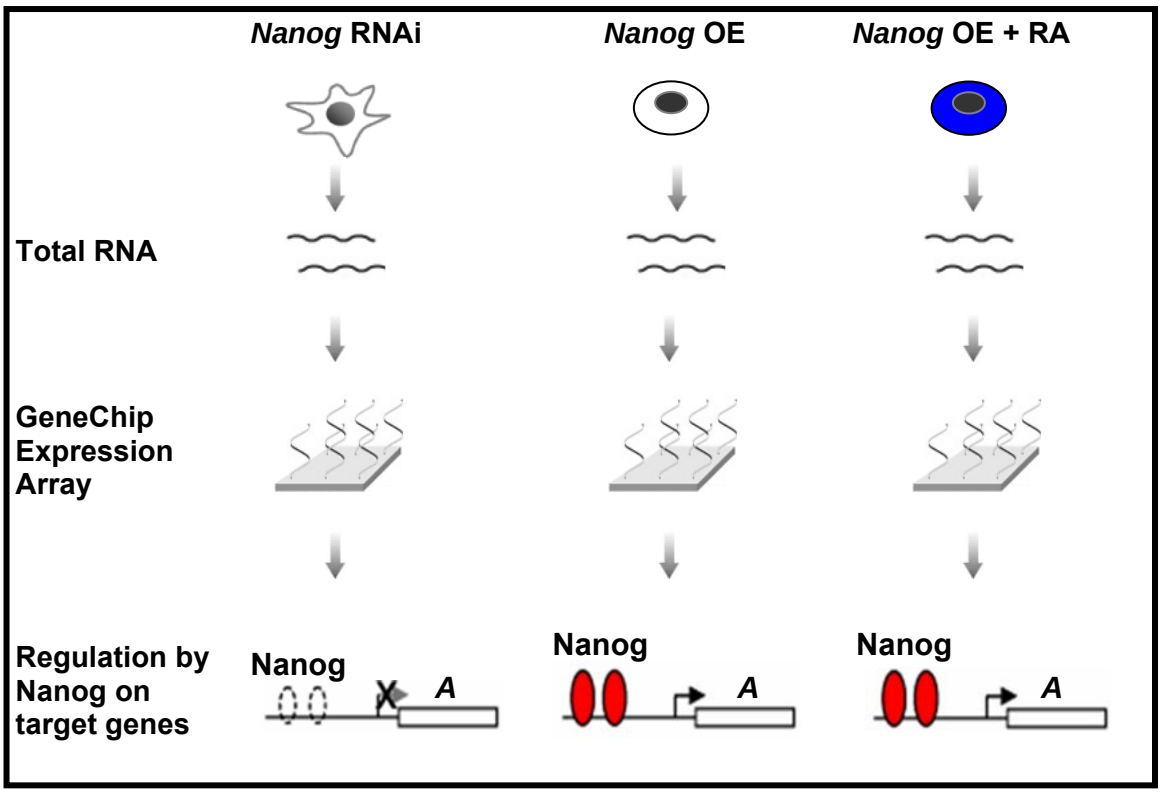
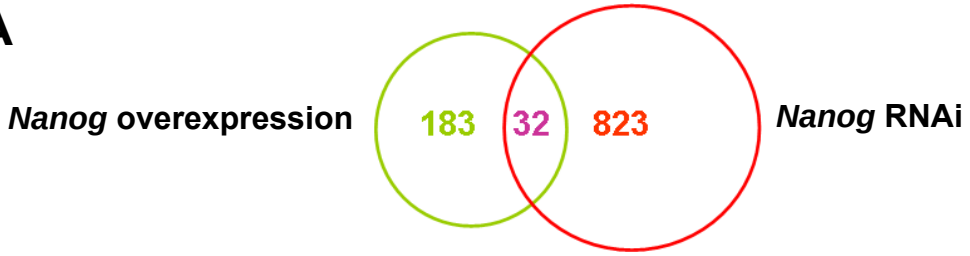


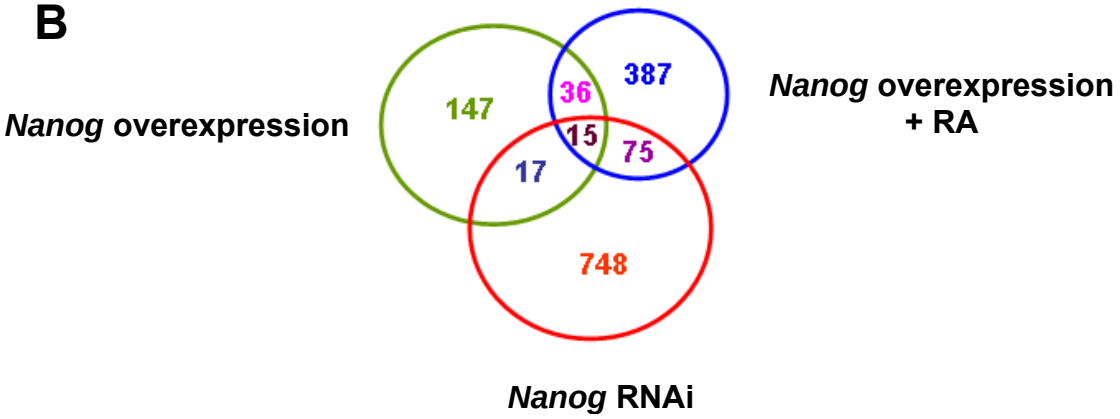
Figure 3.4 Schematic diagram illustrating the strategies undertaken to identify novel regulators of ES cells. The total RNA from the *Nanog*-depleted ES cells, *Nanog*-overexpressed F4 cell line and RA-treated F4 cell line were extracted and analyzed by Affymetrix GeneChip expression microarray. Downstream effector "A" which is positively regulated by Nanog is a candidate for novel regulators that maintain the "stemness" state of the ES cells. The transcript unit of "A" is depicted as a white box. The arrow indicates the direction of gene expression activated by Nanog. OE stands for overexpression.

Since the focus of this project is on transcription factors, 3 genes out of the 15 genes were short-listed due to their potential roles in transcriptional regulation. These three genes include *Estrogen related receptor beta (Esrrb)*, *amino-terminal enhancer of split* (*Esp1/Grg5/Aes*) and *BTB* (*POZ*) *domain containing 1* (*Btbd1*) (Pettersson, *et al*., 1996; Xie *et al.,* 1999; Lu *et al.,* 2001; Sanyal *et al.,* 2002; Brantjes *et al.,* 2001). Among the three candidate genes, only *Esrrb* has been reported to be closely related with pluripotent cells (Pettersson, *et al*., 1996). Pluripotent EC cells and ES cells have an elevated expression of *Esrrb*, and this elevated expression disappears upon differentiation (Pettersson, *et al*., 1996). This implies that Esrrb may be involved in maintaining pluripotent ES cells. To further confirm the regulatory effect of Nanog on *Esrrb*, I performed reverse transcription assay followed by realtime PCR to confirm the microarray results. In *Nanog*-knockdown ES cells, *Esrrb* expression level was reduced to about 30% (Figure 3.5D). Conversely, *Esrrb* transcript level was enhanced by 2 folds in the clone F4 cells in comparison with the clone CAG cells. This induction by Nanog could explain the maintenance of *Esrrb* level in RA-treated F4 cell line (Figure 3.5E). In addition, I found that in the *Oct4*-depleted ES cells, *Esrrb* was also dramatically down-regulated to about 20% (Figure 3.5D). All these findings suggest that the expression of *Esrrb* is closely associated with the pluripotent state of ES cells.

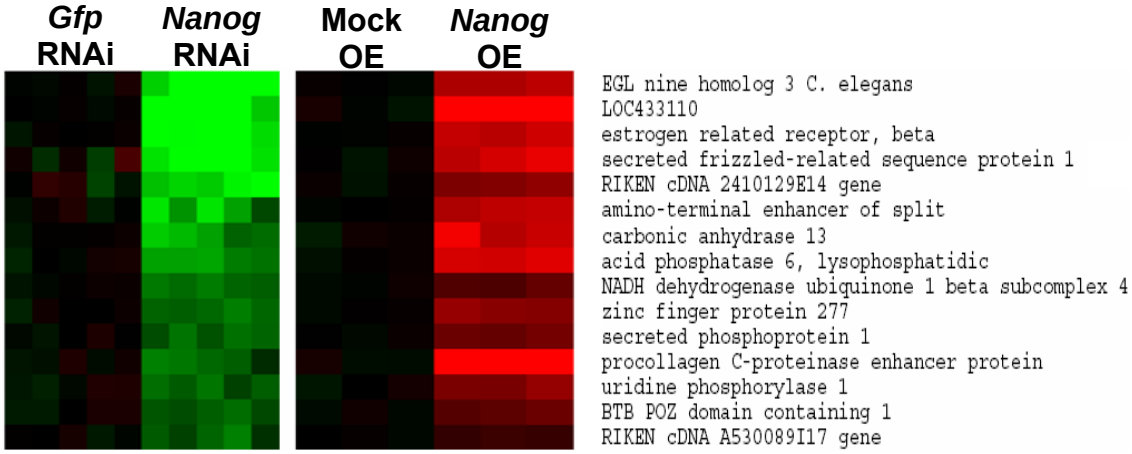
A


Nanog overexpression
183
32
823
Nanog RNAi

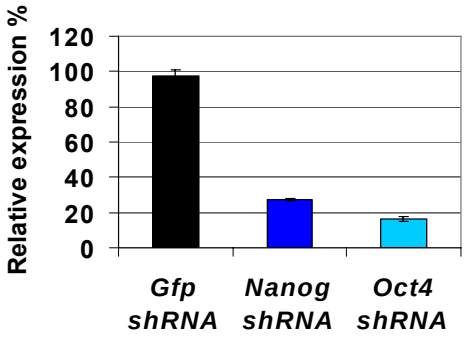
B


Nanog overexpression
147
36
387
15
75
17
748
Nanog overexpression + RA
Nanog RNAi

C


Gfp RNAi
Nanog RNAi
Mock OE
Nanog OE
EGL nine homolog 3 C. elegans
LOC433110
estrogen related receptor, beta
secreted frizzled-related sequence protein 1
RIKEN cDNA 2410129E14 gene
amino-terminal enhancer of split
carbonic anhydrase 13
acid phosphatase 6, lysophosphatidic
NADH dehydrogenase ubiquinone 1 beta subcomplex 4
zinc finger protein 277
secreted phosphoprotein 1
procollagen C-proteinase enhancer protein
uridine phosphorylase 1
BTB POZ domain containing 1
RIKEN cDNA A530089I17 gene

D



E

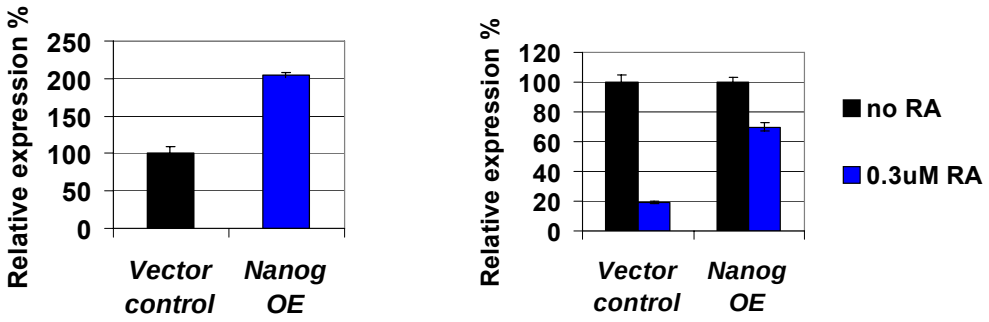
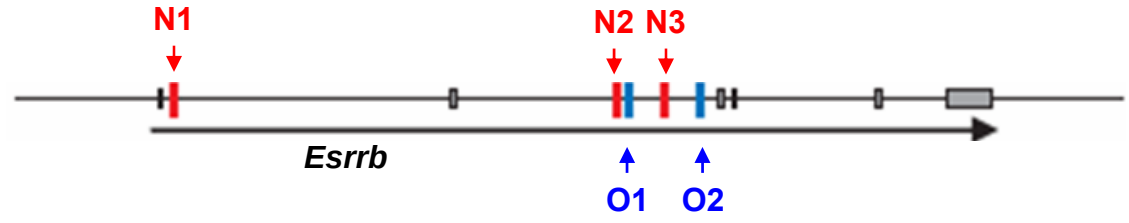


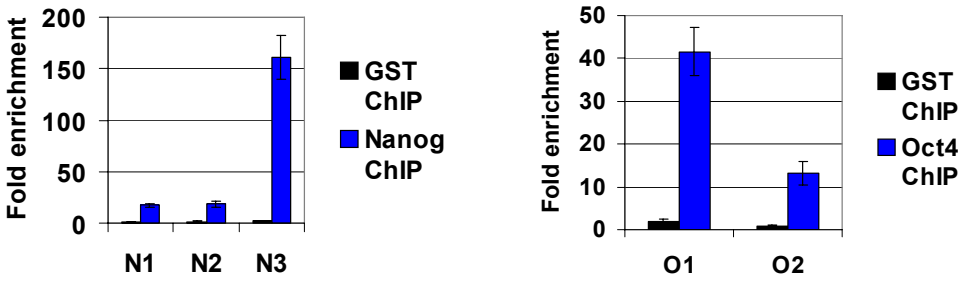
Figure 3.5 *Esrrb* is a downstream target positively regulated by Nanog. **A**. Venn diagram showing overlaps of the genes that were up-regulated in *Nanog*-overexpressed F4 cell line and down-regulated in *Nanog*-depleted ES cells. **B**.Venn diagram showing overlaps of the genes which were up-regulated in *Nanog*-overexpressed F4 cell line, down-regulated in *Nanog*-depleted ES cells, and maintained in the F4 cell line treated with RA. **C**. Heatmap showing the expression profile of the 15 genes which are positively regulated by Nanog. The genes were selected based on the fold changes of their expression levels after *Nanog* depletion and overexpression. They were sorted by the average expression ratio and mean centered. Five replicates were used for the *Nanog* depletion experiment while triplicates were used for the *Nanog* overexpression experiment. **D**. Realtime PCR validation for *Esrrb* level in *Nanog*- and *Oct4*-depleted ES cells. Data are presented as the mean ± SEM. **E**. Realtime PCR validation for *Esrrb* level in *Nanog* overexpression F4 cell line (Left panel) and upon RA-induced differentiation (Right panel). Data are presented as the mean ± SEM.

To validate the direct regulation of *Esrrb* by Nanog, others in our lab have performed chromatin immunoprecipitation (ChIP)-paired end ditag (PET) mapping and identified 3 intronic regions in the *Esrrb* locus where Nanog is bound (Loh *et al.,* 2006) (Figure 3.6A). I confirmed the binding of Nanog by means of ChIP-quantitative PCR (qPCR) (Figure 3.6B). Moreover, I validated the Oct4 binding to the *Esrrb* intronic regions by ChIP-qPCR (Loh *et al.,* 2006) (Figure 3.6 B). By means of electrophoretic mobility shift assay (EMSA), I showed that Nanog and Oct4 directly interact with the *Esrrb* genomic DNA fragments containing either the Nanog or Oct4 binding motif (Figure 3.6C and D). Importantly, the interactions were abolished upon the introduction of mutations in the Nanog or Oct4 binding motif. These findings suggest that *Esrrb* is bound and positively regulated by two transcriptional factors that are known to regulate ES cell pluripotency.

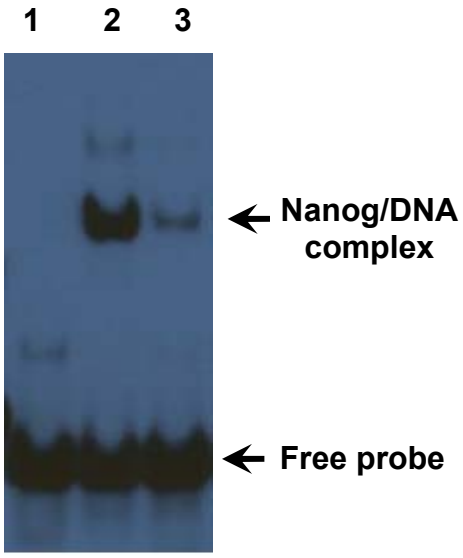
A



B



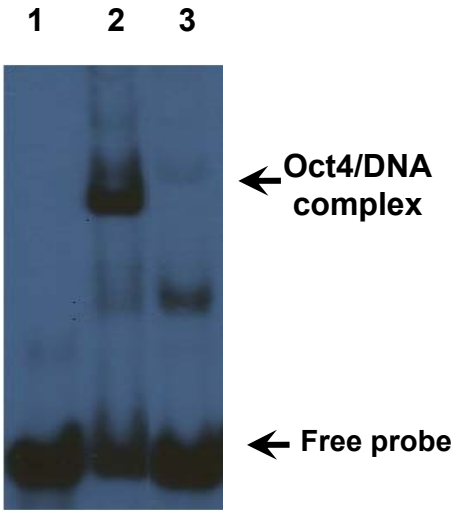
C



WT GTGGGAAGTGTTGCTATTCCATTAAGAAAGAACAGTAGCC

Mut GTGGGAAGTGTTGCTATTCAGGTAAGAAAGAACAGTAGCC

D



WT GATCCCAGTCTTTCAGATGCAAAAATAGCTGATTGCACTG

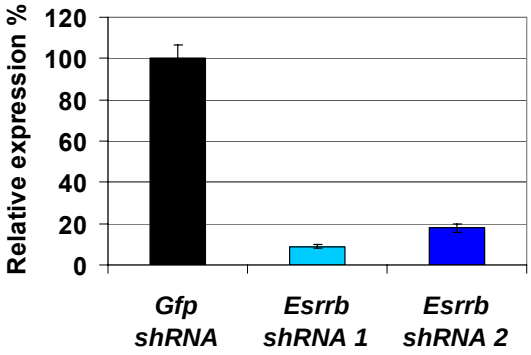
Mut GATCCCAGTCTTTCAGCCCCAAAAATAGCTGATTGCACTG

Figure 3.6 Nanog and Oct4 bind to *Esrrb in vivo* and *in vitro*. **A**. ChIP-PET mapping shows that *Esrrb* is bound by Nanog (red blocks N1-3) and Oct4 (blue blocks O1-2). Exons are depicted as gray boxes. The arrow indicates the direction and body of *Esrrb* gene, extending from first exon to last exon based on the University of California, Santa Cruz mouse genome coordinates. The numbers on the right indicate the window span represented. **B**. ChIP assay using Nanog antibody (left panel) and Oct4 antibody (right panel) confirmed their binding to the *Esrrb* locus. Data are presented as the mean ± SEM. **C.** *In vitro* binding of Nanog to the intronic sequences of *Esrrb*. EMSA was used to analyze the interaction between Nanog and a double-stranded DNA probe (WT) containing a Nanog binding motif labeled in red in the bottom panel (Loh *et al.,* 2006). The N3 Nanog binding site with the highest enrichment was selected for probe design and purified Nanog homeodomain was used for the assay. As compared to the control (No purified protein was added to the interaction mixture with probes) (Lane1), Nanog homeodomain formed a Nanog/DNA complex with the probe used (Lane2). Mutations introduced to the probe (Mut, mutated motif labeled in blue) dramatically suppressed its interaction with the Nanog homeodomain (Lane3). **D**. Oct4 binds to the intronic sequences of *Esrrb*. EMSA was used to analyze the interaction between Oct4 and a double-stranded DNA probe (WT) containing an Oct4 binding motif labeled in red (Loh *et al.,* 2006). The O1 Oct4 binding site with the highest enrichment was selected for probe design and purified Oct4 protein was used to detect the Oct4/DNA interaction. As compared to the control (No purified protein was added to the interaction mixture with probes) (Lane1), Oct4 formed a Oct4/DNA complex with the probe (Lane2). In contrast, the mutant probe (Mut, mutated motif labeled in blue) could not interact with the Oct4 protein (Lane3).

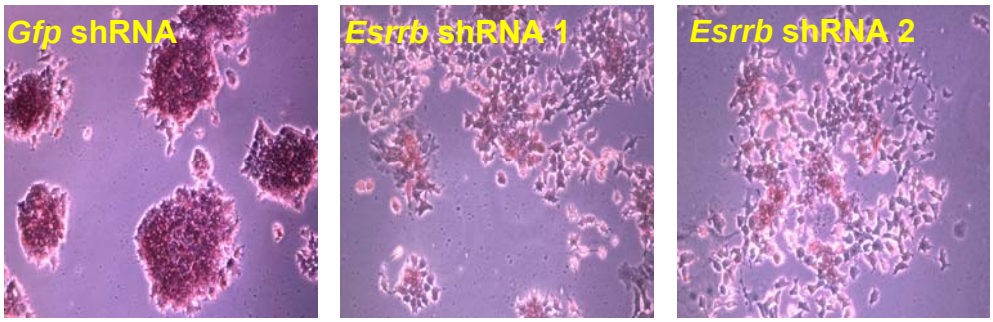
3.3. *Esrrb* plays a role in maintaining undifferentiated ES cells

To investigate the role of Esrrb in regulating ES cells, its expression was depleted using two shRNA constructs that target different sites of the *Esrrb* mRNA. Both constructs efficiently reduced the mRNA level of endogenous *Esrrb* (Figure 3.7A). More strikingly, *Esrrb* knockdown using the shRNA constructs resulted in ES cell differentiation (Figure 3.7B). The knockdown cells could not grow as colonies and became fibroblast-like. Moreover, alkaline phosphatase staining showed that *Esrrb*-depleted cells lost the characteristic red staining signal (Figure 3.7B). I further measured the expression levels of different marker genes, and found that the cellular differentiation induced by *Esrrb* knockdown was accompanied by a corresponding reduction in pluripotency markers and induction of genes associated with differentiation (Figure 3.7C and D). The *Nanog* and *Rex1* expression were decreased by 50% and 40% respectively. The *Oct4* and *Sox2* expressions were reduced subtly. Developmental marker genes, *Msx1*, *Fgf5, Hand1* and *Cdx2*, were induced significantly. However, two endoderm markers, *Sox17* and *Gata4*, were reduced by *Esrrb* depletion. These results indicate that *Esrrb*-depleted ES cells are composed of mixed differentiated cell lineages.

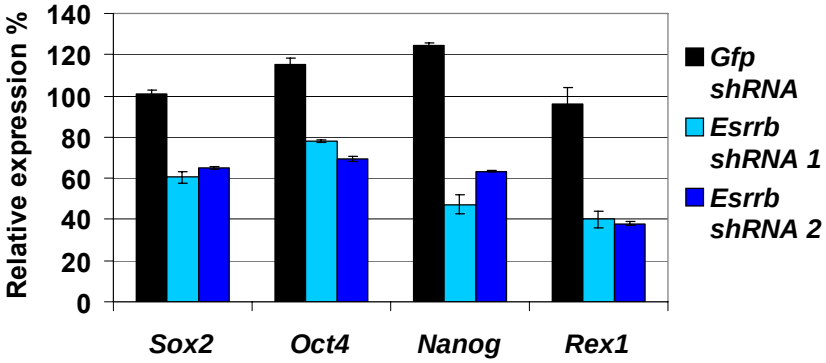
A


Relative expression %
120
100
80
60
40
20
0
Gfp shRNA
Esrrb shRNA 1
Esrrb shRNA 2

B


Gfp shRNA
Esrrb shRNA 1
Esrrb shRNA 2

C


Relative expression %
140
120
100
80
60
40
20
0
Sox2
Oct4
Nanog
Rex1
Gfp shRNA
Esrrb shRNA 1
Esrrb shRNA 2

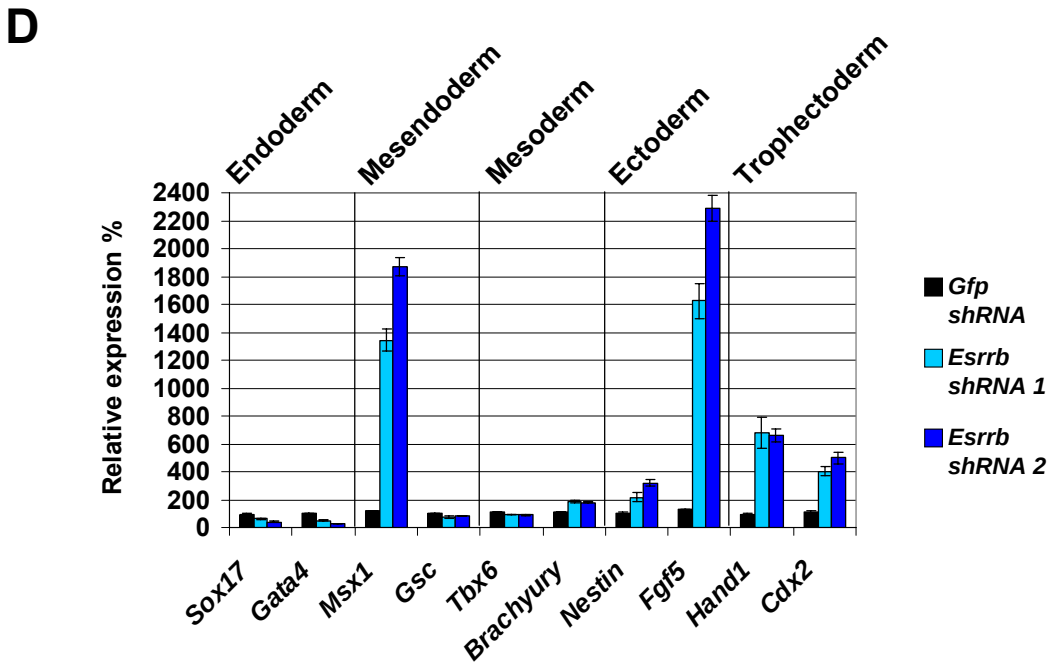
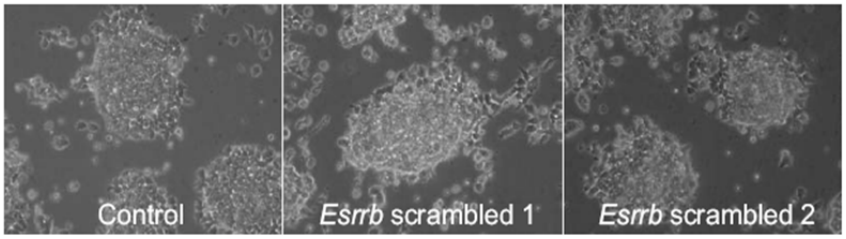


Figure 3.7 shRNA-mediated knockdown of *Esrrb* led to ES cell differentiation. **A**. *Esrrb* was effectively depleted by both of the two shRNA constructs that target different regions of *Esrrb*. The mRNA levels of *Esrrb* were determined by real-time PCR quantification of reverse-transcribed RNAs after 4 days of puromycin selection. Data are presented as the mean ± SEM. **B**. *Esrrb*-depleted ES cells underwent morphological differentiation. Positive alkaline phosphatase staining was used to define undifferentiated cells. As compared to the *Gfp* shRNA-mediated knockdown control, flattened fibroblast-like cells were formed after *Esrrb* knockdown and the red alkaline phosphatase-staining signal was lost. Two *Esrrb* shRNA constructs were used to show RNAi specificity. The cells were stained after 4 days of puromycin selection. **C**. Real-time PCR analysis of ES cell-associated gene expressions in the *Esrrb*-depleted ES cells. Cells were harvested after 4 days of puromycin selection. The levels of the transcripts were normalized against the *Gfp* shRNA-transfected cells. Data are presented as the mean ± SEM. **D**. Real-time PCR analysis of lineage-specific marker gene expressions in the *Esrrb*-depleted ES cells. Cells were harvested after 4 days of puromycin selection. The levels of the transcripts were normalized against the *Gfp* shRNA-transfected cells. Data are presented as the mean ± SEM.

To demonstrate the specificity of the *Esrrb* shRNA constructs, two control experiments were performed. Firstly, the *Esrrb* shRNA sequences were scrambled. Both of the scrambled *Esrrb* shRNA constructs did not lead to cellular differentiation (Figure 3.8A and B). As expected, the scrambled shRNAs constructs did not reduce *Esrrb* expression

(Figure 3.8C). In addition, the down-regulation of *Nanog* by *Esrrb* knockdown was restored, while the induction of *Fgf5* after *Esrrb* knockdown was not observed in the scrambled shRNA-transfected cells (Figure 3.8C). Next, the rescue experiment was performed by using the RNAi-immune *Esrrb* cDNA expression plasmid with two synonymous substitutions at the RNAi target sites. I co-transfected the plasmid containing either the wild-type *Esrrb* ORF or *Esrrb* shRNA1-immune cDNA along with *Esrrb* shRNA constructs into 293T cells. Figure 3.9A shows that both shRNA1 and shRNA2 abolished the expression of the Esrrb protein from the wild-type ORF. However, only shRNA2 was effective in down-regulating the protein level of Esrrb from the *Esrrb* shRNA1-immune cDNA (Figure 3.9A). A similar result was observed when the 293T cells were transfected with the *Esrrb* shRNA2-immune cDNA (data not shown). These results demonstrated the feasibility of using the RNAi-immune cDNAs to rescue the effects of knockdown induced by shRNAs. The *Esrrb* shRNA constructs were next co-transfected along with the RNAi-immune *Esrrb* cDNA expression plasmid into ES cells. Expectedly, the RNAi-immune *Esrrb* could rescue the cellular differentiation induced by knockdown of endogenous *Esrrb* (Figure 3.9B). Like the scrambled shRNA experiment, the knockdown effects of *Esrrb* on *Nanog* and *Fgf5* expression were rescued by the shRNA-immune *Esrrb* cDNA construct (Figure 3.9C). Taken together, the *Esrrb* shRNA constructs are specific in depleting the intended targets and the cellular differentiation induced by the shRNA-mediated *Esrrb* knockdown is not due to aberrant off-targeting effects. On the other hand, *Esrrb*-overexpressed ES cells maintained the morphology similar with the wild-type ES cells (data not shown).

A



B

***Esrrb* scrambled 1S: GAAGGCAGUUUAUUCAGUA**

***Esrrb* shRNA 1: GAUUCGAUGUACAUUGAGA**

***Esrrb* scrambled 2S: GCCCUACCUGAAAUGUUAU**

***Esrrb* shRNA 2: GAUCGUCUCGAAUCUACUA**

C

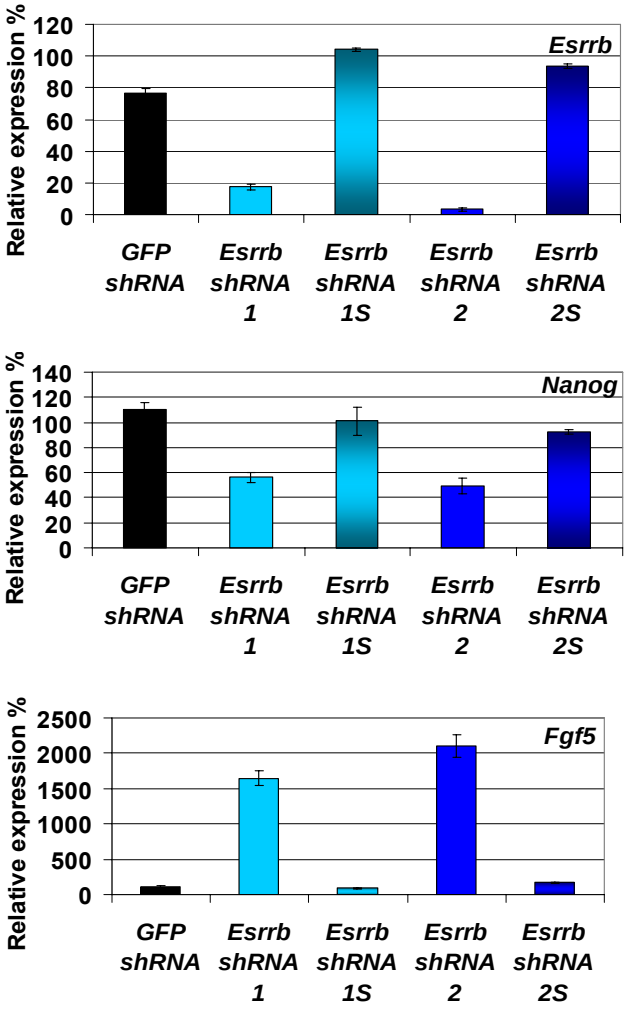
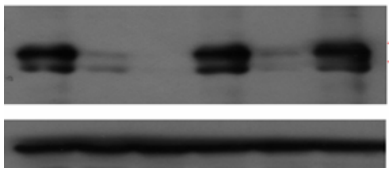


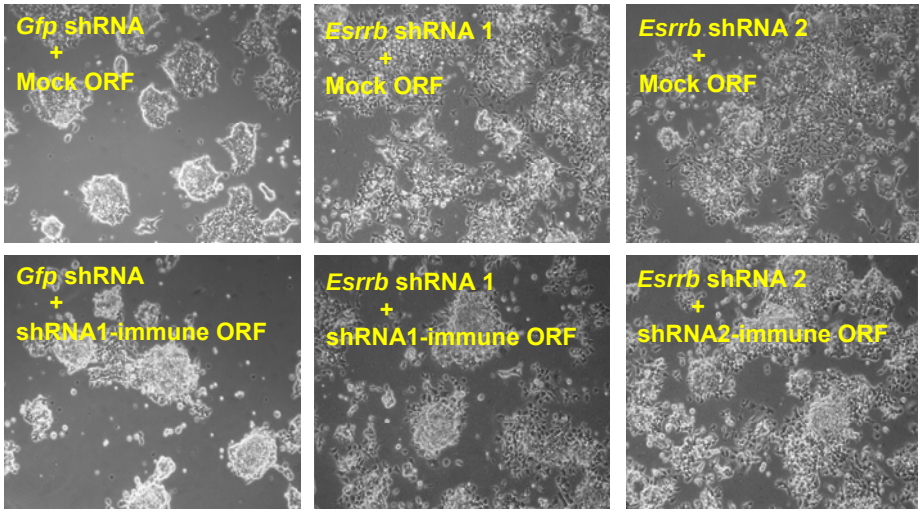
Figure 3.8 Scrambled *Esrrb* shRNAs did not lead to ES cell differentiation. **A**. ES cells expressing scrambled *Esrrb* shRNA sequences retained undifferentiated cell morphology. **B**. Sequences of *Esrrb* shRNAs and scrambled shRNAs. **C**. The scrambled shRNAs rescued the *Esrrb* knockdown effects on *Esrrb*, *Nanog* and *Fgf5* levels. The levels of the transcripts were normalized against the *Gfp* shRNA-transfected cells. Data are presented as the mean ± SEM.

A

Esrrb ORF	+	+	+	-	-	-
Esrrb shRNA1-immune ORF	-	-	-	+	+	+
Esrrb shRNA 1	-	-	+	-	-	+
Esrrb shRNA 2	-	+	-	-	+	-



B



Esrrb shRNA1 target	AAG	ATC	GTC	TCG	AAT	CTA	CTA
shRNA1-immune ORF	AAG	ATT	GTT	TCT	AAC	CTG	CTC
Amino Acid	K	I	V	S	N	L	L

Esrrb shRNA2 target	GAT	TCG	ATG	TAC	ATT	GAG	AAC
shRNA1-immune ORF	GAC	TCC	ATG	TAT	ATC	GAA	AAC
Amino Acid	D	S	M	Y	I	E	N

C

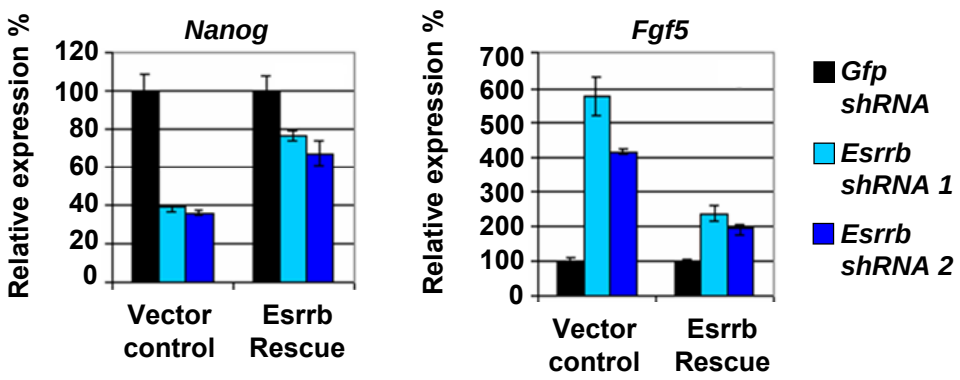


Figure 3.9 Rescue experiment demonstrating the specificity of the *Esrrb* shRNA constructs. **A**. Specificity of the *Esrrb* shRNA constructs was tested by co-transfection of the vector inserted with *Esrrb* ORF or the *Esrrb* shRNA1-immune ORF with *Esrrb* shRNA 1 or *Esrrb* shRNA 2 into 293T cells. The cell lysates were analyzed by western blot with anti-HA tag. ß tubulin served as a loading control. The red arrows indicate the HA-tagged Esrrb. **B**. Silent mutations were introduced into the *Esrrb* ORF. The mutated regions were no longer fully complementary to the shRNAs, which rendered the transcript resistant to RNAi. This RNAi-immune *Esrrb* expression plasmid was co-transfected along with the *Esrrb* shRNA constructs. *Gfp* shRNA construct and the vector without *Esrrb* cDNA insertion were used as controls. The pictures were taken after 3 days of selection. The lower panel shows the two *Esrrb* shRNA target sequences (in blue highlight) and the residues with silent mutations (in red). **C**. *Esrrb* shRNA-immune cDNA construct rescued the *Esrrb* knockdown effects on *Nanog* and *Fgf5* levels. The levels of the transcripts were normalized against the *Gfp* shRNA-transfected cells. Data are presented as the mean ± SEM.

3.4. Genome-wide mapping of Esrrb targets in ES cells

Since Esrrb is a transcription factor, it could exert its functions in ES cells by regulating the expressions of genes. As such, I sought to identify the targets of Esrrb in ES cells by using the newly developed ChIP-sequencing technology (Solexa, Illumina Inc).

3.4.1. Generation of Esrrb antibody for ChIP-sequencing assay

To identify the genomic locations of Esrrb binding, antibodies against Esrrb were generated. For the generation of antigens, I first performed Blast analysis and identified a region (aa 1-200) specific to Esrrb. The cDNA that encodes this fragment was cloned into the PET42b (+) vector which contains a GST tag. The expressed recombinant Esrrb protein was purified using GSH-sepharose beads. Protein SDS-PAGE electrophoresis followed by Commassie blue staining showed that only one specific band with the correct

size was detected (Figure 3.10A). The purified proteins were then used to immunize rabbits for the production of polyclonal antibodies.

To examine the specificity of the Esrrb antibody generated, western blotting was performed with lysates obtained from the *Esrrb*-overexpressed or *Esrrb*-depleted ES cells. A major band of the expected Esrrb size was detected in the control lysates (Mock OE and *Gfp* knockdown control). Notably, the intensity of this band was increased in the *Esrrb*-overexpressed ES cells. In contrast, the depletion of *Esrrb* mediated by both shRNA constructs abolished the signal (Figure 3.10B and C). These results demonstrate that the generated antibody is specific for Esrrb.

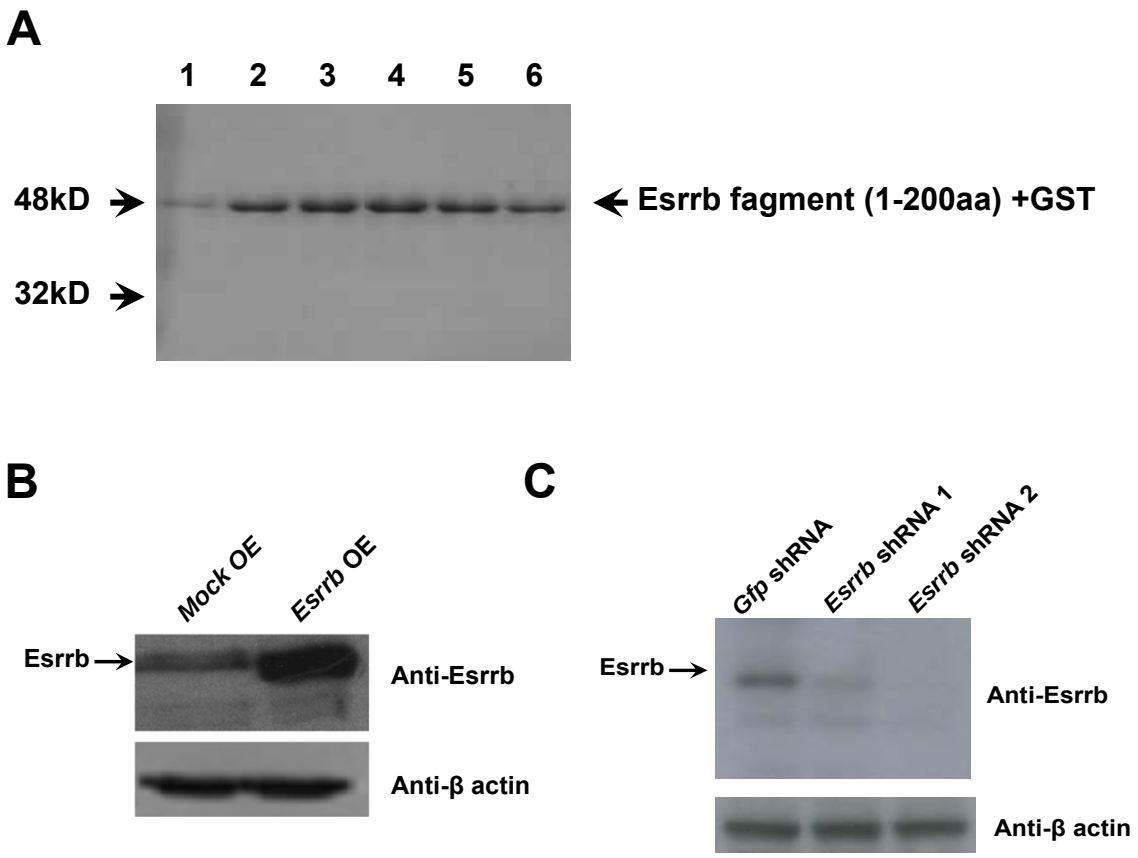


Figure 3.10 Generation and specificity of anti-Esrrb antibody for ChIP-sequencing. **A**. Commassie blue staining shows the recombinant Esrrb fragment (1-200aa) purified by using GSH-sepharose beads. 20mM GSH elution buffer was used to release protein from the beads, and the elution was collected into six 1.5ml-eppendoff tubes. 10ul from the six tubes of 1ml collections were loaded to SDS-PAGE protein gel for Commassie blue staining (Lane 1-6). **B.** Western blotting shows the induction of Esrrb protein level in *Esrrb*-overexpressed ES cells. The lysates were harvested 3 days after transfection and analyzed by with anti-Esrrb or anti-ß actin antibodies. **C**. Western blotting shows the depletion of Esrrb protein level in RNAi treated ES cells. The lysates were harvested 4 days after transfection and analyzed with anti-Esrrb or anti-ß actin antibodies.

Before embarking on the ChIP-sequencing assay, control experiments were performed to ensure that the antibody was able to specifically capture Esrrb-DNA interactions. Previous work by Wang *et al*. suggested that Esrrb is a putative interaction partner of

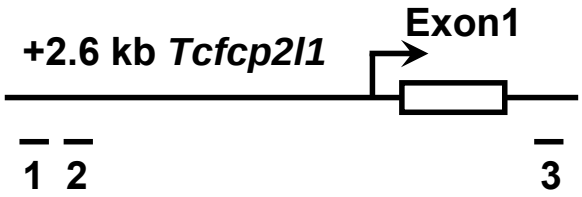
Nanog (Wang, *et al.,* 2006). Therefore, Esrrb may bind and regulate genes that are similarly regulated by Nanog in ES cells. To test the hypothesis, two Nanog target genes, *Tcfcp2l1* and *Rif1,* were selected. For *Tcfcp2l1*, I scanned the 2.6kb promoter region and found regions of Esrrb binding (Figure 3.11A and B). The specificity of the Esrrb binding is demonstrated by the lack of fold enrichment when ChIP was performed on the *Esrrb*-depleted ES cell chromatin (Figure 3.11C).

Diethylstilbestrol (DES) is an Esrrb antagonist that inhibits its transcriptional activity (Tremblay *et al.,* 2001). ES cells began to differentiate after three days of DES treatment (Figure 3.11D), presumably due to the prohibition of the Esrrb activity. The level of *Esrrb* mRNA remained constant throughout the three days of DES treatment, which suggests that DES functions by repressing the Esrrb activity at the protein level (Figure 3.11E). This provided an alternative way to confirm the specificity of the Esrrb antibody to detect the Esrrb binding. Since the onset of differentiation was later than two days after DES treatment, I harvested the chromatin on the second day of the DES treatment for ChIP assay. As expected, the DES treatment suppressed the binding of Esrrb to the *Tcfcp2l1* promoter (Figure 3.11F). For *Rif1,* I scanned its 760bp promoter sequence and identified regions with fold enrichment of Esrrb occupancy (Figure 3.12A and B). Similarly, the *Esrrb* knockdown and DES treatment reduced the enrichment of Esrrb binding to the *Rif1* promoter (Figure 3.12C and D).

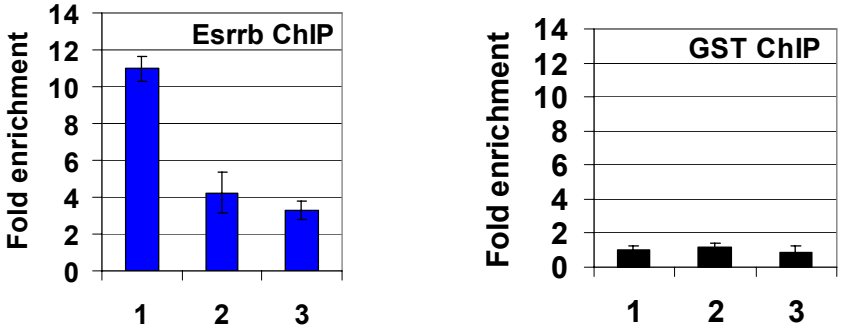
It can thus be concluded that the antibody raised against Esrrb is well-suited for ChIP experiments with the aim of identifying the downstream target genes of Esrrb. Moreover,

the ChIP results from the *Esrrb*-depleted and DES-treated ES cell samples further highlighted the specificity of the Esrrb antibody. Having obtained these results, I proceeded to perform genome-wide identification of Esrrb binding sites using the ChIP-sequencing technology.

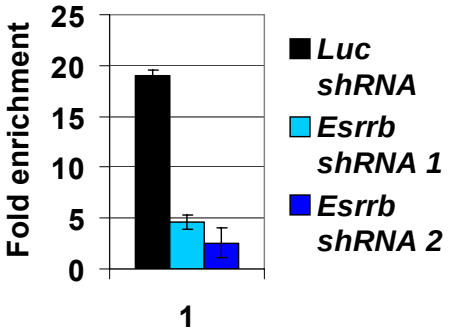
A


+2.6 kb *Tcfcp2l1*
Exon1
1
2
3

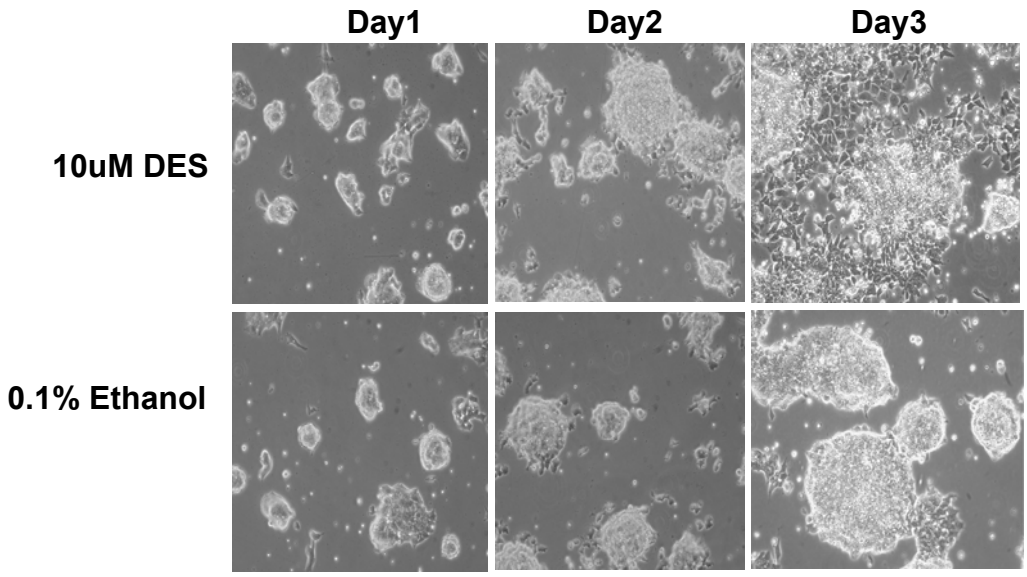
B


Esrrb ChIP
Fold enrichment
14
12
10
8
6
4
2
0
1
2
3
GST ChIP
Fold enrichment
14
12
10
8
6
4
2
0
1
2
3

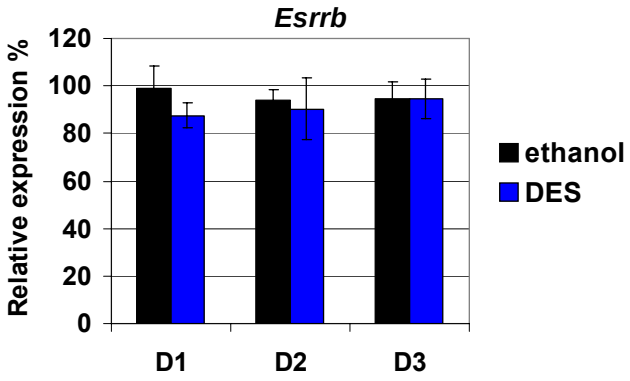
C


Fold enrichment
25
20
15
10
5
0
1
Luc shRNA
Esrrb shRNA 1
Esrrb shRNA 2

D


Day1
Day2
Day3
10uM DES
0.1% Ethanol

E


Esrrb
Relative expression %
120
100
80
60
40
20
0
D1
D2
D3
ethanol
DES

F

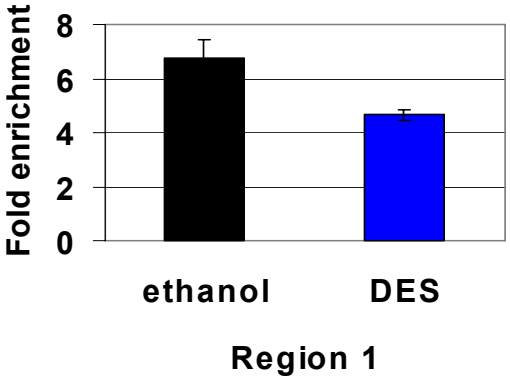

Fold enrichment
8
6
4
2
0
ethanol
DES
Region 1

Figure 3.11 *Tcfcp2l1* is a target of Esrrb. **A**. Schematic diagram showing the location of the amplicons (black bars labeled 1-3) used to detect ChIP-enriched fragments over the 2.6 kb *Tcfcp2l1* locus. Amplicons are numbered in order relative to their sites along the gene. The open box represents an exon. **B**. Real-time PCR detection of enriched fragments from the ChIP assay using Esrrb antibody. Fold enrichment is the relative abundance of DNA fragments at the amplified region over a control amplified region. GST antibody was used as a control. Data are presented as the mean ± SEM. **C**. Knockdown of *Esrrb* reduced the ChIP signal derived from anti-Esrrb antibody. ES cells were transfected with control *Luc* shRNA, *Esrrb* shRNA 1 or *Esrrb* shRNA 2. Primers for *Tcfcp2l1* promoter region 1 were used. Data are presented as the mean ± SEM. **D**. DES treatment resulted in ES cell differentiation. ES cells were treated with 10uM DES dissolved in 0.1% ethanol or control 0.1% ethanol for 3 days. Morphologies were examined by microscopy on each of the day. **E**. Expression level of *Esrrb* was maintained in DES-treated ES cells. ES cells were treated with control 0.1% ethanol or 10uM DES for three days. RNA was harvested on each of the day. Data are presented as the mean ± SEM. **F**. Esrrb binding to the *Tcfcp2l1* promoter region (region 1) was reduced upon DES treatment. ChIP assay was performed with the samples of ES cells treated with 10uM DES or control 0.1% ethanol for two days. Data are presented as the mean ± SEM.

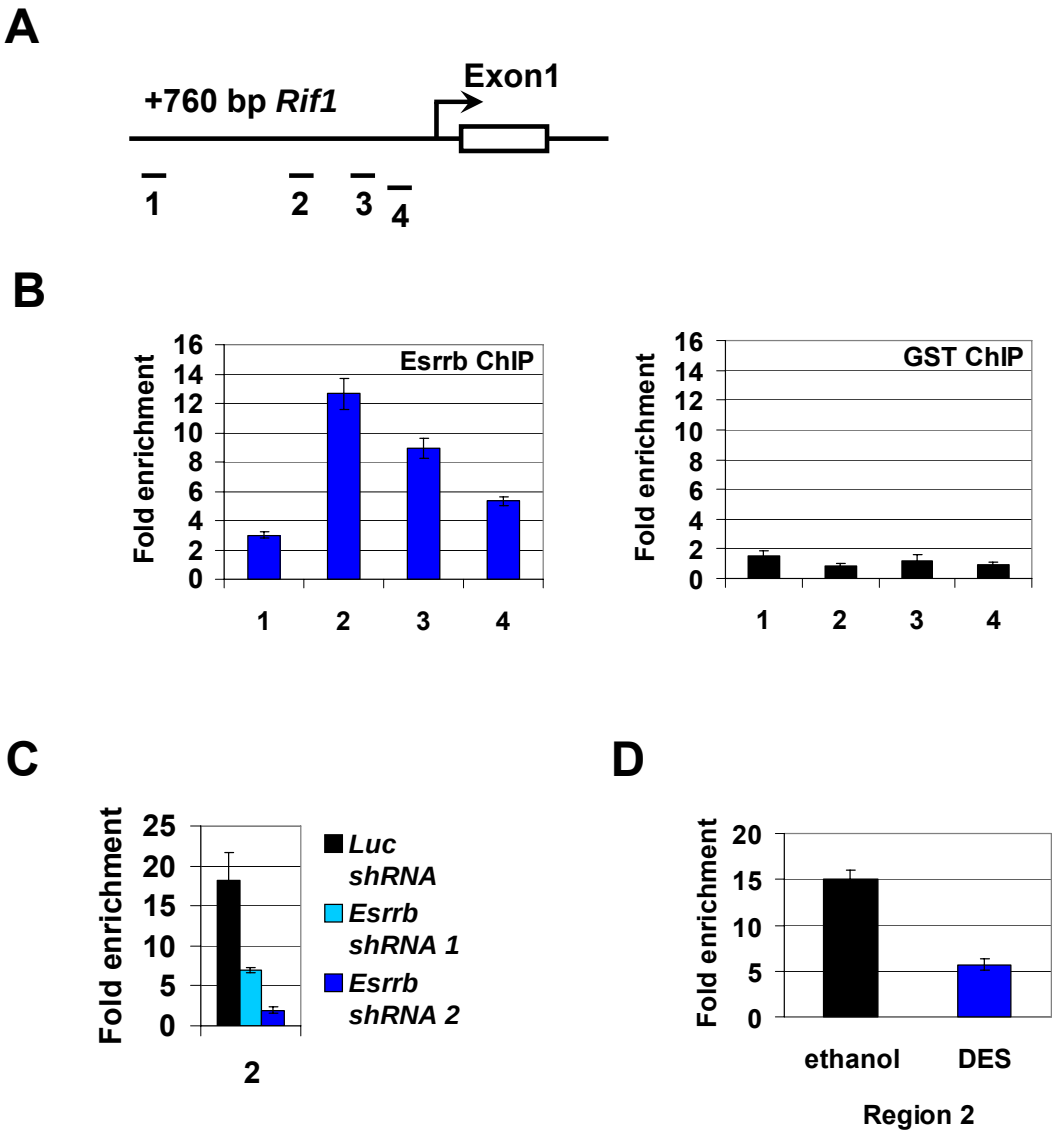


Figure 3.12 *Rif1* is a target of Esrrb. **A**. Schematic diagram showing the location of the amplicons (black bars labeled 1-4) used to detect ChIP-enriched fragments over the 760 kb *Rif1* promoter. Amplicons are numbered in order relative to their sites along the gene. The open box represents an exon. **B**. Real-time PCR detection of enriched fragments from ChIP assays in ES cells using anti-Esrrb antibody. Fold enrichment is the relative abundance of DNA fragments at the amplified region over a control amplified region. GST antibody was used as a mock ChIP control. Data are presented as the mean ± SEM. **C**. Knockdown of *Esrrb* abolished the ChIP signal derived from anti-Esrrb antibody. ES cells were transfected with control *Luc* shRNA, *Esrrb* shRNA 1 or *Esrrb* shRNA 2. Primers for *Rif1* promoter region 2 were used. Data are presented as the mean ± SEM. **D**. Esrrb binding to the *Rif1* promoter region 2 was reduced upon DES treatment. ChIP assay was performed with the samples of ES cells treated with 10uM DES or control 0.1% ethanol for two days. Data are presented as the mean ± SEM.

3.4.2. Genome-wide mapping of Esrrb binding sites

Figure 3.13 outlines the major procedural steps of the ChIP-sequencing technology. Esrrb-ChIP DNA fragments were generated for the Solexa sequencer. Mock GFP ChIP DNA sample was also generated to serve as a negative control. A total of 10,702,372 sequencing reads were obtained for the Esrrb ChIP-sequencing library. After filtering out the noise obtained from the GFP-ChIP mapping, 7,937,535 sequencing reads were uniquely mapped to the mouse genome (mm8), among which 3,609,843 reads mapped to Chromosome1-22 and X were further processed for analysis. To define the specific binding sites of Esrrb, false discovery rate (FDR) analysis and normalization against a background peak generated in the GFP ChIP-sequencing library were performed. The clusters are defined as Esrrb binding sites only when they satisfy the criteria of FDR<0.05 and peak heights ≥ 5. As such, 76,384 clusters representing potential Esrrb binding were identified (*www.t2g.bii.a-star.edu.sg*.).

To further confirm the mapping results, I randomly picked 100 enriched loci and performed Esrrb ChIP-qPCR. Out of the 100 tested loci, 94 were validated as enriched sites for Esrrb (Appendix 3) (Figure 3.14 A and B). Incidentally, the remaining six loci contained low enrichment detected by the ChIP-sequencing technology and thereby assigned as false positives. In addition, loci with higher peak heights detected by the ChIP-sequencing technology had corresponding higher fold enrichment as detected by the ChIP-qPCR approach. This indicates a close correlation between the two assays used in the detection of the ChIP-enriched DNA fragments. Based on the validation results, I

set peak height of 13 as the cutoff for the ChIP-sequencing results to ensure a mapping accuracy of 100%. 17,548 Loci with peak height of 13 or higher were identified to be high confidence sites of Esrrb binding. To further establish the specificity of Esrrb binding, ChIP followed by quantitative PCR for the top 20 loci with the highest fold enrichment was performed by using the normal or *Esrrb*-depleted ES cell chromatin. Results showed that the Esrrb occupancies on all the tested loci were significantly reduced upon *Esrrb* knockdown (Figure 3.15). The loci of *Rif1* and *Tcfcp2l,* which were previously identified as Esrrb targets from ChIP-qPCR assay, were also detected by the ChIP-sequencing (Solexa) method (Figure 3.16A and B). For *Tcfcp2l1*, ChIP-sequencing (Solexa) also uncovered additional Esrrb binding sites at its intronic regions.

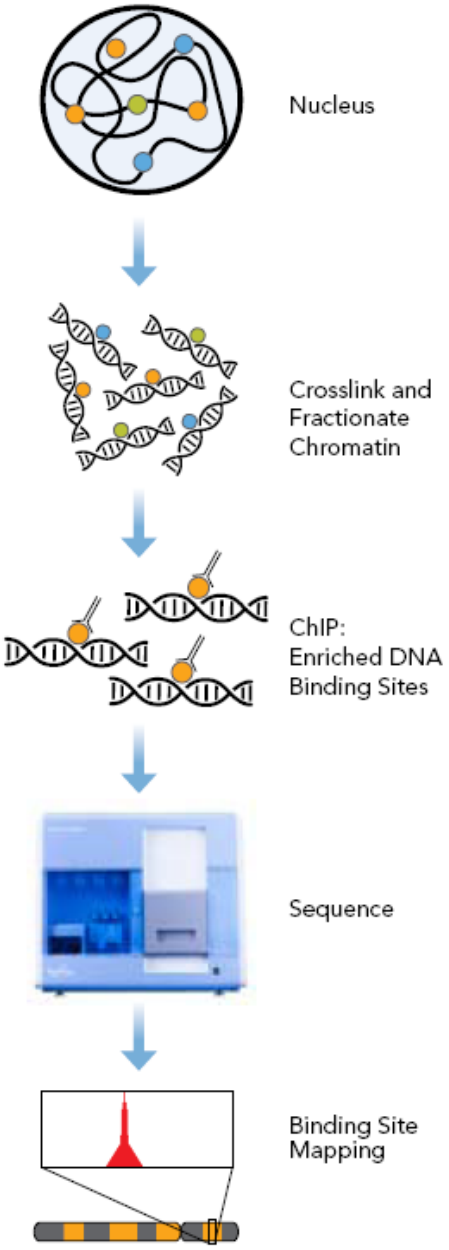


Figure 3.13 Workflow of ChIP-sequencing assay. The ChIP assay is processed to generate protein-DNA fragments by using the antibodies against Esrrb and GFP proteins. Oligonucleotide adapters are then linked to the small stretches of ChIP DNA to enable massively parallel sequencing. After size selection (150-300bp), all the resulting ChIP DNA fragments are sequenced simultaneously using the Genome Analyzer and Solexa sequencing technology. The resulting sequences are mapped back to the reference genome (mouse genome mm8), whereby the most frequently sequenced fragments formed peaks at specific genomic regions (Adapted from the Illunima ChIP-sequencing datasheet).

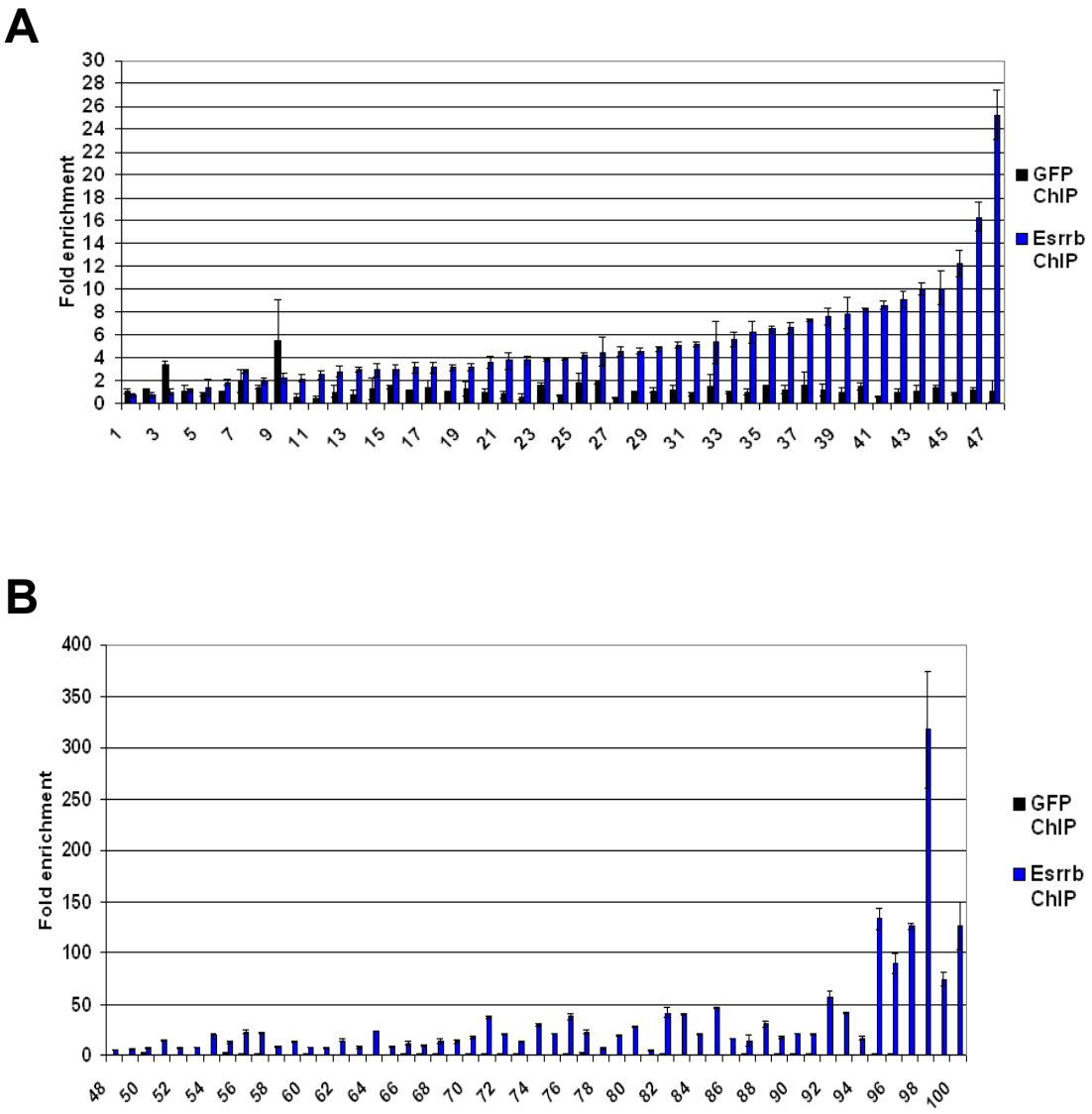


Figure 3.14 ChIP-quantitative PCR (ChIP-qPCR) validation of Esrrb binding sites identified from the ChIP-sequencing dataset. **A.** Validation results for the loci with low and moderate peak heights for Esrrb ChIP (<30). The locus coordinates were labeled as number 1-47 (Appendix 3). The first six loci show less than 2 folds of enrichment, suggesting that they are false positives. qPCR was performed with the Esrrb-ChIP DNA fragment, and the results were normalized against input DNA and a negative region without Esrrb binding. Data are presented as the mean ± SEM. **B.** Validation results for the loci with high peak heights (≥30). The locus coordinates were labeled as number 48-100 (Appendix 3). qPCR was performed with the Esrrb-ChIP DNA fragment, and the results were normalized against input DNA and a negative region without Esrrb binding. Data are presented as the mean ± SEM.

Table 3.1 20 loci with high peak heights were chosen for validation by Esrrb ChIP-quantitative PCR with the *Esrrb*-depleted ES cell chromatin.

Locus	Peak height
chrX:67690979-67690988	39
chr3:88732281-88732284	40
chr13:52234757-52234757	48
chr5:149051358-149051359	55
chr4:62942115-62942120	58
chr2:92225592-92225596	60
chr11:90046146-90046146	61
chr5:136386836-136386836	64
chr7:83779281-83779285	66
chr15:96161593-96161593	67
chr14:23086460-23086468	69
chr6:54628678-54628678	72
chrX:12618787-12618788	76
chr16:6997220-6997221	77
chr12:87391144-87391147	148
chr9:20765401-20765401	245
chr2:5451939-5451940	245
chr8:73576520-73576520	254
chr2:19298400-19298400	259
chr18:35169414-35169414	287

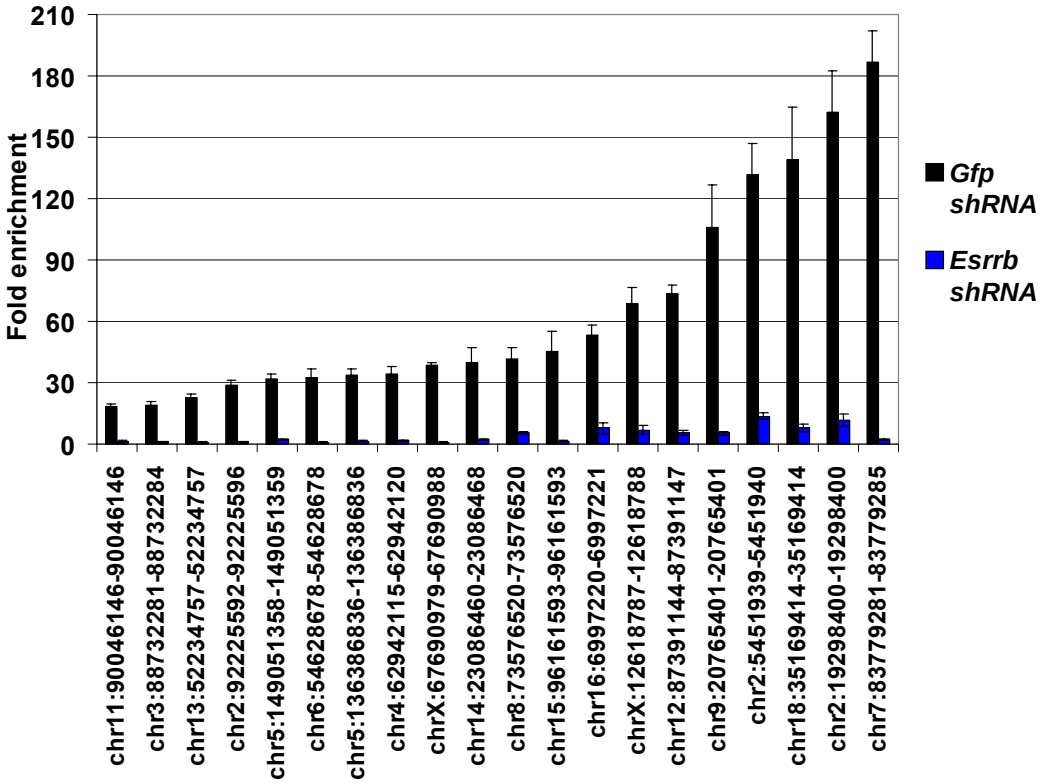


Figure 3.15 *Esrrb* depletion led to the abolishment of Esrrb occupancy. The top 20 loci with the highest fold enrichments were selected from the 94 validated Esrrb-bound sites. Esrrb ChIP-qPCR was performed and the results were normalized against the input DNA and a negative region without Esrrb binding. Chromatin was harvested from ES cells transfected with *Esrrb* shRNA 1 after 4 days of selection. *Gfp* shRNA-transfected ES cells were used as a negative control. Data are presented as the mean ± SEM.

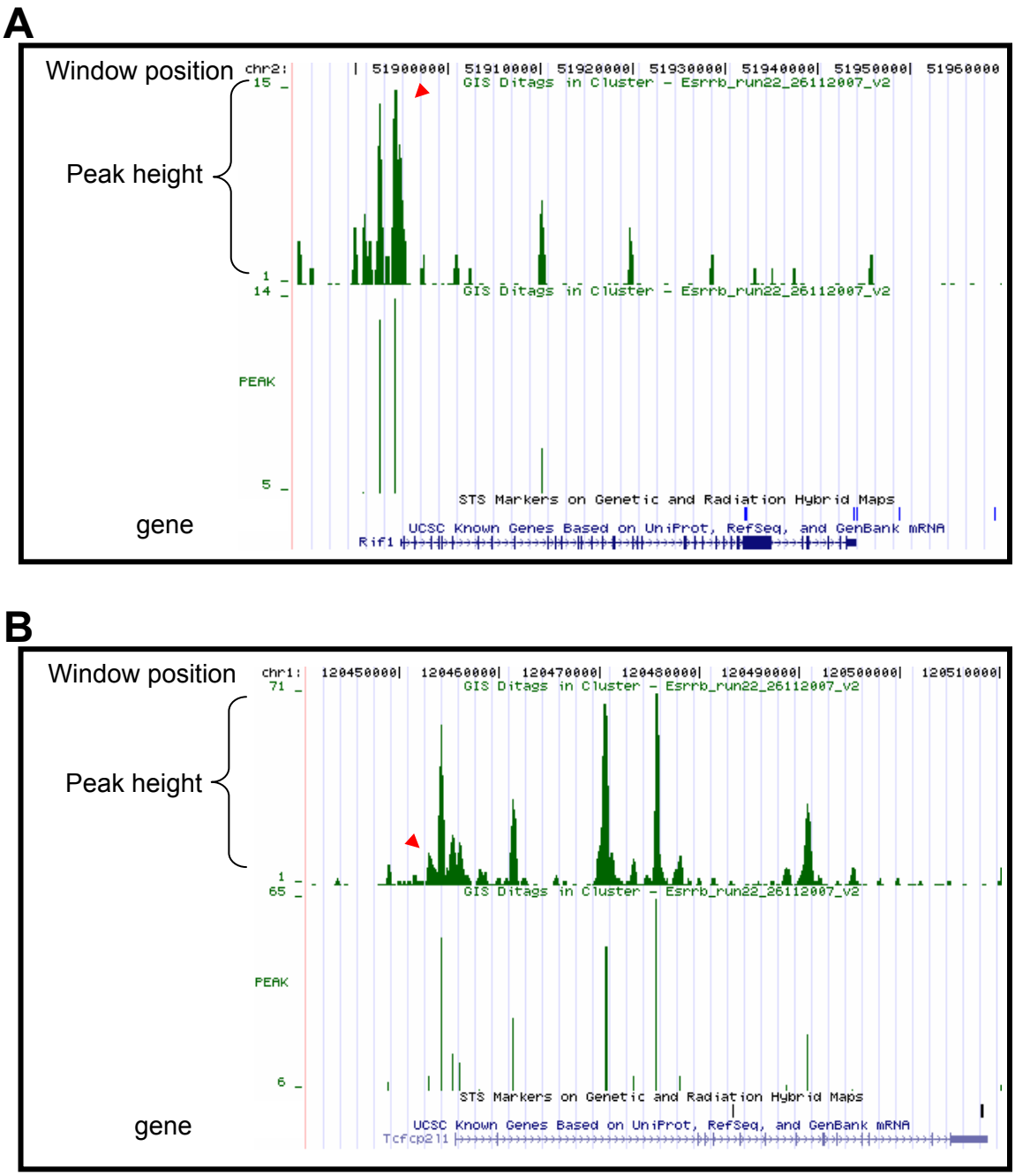
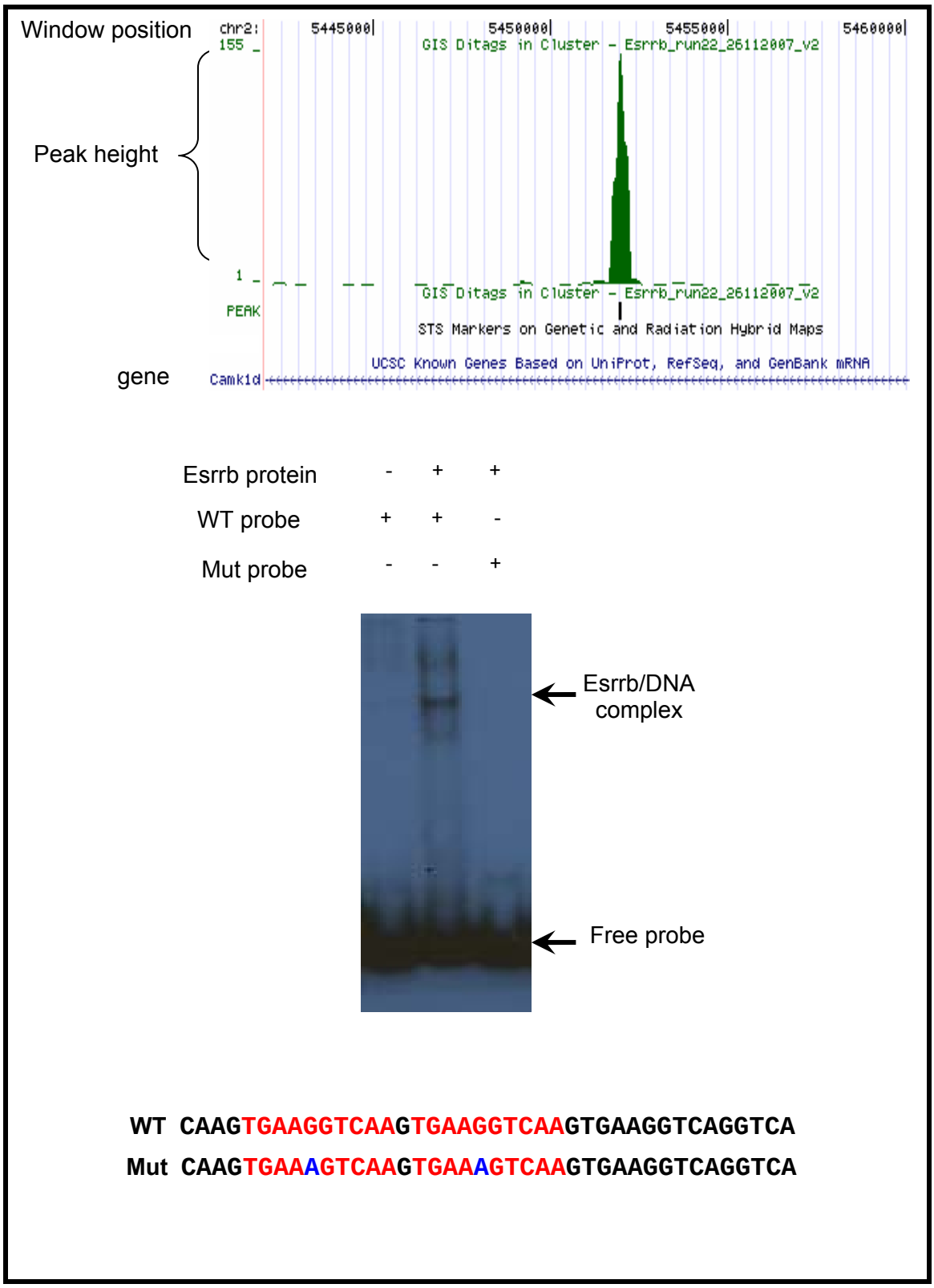


Figure 3.16 The screen shot of the T2G browser showing the binding profiles of Esrrb on *Tcfcp2l1* and *Rif1* loci detected by ChIP-sequencing assay. **A**. Peaks corresponding to Esrrb binding at the locus of *Rif1*. A sharp peak at the *Rif1* promoter as indicated by the red arrow was detected previously by ChIP-qPCR (Fig 3.12). **B**. Peaks corresponding to Esrrb binding at the locus of *Tcfcp2l1*. A sharp peak at the *Tcfcp2l1* promoter as indicated by the red arrow was detected previously by ChIP-qPCR (Fig 3.11). Additionally, four novel peaks were detected in the intronic region of *Tcfcp2l1* by ChIP-sequencing.

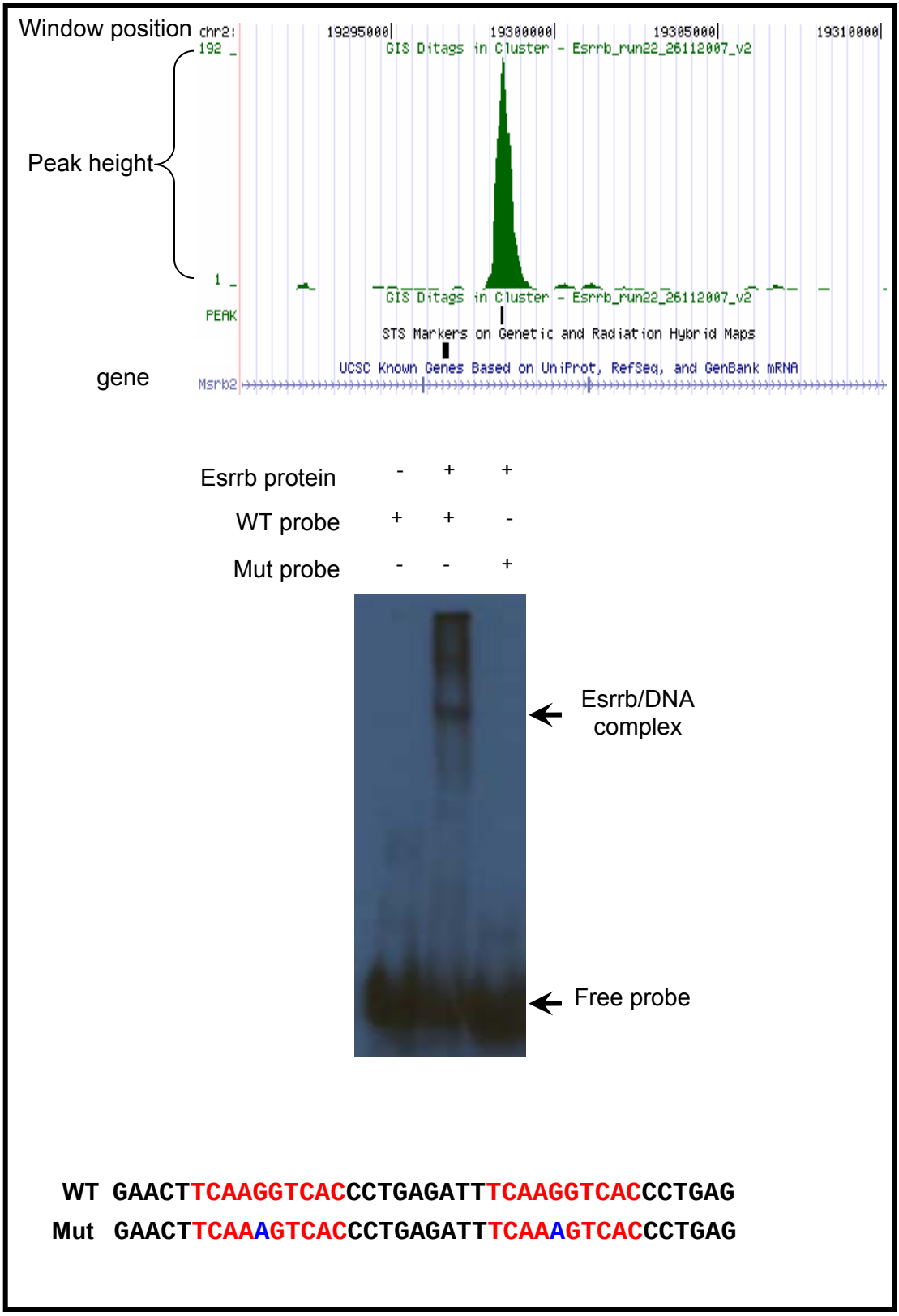
Analysis of the dataset that contains the genome-wide bound sequences of Esrrb has generated a consensus *cis* element of Esrrb binding. Interesting, this element matched the enhancer element ERRE (estrogen related receptor element) that was previously described as the motif mediating Esrrb-DNA interaction (Figure 3.17) (See Discussion). I further performed an EMSA assay to confirm the *in vitro* binding activity of Esrrb to the consensus element. Five regions with high enrichment of Esrrb binding were randomly selected and the probes designed for these regions contain the ERRE motif. The EMSA results showed that all the five probes were shifted by recombinant Esrrb protein, and these interactions were abolished by the introduction of mutations to the probes in the ERRE motif (Figure 3.18). Based on these results, I concluded that Esrrb can interact directly with DNA via the ERRE motif.

Figure 3.17 The *cis*-element mediating Esrrb-DNA interaction identified from the Esrrb ChIP-sequencing dataset. This work was done by Han Xu from the Bioinformatics group at the Genome Institute of Singapore.

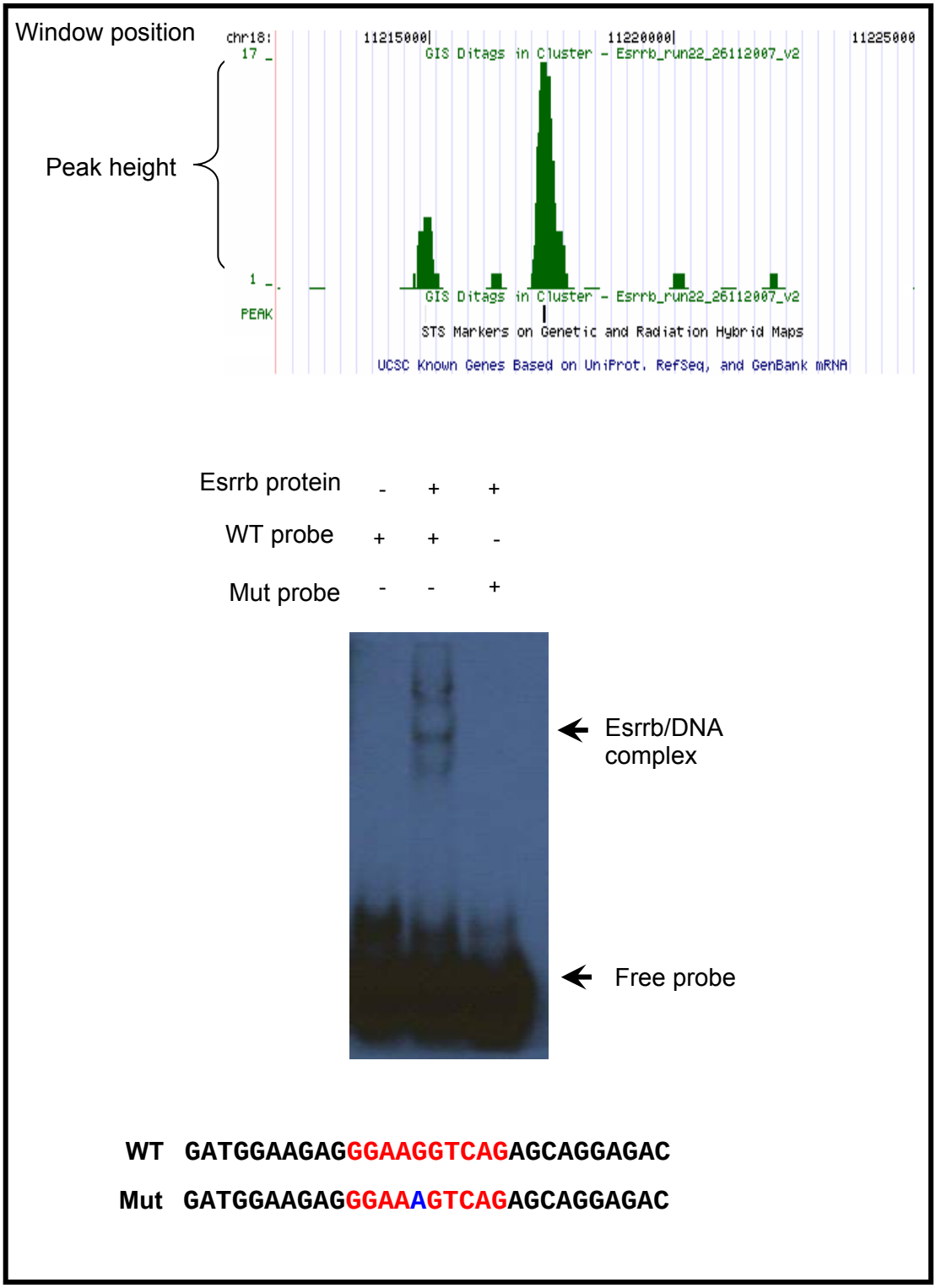
A


Window position
chr2:
5445000
5450000
5455000
5460000
155
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
Peak height
1
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
PEAK
STS Markers on Genetic and Radiation Hybrid Maps
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
gene
Camk1d
Esrrb protein - + +
WT probe + + -
Mut probe - - +
Esrrb/DNA complex
Free probe
WT CAAGTGAAGGTCAAGTGAAGGTCAAGTGAAGGTCAGGTCA
Mut CAAGTGAAAGTCAAGTGAAAGTCAAGTGAAGGTCAGGTCA

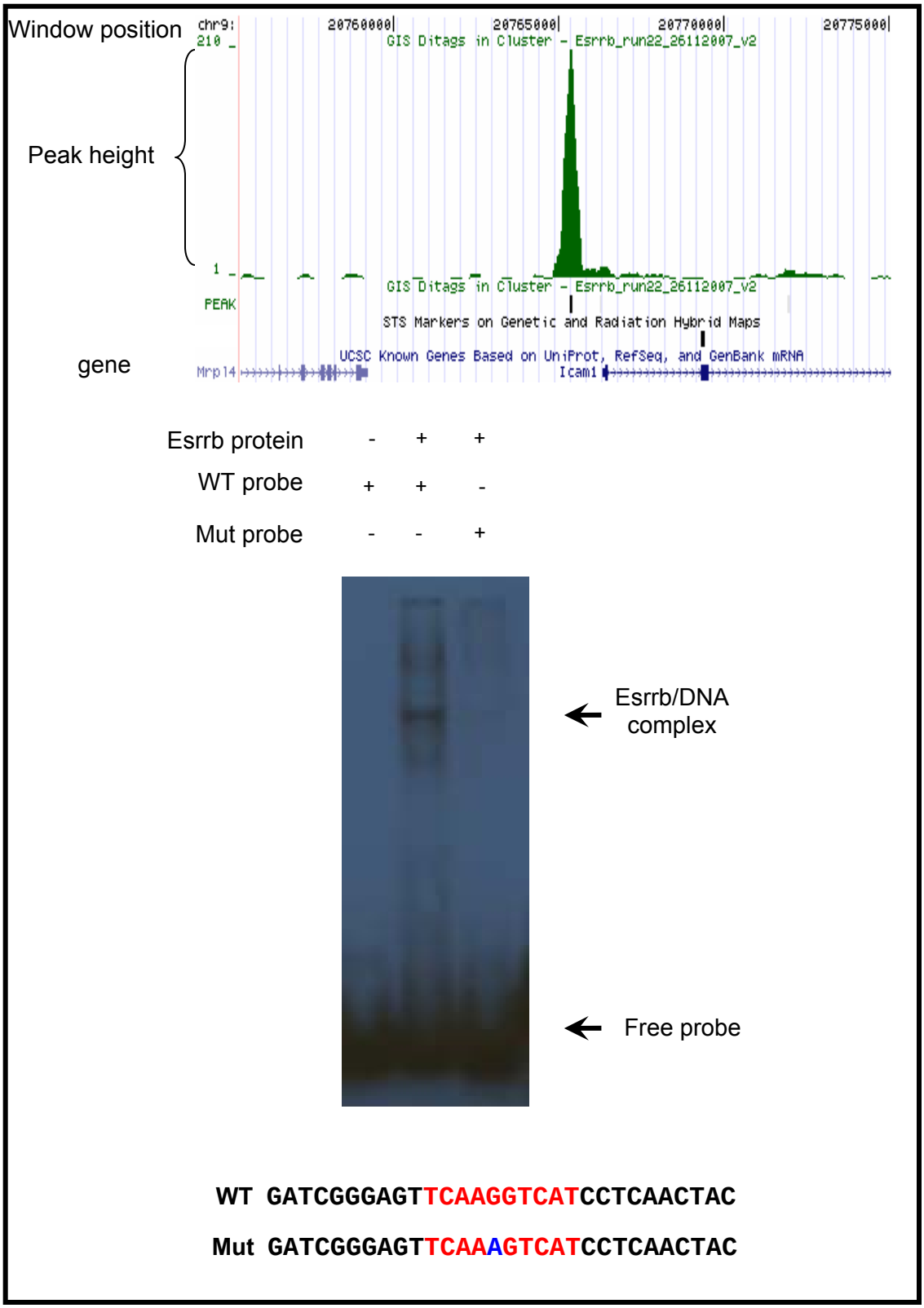
B


Window position
chr2:
19295000
19300000
19305000
19310000
192
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
Peak height
1
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
PEAK
STS Markers on Genetic and Radiation Hybrid Maps
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
gene
Msrb2
Esrrb protein - + +
WT probe + + -
Mut probe - - +
Esrrb/DNA complex
Free probe
WT GAACTTCAAGGTCACCCTGAGATTTCAAGGTCACCCTGAG
Mut GAACTTCAAAGTCACCCTGAGATTTCAAAGTCACCCTGAG

C


Window position
chr18:
11215000
11220000
11225000
17
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
Peak height
1
PEAK
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
STS Markers on Genetic and Radiation Hybrid Maps
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
Esrrb protein - + +
WT probe + + -
Mut probe - - +
Esrrb/DNA complex
Free probe
WT GATGGAAGAGGGAAGGTCAGAGCAGGAGAC
Mut GATGGAAGAGGGAAAGTCAGAGCAGGAGAC

D


Window position
chr9:
20760000
20765000
20770000
20775000
210
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
Peak height
1
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
PEAK
STS Markers on Genetic and Radiation Hybrid Maps
gene
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
Mrp14
Icam1
Esrrb protein - + +
WT probe + + -
Mut probe - - +
Esrrb/DNA complex
Free probe
WT GATCGGGAGTTCAAGGTCATCCTCAACTAC
Mut GATCGGGAGTTCAAAGTCATCCTCAACTAC

E

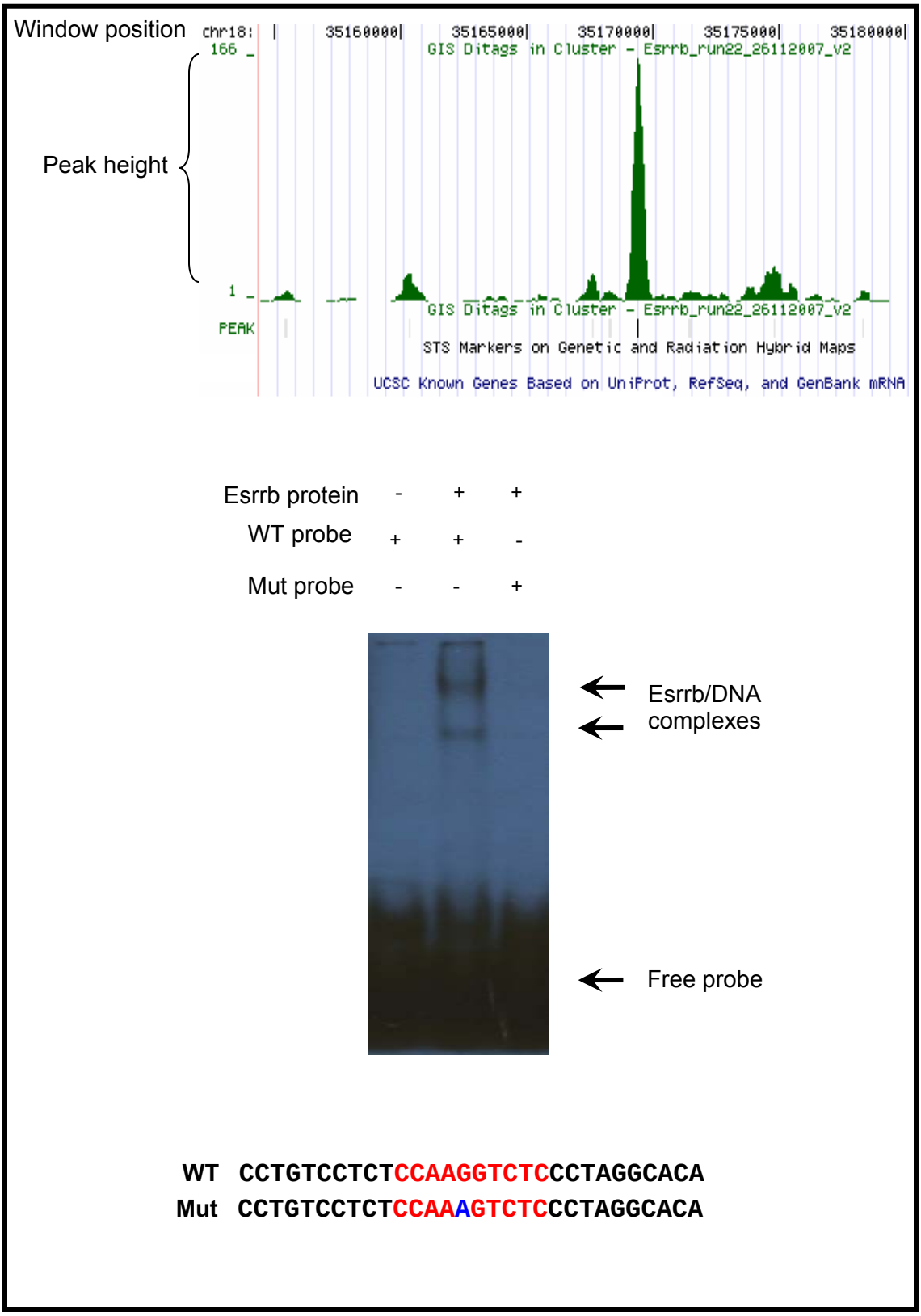

Window position
chr18:
35160000
35165000
35170000
35175000
35180000
166
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
Peak height
1
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
PEAK
STS Markers on Genetic and Radiation Hybrid Maps
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
Esrrb protein - + +
WT probe + + -
Mut probe - - +
Esrrb/DNA complexes
Free probe
WT CCTGTCCTCTCCAAGGTCTCCCTAGGCACA
Mut CCTGTCCTCTCCAAAGTCTCCCTAGGCACA

Figure 3.18 Esrrb can directly interact with double-stranded DNA sequences that contain the Esrrb binding motif. EMSA was performed to detect the interaction between purified Esrrb proteins and the respective DNA probes. Five regions (upper pannels in **A-E**) with high peak heights of Esrrb occupancy were chosen for probe design. The upper panels in **A-E** are the screen shots of the T2G browser showing the binding profiles of Esrrb at these regions. The middle panels in **A-E** show the Esrrb/DNA complex as detected by EMSA. All of these probes can form complexes with Esrrb proteins. Mutant probes on the other hand could not interact with Esrrb proteins. The sequences of the DNA probes and its respective mutants were listed in the bottom panels. The Esrrb binding motif was highlighted in red and the mutant nucleotide in blue.

3.4.3. Distribution of Esrrb binding and gene expression profiling

In order to characterize the targets of Esrrb, the binding clusters were mapped to annotated genes if they fall within 100kb upstream or downstream of a transcript unit (Figure 3.19A) (Wei *et al.,* 2006; Loh *et al.,* 2006). Binding sites within 10kb upstream or downstream of the nearest transcript units were defined as 5'proximal upstream and 3'proximal downstream of genes respectively. The remaining binding sites outside the proximal regions and within the 100kb range were defined as 5'distal upstream or 3'distal downstream of genes (Loh *et al.,* 2006). Binding sites which are located further than 100kb away from the nearest genes were defined as gene desert regions.

From the Esrrb ChIP-seq dataset, most of the Esrrb binding clusters (92.7%) could be annotated to known genes while only 1,285 binding sites (7.3%) were located in the gene desert regions (Figure 3.19B). Among the binding sites annotated to known genes, 6,371 Esrrb binding clusters (36.3%) were located in the introns of transcript units, whereas only 1,421 binding sites (8.1%) were found in exonic regions. 2,286 (13%) and 2,579 (14.7%) binding sites were located respectively in the 5' proximal and distal upstream

regions of genes. The 3' proximal and distal downstream regions contained 1,492 (8.5%) and 2,114 (12%) Esrrb binding clusters respectively. By using this criterion, a total of 8,092 genes were defined as Esrrb target genes.

To investigate the regulation of Esrrb on its target genes, I performed global gene expression profiling of the *Esrrb*-knockdown ES cells. RNA was harvested every two days in a 6-day experiment and was used to interrogate the Illunima microarray. Interestingly, 64% (5,196 genes) of the Esrrb target genes were differentially expressed upon *Esrrb* depletion in the 6-day interval (q value<0.05) (Figure 3.20). Among these differentially expressed Esrrb targets, various groups of genes involved in ES cell biology showed a reduction, no change or an induction in their transcription levels, which will be discussed in detail in the following sections. It was observed that Esrrb regulates its target genes in five regulation patterns (Figure 3.21). In regulation pattern 1, genes were down-regulated on the 2nd day after *Esrrb* depletion, and this down-regulation persisted throughout the 6 days of the experiment (such as *Kdelr3*). Esrrb may play a key role in activating these genes. For regulation pattern 2, genes were down-regulated on the 2nd day after *Esrrb* depletion but this reduction was reversed in the following days (such as *Edn1*). The down-regulation on the 2nd day reflects a direct activating effect of Esrrb binding on this group of genes. The further reduction may be due to a redundancy effect from other activators of these genes. For regulation pattern 3, genes were induced throughout the 6-day experiment (such as *E430036104Rik*). These genes represent the repressed gene targets of Esrrb. For regulation pattern 4, genes were induced on the 2nd day of *Esrrb* depletion, but this induction was reversed and the gene expression was

reduced in the following days (such as *Hoxb13*). Regulatory pattern 2 and 4 may be the manifestation of indirect effects from *Esrrb* depletion over the 6-day experiments. Alternatively, other factors that also bind to these genes could modulate and affect the regulatory effects of Esrrb on its targets. For regulation pattern 5, *Esrrb* depletion did not result in expression changes (such as *4930451A13Rik*). This could be explained by redundancy effects or the non-functional binding of Esrrb on these genes.

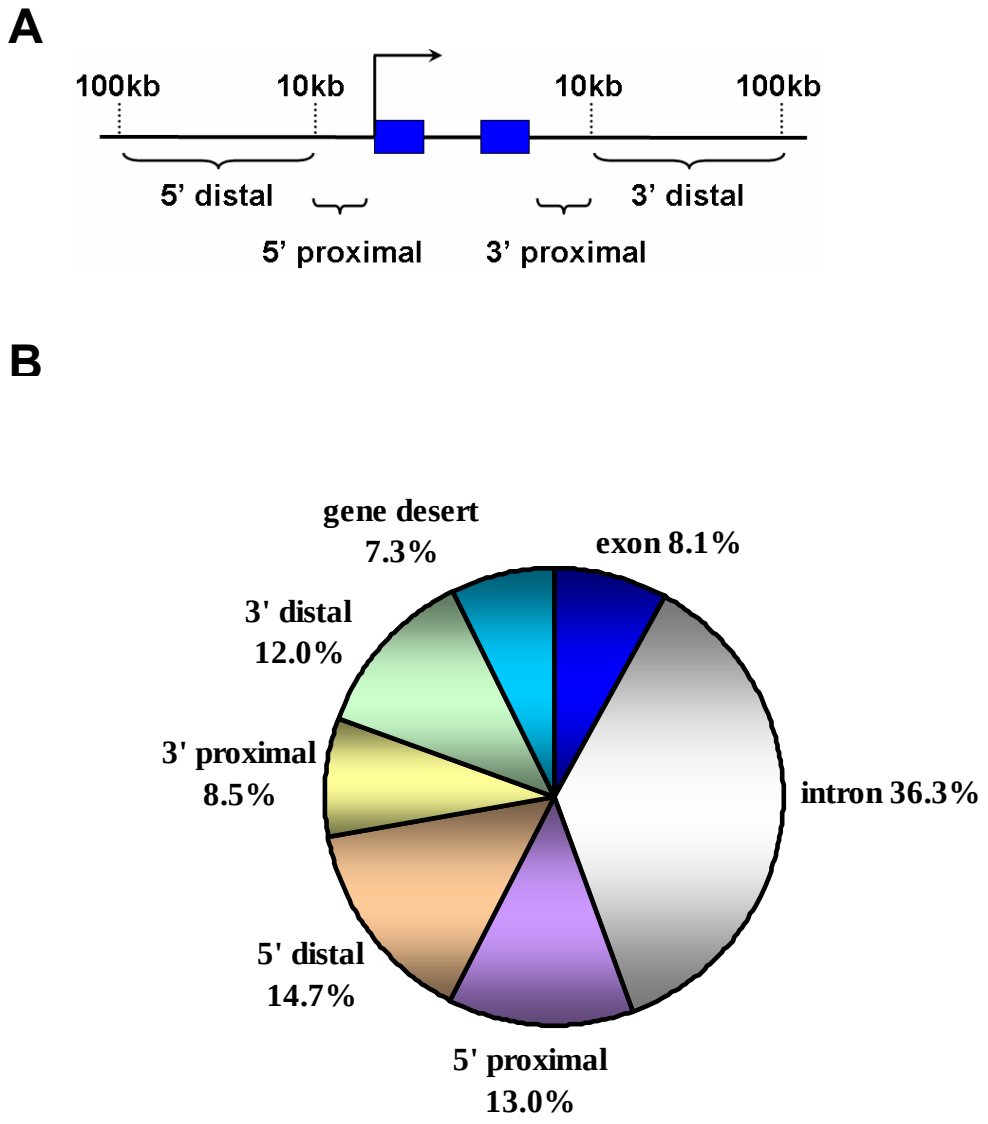


Figure 3.19 Distribution of Esrrb binding clusters was defined by their locations relative to a gene structure. **A.** Schematic diagram shows the locus annotation based on the relative positions in a transcription unit. Exons are depicted as blue boxes. The arrow indicates the direction of gene transcription. The loci within 10kb upstream or downstream of the nearest transcript units were defined as 5'proximal upstream or 3'proximal downstream of genes respectively. The loci within the range of 10-100kb upstream or downstream of the nearest transcript units were respectively defined as 5'distal upstream or 3'distal downstream of genes. The loci out of 100kb away from the nearest genes were defined as gene desert regions. **B.** The distribution of Esrrb binding clusters relative to the nearest transcript units. The ratios of the binding sites in respective locations to the total binding sites were shown.

Figure 3.20 Venn diagram showing the overlap between the Esrrb target genes and the differentially expressed genes during the 6-day interval after *Esrrb* knockdown (q value<0.05).

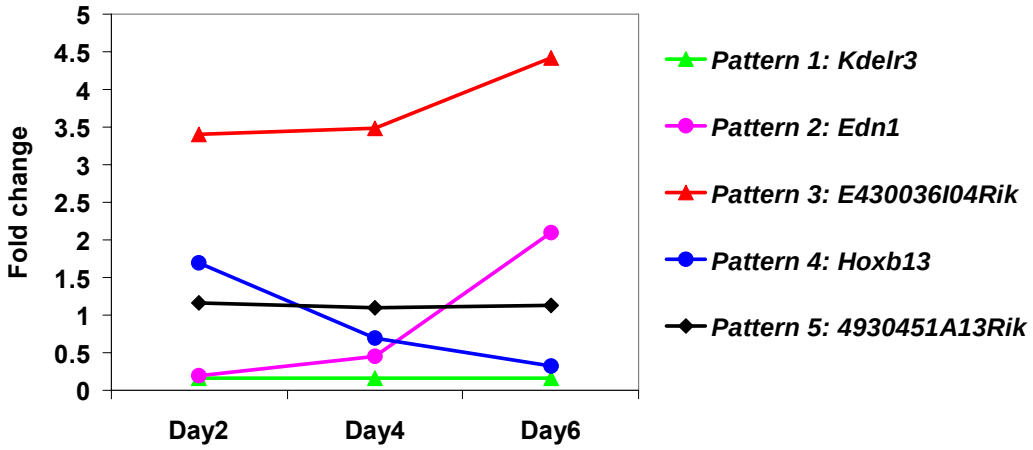


Figure 3.21 Different regulation patterns of Esrrb on its target genes. The x-axis represents the 6-day interval of *Esrrb* depletion. The y-axis represents the fold changes in gene expression levels which are computed as the ratio of the mRNA level in *Esrrb*-depleted ES cells to the *Gfp* shRNA-transfected control cells. Five genes were presented as an example of each pattern. For regulation pattern 1, genes were down-regulated on the 2nd day after *Esrrb* knockdown and this reduction in expression level lasted throughout the 6-day interval (such as *Kdelr3*). For regulation pattern 2, genes were down-regulated on the 2nd day after *Esrrb* knockdown but this reduction was reversed in the following days (such as *Edn1*). For regulation pattern 3, genes were induced throughout the 6-day interval (such as *E430036104Rik*). For regulation pattern 4, genes were induced on the 2nd day of *Esrrb* knockdown. Unlike regulation pattern 3, this induction was reversed and the gene expression was reduced in the following days (such as *Hoxb13*). For regulation pattern 5, *Esrrb* knockdown did not result in any expression change of the genes (such as *4930451A13Rik*).

3.4.4. Functional relevance of the target genes

To further examine the functional relevance of the Esrrb target genes in ES cell biology, I performed gene ontology (GO) analysis (http://www.pantherdb.org./) and found that the target genes of Esrrb are involved in various functional processes (Table 3.2). In particular, I focused on the regulatory effects of Esrrb on the genes which were previously identified to be involved in ES cell biology.

Table 3.2 Gene Ontology (GO) analysis was performed for functional annotation of Esrrb target genes (p value<0.01).

		P-value
Biological Process	Developmental processes	7.42E-07
	mRNA transcription regulation	7.71E-07
	mRNA transcription	1.19E-06
	Protein phosphorylation	1.37E-05
	Nucleoside, nucleotide and nucleic acid metabolism	3.04E-05
	Intracellular signaling cascade	1.34E-04
	Protein modification	1.53E-04
	Cell structure and motility	6.70E-04
	Mesoderm development	3.35E-03
	Stress response	4.96E-03
	Oxidative phosphorylation	5.84E-03
	Neurogenesis	9.49E-03
Molecular Function	Kinase	2.94E-06
	Transcription factor	2.66E-05
	Protein kinase	7.87E-04

3.4.4.1. Esrrb binds to ES cell-associated genes

The down-regulation of *Esrrb* using RNAi resulted in the differentiation of ES cells. I therefore speculated that one of the key mechanisms underlying the important role of Esrrb is its regulation of ES cell-associated genes. With reference to previous studies, I selected genes which play critical roles in ES cell maintenance or are preferentially expressed in pluripotent ES cells (Table 3.3) (Ramalho-Santos *et al*., 2002; Ivanova *et al*., 2002; Mitsui *et al*., 2003; Loh *et al.,* 2007). Interestingly, 20 out of the 25 selected genes are physically bound by Esrrb (Table 3.3). Only 5 genes, which include *Ecat8*, *Dnmt3I*, *Nrob1*, *Pem* and *Pbmxrt,* do not show Esrrb binding (Table 3.3). Most of the ES cell-associated genes bound by Esrrb have multiple binding sites, while *Hesx1*, *Gdf3*, *Eras*, and *Stat3* only have a single Esrrb binding site (Table 3.3). I randomly picked three genes, *Esrrb*, *Oct4* and *Sall4*, to validate Esrrb binding by using the ChIP-qPCR method. A sharp peak with peak height of over 100 was observed in the intronic region of the *Esrrb* gene. Importantly, the binding can be validated by ChIP-qPCR. This suggests that Esrrb regulates the expression of its own encoding gene through an auto-regulatory loop (Figure 3.22). *Sall4* is bound by Esrrb at multiple sites including the promoter and intronic regions (Figure 3.23). Two regions with the highest peak height were picked for validation using the ChIP-qPCR method. Interestingly, Sall4 has recently been reported to regulate the self-renewal of ES cells through the activation of *Oct4* gene or as an interaction partner of Nanog (Zhang *et al.,* 2006; Wu *et al.,* 2006). In addition, four Esrrb binding clusters were detected in the promoter and enhancer regions of the *Oct4* locus, all of which were validated by the ChIP-qPCR method (Figure 3.24).

Since the binding of a transcription factor on genes does not necessarily imply transcriptional regulation, I examined the expression profiles of these ES cell-associated genes in the *Esrrb*-depleted ES cells which have undergone shRNA treatments for 2, 4 or 6 days. Most of the genes were significantly downregulated in the *Esrrb*-depleted ES cells, suggesting that Esrrb plays a positive regulatory role on their expression (Figure 3.25). However, there are also ES cell-associated genes that were not affected in expression when *Esrrb* was depleted. Therefore, the binding of Esrrb to its gene targets could have either positive, negative or no regulatory roles.

Table 3.3 Summary of Esrrb binding to ES cell-associated genes.

Genes	Location of Esrrb binding sites	Number of Esrrb binding sites
2410004A20Rik/ecat1	5' proximal	1
	3' proximal	1
2410039E07Rik/ecat6	5' proximal	3
	5' distal	1
Eras	in intron	1
Esrrb	in intron	11
Fbxo15/ecat3	5' proximal	2
	3' proximal	1
Gdf3	in exon	1
Hesx1	3' proximal	1
Nanog	5' proximal	2
Pou5f1	5' proximal	3
	5' distal	1
Sall4	in intron	1
	5' proximal	2
	5' distal	2
Sox2	5' proximal	1
	5' distal	1
	3' proximal	1
	3' distal	3
Stat3	5' proximal	1
Tbx3	in intron	1
	5' distal	8
	3' distal	1
Tcfcp2l1	in intron	4
	3' distal	2
Tcl1	5' proximal	1
	5' distal	1
Tdgf1/cripto	5' proximal	2
Utf1	in exon	1
	3' proximal	2
Zfp296	5' proximal	2
Zfp42	5' proximal	1
	5' distal	3
Zfp57	5' proximal	2
2410004F06Rik/ecat8	—	—
Dnmt3l	—	—
Nrob1	—	—
Pem	—	—
Rbmxrt	—	—

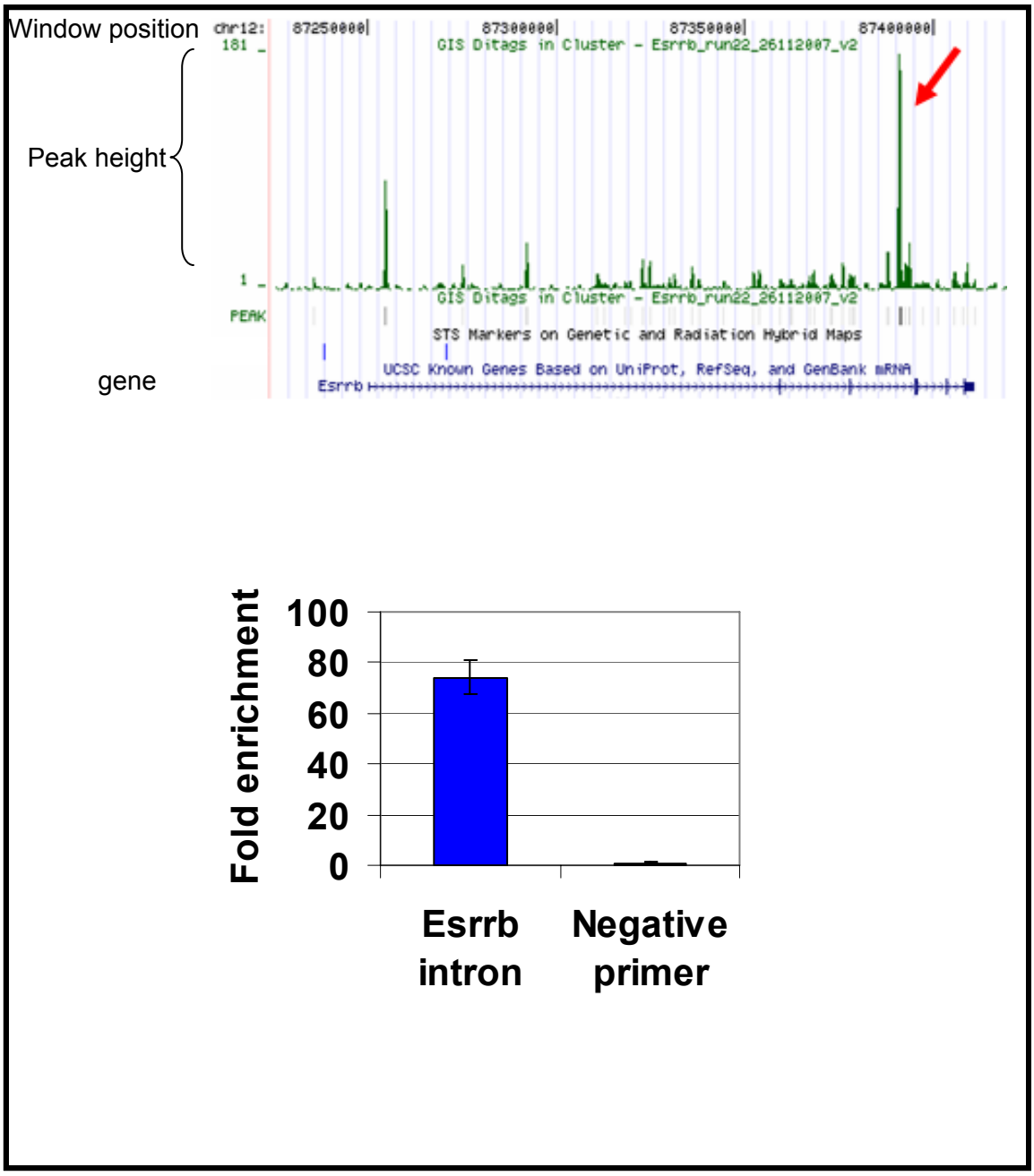


Figure 3.22 Esrrb binds to its encoding gene in ES cells. The upper panel shows the screen shot of the T2G browser presenting the binding profiles of Esrrb on the intron regions of its encoding gene. The red arrow shows the region picked for validation by ChIP-qPCR which was shown in the lower panel. A primer pair for this region without Esrrb binding was chosen as a negative control.

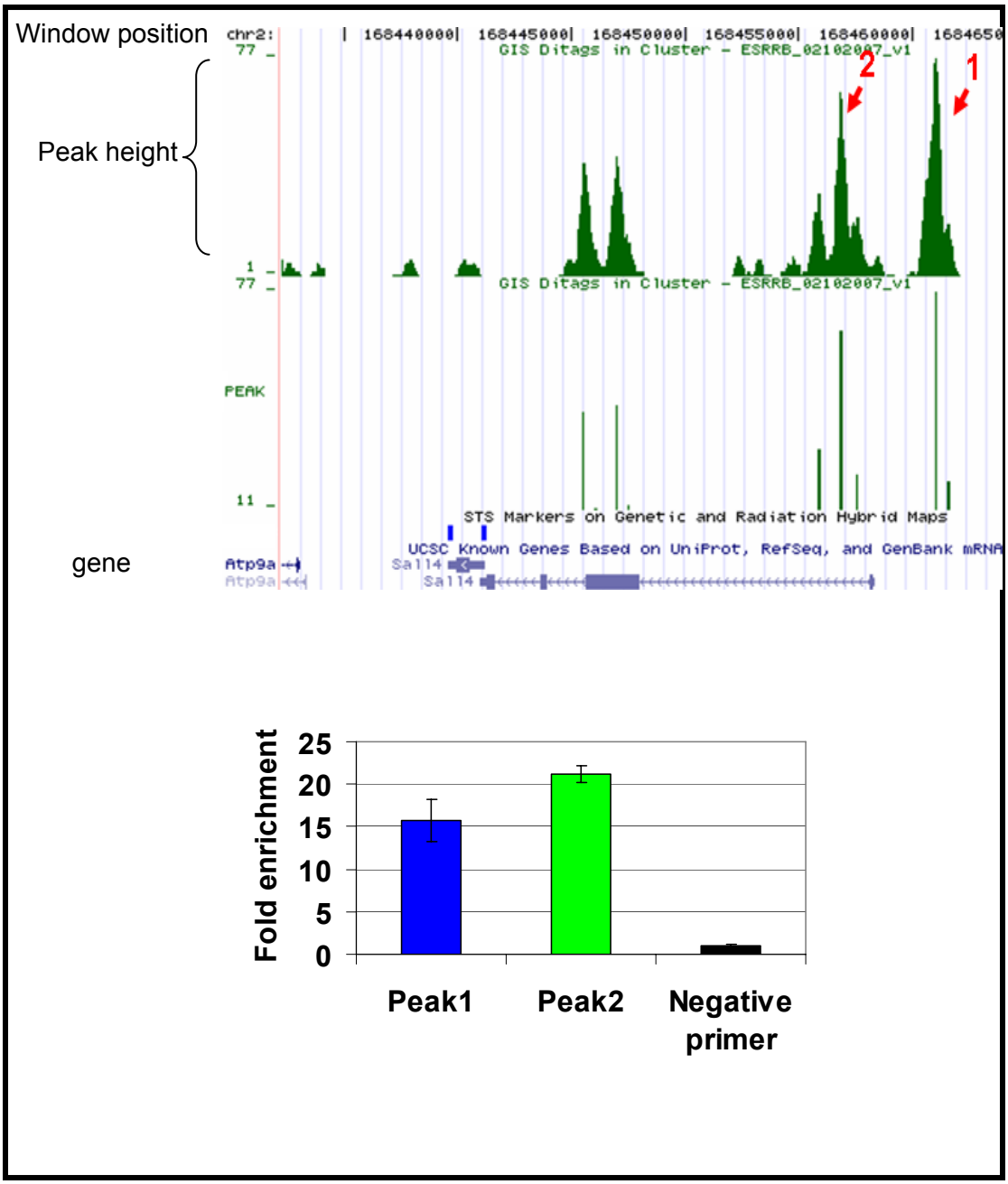


Figure 3.23 Esrrb binds to the promoter and intronic regions of the *Sall4* gene in ES cells. The upper panel shows the screen shot of the T2G browser presenting the binding profiles of Esrrb on the promoter region of *Sall4* encoding gene. The red arrows (1-2) show the regions picked for validation by ChIP-qPCR which are shown in the lower panel. A primer pair for this region without Esrrb binding was chosen as a negative control.

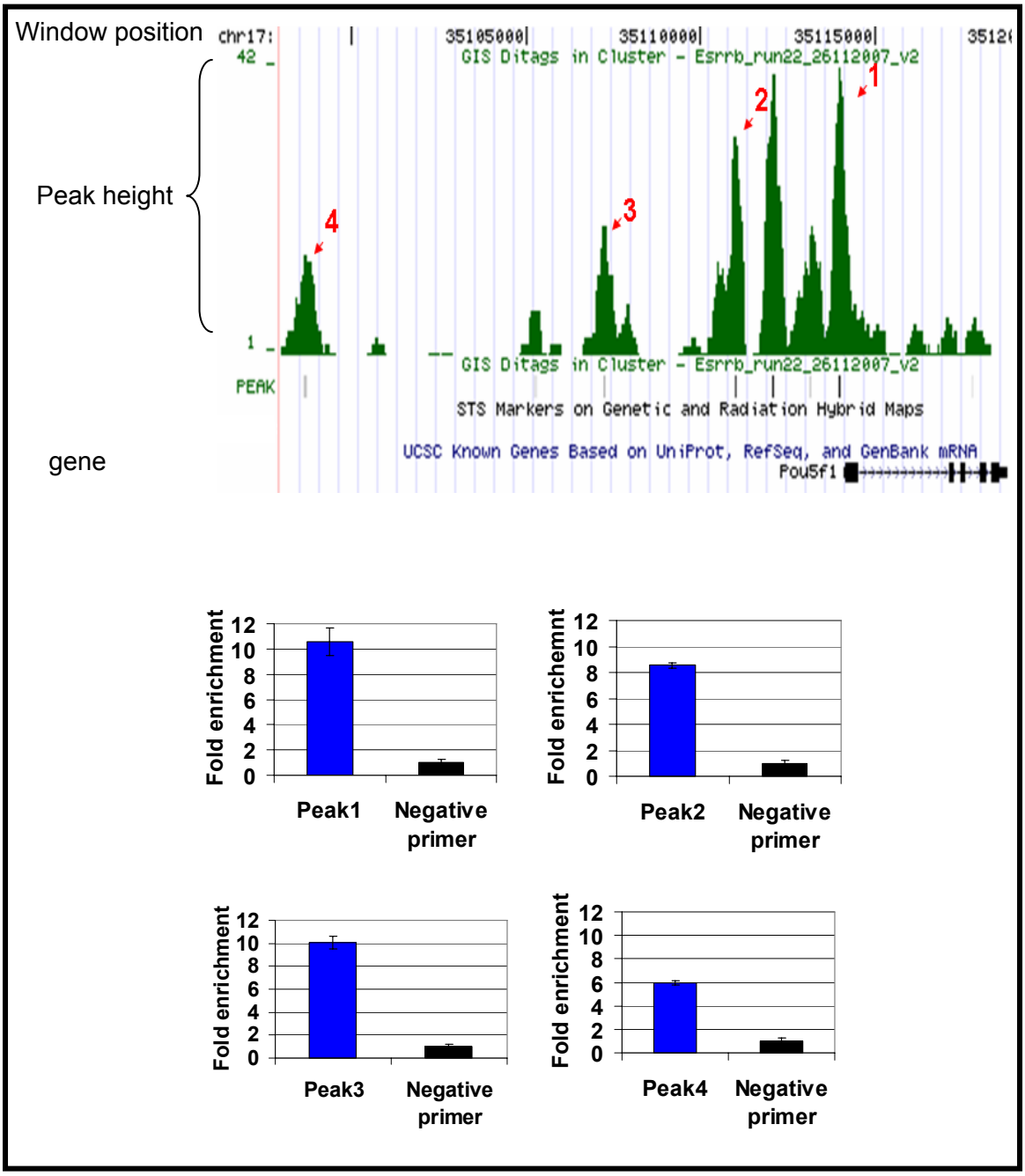


Figure 3.24 Esrrb binds to *Oct4* gene in ES cells. The upper panel shows the screen shot of the T2G browser representing the binding profiles of Esrrb on the promoter region of *Oct4* encoding gene. The red arrows (1-4) show the regions picked for validation by ChIP-qPCR which are shown in the lower panel. A primer pair for this region without Esrrb binding was chosen as a negative control.

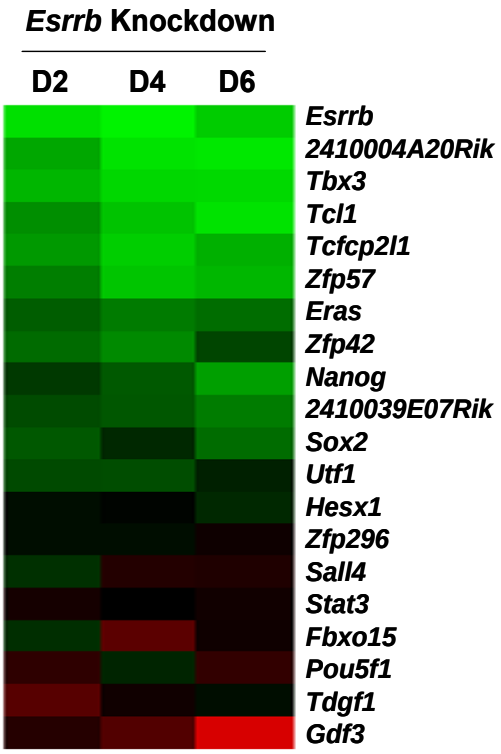


Figure 3.25 The regulation of Esrrb on the ES cell-associated genes. Heatmap shows the expression profile of the ES cell-associated genes which are bound by Esrrb. The genes are presented based on the fold changes of their expression levels after different days of *Esrrb* depletion (D2-6) compared with the *Gfp* shRNA-transfected control cells. They were sorted by the average expression ratio and mean centered.

Notably, among the ES cell-associated genes, *Oct4*, *Sox2* and *Nanog* have been shown to be reprogramming factor encoding genes (Takahashi and Yamanaka, 2006; Maherali *et al.,* 2007; Okita *et al.,* 2007; Wernig *et al.,* 2007; Yu *et al.,* 2007). I am therefore particularly interested in examining the binding of Esrrb to the genes that encode other reprogramming factors. In addition to *Oct4*, *Sox2* and *Nanog*, reprogramming factor encoding genes *Klf2, Klf4, Klf5*, *Myc* and *Mycn* were also identified to be bound by Esrrb (Table 3.4) (Maherali *et al.,* 2007; Okita *et al.,* 2007; Wernig *et al.,* 2007; Yu *et al.,* 2007; Nakagawa *et al.,* 2007). I further checked the regulatory effect of Esrrb on their

expression (Figure 3.26). The microarray results showed that the expression of *Klf4* and *Klf5* were sensitive to the Esrrb level. In the *Esrrb*-depleted ES cells, these two genes were significantly and persistently downregulated during the entire six-day observational period. However, the expression level of *Klf2* was unaltered upon the *Esrrb* knockdown, indicating the non-functional binding of Esrrb on this gene. For *Myc*, its expression was reduced by about 30% on the 2nd and 4th day of *Esrrb* depletion. The restoration of Myc expression levels on the 6th day indicates a possible compensation of activation by other regulators. Unlike Myc, I did not detect a differential expression of *Mycn* after the *Esrrb* depletion (Data not shown).

In summary, Esrrb binds to many ES cell-associated genes and reprogramming factors. Since the majority of these bound genes are positively regulated by Esrrb, I speculated that the key mechanism by which Esrrb maintains the ES cell state is by sustaining the high expression levels of the ES cell-associated genes.

Table 3.4 Summary of Esrrb binding to reprogramming factor encoding genes.

Genes	Location of Esrrb binding sites	Number of Esrrb binding sites
Klf2	5' distal	3
	3' proximal	1
Klf4	3' proximal	1
	3' distal	3
Klf5	3' distal	3
Myc	3' distal	1
Mycn	in intron	1
	3' proximal	1
	3' distal	3

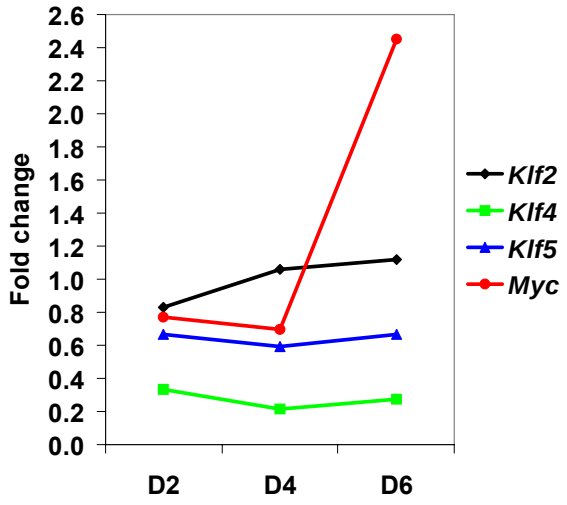


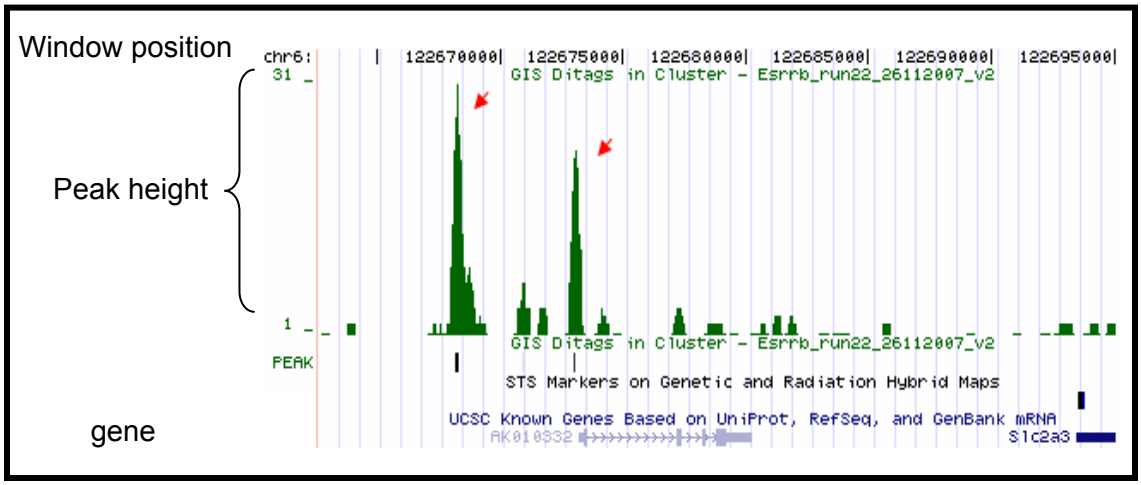
Figure 3.26 Expression profiles of reprogramming factors after *Esrrb* knockdown. The x axis represents different days of *Esrrb* knockdown and the y axis represents the gene expression fold change after different days of *Esrrb* knockdown (D2-6) compared to the knockdown contol.

3.4.4.2. Regulatory relationship between Esrrb and Nanog

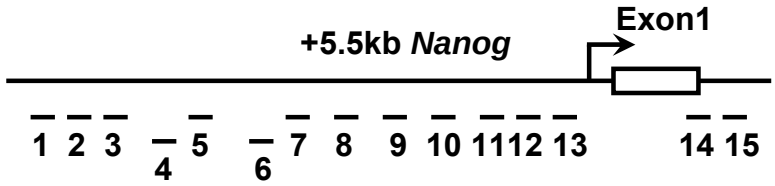
The ChIP-sequencing mapping results revealed that Esrrb occupies two binding sites, located at the distal enhancer and proximal promoter regions of the *Nanog* gene (Figure 3.27A). This observation is particularly interesting because *Esrrb* is also a target of Nanog. Moreover, the depletion of either *Esrrb* or *Nanog* results in the down-regulation of both corresponding genes (Figure 3.5D and 3.7C). I hypothesized that *Nanog* may be a key downstream effector of Esrrb's function.

First of all, I designed a series of primers along the 5.5 kb region upstream of the *Nanog* gene to validate the ChIP-sequencing results by the ChIP-qPCR method (Figure 3.27B and C). The depletion of *Esrrb* by shRNA and DES treatment led to the reduction of Esrrb binding at the *Nanog* loci, suggesting that the detected bindings are specific (Figure 3.27D and E). Since *Esrrb* knockdown resulted in a decrease in the *Nanog* expression level by 50%, this indicates that Esrrb exerts a positive regulatory effect on *Nanog* expression. To further confirm the activating effect of Esrrb on the *Nanog* gene, I cloned the 5.5kb promoter region of the *Nanog* gene into the luciferase reporter vector (Figure 3.28A). Co-transfection of this *Nanog* reporter construct with the *Esrrb* shRNA resulted in a strong reduction of the luciferase activity (Figure 3.28A). Conversely, when the *Esrrb* was overexpressed in either 293T or ES cells, the reporter activity driven by the *Nanog* promoter was induced significantly (Figure 3.28B and 3.28C). Importantly, overexpression of *Esrrb* in ES cells also induced the level of endogenous *Nanog* expression by at least 2 folds, while the expressions of *Oct4* or *Sox2* were not alerted by *Esrrb* overexpression (Figure 3.28D).

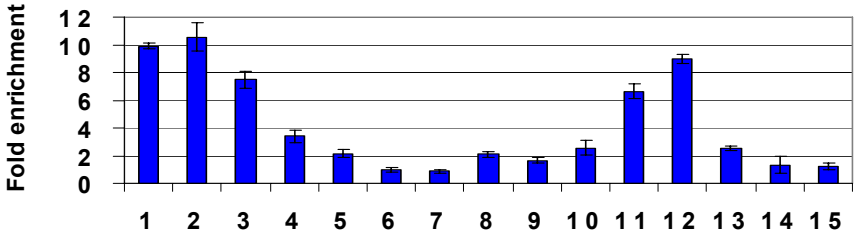
A


Window position
Peak height
gene
chr6:
122670000
122675000
122680000
122685000
122690000
122695000
GIS Ditags in Cluster - Esrrb_run22_26112007_v2
PEAK
STS Markers on Genetic and Radiation Hybrid Maps
UCSC Known Genes Based on UniProt, RefSeq, and GenBank mRNA
Slc2a3

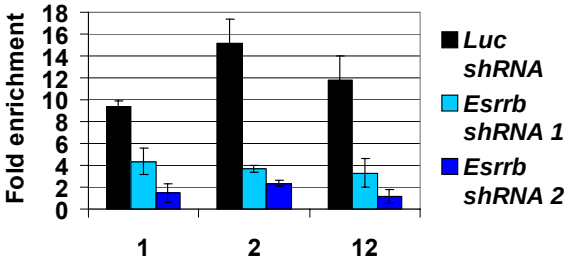
B


+5.5kb *Nanog*
Exon1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

C


Fold enrichment
0
2
4
6
8
10
12
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

D


Fold enrichment
0
2
4
6
8
10
12
14
16
18
1
2
12
Luc shRNA
Esrrb shRNA 1
Esrrb shRNA 2

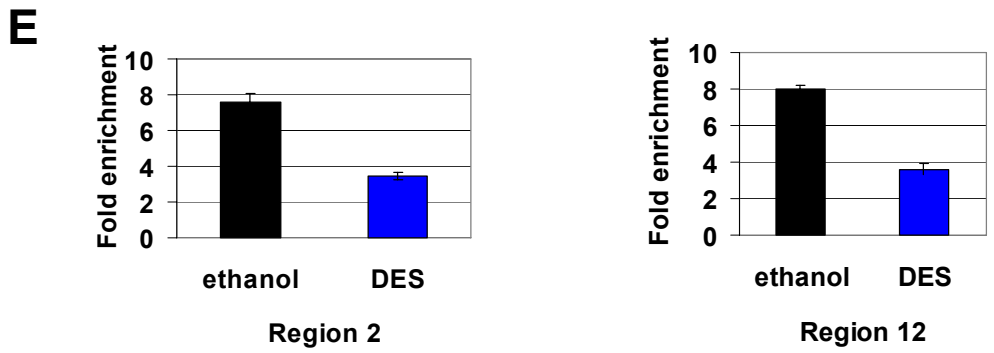


Figure 3.27 Esrrb binds to *Nanog* gene. **A**. The screen shot of the T2G browser showing the binding profile of Esrrb on *Nanog* (*AK010332*) locus. The two peaks highlighted by red arrows correspond to the *Nanog* enhancer and proximal promoter. **B**. Schematic diagram showing the locations of the amplicons (black bars labeled 1-15) that were used to detect ChIP-enriched fragments over the 5.5 kb *Nanog* promoter. Amplicons are numbered in order relative to their sites along the gene. The open box represents an exon. **C**. Real-time PCR validation of Esrrb binding to the *Nanog* gene. Fold enrichment is computed as the relative abundance of DNA fragments at the amplified region over a control amplified region. Data are presented as the mean ± SEM. **D**. Knockdown of *Esrrb* abolished the ChIP signal derived from anti-Esrrb antibody. Region 1, 2 and 12 were selected based on their high-fold enrichments of Esrrb binding. ES cells were transfected with control *Luc* shRNA, *Esrrb* shRNA 1 or *Esrrb* shRNA 2. Data are presented as the mean ± SEM. **E**. Esrrb bindings to the *Nanog* enhancer and proximal promoter regions were abolished upon DES treatment. Region 2 and 12 were selected for q-PCR. ChIP assay was performed with the samples of ES cells treated with 10uM DES or control 0.1% ethanol for two days. Data are presented as the mean ± SEM.

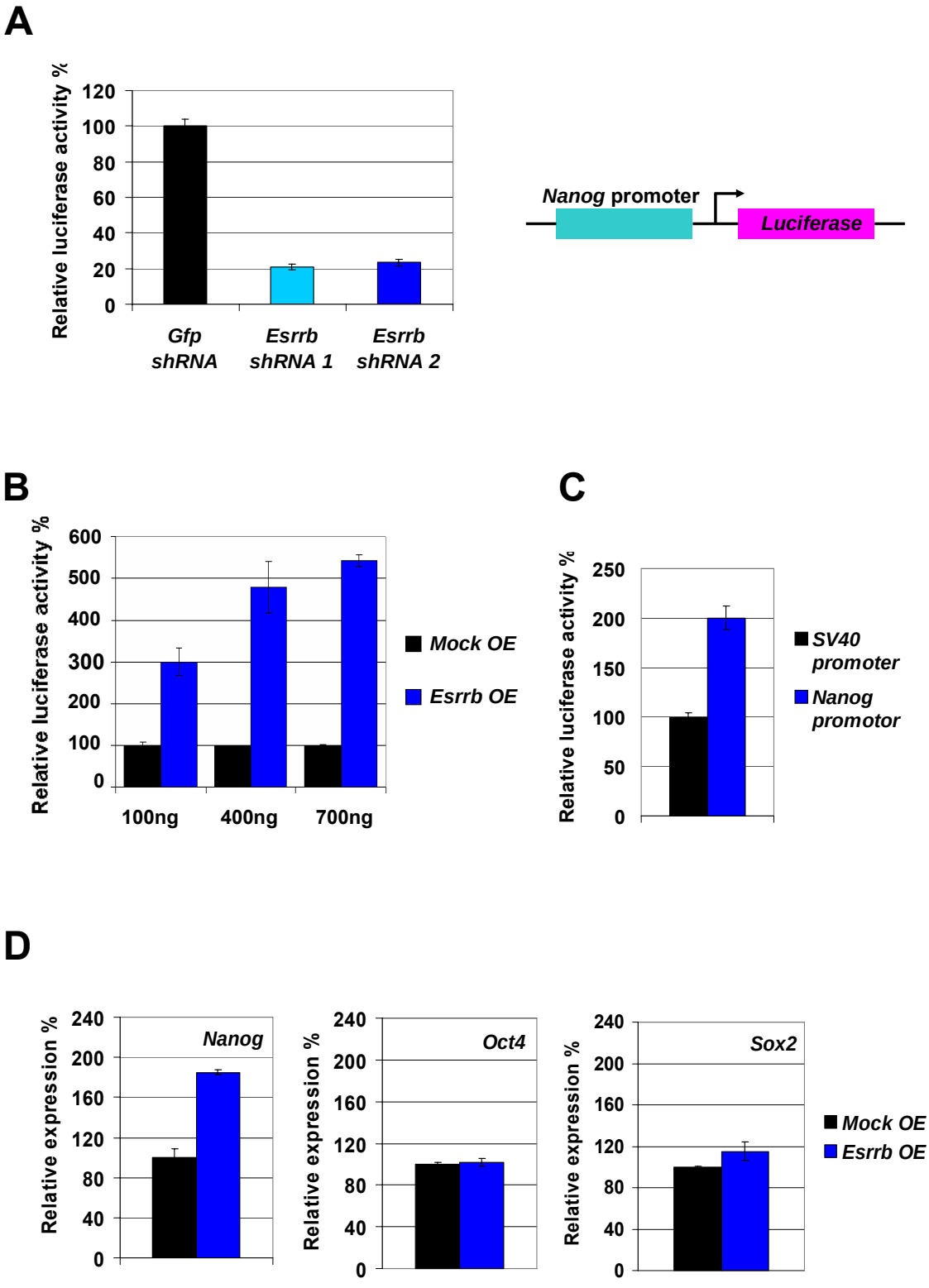

A
Relative luciferase activity %
120
100
80
60
40
20
0
Gfp shRNA
Esrrb shRNA 1
Esrrb shRNA 2
Nanog promoter
Luciferase
B
Relative luciferase activity %
600
500
400
300
200
100
0
100ng
400ng
700ng
Mock OE
Esrrb OE
C
Relative luciferase activity %
250
200
150
100
50
0
SV40 promoter
Nanog promotor
D
Relative expression %
240
200
160
120
80
40
0
Nanog
Oct4
Sox2
Mock OE
Esrrb OE

Figure 3.28 Esrrb activates *Nanog* expression. **A**. The right panel shows the schematic digram of the luciferase reporter driven by the 5.5kb *Nanog* promoter. The effect of *Esrrb* RNAi on the *Nanog* promoter activity was tested by co-transfecting each *Esrrb* shRNA construct along with the *Nanog* promoter-*Luc* construct into ES cells. Luciferase activity was analyzed 2 days after transfection and is presented relative to the *Gfp* shRNA knockdown control. Data are presented as the mean ± SEM. **B**. *Esrrb* overexpression (OE) induced the *Nanog* promoter activity in a dosage-dependent manner in 293T cells. Data are presented as the mean ± SEM. **C**. *Esrrb* OE induced the *Nanog* promoter activity in ES cells. The *SV40* promoter was used as a negative control. Data are presented as the mean ± SEM. **D**. In ES cells, *Esrrb* overexpression induced the expression of endogenous *Nanog* but not *Oct4* or *Sox2*. The vector without *Esrrb* insertion was used as a mock overexpression control. Data are presented as the mean ± SEM.

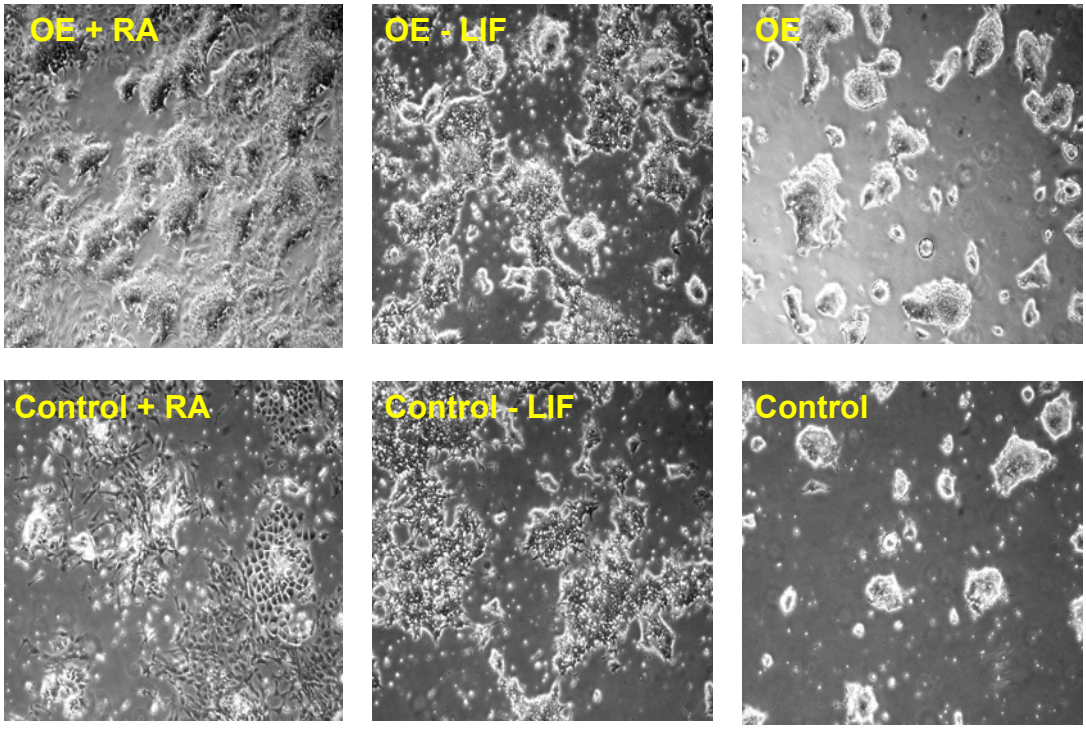
Next, to address the question of whether Nanog is the key downstream effector of Esrrb in ES cell maintenance, I examined if an elevation of *Nanog* expression could rescue the differentiation induced by *Esrrb* knockdown. Since the *Esrrb* shRNA constructs confer puromycin resistance, I established a hygromycin-resistant *Nanog* overexpression construct. I generated an ES cell line that was transfected with this *Nanog* overexpression construct and maintained the cells in culture with hygromycin selection. The hygromycin-selected *Nanog* overexpression cell line displayed a great similarity to the F4 cell line. Moreover, it has ability to resist differentiation induced by either RA-treatment or LIF-withdrawal (Figure 3.29A).

Interestingly, most of the cells maintained the colony morphology of ES cells when *Esrrb* was depleted in the *Nanog*-overexpressed cell line. In contrast, *Esrrb* knockdown in the mock overexpression cell line led to expected cell differentiation (Figure 3.29B). The *Esrrb* mRNA levels in both the cell lines were similar after *Esrrb* knockdown, thus the dissimilarity in cell morphology is not due to differences in knockdown efficiency

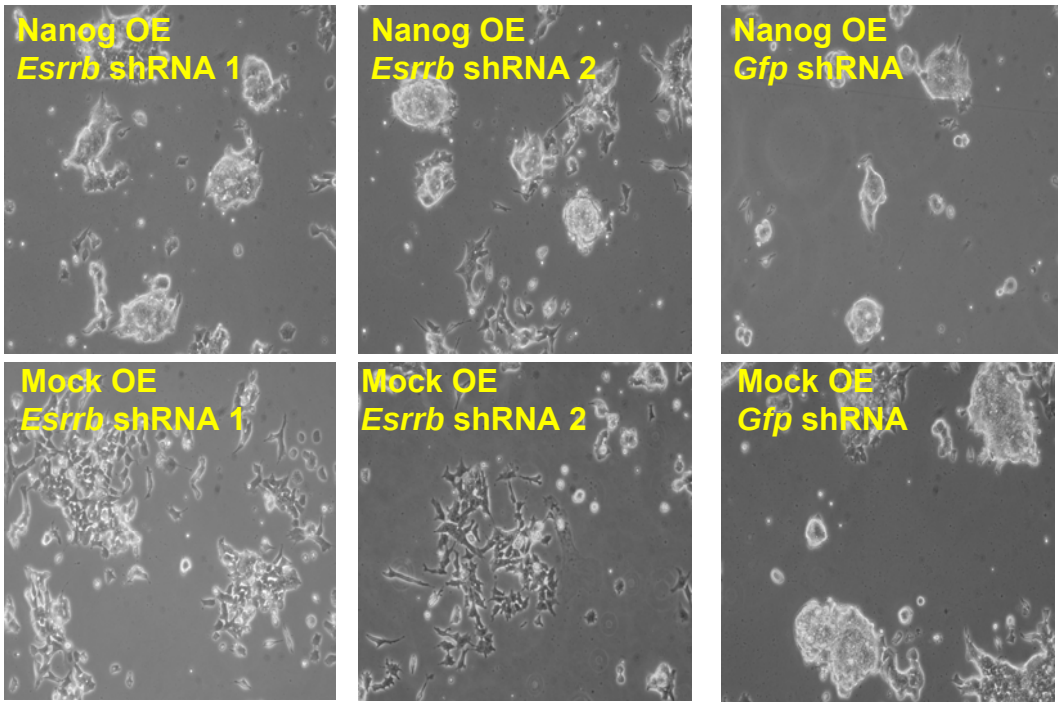
(Figure 3.29C). At the molecular level, overexpression of *Nanog* effectively rescued the marker changes induced by the *Esrrb* knockdown (Figure 3.29D). The reduction in *Rex1* and *Sox17* transcript levels were restored, while the induction of *Fgf5* was suppressed. These results demonstrated that *Nanog* overexpression can rescue ES cell differentiation induced by *Esrrb* depletion. Through comparing the differentially regulated genes by Esrrb and Nanog through using the databases reported by a previous study, 436 genes were identified to be regulated by both of these two factors, supporting the overlapped roles in their functional pathways in ES cells (Appendix 2) (Ivanova et al., 2006)

In summary, *Nanog* is the downstream target of Esrrb and its expression could be directly activated by Esrrb. In addition, overexpression of *Nanog* could partially rescue the differentiation induced by *Esrrb* depletion. Therefore, I concluded that the regulation of *Nanog* is a key function of Esrrb in ES cell maintenance.

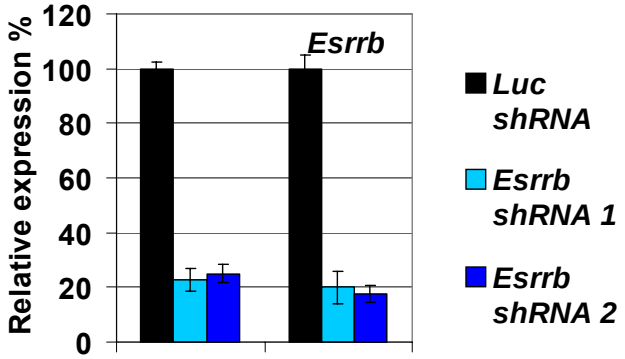
A


OE + RA
OE - LIF
OE
Control + RA
Control - LIF
Control

B


Nanog OE
Esrrb shRNA 1
Nanog OE
Esrrb shRNA 2
Nanog OE
Gfp shRNA
Mock OE
Esrrb shRNA 1
Mock OE
Esrrb shRNA 2
Mock OE
Gfp shRNA

C


Relative expression %
120
100
80
60
40
20
0
Esrrb
Luc shRNA
Esrrb shRNA 1
Esrrb shRNA 2

D

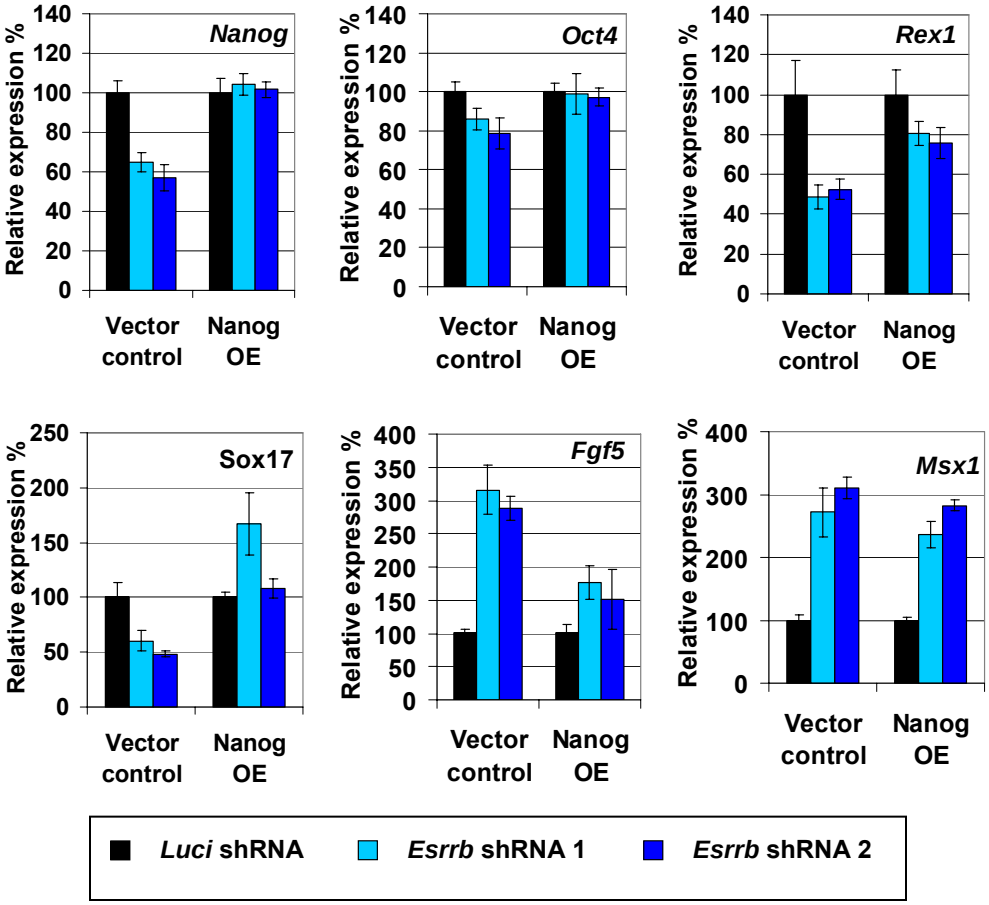

Nanog
Oct4
Rex1
Sox17
Fgf5
Msx1
Relative expression %
Vector control
Nanog OE
Luci shRNA
Esrrb shRNA 1
Esrrb shRNA 2

Figure 3.29 Overexpression of *Nanog* can rescue the differentiation induced by *Esrrb* knockdown. **A**. Characterization of the hygromycin resistance *Nanog* overexpression cell line. *Nanog*-overexpressed ES cells (OE) and control ES cells were grown in the presence 0.3 μM RA (+RA) or no RA (-RA) for 48 hours, or in the presence of LIF (+LIF) or absence of LIF (-LIF) for 60 hours. **B**. ES cells with constitutive *Nanog* overexpression were challenged with shRNA-directed *Esrrb* knockdown for 3 days. **C**. *Esrrb* was similarly depleted both in *Nanog*-overexpressed (Nanog OE) and vector control ES cells. Data are presented as the mean ± SEM. **D**. *Nanog* overexpression reduced the down-regulation of *Nanog, Oct4, Rex1* and *Sox17* upon the *Esrrb* depletion. On the other hand, it compensated for the loss of Esrrb function by reducing the induction of differentiation markers *Fgf5*, but not *Msx1*. The levels of the transcripts were normalized against control plasmid transfected cells. Data are presented as the mean ± SEM.

3.4.4.3. Binding of Esrrb to the genes encoding for developmental regulators

It was observed that the *Nanog* rescue on *Esrrb* depletion was only partial, as some differentiated cells were still observed. Thus the regulation of ES cell self-renewal and pluripotency by Esrrb is also mediated by some other Nanog-independent mechanisms. I next asked if Esrrb is also involved in the regulation of genes implicated in differentiation or development. Several gene families which have been reported for their roles in regulating developmental processes were selected for further analysis. These families include Forkhead box (Fox) proteins, Homeobox containing protein family (Hox), Myogenic basic domain (Myo) proteins, NK transcription factor related proteins (Nkx), Paired Box and paired-like proteins (Pax), Pou transcription factor family (Pou), SRY-related HMG Box proteins (Sox), and T-box proteins (Tbx). Hox proteins function as transcriptional activators or repressors and are involved in numerous cellular processes including organogenesis, cellular differentiation, cell adhesion and migration, cell cycle and apoptosis (Svingen and Tonissen, 2006). Sox proteins are widely involved in developmental processes such as neuronal development, sex determination and

lymphocyte development (Kiefer, 2007). Pax proteins are tissue specific transcription factors containing a PAIRED domain and a homeodomain, which are important in early tissue specification of embryos (Lang *et al.*, 2007). They are also related with limb regeneration (Lang *et al.,* 2007). Unexpectedly, Esrrb binds to diverse classes of developmental genes mentioned above (Table 3.5). One notable example is the binding of Esrrb to the *Sox17* gene (Figure 3.30). The validation by ChIP-qPCR confirmed the interaction of Esrrb with the *Sox17* gene (Figure 3.30). *Sox17* is a well-established endoderm marker gene. Another endoderm gene *Gata6* was also bound by Esrrb as shown in the ChIP-sequencing and ChIP-qPCR results (Figure 3.31). These results imply the possible regulatory effect of Esrrb on the endoderm differentiation.

Table 3.5 Summary of Esrrb binding to developmental genes.

Hox	*Fox*	*Tcf*	*Sox*	*Myo*	*Tbx*	*Nkx*	*Pax*	*Pou*
Hoxa1	*Foxa1*	*Tcf12*	*Sox11*	*Myo10*	*Tbx1*	*Nkx1-2*	*Pax1*	*Pou2f2*
Hoxa10	*Foxa2*	*Tcf15*	*Sox12*	*Myo15*	*Tbx2*	*Nkx2-3*	*Pax2*	*Pou2f3*
Hoxa11	*Foxa3*	*Tcf19*	*Sox13*	*Myo18a*	*Tbx20*	*Nkx6-1*	*Pax6*	*Pou4f2*
Hoxa13	*Foxb2*	*Tcf2*	*Sox17*	*Myo1b*	*Tbx21*	*Nkx6-2*	*Pax7*	*Pou5f1*
Hoxa5	*Foxc2*	*Tcf21*	*Sox18*	*Myo1e*	*Tbx3*		*Pax8*	
Hoxa6	*Foxd1*	*Tcf3*	*Sox2*	*Myo1f*	*Tbx4*		*Pax9*	
Hoxb1	*Foxd2*	*Tcf4*	*Sox3*	*Myo1g*	*Tbx5*			
Hoxb13	*Foxd3*	*Tcf7*	*Sox4*	*Myo5b*				
Hoxb4	*Foxd4*	*Tcf7l2*	*Sox5*	*Myo6*				
Hoxb5	*Foxf1a*		*Sox6*	*Myo7a*				
Hoxb9	*Foxi1*		*Sox8*	*Myocd*				
Hoxc12	*Foxj1*		*Sox9*	*Myog*				
Hoxc13	*Foxj2*			*Myom2*				
Hoxc4	*Foxj3*			*Myot*				
Hoxd10	*Foxk1*			*Myoz2*				
Hoxd12	*Foxm1*							
	Foxn2							
	Foxo1							
	Foxo3a							
	Foxo6							
	Foxp1							
	Foxp4							
	Foxq1							
	Foxr1							

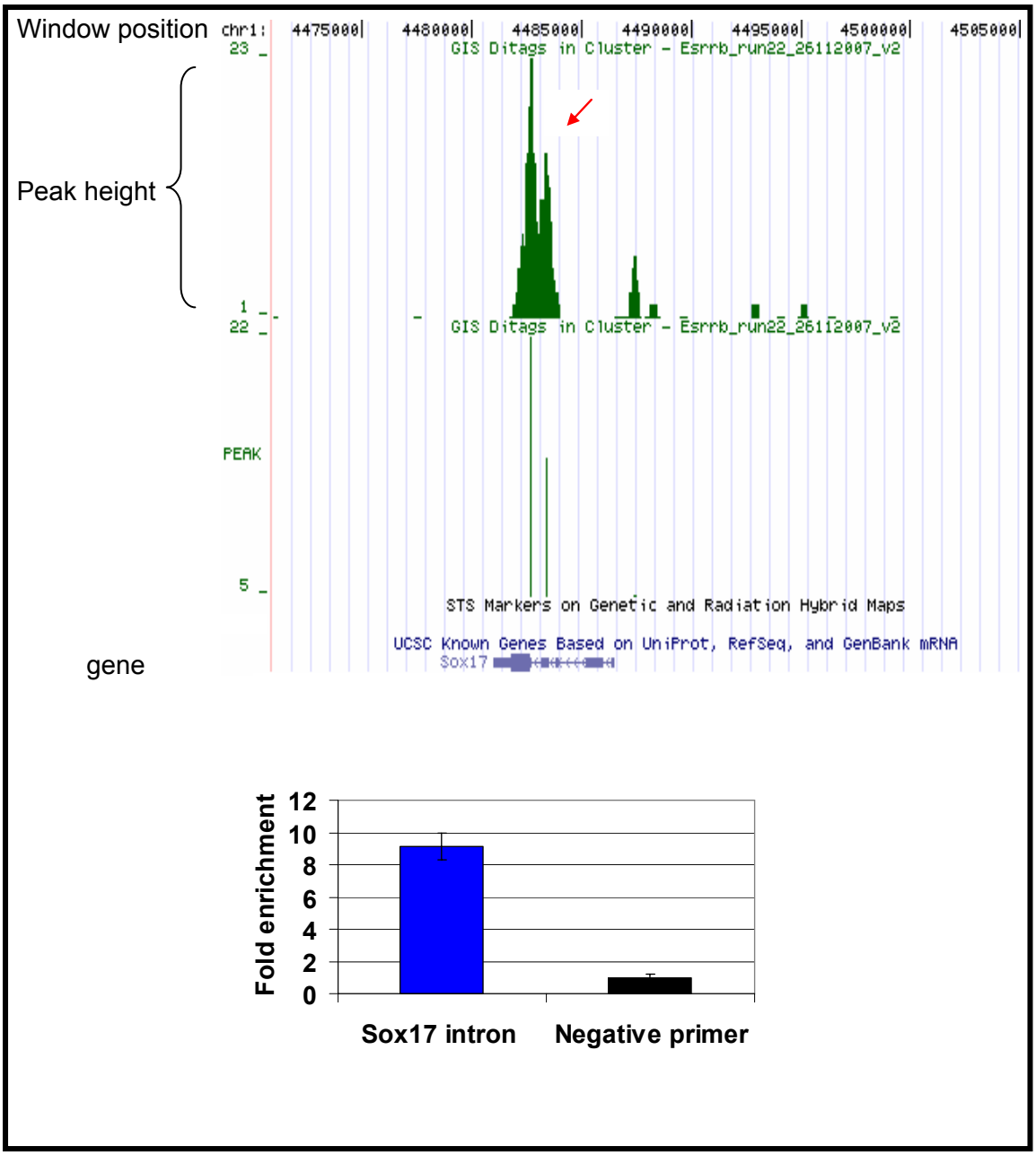


Figure 3.30 Esrrb binds to the intronic region of *Sox17* gene in ES cells. The upper panel is the screen shot of the T2G browser showing the binding profiles of Esrrb on the promoter region of *Sox17* gene. The red arrow indicates the region picked for validation by ChIP-qPCR which is shown in the lower panel. A primer pair for this region without Esrrb binding was chosen as a negative control.

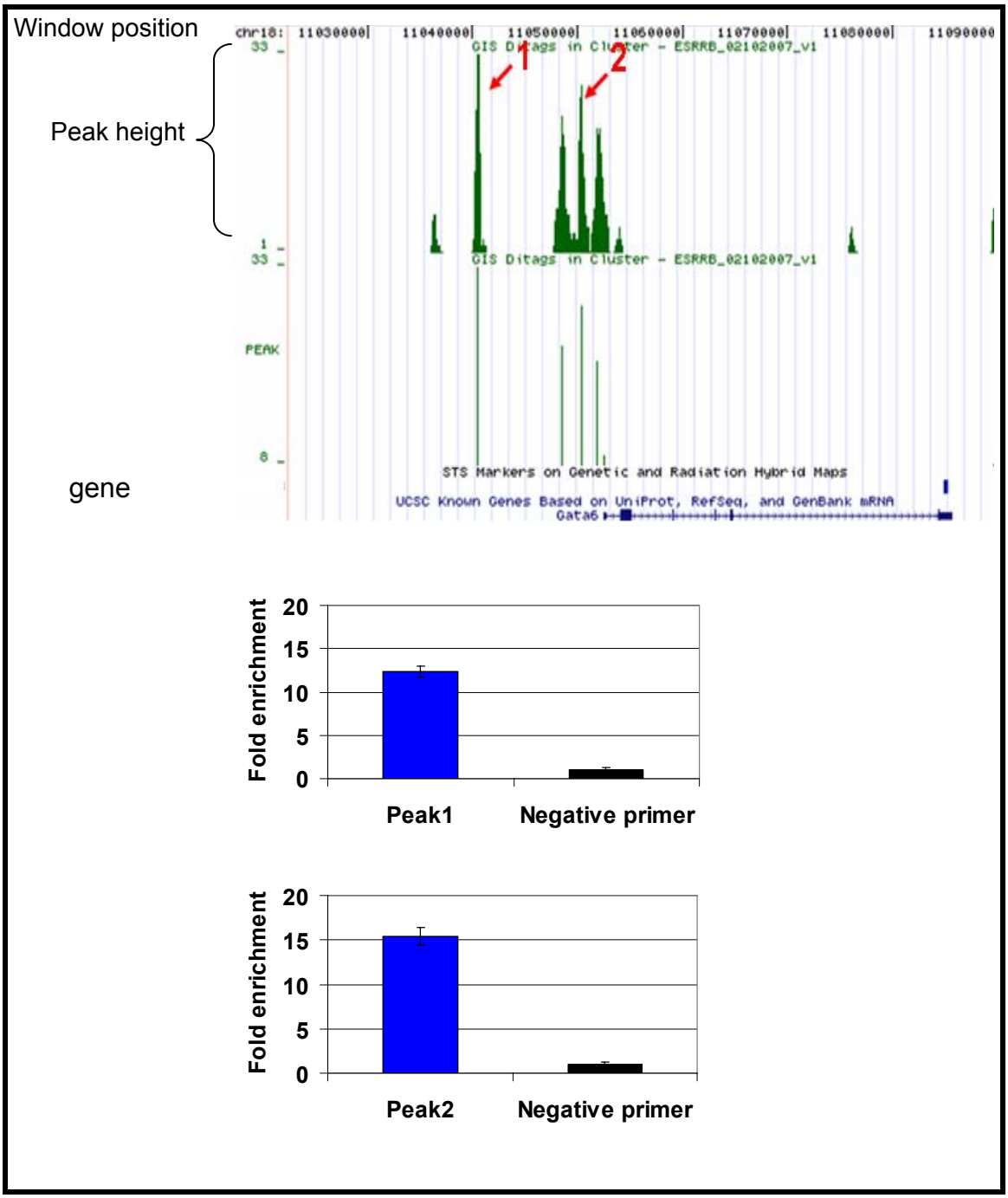


Figure 3.31 Esrrb binds to *Gata6* gene in ES cells. The upper panel is the screen shot of the T2G browser showing the binding profiles of Esrrb on the promoter region of *Gata6*. The red arrows indicate the peak sites used for validations. The binding of Esrrb to *Gata6* gene was validated by Esrrb ChIP-qPCR (lower panel). Two peak regions indicated by 1 and 2 were picked for ChIP-qPCR. A primer pair for this region without Esrrb binding was chosen as a negative control.

I further checked the regulatory effect of Esrrb on these developmental genes by expression profiling. I found that Esrrb exerts different forms of transcriptional regulations on these genes (Figure 3.32). For example, *Esrrb* knockdown led to the reduction in expression of *Hox* genes (*Hoxa7*, *Hoxb13, Hoxc13* and *Hoxc4*) and *Myo* genes (*Myo18a, Myo1f, Myo1g, Myo7a and Myom2*) while it resulted in an induction of *Fox* gene expression (*Foxa2, Foxd4 and Foxo1*) and *Sox* gene family (*Sox11, Sox13, Sox18, Sox4 and Sox9*). In some cases, members of the same gene family can be regulated differentially by Esrrb. One notable example is *Tbx* gene family where Esrrb activates the expression of *Tbx3* but represses *Tbx4* expression. On the other hand, for the *Tcf, Nkx* and *Pou* gene families, *Esrrb* depletion did not lead to any obvious changes in their expression, indicating the "non-functional binding" of Esrrb on these genes.

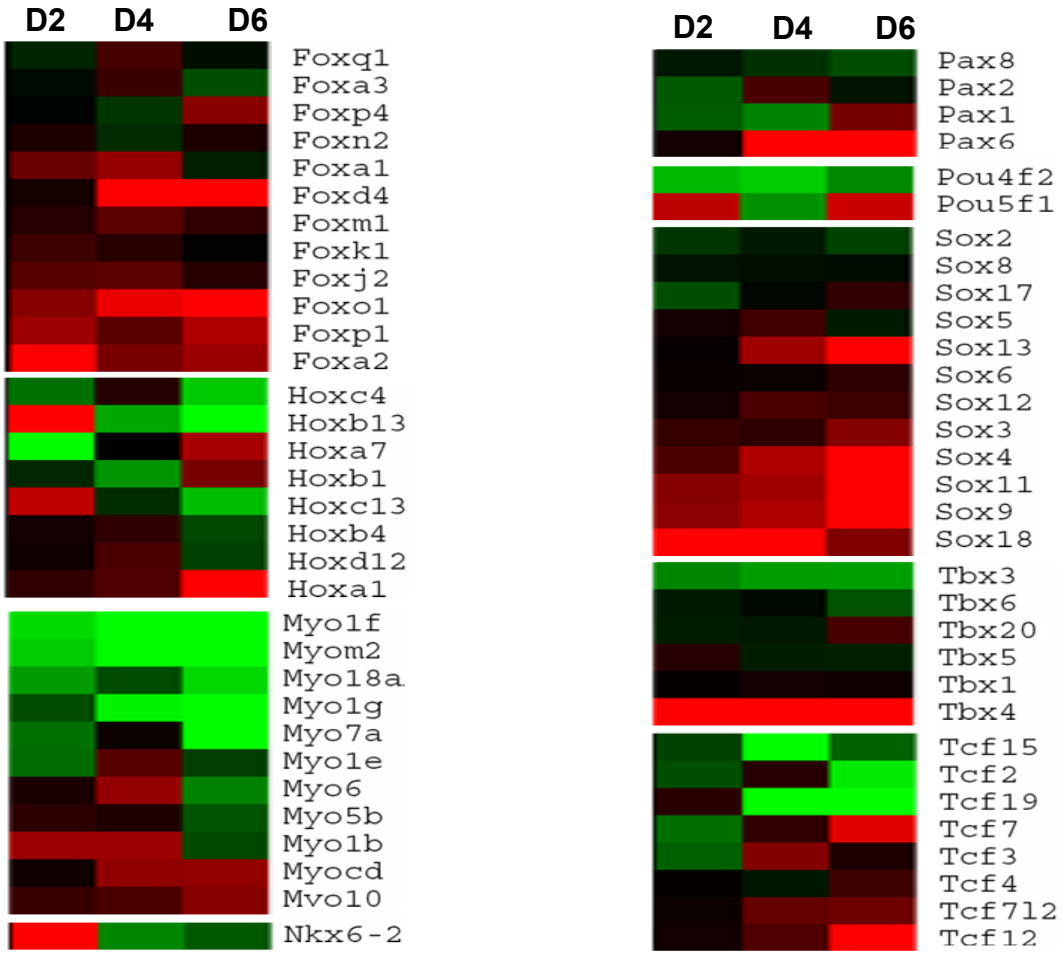


Figure 3.32 The regulation of Esrrb on developmental genes. Heatmaps showing the expression profile of the developmental genes which were bound by Esrrb. The genes were presented based on the fold changes of their expression levels after different days (D2-6) of *Esrrb* depletion compared with the *Gfp* shRNA-transfected cell control. They were sorted by the average expression ratio and mean centered.

Importantly, endoderm genes (*Sox17* and *Gata6*) (Figure 3.30 and 3.31), Mesendoderm markers (*Msx1*), ectoderm marker (*Fgf5*) and trophectoderm regulators (*Hand1* and *Cdx2*) were all occupied by Esrrb (Table 3.6). Thus Esrrb has an important regulatory role in controlling the expression of key lineage genes, thereby preventing differentiation and maintaining the pluripotent state of ES cells (Table 3.6; Figure 3.7D).

Table 3.6 Summary of Esrrb binding to lineage marker genes.

Genes	Location of Esrrb binding sites	Number of Esrrb binding sites
Msx1	5' distal	1
Fgf5	5' distal	1
Hand1	5' proximal	1
	5' distal	3
Cdx2	in intron	1
	3' proximal	1

3.4.4.4. Esrrb binds to the genes encoding for epigenetic modifiers

Due to the importance of epigenetic processes in ES cell maintenance (Montgomery *et al.,* 2005; Bernstein *et al.,* 2006; Azuara *et al.,* 2006; Lee *et al.,* 2006; Pasini *et al.,* 2007), I was interested to know whether Esrrb is involved in regulating the epigenetic processes via the control of expression of epigenetic regulators.

Firstly, I sought to identify the binding of Esrrb to epigenetic regulator encoding genes which have been shown to be critical for ES cell derivation from the embryo or the maintenance of the ES cell state (summarized by Surani et al, 2007). These genes include *Brg1* (Bultman et al., 2000; Bultman *et al.,* 2006), *CAF-1* (Houlard *et al.,* 2006), *Dicer1* (Bernstein *et al.,* 2003), *Dnmt3L* (Bourc'his *et al.,* 2001; Okano *et al.,* 2002), *Eed* (Shumacher *et al.,* 1996), *Ezh2* (O'Carroll *et al.,* 2001), *Hdac1* (Lagger *et al.,* 2002), *Gcn5* (Xu *et al.,* 2002), *Mll/All-1* (Yu *et al.,* 1995; Yagi *et al.,* 1998; Glaser *et al.,* 2006),

Mbd3 (Hendrich *et al.,* 2001; Kaji *et al.,* 2006), *Nasp* (Richardson *et al.,* 2006), *Npm2* (Burns *et al.,* 2003), *Setdb1* (Dodge *et al.,* 2004), *Smarcb1* (Klochendler-Yeivin *et al.,* 2000), *Suz12* (Pasini *et al.,* 2004), *YY1* (Donohoe *et al.,* 1999). Interestingly, 6 of them, *Dicer, Eed, Hdac1*, *Mbd3, Suz12* and *Smarcb1* were targeted by Esrrb. However, despite the binding, none of them were downregulated alongside with the knockdown of *Esrrb*.

Secondly, I focused on the JmjC domain-containing proteins. Several of JmjC domain-containing proteins have been reported to be function as histone demethylases (Yamane *et al.,* 2006; Tsukada *et al.,* 2006; Klose *et al.,* 2006; Klose and Zhang, 2007). Recently, JmjC domain-containing protein encoding genes, *Jmjd1a* and *Jmjd2c* have been reported to be critical for the self-renewal of ES cells as the downstream targets of Oct4 (Loh *et al*., 2007). Even though no binding of Esrrb was found on these two genes, several genes that encode JmjC domain-containing proteins were targeted by Esrrb. They include *Jarid1b, Jarid2*, *Jmjd2b*, *Jmjd3* and *Jmjd4* (Table 3.7). Although the expression profiling indicates that *Jmjd2b* and *Jarid1b* were only down-regulated on the 6th day of *Esrrb* knockdown (Data not shown), *Jmjd3* appeared to be sensitive to the Esrrb level (Figure 3.33). Its expression was reduced on the 2nd day of *Esrrb* depletion and this reduction lasted in the following days, suggesting the activating effect of Esrrb on this gene. *Jmjd3* is a novel histone H3K27 demethylase that participates in cell fate decisions and developmental processes by antagonizing polycomb protein-mediated gene silencing (Hong *et al.,* 2007; Jepsen *et al.,* 2007; Xiang *et al.,* 2007; Lan *et al.,* 2007; De Santa *et al.,* 2007; Agger *et al.,* 2007). However, the role of *Jmjd3* in ES cells remains unclear. If *Jmjd3* could antagonize the polycomb protein-mediated gene silencing in ES cells (Lan *et al.,* 2007;

De Santa *et al.,* 2007; Agger *et al.,* 2007), it could potentially be involved in pluripotency decisions through its regulation on lineage-related genes. Thus as an activator of *Jmjd3*, Esrrb may indirectly maintain the balance of gene expression in ES cells through epigenetic mechanisms.

Table 3.7 Summary of Esrrb binding to epigenetic regulator encoding genes.

Genes	Location of Esrrb binding sites	Number of Esrrb binding sites
Jarid1b	in intron	3
	5' proximal	1
	5' distal	1
	3' proximal	1
Jarid2	in intron	5
	5' proximal	1
	5' distal	7
Jmjd2b	in exon	1
Jmjd3	in intron	2
Jmjd4	in exon	1
	3' proximal	1

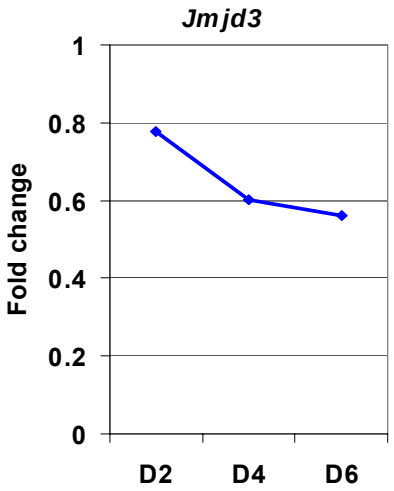


Figure 3.33 Expression profile of *Jmjd3* after *Esrrb* depletion. The x axis represents different days of *Esrrb* depletion (D2-6) and the y axis represents the gene expression fold change based on the microarray data after *Esrrb* depletion as compared to the knockdown control.

3.4.4.5. Esrrb, Nanog and Oct4 co-occupy common target genes

To further examine the roles of Esrrb in the pluripotent regulatory network, I compared the overlaps between the Esrrb target genes with the targets of Nanog and Oct4 identified previously from the ChIP-PET study (Loh *et al.,* 2006). A comparison of the genes bound by Esrrb, Nanog and Oct4 revealed that the three transcription factors have overlapping yet distinct downstream targets. 51% of Oct4 target genes and 41% of Nanog target genes were also targeted by Esrrb. A total of 253 genes were found to be common targets of all the three transcription factors (Figure 3.34; Appendix 4). Among these 253 genes are ES cell-associated genes like *Oct4, Sox2, Nanog, Esrrb, Tcfcp2l1*, and *Tbx3*. The co-binding of Esrrb, Nanog and Oct4 on these self-renewal genes suggests the importance of maintaining their high expression levels in ES cells. In addition, some developmental gene family members, such as *Fgfr1*, *Foxn2*, *Myo10* and *Nkx6,1*, were also observed to

be bound by these three transcription factors. Interestingly, two epigenetic regulators, *Suz12* and *Jarid2* were co-occupied by Esrrb, Oct4 and Nanog. This potentially implies the crosstalk between genetic and epigenetic regulators in ES cell maintenance. Strikingly, 12 common target genes of Oct4 and Nanog previously found to be conserved in human and mouse ES cells were occupied by Esrrb (Loh *et al.,* 2006). These 12 genes are *Nanog*, *Sox2*, *Zic3*, *Eomes*, *Rif1*, *Rest*, *Cdyl*, *Jarid2*, *Gsh2*, *Smarcad1*, *Atbf1* and *Sall1*. This interesting result suggests that Esrrb is intricately involved in regulating genes that are also bound by the "master regulators" Oct4 and Nanog. Furthermore, through comparing the differentially expressed genes in *Esrrb*-depleted, *Nanog*-depleted and *Oct4*-depleted ES cells, 10 common target genes of Esrrb, Nanog and Oct4 were found to be functionally regulated by all these three transcription factors. These 10 genes include *5033405K12Rik*, *9630008K15Rik*, *Btbd11*, *Dkk1*, *Egln3*, *Fzd5*, *Jam2*, *Lpp*, *Manba* and *Otx2* (Ivanova et al., 2006). The results indicate that the three factors may function in distinct pathways but operate inter-dependently to coordinate the specification of the ES cell state. To further uncover the functional relevance of these overlapped targets shared among Esrrb, Oct4 and Nanog, I performed GO analysis for these genes (Table 3.8). Notably, these targets are significantly over-represented in the cellular processes of transcription regulation and developmental processes.

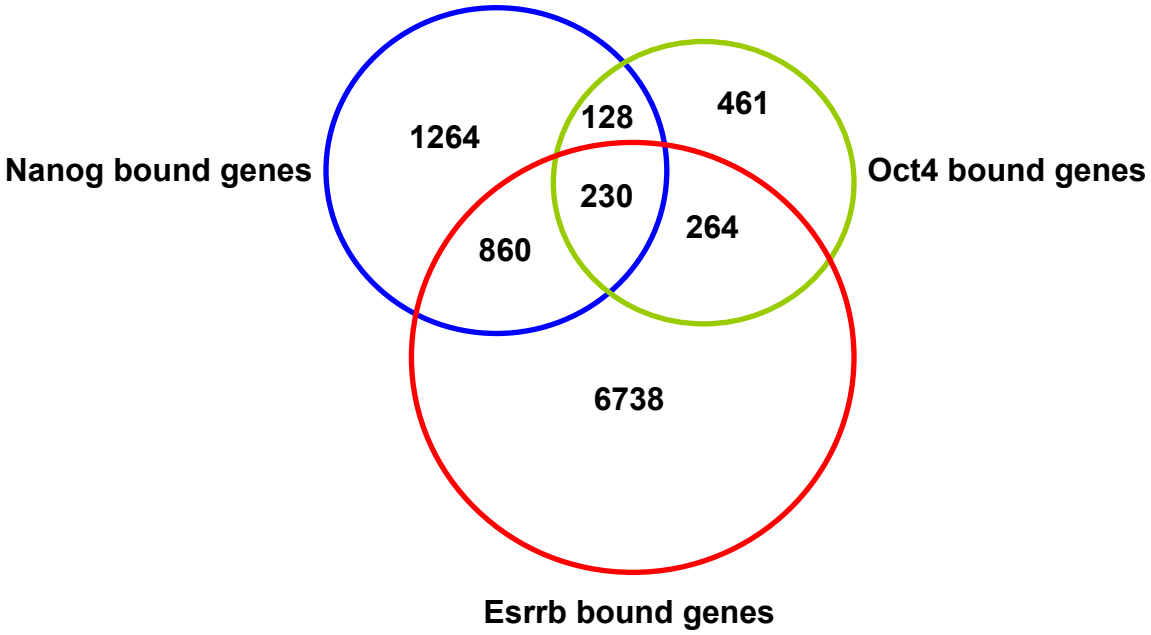


Figure 3.34 Venn diagram showing the overlaps of target genes bound by Esrrb, Oct4 or Nanog.

Table 3.8 Gene Ontology (GO) analysis was performed for functional annotation of the overlapped genes bound by Esrrb, Oct4 and Nanog (p value<0.01).

		P-value
Biological Process	Developmental processes	5.99E-08
	mRNA transcription	2.16E-05
	mRNA transcription regulation	8.60E-05
	Nucleoside, nucleotide and nucleic acid metabolism	3.98E-04
Molecular Function	Transcription factor	4.96E-07

CHAPTER FOUR

DISCUSSION

Chapter IV. Discussion

In this project, *Esrrb* was identified as a novel downstream target of the ES cell regulator Nanog. The *Esrrb*-depleted ES cells differentiated into fibroblast-like cells, along with the reduction in pluripotent gene expression and the induction of differentiation marker gene expression. Through genome-wide mapping of Esrrb binding sites and expression profiling of the *Esrrb* knockdown cells, the roles of Esrrb in maintaining the undifferentiated state of ES cells were dissected. Hence, this project provides greater understanding into the mechanisms that underlie the important roles of Nanog and Esrrb in ES cell maintenance.

4.1. Nanog target genes as candidate regulators of the self-renewal and pluripotency of ES cells

In this project, the genes positively regulated by Nanog were identified as candidate factors that may function in maintaining the pluripotent state of ES cells. The rationale behind this strategy was based on the critical roles of Nanog in maintaining ES cells and in early embryonic development (Mitsui *et al.*, 2003). Independent studies by other groups have also confirmed the role of Nanog in self-renewal, where the down-regulation of *Nanog* resulted in extensive differentiation of ES cells (Ivanova *et al.,* 2006).

The functional significance of Nanog was re-evaluated in a recent study by Chambers *et al.* (2007). They demonstrated that *Nanog*-/- ES cells can remain undifferentiated while maintaining the capacity to differentiate *in vitro* and contribute to *in vivo* lineages via the formation of chimeras. Hence, the report concluded that Nanog was dispensable for

somatic pluripotency (Chambers *et al*., 2007). However, several questions arise from the study of Chambers *et al.* (2007). Firstly, the previous work by Mitsui *et al.* (2003) has demonstrated the requirement of Nanog for ES cell derivation from the ICM. Importantly, homozygous *Nanog* knockout ES cells exhibited phenotypic differentiation, characteristic of endoderm lineage. Molecular marker analysis also showed down-regulation of pluripotent marker expression (such as *Oct4*) and up-regulation of differentiation marker expression. Notably, Mitsui *et al*. (2003) used the RF8 mouse ES cell line and targeted the second exon of the *Nanog* gene. On the other hand, Chambers *et al.* (2007) used the E14 mouse ES cell line and performed conditional knockout through targeting the region from intron 1 to the 3' UTR of the *Nanog* locus under the tamoxifen-inducible control. Whether the disparities in the results between the two studies can be accounted for by the use of different ES cell lines and knockout strategies remains to be answered. Secondly, even though the heterogeneous expression of *Nanog* in ES cell culture as shown by Chambers *et al*. (2007) is supported by the work from Singh *et al* (2007), the two studies arrived at a very different conclusion on the regulation of downstream genes. In the study of Chambers *et al*. (2007), it was found that *Nanog*-/- ES cells expressed normal level of *Oct4*, *Sox2* and *Rex1*. Remarkably, Singh *et al* (2007) reported the dramatic reduction of *Oct4*, *Sox2* and *Rex1* expression in ES cell population that were sorted for low *Nanog* expression. This raised a question whether the *Nanog*-null ES cells of Chambers *et al*. (2007) have acquired compensatory changes or mutations during the selection process. Thus it will be necessary to further characterize these *Nanog*-null ES cells for normal karyotype and genomic integrity. Nevertheless, the study of Chambers *et al*. (2007) showed great consistency in revealing the role of Nanog in maintaining the

undifferentiated state of the ES cells. *Nanog*-/- ES cells have reduced self-renewal capacity and a propensity to differentiate spontaneously in culture (Chambers *et al*., 2007). Furthermore, *Nanog*-/- ES cells are impaired in germline transmissions. Together, this suggests that Nanog has a role in maintaining the undifferentiated state of ES cells and is required for the ES cell pluripotency.

Besides *Esrrb*, other downstream targets of Nanog have previously been shown to play roles in ES cell maintenance. These genes include *FoxD3, Rif1, Zic3* and *Sall4* (Hanna *et al*, 2002; Guo *et al*, 2002; Loh *et al*, 2006; Lim *et al*, 2006; Wu *et al*, 2006; Zhang *et al*, 2006). FoxD3 generally functions as a repressor and is critical for the ICM and epiblast formation *in vivo* and for the pluripotent stem cell maintenance *in vitro* (Hanna *et al*, 2002; Guo *et al*, 2002). Rif1 and Zic3 are indispensable for pluripotent ES cells, as shRNA-mediated depletion of *Rif1* or *Zic3* leads to ES cell differentiation (Loh *et al.*, 2006; Lim *et al.*, 2006). Two studies demonstrated the importance of Sall4 in ES cells (Zhang *et al.*, 2006; Wu *et al.*, 2006). Notably Sall4 activates *Oct4* expression (Zhang *et al.*, 2006) and at the same time interacts with Nanog to co-regulate many downstream target genes (Wu *et al.*, 2006). Hence, the Nanog transcriptional network serves as a useful guide for identifying additional components involved in the regulation of the self-renewal, pluripotency and differentiation of ES cells. The identification of the role of *Esrrb* in ES cell maintenance validates the importance of the Nanog network for the purpose of finding novel ES cell regulators.

4.1.1. Esrrb is a nuclear receptor protein that is critical for ES cell maintenance.

Esrrb is a nuclear receptor protein, belonging to the estrogen related receptor protein (ERR) family. Because natural ligands of this class of nuclear receptor proteins have not been identified, they are also known as orphan nuclear receptors (Giguère, 1999). Sequence comparison analysis revealed that ERRs are close relatives of the estrogen receptors (ERs) (Giguère, 2002). *ER* genes include three members, *ER*α, *ERβ* and *ERγ* (Xia *et al.*, 1999; Green *et al.,* 1986). ERs contain the conserved domains A to F common to nuclear receptors and regulate gene expression through recognition of a specific enhancer known as the estrogen responsive element (ERE) with the consensus sequence of 5'-AGGTCAnnnTGACCT-3' (Beato *et al.*, 1989; Rollerova *et al.,* 2000). In contrast to ERRs, ERs are responsive to the natural ligand, estradiol (Green *et al.*, 1986; Tremblay *et al.*, 1999). *ERRs* consist of *ERRα, ERRβ/Esrrb* and *ERRγ* (Giguère *et al.*, 1988; Heard *et al.*, 2000). Similar to most nuclear receptors, ERRs possess a well-conserved DNA-binding domain (C domain), ligand-binding domain (E domain) and the less conserved transactivation domain (N-terminal domain) (Giguere, 2002). Although they do not respond to estradiol, ERRs can recognize the ERE or estrogen related responsive element (ERRE) (5'-TCAAGGTCA-3') and bind either as dimers or monomers (Hong *et al.*, 1999; Bonnelye *et al.*, 1997; Vanacker *et al.*, 1998; Lu *et al*., 2001; Vanacker *et al.*, 1999; Xie *et al.*, 1999; Zhang and Teng, 2000; Yang *et al.*, 1998; Sanyal *et al.*, 2002; Pettersson *et al.*, 1996; Sladek *et al.*, 1997; Johnston *et al.*, 1997; Yang *et al.*, 1996; Vanacker *et al.*, 1999). Apart from the ERRs and ERs, other nuclear receptor, such as steroidogenic factor-1 (SF-1; NR5A1) can also bind to the ERRE (Wilson *et al.*, 1993; Luo *et al.*, 1994). Therefore extensive crosstalk exists between ERs, ERRs and other classes of nuclear

receptors in the transcriptional regulation of downstream target genes (Yang *et al.*, 1996; Vanacker *et al.*, 1999). In mouse embryonic stem cells, Esrrb (also known as ERRβ) is a predominant estrogen receptor related protein that is highly expressed, whereas the expressions of *ERRα* and *ERRγ* are very low.

Expression of *Esrrb* in the mouse embryo is restricted to the extra-embryonic ectoderm and the early stages of chorion formation (Luo *et al.*, 1997). *Esrrb* transcript is first detected in the extra-embryonic ectoderm at E5.5. At E6.5, its expression becomes more prominent. By E8.5, however, *Esrrb* expression starts to diminish and becomes undetectable at the later stages of embryonic development (Luo *et al.*, 1997). Expression of *Esrrb* beyond the embryonic stages is not well-studied and remains an area for further exploration. Knockout of *Esrrb* in mice results in abnormal chorion formation, failure in trophoblast proliferation, and consequentially embryonic lethality at E10.5 due to abnormalities of the placenta (Luo *et al.*, 1997). Failure of the *Esrrb* knockout embryo to proceed beyond early developmental stages provides evidence of its functional importance in embryonic development. A recent study showed that *Esrrb* expression persisted in embryonic germ cells at E11.5 (Mitsunaga *et al*., 2004). Mitsunaga *et al*. have also shown that conditional *Esrrb* mutant embryos gave rise to fewer germ cells (Mitsunaga *et al*., 2004). Germ cells are a highly specialized cell type that closely resembles the pluripotent ES cells. Embryonic germ cells derived from the primordial germ cells have great developmental potential and can contribute efficiently to chimeras (Labosky *et al*., 1994, Stewart *et al*., 1994).

In ES cells, the expression of *Esrrb* is up-regulated (Wei *et al*., 2005; Pettersson *et al.*, 1996). Its expression is inhibited during differentiation induced by LIF withdrawal (Palmqvist *et al*., 2005). Hence, the expression of *Esrrb* is positively correlated with the pluripotent state of the ES cells. During the shRNA-mediated *Esrrb* knockdown, ES cells became flattened and adopted the appearance of fibroblasts with a loss of alkaline phosphatase staining. Consistent with this differentiation phenotype, expression analysis of marker genes revealed that the expression of ES cell-specific genes was reduced in *Esrrb*-depleted cells, whereas that of multiple lineage markers was induced. These results suggested a previously uncharacterized role of Esrrb in maintaining the self-renewing state of ES cells.

Previously, Ivanova *et al*. performed stable knockdown of *Esrrb* using RNAi and confirmed its role as a regulator of the ES cell self-renewal (Ivanova *et al*., 2006). Using a fluorescence-based competition assay, they demonstrated that *Esrrb* is a gene important for self-renewal. Consistent with my findings, Ivanova *et al*. showed that depletion of *Esrrb* led to ES cell differentiation and resulted in the induction of various lineage marker expressions (Ivanova *et al.*, 2006). However, in their study, the expression of trophectoderm lineage markers was not affected during the 8-day *Esrrb* depletion experiment. Instead, *Esrrb* depletion resulted in the induction of endoderm marker expression. These observations contradict the findings in this thesis, where trophectoderm markers were activated and endoderm genes were suppressed in the *Esrrb*-depleted ES cells. It is noteworthy that the mock knockdown control cells used in the study of Ivanova *et al*. seem to induce spontaneous perturbation of lineage markers, including endoderm

markers (Ivanova *et al.*, 2006). Thus it is conceivable that the differences in the two studies pertaining to lineage marker expression levels could be due to the dissimilarity in the behaviors of the control cells used. To further elucidate and characterize the mechanisms underlying the role of Esrrb in pluripotency, a microarray time course experiment (1 to 7 days) was performed on the *Esrrb* knockdown ES cells (Ivanova *et al*., 2006). From the gene expression profiles of the shRNA-mediated knockdown cells, they hypothesized that the mechanisms by which Esrrb maintains self-renewal may include the direct repression of differentiation-promoting genes (Ivanova *et al*., 2006). This corroborated well with my findings that Esrrb binds extensively to developmental genes and regulates their expression. The genome-wide analysis in this project also uncovers many previously unknown regulatory nodes that link Esrrb to genes with diverse cellular functions including self-renewal, chromatin modifications, and reprogramming process. These will be further discussed in the following sections. Recently, by using a computational approach, Zhou *et al.* identified a list of 15 core regulators in mouse ES cells which include Oct4, Nanog, and Esrrb. This further highlights the important role of Esrrb in the transcriptional network regulating the maintenance of ES cell self-renewal and pluripotency (Zhou *et al.*, 2007).

4.2. Relationship between Esrrb and the key ES cell regulators

As a downstream target of Nanog and Oct4, *Esrrb* is directly bound and positively regulated by these two key transcription factors (Loh *et al*., 2006). It was previously reported that Oct4 regulates *Nanog* expression in ES cells (Rodda *et al*., 2005; Kuroda *et al*., 2005). Knockdown of *Oct4* also reduced the reporter expression driven by the *Nanog*

promoter (Rodda *et al*., 2005). Nanog is also functionally linked to *Oct4* through regulating their common downstream targets to prevent ES cells from differentiation (Loh *et al*., 2006). In addition, *Nanog* knockdown resulted in the drastic reduction of *Oct4* expression. Moreover, overexpression of *Nanog* in ES cells can sustain the level of *Oct4* expression under RA-induced differentiation condition. Interestingly, the Esrrb ChIP-sequencing results uncovered a novel link between Esrrb and the *Nanog* and *Oct4* genes (Figure 4.1). In particular, depletion of *Esrrb* can reduce *Nanog* expression while forced expression of *Esrrb* induced *Nanog* promoter activity. The close relationship between Esrrb, Nanog and Oct4 extends to their overlapped target genes. The combinatorial regulation of common targets may provide robust and rapid response to developmental signals in the processes of embryonic development.

Two regulatory motifs can be use to define the transcriptional network formed by Esrrb, Nanog and Oct4. The feed-forward loop describes the observation that Oct4 and Nanog converge on *Esrrb* and regulate its expression, which in turn act in concert with these two factors to control downstream target genes. The feed-forward loop motif may confer stability to the system and allow for rapid response to developmental switching depending on the activities of the individual factors in the loop. Esrrb, Nanog and Oct4 bind to the regulatory elements of their own genes (Figure 4.1) (Loh *et al.*, 2006), thus forming an interconnected auto-regulatory loop. In this auto-regulatory loop, the expression and function of these three key stem cell factors are linked to one another. Auto-regulation is thought to provide several advantages, including reduced response time to environmental stimuli and increased stability of gene expression.

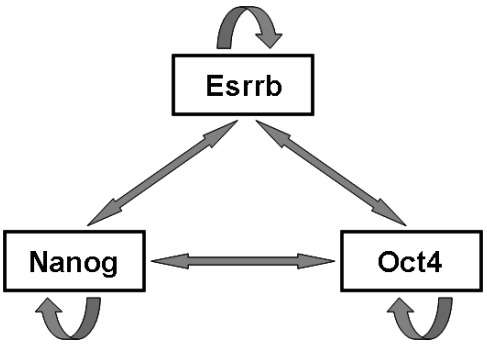


Figure 4.1 The interconnected regulatory loop formed by Esrrb, Nanog and Oct4. Esrrb, Nanog and Oct4 can target each other's encoding gene, while each of them auto-regulates their own expression (Loh *et al.*, 2006).

The close inter-connected relationship between Esrrb, Nanog and Oct4 is supported by recent studies. Wang *et al.* used affinity purification coupled with mass-spectrometry to identify Esrrb as a member of the Nanog interactome. Sall4 is also a member of the Nanog interactome (Wang *et al.*, 2006). Using ChIP analysis, Wu *et al*. demonstrated that there are considerable overlap in the binding sites of Nanog and Sall4, suggesting co-regulation of downstream target genes (Wu *et al.,* 2006). Thus it is reasonable to envisage that Esrrb and Nanog may function as partners in regulating common downstream targets in ES cells. Another key observation in the study of Wang *et al*. is that members of the interactomes are also targets of Nanog and/or Oct4 (Wang *et al.*, 2006). What emerges is a pluripotency network that is tightly inter-regulated and intricately connected by protein-protein interactions. Interestingly, I found that many genes in the interactome are also targeted by Esrrb, such as *Nanog, Oct4, Sall4, Rif1, Rest, Esrrb* and *Dax1*. Thus the findings here place Esrrb as an important member in the tight protein interactome/regulatory networks that govern ES cell self-renewal. In summary, the tight

interconnected circuitry formed by Esrrb, Oct4 and Nanog suggests the robust and concerted regulatory effects of these three factors in maintaining undifferentiated ES cells.

4.3. The Esrrb network is highly enriched in self-renewal and developmental genes

The work by Ivanova *et al.* has identified several self-renewal genes including *Nanog, Oct4, Sox2, Tbx3, Esrrb, Tcl1, Dppa4, Mm.343880*, *Mm.276044* and *Mm.219358* (Ivanova *et al*., 2006). In particular, knockdown of *Nanog, Oct4, Sox2, Tbx3, Esrrb, Tcl1* or *Dppa4* led to the loss of alkaline phophatase staining suggesting that these genes are required for the maintenance of ES cell state. Interestingly, these factors were identified by this study as targets of Esrrb. Moreover, from expression profile analysis, *Nanog, Sox2, Tbx3 and Tcl1* were found to be positively regulated by Esrrb. Thus Esrrb plays a positive role in the regulation of the newly identified self-renewal regulators. In their study, Ivanova *et al.* have only identified a role for Esrrb in regulating the self-renewal of ES cells. The findings in this thesis provide novel functional mechanisms for Esrrb to regulate the expression of critical downstream self-renewal regulators, thereby maintaining the pluripotency of ES cells.

Two groups have recently reported the construction of the Oct4 and Nanog transcriptional regulatory networks in human and mouse embryonic stem cells respectively (Boyer *et al.*, 2005; Loh *et al.*, 2006). These studies have found that both Oct4 and Nanog bind extensively to genes involving in developmental processes (Boyer *et al.*, 2005; Loh *et al.*, 2006). Expression profile analysis in the study of Ivanova *et al*. revealed that several developmental genes are preferentially up-regulated in the *Oct4* and

Nanog knockdown cells (Ivanova *et al*., 2006). These genes include *Otx2* and *Pitx2* which induced ES cells differentiation when over-expressed (Ivanova *et al*., 2006). Interestingly, *Otx2* was previously identified to be a gene target of both Nanog and Oct4, whereas *Pitx2* seem to be regulated by Nanog (Boyer *et al.*, 2005; Loh *et al.*, 2006). Thus one strategy that Oct4 and Nanog may maintain pluripotency is by promoting the expression of downstream self-renewal genes while simultaneously repressing the activity of differentiation-promoting genes.

Similar with Oct4 and Nanog, many target genes of Esrrb were uncovered to have functions in differentiation and developmental processes. Esrrb binds to the endoderm gene, *Gata6*, and trophectoderm regulator, *Cdx2*. Previous studies performed by forced-expression of sole transcription factors such as *Gata6* or *Cdx2* in ES cells (Fujikura *et al*., 2002; Niwa *et al*., 2005) demonstrate that they are sufficient to induce endodermal and trophectodermal differentiation respectively. Forced expression of *Otx2*, *Pitx2, Sox18*, *Snai1*, *Ets2*, *Irx3* and *Sox9* in ES cells resulted in differentiated phenotypes (Ivanova *et al*., 2006). Interestingly, my study identified the binding of Esrrb to all of these developmental genes and controlling their expression (Table 4.1; Figure 4.2). Thus, Esrrb appears to regulate differentiation-associated and lineage marker gene expressions in ES cells. This suggests that the mechanisms by which Esrrb maintains pluripotency may include the direct repression of differentiation-promoting genes.

Table 4.1 Summary of Esrrb binding to differentiation-related genes.

Genes	Location of Esrrb binding sites	Number of Esrrb binding sites
Otx2	3' proximal	2
Pitx2	in intron	1
Sox18	3' proximal	1
Snai1	in exon	1
	5' proximal	1
	3' proximal	1
	3' distal	1
Ets2	5' distal	1
	3' distal	2
Irx3	5' distal	1
Sox9	5' proximal	1

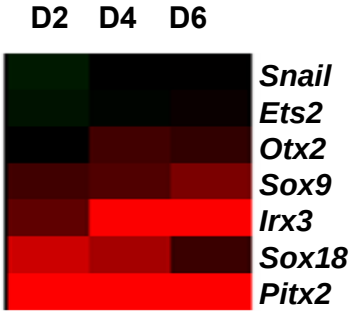


Figure 4.2 The regulation of differentiation-associated genes by Esrrb. Heatmap showed the expression profile of these genes which are bound by Esrrb. The genes were presented based on the fold changes of their expression levels after different days (D2-6) of *Esrrb* depletion compared with the *Gfp* shRNA-transfected control cells. They were sorted by the average expression ratio and mean centered.

4.4. The regulation of the ES cell chromatin structures by Esrrb

The recent study by Loh *et al.* presented a novel mechanism whereby Oct4 can activate its target genes indirectly through regulating the expression of chromatin modifiers. These chromatin modifiers maintain the gene expression by ensuring the accessible

chromatin conformation at the sites of the Oct4 target genes. For example, *Tcl1* is regulated by a histone H3K9 demethylase Jmjd1a (JmjC domain containing protein 1a) which in turn is transcriptionally regulated by Oct4 (Loh *et al*, 2007). This study connects the ES cell transcription circuitry to epigenetic modifiers to specify the pluripotent epigenetic landscape.

In finding the answers to whether Esrrb may also regulate chromatin modifier encoding genes that could potentially specify the epigenetic processes in the ES cells, I uncovered the Esrrb bound genes that encode for chromatin modifying enzymes, *Dicer, Eed, Hdac1, Mbd3*, *Suz12* and *Smarcb1*. These genes have previously been ascribed roles in regulating ES cell self-renewal or pluripotency (Bernstein *et al.*, 2003; Shumacher *et al.*, 1996; Lagger *et al.*, 2002; Hendrich *et al*., 2001; Kaji *et al*., 2006; Pasini *et al.*, 2004; Klochendler-Yeivin *et al.*, 2000; Boyer *et al.*, 2006). Although retaining the ability to self-renew, the *Mbd3*-null ES cells lost the capability to differentiate into somatic lineage cells (Kaji *et al.*, 2006). Similarly, *Eed*-null or *Suz12*-null ES cells appear to retain normal self-renewing capacity, but are impaired in their ability to differentiate (Montgomery *et al*., 2005; Azuara *et al*., 2006; Lee *et al.,* 2006). Eed and Suz12 are part of the Polycomb repressive complex (PRC) that appear to play an important role in specifying the epigenetic landscape of pluripotent ES cells (Boyer *et al.,* 2006; Lee *et al*., 2006). However, expression profiling suggests that the binding of Esrrb to these sites may not be functional. One possible explanation is that the expression of these epigenetic regulators is critical for the ES cell self-renewal and pluripotency; therefore they are under tight regulatory control. Thus the removal of a single factor (such as Esrrb) may be

insufficient to perturb their expression. Multiple transcription factors that bind to a single gene may have redundant effects in regulating the expression of the target. It is noteworthy that similar with Esrrb, Oct4 and Nanog only functionally regulate a subset of their target genes (Loh *et al.*, 2006). Since Esrrb shares a significant subset of target genes with Nanog and/or Oct4, it will be interesting to determine if some of these common targets that have chromatin modifying functions will be perturbed when the ES cells are treated with the collective down-regulation of two or more of the key factors.

Importantly, I identified *Jmjd3* which encodes a histone H3K27 demethylase to be a downstream target and actively regulated by Esrrb. Jmjd3 has been reported to be involved in both cell fate decision and developmental processes (Hong *et al.*, 2007; Jepsen *et al.*, 2007; Xiang *et al.*, 2007; Lan *et al.*, 2007; De Santa *et al.*, 2007; Agger *et al.*, 2007). It was suggested that Jmjd3 functions through antagonizing the polycomb protein-mediated gene silencing (Lan *et al.*, 2007; De Santa *et al.*, 2007; Agger *et al.*, 2007). However, the function of Jmjd3 in ES cells remains unclear. Matoba *et al.* showed that *Jmjd3* was up-regulated after the repression of *Oct4* in ES cells (Matoba *et al.*, 2006). Thus, it is hypothesized that Jmjd3 could play a positive role in the regulation of developmental genes during the processes of lineage commitment. Hence, Jmjd3 could potentially regulates the "poised" state of the lineage genes, and in turn contribute to the pluripotent properties of the ES cells (Bernstein *et al.*, 2006; Boyer *et al.*, 2006).

Esrrb was previously identified to be a member of the pluripotent interactome which was found to be enriched in chromatin modifiers such as histone deacetylase NuRD (p66b and

HDAC2), polycomb proteins (YY1, Rnf2 and Rybp), SWI/SNF chromatin remodeling (BAF155) complexes and co-repressor Tif1β (Wang *et al.*, 2006). Significantly, Esrrb may interact either directly or indirectly with these chromatin modifiers. Some of these chromatin modifiers, for example Rnf2 and KAP1 are known to play important role in early development and knockout mice die between E6.5-8.5 days (Voncken *et al.,* 2003; Cammas *et al.,* 2000). Thus, besides the regulation of chromatin modifiers, Esrrb may govern the chromatin state of pluripotent ES cells by its interaction with the chromatin modifiers (Wang *et al*., 2006).

4.5. Regulation of the reprogramming circuitry by Esrrb

Seminal work by Yamanaka's group has demonstrated that fully differentiated somatic cells can be induced into a pluripotent state by the introduction of only four transcription factors, Oct4, Sox2, Klf4 and Myc (Takahashi and Yamanaka, 2006). However, these induced pluripotent stem (iPS) cells that were selected based on expression of *Fbx15* could not be considered fully pluripotent because no live chimeric mice were obtained from the introduction of the iPS cells into blastocysts. Improvement of the methods using selection based on *Oct4* or *Nanog* expression gave rise to iPS cells that had a global transcriptional profile and epigenetic pattern more similar to mouse ES cells. Importantly, Nanog iPS cells and Oct4 iPS cells could form viable chimeras and are transmitted through the germ line (Maherali *et al.*, 2007; Okita *et al.*, 2007; Wernig *et al.*, 2007). More recently, several groups have shown that pluripotency can be restored when defined factors are introduced into somatic cells of human origin (Takahashi *et al.*, 2007; Yu *et al.*, 2007; Park *et al.*, 2007). Takahashi *et al*. (2007) and Park *et al.* (2007) have

demonstrated that application of the same four factors, OCT4, SOX2, KLF4 and MYC, can also induce human iPS cells from fetal, neonatal and adult somatic cells. Using a slightly different combination of genes, *OCT4, SOX2, NANOG and LIN28*, Yu *et al.* (2007) are able to reprogram cells originating from human fetal skin and foreskin of a newborn to pluripotent iPS cells. The prospect of deriving patient-specific pluripotent stem cells from reprogrammed somatic cells has moved the concept of regenerative medicine and cell therapy closer to reality.

However, despite these promising results, inherent problems could limit the application of these reprogrammed cells in replacement therapy. For example, there is a high frequency of tumor formation due to re-activation of retroviral *Myc*. Nakagawa *et al.* (2007) have recently reproduced the iPS cell derivation by introducing only *Oct4*, *Sox2* and *Klf4*. This circumvents the tumorigenicity provoked by *Myc* re-activation that commonly affects chimeric mice generated from the iPS cells. The low efficiency of iPS cell generation is another key issue that needs to be resolved. Consequentially, the search for alternative factors for inducing reprogramming in somatic cells is an area of great interest. As a target of Oct4 and Nanog, *Esrrb* is positively regulated by these two key reprogramming factors. This suggests that endogenous *Esrrb* may be activated by Oct4 and Nanog and participates in the process of reprogramming.

The findings of this project uncover the binding of Esrrb within the loci of the known reprogramming factors, *Oct4, Sox2, Klf2, Klf4, Klf5, Myc, Mycn* and *Nanog*. Notably, the bindings, except for *Klf2* and *Myc*, are functional as depletion of *Esrrb* resulted in the

corresponding perturbation of the expression of these reprogramming factor encoding genes. Incidentally, Nakagawa *et al.* (2007) and Blelloch *et al.* (2007) have recently indicated that *mycn* can substitute for *myc* in iPS induction. Nakagawa *et al.* (2007) even suggests that myc proteins are dispensable for reprogramming which can be activated by Oct4, Sox2 and Klf4.

It is of interest to note that in both studies, a limited array of genes for reprogramming functions were screened, and Esrrb was not among the candidates tested (Takahashi *et al.*, 2006; Yu *et al.*, 2007). Thus the potential of Esrrb as a reprogramming factor is not fully explored. Therefore it will be of tremendous interest to validate such roles of Esrrb for future work.

CHAPTER FIVE

CONCLUSIONS

Chapter V. Conclusions

Embryonic stem (ES) cells are remarkable in their ability to propagate indefinitely in culture, while retaining the potential to generate every cell type of the organism. Understanding the molecular mechanisms that underlie these unique properties will be useful for harnessing ES cells for application in regenerative medicine. The aim of this project is to identify transcription factors that are downstream of Nanog and play important roles in maintaining the undifferentiated state of ES cells.

This project has confirmed the important role of Nanog in maintaining the self-renewal state of the ES cells. shRNA-mediated depletion of *Nanog* led to the differentiation of ES cells which was accompanied by the reduced expression of pluripotent markers and elevation of differentiation genes. On the other hand, the *Nanog* overexpression cell line was shown to partially resist the differentiation induced by the addition of RA or withdrawal of LIF. This highlights the essential role of Nanog in preventing the differentiation of ES cells. Hence, these results are consistent with the findings from previous studies that reported the importance of Nanog in ES cells maintenance (Chamber *et al*., 2003; Mitsui *et al*., 2003; Ivanova *et al.,* 2006).

Using genome-wide affymetrix microarray analysis, 15 genes were identified to be putative targets that are positively regulated by Nanog. *Esrrb* is of special interest for several reasons. Firstly, Esrrb is a transcription factor (Pettersson *et al*., 1996; Xie *et al*., 1999; Lu *et al*., 2001; Sanyal *et al*., 2002). Secondly, Esrrb has important roles in early development and is highly expressed in pluripotent stem cells (Luo *et al*., 1997;

Pettersson *et al*., 1996). Interestingly, *Esrrb* was also found to be positively regulated by Oct4, another critical regulator of ES cells. Thus, this project has identified *Esrrb* as a direct target of two transcription factors that are critical for ES cell self-renewal.

The novel role of *Esrrb* in maintaining pluripotent ES cells was uncovered by shRNA-mediated knockdown. Strikingly, upon *Esrrb* depletion, ES cells lost their distinctive colony morphology and adopted fibroblast phenotype with the loss of alkaline phosphatase staining signal. Consistent with this differentiation phenotype, the expressions of pluripotent marker genes were down-regulated, while those of lineage marker genes were up-regulated, including endoderm, mesoderm, ectoderm and trophectoderm. These findings strongly suggest the indispensable role of Esrrb in the maintenance of the undifferentiated state of ES cells. Previous studies have reported the roles of Esrrb in early embryonic and germ cell development (Luo *et al*., 1997; Mitsunaga *et al*., 2004). Several other studies have also reported the positive expression of *Esrrb* in pluripotent ES cells (Wei *et al*., 2005; Palmqvist *et al*., 2005). However, no previous studies have focused on the functions of *Esrrb* in ES cell biology. Hence, this thesis reports the first instance of the critical roles that Esrrb plays in ES cell self-renewal.

To explore the mechanism underlying the function of Esrrb in ES cells, strategies involving genome-wide mapping of Esrrb binding and global expression profiling were employed. Over eight thousand genes were found to be bound by Esrrb, 64% of which showed differential expression upon *Esrrb* depletion, suggesting a regulatory effect of Esrrb on these target genes. It is noteworthy that Esrrb positively regulates many self-

renewal genes (such as *Nanog*, *Oct4*, *Sox2*, *Esrrb*, *Tbx3* and *Tcl1*) but represses many developmental genes (such as genes from the *Hox*, *Sox*, *Fox*, *Pax*, *Nkx*, *Pou*, *Tcf*, *Myo* and *Tbx* families). Hence, one strategy that Esrrb maintains pluripotency is by promoting the expression of downstream self-renewal genes while simultaneously repressing the activity of lineage specific genes. Furthermore, Esrrb binds to the genes that encode for epigenetic modifiers (*Jmjd3*, *Jmjd2b* and *Jarid1b*) and reprogramming factors (*Nanog, Oct4, Sox2, Klf2, Klf4, Klf5, Myc* and *Mycn*). This suggests a putative role of Esrrb in epigenetic processes and the restoration of pluripotency during the reprogramming of somatic cells. Moreover, comparison analysis of the genes bound by Esrrb, Nanog and Oct4 has identified the overlapped target genes shared by these three transcription factors. Interestingly, the function of these common target genes was highly enriched in transcriptional regulation. Future efforts should be directed toward the understanding of the different mode of regulations on common core targets by each of the three regulators.

Based on these findings, a model was generated to describe the function of Esrrb in maintaining the undifferentiated state of ES cells (Figure 5.1). In this model, Esrrb is part of the ES cells circuitry network that also consists of other key transcriptional regulators, Oct4 and Nanog. Through its regulation of ES cell-associated and developmental genes, Esrrb maintains the transcriptional landscape that specifies the self-renewing and pluripotent state of ES cells. In addition, by occupying epigenetic regulator encoding genes, Esrrb may also be involved in ensuring the epigenetic modifications, conductive for the self-renewal of ES cells.

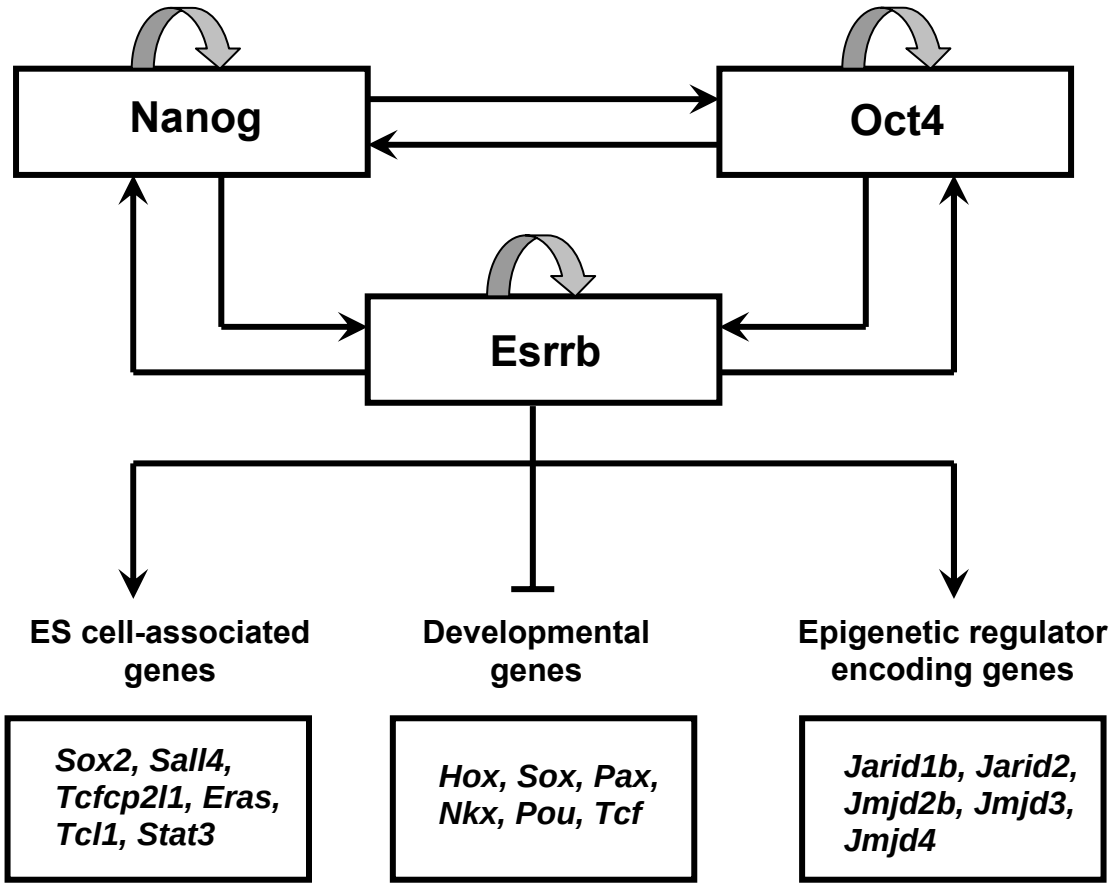


Figure 5.1 Model for the role of Esrrb in gene regulation in pluripotent ES cells.

In conclusion, this project has made important contributions toward elucidating the complexity of cell fate determination and identifying mechanisms for the stable propagation of a pluripotent ES cell state. The elucidation of the molecular mechanisms governing interactions between the nodes in the transcriptional network have provided critical insight into how pluripotency is established and maintained in embryonic stem cells. The knowledge gained from my study will thus aid in realizing the therapeutic

objective of ES cells and contribute towards harnessing their potential in regenerative medicine.

BIBLIOGRAPHY

Bibliography

Abeyta, M.J., Clark, A.T., Rodriguez, R.T., Bodnar, M.S., Pera, R.A. and Firpo, M.T. (2004). Unique gene expression signatures of independently derived human embryonic stem cell lines. *Hum Mol Genet*. **13**, 601–608.

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., *et al*. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–1656.

Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E. and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*. **449**, 731-734.

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.

Amit, M., Shariki, C., Margulets, V. and Itskovitz-Eldor, J. (2004). Feeder layer- and serum-free culture of human embryonic stem cells. *Biol Reprod*. **70**, 837–845.

Andrews, P.W., Damjanov, I. Simon, D., Banting, G.S., Carlin, C., Dracopoli, N.C. and Føgh, J. (1984). Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Differentiation in vivo and in vitro. *Lab Invest.* **50**, 147-162.

Anisimov, S.V., Tarasov, K.V., Tweedie, D., Stern, M.D., Wobus, A.M. and Boheler, K.R. (2002). SAGE identification of gene transcripts with profiles unique to pluripotent mouse R1 embryonic stem cells. *Genomics* **79**,169–176.

Annicotte, J. Fayard, S., Swift, E.G.H., Selander, L., Edlund, H., Tanaka, T., Kodama, T., Schoonjans, K. and Auwerx, J. (2003). Pancreatic-duodenal homeobox 1 regulates expression of liver receptor homolog 1 during pancreas development. *Mol. Cell. Biol.* **23**, 6713–6724.

Aubin, J., Davy, A. and Soriano, P. (2004). In vivo convergence of BMP and MAPK signaling pathways: impact of differential Smad1 phosphorylation on development and homeostasis. *Genes Dev*. **18**, 1482–1494

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., Jørgensen, 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.

Barak, Y., Nelson, M. C., Ong, E. S., Jones, Y. Z., Ruiz-Lozano, P., Chien, K. R., Koder, A. and Evans, R. M. (1999). PPAR □ is required for placental, cardiac, and adipose tissue development *Mol Cell* **4**, 585-95.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837.

Beato, M. (1989). Gene regulation by steroid hormones. *Cell* **56**, 335-344.

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

Behrens, J. (2000). Cross-regulation of the Wnt signalling pathway: a role of MAP kinases. *J. Cell Sci*. **113**, 911 – 919.

Ben-Shushan, E., Pikarsky, E., Klar, A. and Bergman, Y. (1993). Extinction of Oct-3/4 gene expression in embryonal carcinoma x fibroblast somatic cell hybrids is accompanied by changes in the methylation status, chromatin structure, and transcriptional activity of the Oct-3/4 upstream region. *Mol Cell Biol.* **13**, 891-901.

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.

Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V. and Hannon, G.J. (2003). Dicer is essential for mouse development. *Nat. Genet*. **35**, 215–217.

Bhattacharya, B., Cai, J., Luo, Y., Miura, T., Mejido, J., Brimble, S.N., Zeng, X., Schulz, T.C., Rao, M.S. and Puri, R.K. (2005). Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies. *BMC Dev Biol*. **5**, 22.

Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S., Amit, M., Lyons, I., Condie, B.G., Itskovitz-Eldor, J., Rao,

M.S. and Puri, R.K. (2004). Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood*, **103**, 2956–2964.

Blelloch, R., Venere, M., Yen, J. and Ramalho-Santos, M. (2007). Generation of Induced Pluripotent Stem Cells in the Absence of Drug Selection. *Cell Stem Cell* **1**, 245-247.

Blelloch, R.H., Hochedlinger, K., Yamada, Y., Brennan, C., Kim, M., Mintz, B., Chin, L. and Jaenisch, R. (2004). Nuclear cloning of embryonal carcinoma cells. *Proc. Natl Acad. Sci. USA* **101**, 13985–13990.

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

Bonnelye, E., Vanacker, J.M., Dittmar, T., Begue, A., Desbiens, X., Denhardt, D.T., Aubin, J.E., Laudet, V. and Fournier, B. (1997). The ERR-1 orphan receptor is a transcriptional activator expressed during bone development. *Mol. Endocrinol.* **11**, 905–916.

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

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539.

Bourguet, W., Ruff, M., Bonnier, D., Granger, F., Boeglin, M., Chambon, P., Moras, D. and Gronemeyer, H. (1995). Purification, functional characterization, and crystallization of the ligand binding domain of the retinoid X receptor. *Protein Expr. Purif.* **6**, 604-608.

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., 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. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353.

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

Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J. and Rao, M. (2004). MPSS profiling of human embryonic stem cells. *BMC Dev Biol*. **4**, 10.

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.

Brantjes, H., Roose, J., van de Wetering, M. and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res*. **29**, 1410–1419.

Brehm, A., Ohbo, K. and Scholer, H. (1997). The carboxy-terminal transactivation domain of Oct4 acquires cell specificity through the POU domain. *Mol Cell Biol.* **17**, 154-162.

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.

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

Brtko, J. (1994). Accurate determination and physicochemical properties of rat liver nuclear retinoic acid (RA) receptors. *Biochem. Biophys. Res. Commun*. **204**, 439-445.

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.

Bultman, S.J., Gebuhr, T.C., Pan, H., Svoboda, P., Schultz, R.M. and Magnuson, T. (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes Dev.* **20**, 1744–1754.

Bultman, S.J., Gebuhr, T.C., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G. and Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell*. **6**, 1287–1295.

Burns, K.H., Viveiros, M.M., Ren, Y., Wang, P., DeMayo, F.J., Frail, D.E., Eppig, J.J. and Matzuk, M.M. (2003). Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos. Science **300**, 633–636.

Cammas, F., Mark, M., Dollé, P., Dierich, A., Chambon, P. and Losson, R. (2000). Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development*. **127**, 2955-2963.

Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature Rev. Genet.* **6**, 507–512.

Capo-Chichi, C.D., Rula, M.E., Smedberg, J.L., Vanderveer, L., Parmacek, M.S., Morrisey, E.E., Godwin, A.K. and Xu, X.X. (2005). Perception of differentiation cues by GATA factors in primitive endoderm lineage determination of mouse embryonic stem cells. *Dev Biol*. **286**, 574-586.

Cardigan, K. M. and Nusse. R. (1997). Wnt signaling: a common theme in animal development. *Genes Dev*. **11**, 3286-3305.

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., Schöler, 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.

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

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

Chen, D., Zhao, M. and Mundy, G.R. (2004). Bone morphogenetic proteins. *Growth factors* **22**, 233-241.

Chen, W. S., Manova, K., Weinstein, D. C., Duncan, S. A., Plump, A. S., Prezioso, V. R., Bachvarova, R. F. and Darnell, J. E., Jr. (1994). Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos *Genes Dev* **8**, 2466-2477.

Chen S., Zhang Q., Wu X., Schultz P.G. and Ding S. (2004). Dedifferentiation of lineage-committed cells by a small molecule. *J Am Chem Soc.* **126**, 410-411.

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., Katz, D., Pereira, F.A., Jackson, K.J., DeMayo, F.J., Cooney, A.J. and O'Malley, B.W. (2001). Loss of orphan receptor germ cell nuclear factor function results

in ectopic development of the tail bud and a novel posterior truncation. *Mol. Cell Biol*. **21**, 663-677.

Clipsham, R.C., Niakan, K.K. and McCabe, E.R.B. (2004). Nr0b1 and its network partners are expressed early in murine embryos prior to steroidogenic axis organogenesis. *Gene Expr. Patterns.* **4**, 3–14.

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., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J. and Smith, A. (1998). Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev. Biol*. **203,** 149-162.

De Santa, F., Totaro, M.G., Prosperini, E., Notarbartolo, S., Testa, G. and Natoli, G. (2007). The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell*. **130**, 1083-1094.

Dechering, K., Boersma, C. and Mosselman, S. (2000). Estrogen receptors alpha and beta: two receptors of a kind? *Curr. Med. Chem*. **7**, 561-576.

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

DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686.

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* **93**, 6025–6030.

Dickins, R.A., Hemann, M.T., Zilfou, J.T., Simpson, D.R., Ibarra, I., Hannon, G.J. and Lowe, S.W. (2005). Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat. Genet*. **37**, 1289–1295.

Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H. and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Mol. Cell. Biol. **24**, 2478–2486.

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.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies O. (1987). Targetted correction of a mutant *HPRT* gene in mouse embryonic stem cells. *Nature* **330**, 576–578.

Dolci, S., Williams, D.E., Ernst, M.K., Resnick, J.L., Brannan, C.I., Lock, L.F., Lyman, S.D., Boswell, H.S. and Donovan, P.J. (1991). Requirement for Mast Cell Growth Factor for Primordial Germ Cell Survival in Culture; *Nature* **352**, 809-811.

Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E. and Shi, Y. (1999). Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell. Biol*. **19**, 7237–7244.

Duncan, S. A., Nagy, A. and Chan, W. (1997). Murine gastrulation requires HNF-4 regulated gene expression in the visceral endoderm: tetraploid rescue of Hnf-4(-/-) embryos *Development* **124**, 279-287.

Durcova-Hills, G., Adams, I.R., Barton, S.C., Surani, M.A. and McLaren, A. (2006). The role of exogenous fibroblast growth factor-2 on the reprogramming of primordial germ cells into pluripotent stem cells. *Stem Cells* **24**, 1441–1449.

Durston, A. J., Timmermans, J. P., Hage, W. J., Hendriks, H. F., de Vries, N. J., Heideveld, M. and Nieuwkoop, P. D. (1989). Retinoic acid causes an anteroposterior transformation in the developing central nervous system *Nature* **340**, 140-144.

Dykxhoorn, D.M., Novina, C.D. and Sharp, P.A. (2003). Killing the messenger: short RNAs that silence gene expression. *Nat. Rev. Mol. Cell Biol*. **4**, 457–467.

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

Evans, M.J. and Kaufman, M. (1983). Pluripotential cells grown directly from normal mouse embryos. *Cancer Surv.* **2**, 185–208.

Evans, R.M. (1988). The steroid and thyroid hormone receptor superfamily. *Science*. **240**, 889-895.

Feldman, B., Poueymirou, W., Papaioannou, V.E., DeChiara, T.M. and Goldfarb, M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science*. **267**, 246-249.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H. and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of *Oct-3/4* during early embryogenesis. *Nat Cell Biol.* **8**, 188-194.

Fenderson, B. A., Andrews, P. W., Nudelman, E., Clausen, H. and Hakomori, S. (1987). Glycolipid core structure switching from globo- to lacto- and ganglio-series during

retinoic acid-induced differentiation of TERA-2-derived human embryonal carcinoma cells. *Dev. Biol.* **122**, 21–34.

Fuhrmann, G., Chung, A.C., Jackson, K.J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Schöler, 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, Ji. J. and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* **16**, 784-789.

Fujiwara, T., Dunn, N. R. and Hogan, B. L. (2001). Bone morphogenetic protein 4 in the extraembryonic mesoderm is required for allantois development and the localization and survival of primordial germ cells in the mouse. *Proc. Natl Acad. Sci. USA* **98**, 13739–13744.

Gassmann, M., Donoho, G. and Berg, P. (1995). Maintenance of an extrachromosomal plasmid vector in mouse embryonic stem cells. *Proc Natl Acad Sci U S A*. **92**,1292-1296.

Gerami-Naini, B., Dovzhenko, O.V., Durning, M., Wegner, F.H., Thomson, J.A. and Golos, T.G.. (2003). Trophoblast Differentiation in Embryoid Bodies Derived from Human Embryonic Stem Cells. *Endocrinology* **145**, 1517-1524.

Giguère, V. (1999) Orphan nuclear receptors: from gene to function. *Endocr. Rev.* **20**, 689–725.

Giguère, V. (2002). To ERR in the estrogen pathway. *Trends Endocrinol Metab.* **13**, 220-225.

Giguère, V., Yang, N., Segui, P. and Evans, R.M. (1988). Identification of a new class of steroid hormone receptors. *Nature* **331**, 91-94.

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J. and Rao, M.S. (2004). Differences between human and mouse embryonic stem cells. *Dev Biol* **269**, 360–380.

Gitton, Y., Dahmane, N., Baik, S., Altaba, A., Neidhardt, L., Scholze, M., Herrmann, B.G., Kahlem, P., Benkahla, A., Schrinner, S., Yildirimman, R., Herwig, R., Lehrach, H. and Yaspo, M.L. (2002). HSA21 expression map initiative.A gene expression map of human chromosome 21 orthologues in the mouse. *Nature* **420**, 586–590.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.L. and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development **133**, 1423–1432.

Glass, C.K. (1994). Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr. Rev*. **15**, 391–407.

Green, S. and Chambon, P. (1998). Nuclear receptors enhance our understanding of transcription regulation. *Trends Genet*. **4**, 309–314.

Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.M., Argos, P. and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature* **320**, 134-139.

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. (2005). Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell Biol.* **25**, 3492-3505.

Guo, Y., Costa, R., Ramsey, H., Starnes, T., Vance, G., Robertson, K., Kelley, M., Reinbold, R., Scholer, H. and Hromas, R. (2002). The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. *Proc. Natl. Acad. Sci. USA*. **99**, 3663–3667.

Haegele, L., Ingold, B., Naumann, H., Tabatabai, G., Ledermann, B. and Brandner, S. (2003). Wnt signalling inhibits neural differentiation of embryonic stem cells by controlling bone morphogenetic protein expression. *Mol. Cell. Neurosci.* **24**, 696–708.

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.

Hart, A. H., Hartley, L., Ibrahim, M. and Robb, L. (2004). Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev Dyn.* **230**, 187-198.

Hatano, S.Y., Tada, M., Kimura, H., Yamaguchi, S., Kono, T., Nakano, T., Suemori, H., Nakatsuji, N. and Tada, T. (2005). Pluripotential competence of cells associated with Nanog activity. *Mech. Dev.* **122**, 67–79.

Hatoya, S., Torii, R., Kondo, Y., Okuno, T., Kobayashi, K., Wijewardana V., Kawate N., Tamada H., Sawada T., Kumagai D., Sugiura K. and Inaba T. (2006). Isolation and characterization of embryonic stem-like cells from canine blastocysts. *Mol. Reprod. Dev.* **73**, 298–305.

He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B. and Kinzler, K.W. (1998). Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512.

Heard, D.J., Norby, P.L., Holloway, J. and Vissing, H. (2000). Human ERRgamma, a third member of the estrogen receptor-related receptor (ERR) subfamily of orphan nuclear receptors: tissue-specific isoforms are expressed during development and in the adult. *Mol Endocrinol.* **14**, 382-392.

Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Muller-Newen, G. and Schaper, F. (2003). Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J*. **374,** 1–20.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A. and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev*. **15**, 710–723.

Hogan, B., Fellous, M., Avner, P. and Jacob, F (1977). Isolation of a human teratoma cell line which expresses F9 antigen. *Nature* **270**, 515–518.

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

Hong, H., Yang, L. and Stallcup, M.R. (1999). Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. *J. Biol. Chem.* **274**, 22618–22626.

Hong, S., Cho, Y.W., Yu, L.R., Yu, H., Veenstra, T.D. and Ge, K. (2007). Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. *Proc Natl Acad Sci U S A*. **104**, 18439-18444.

Hooper, M.L., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987). *Nature*, **326**, 292–295.

Houlard, M., Berlivet, S., Probst, A.V., Quivy, J.P., Hery, P., Almouzni, G. and Gerard, M. (2006). CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. *PLoS Genet*. **2**, e181.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C. and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. *J Cell Biol* **148**, 567–578.

Hughes, T.R. and Shoemaker, D.D. (2001). DNA microarrays for expression profiling. *Curr Opin Chem Biol*. **5**, 21-25.

Hummelke, G.C. and Cooney, A.J. (2001). Germ cell nuclear factor is a transcriptional repressor essential for embryonic development. *Front Biosci.* **6**, D1186-1191.

Illmensee, K. and Mintz, B. (1976). Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts. *Proc. Natl Acad. Sci. USA* **73**, 549–553.

Ishitani, T., Kishida, S., Hyodo-Miura, J., Ueno, N., Yasuda, J., Waterman, M., Shibuya, H., Moon, R.T., Ninomiya-Tsuji, J. and Matsumoto, K. (2003). The TAK1–NLK mitogen-activated protein kinase cascade functions in the Wnt-5a/Ca(2+) pathway to antagonize Wnt/beta-catenin signaling. *Mol. Cell Biol*. **23**, 131–139.

Ishitani, T., Ninomiya-Tsuji, J., Nagai, S., Nishita, M., Meneghini, M., Barker, N., Waterman, M., Bowerman, B., Clevers, H., Shibuya, H. and Matsumoto, K. (1999). The TAK1–NLK–MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. *Nature* **399**, 798– 802.

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.

Ivanova, N.B., 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.

Jaenisch, R. (1997). DNA methylation and imprinting: why bother? *Trends Genet*. **13**, 323-329.

Jepsen, K., Solum, D., Zhou, T., McEvilly, R.J., Kim, H.J., Glass, C.K., Hermanson, O. and Rosenfeld, M.G.. (2007). SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. *Nature*. **450**, 415-419.

Jiang G., Lee U. and Sladek F.M. (1997). Proposed mechanism for the stabilization of nuclear receptor DNA binding via protein dimerization. *Mol. Cell Biol.* **17**, 6546-6554.

Johnson, D.S., Mortazavi, A., Myers, R.M. and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502.

Johnston, S.D, Liu, X., Zuo, F., Eisenbraun, T.L., Wiley, S.R., Kraus, R.J. and Mertz, J.E. (1997) Estrogen-related receptor-α1 functionally binds as a monomer to extended half-site sequences including ones contained within estrogen-response elements.*Mol. Endocrinol.* **11**, 342–352.

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A. and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat. Cell Biol*. **8**, 285–292.

Kannagi, R., Cochran, N.A., Ishigami, F., Hakomori, S., Andrews, P.W., Knowles, B.B. and Solter, D. (1983). Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of

a unique globo-series ganglioside isolated from human teratocarcinoma cells. *EMBO J.* **2**, 2355-2361.

Kastner, P., Grondona, J. M., Mark, M., Gansmuller, A., LeMeur, M., Decimo, D., Vonesch, J. L., Dolle, P. and Chambon, P. (1994). Genetic analysis of RXR developmental function: convergence of RXR and RAR signaling pathways in heart and eye morphogenesis *Cell* **78**, 987-1003.

Kato, K. (1997). Adaptor-tagged competive PCR: a novel method for measuring relative gene expression. *Nucl. Acids Res.* **25**, 4694-4696.

Kato, Y., Rideout, W.M. 3rd, Hilton, K., Barton, S.C., Tsunoda, Y. and Surani, M.A. (1999). Developmental potential of mouse primordial germ cells. Development. **126**, 1823-1832.

Keegan, C.E. and Hammer, G.D. (2002). Recent insights into organogenesis of the adrenal cortex. *Trends. Endocrinol. Metab*. **13**, 200–208.

Kiefer, J.C. (2007). Back to basics: Sox genes. *Dev Dyn.* **236**, 2356-2366.

Kim, J.H., Auerbach, J.M., Rodríguez-Gómez, J.A., Velasco, I., Gavin, D., Lumelsky, N., Lee, S.H., Nguyen, J., Sánchez-Pernaute, R., Bankiewicz, K. and McKay, R. (2002). Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature* **418**, 50–56.

Kleinsmith, L.J. and Pierce, G.B.Jr. (1964). Multipotentiality of single embryonal carcinoma cells. *Cancer Res.* **24**, 1544–1551.

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C. and Yaniv, M. (2000). The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. *EMBO Rep*. **1**, 500–506.

Klose, R.J., Yamane, K., Bae, Y. and Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **442**, 312-316.

Klose, R.J. and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol.* **8**, 307-318.

Knoblauch, R. and Garabedian, M.J. (1999). Role for Hsp90-associated cochaperone p23 in estrogen receptor signal transduction. *Mol. Cell Biol.* **19**, 3748-3759.

Komiya, T., Tanigawa, Y. and Hirohashi, S. (1997). A large-scale in situ hybridization system using an equalized cDNA library. *Anal Biochem*. **254**, 23–30.

Kretzschmar, M., Doody, J. and Massague, J. (1997a). Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1.*Nature* **389**, 618–622.

Kretzschmar, M., Liu, F., Hata, A., Doody, J. and Massague, J. (1997b). The TGF-beta family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. *Genes Dev.* **11**, 984–995.

Kuehn, M.R., Bradley, A., Robertson, E.J. and Evans, M.J. (1987). A potential animal model for Lesch–Nyhan syndrome through introduction of *HPRT* mutations into mice. *Nature* **326**, 295–298.

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.

Kyba, M., Perlingeiro, R.C. and Daley, G.Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell*. **109**, 29-37.

Labosky, P.A., Barlow, D.P. and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulinlike growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197–3204.

Lagger, G., O'Carroll, D., Rembold, M., Khier, H., Tischler, J., Weitzer, G., Schuettengruber, B., Hauser, C., Brunmeir, R., Jenuwein, T. and Seiser, C. (2002). Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J*. **21**, 2672–2681.

Lai, E.C. (2004). Notch signaling: control of cell communication and cell fate. *Development* **131**, 965–973.

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., Roberts, T.M., Chang, H.Y. and Shi, Y. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature*. **449**, 689-694.

Lang, D., Powell, S.K., Plummer, R.S., Young, K.P. and Ruggeri, B.A. (2007). PAX genes: roles in development, pathophysiology, and cancer. *Biochem Pharmacol*. **73**, 1-14.

Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversiWcation from an ancestral orphan receptor. *J. Mol. Endocrinol.***19,** 207–226.

Laudet, V., Hänni, C., Coll, J., Catzeflis, F. and Stéhelin, D. (1992). Evolution of the nuclear receptor gene superfamily. *EMBO J.* **11**,1003-1013.

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.

Levinson-Dushnik, M. and Benvenisty, N. (1997). Involvement of hepatocyte nuclear factor 3 in endoderm differentiation of embryonic stem cells. *Mol. Cell Biol.* **17**, 3817–3822.

Li, H., Roblin, G., Liu, H. and Heller, S. (2003). Generation of hair cells by stepwise differentiation of embryonic stem cells. *Proc. Natl Acad. Sci. USA* **100**, 13495–13500.

Li, J., Pan, G., Cui, K., Liu, Y., Xu, S. and Pei, D. (2007). A dominant negative form of mouse Sox2 induces trophectoderm differentiation and progressive polyploidy in mouse ES cells. *J Biol Chem*. **282**, 19481-19492.

Li, M., Pevny, L., Lovell-Badge, R. and Smith, A. (1998). Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr. Biol.* **8**, 971–977.

Li, X., Zhou, S. G., Imreh, M.P.,Ahrlund-Richter, L. and Allen,W. R. (2006). Horse embryonic stem cell lines from the proliferation of inner cell mass cells. *Stem Cells Dev.* **15**, 523–531.

Liang, P. and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971.

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.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. *Nat Genet*. **21**, 20-24.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675-1680.

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.

Loh, Y.H., Zhang, W. Chen, X., George, J. and Ng, H.H.. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev*. **21,** 2545-2557.

Lohnes, D., Mark, M., Mendelsohn, C., Dolle, P., Dierich, A., Gorry, P., Gansmuller, A. and Chambon, P. (1994) Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants *Development* **120**, 2723-48.

Lowell, S., Benchoua, A., Heavey, B. and Smith, A.G. (2006). Notch promotes neural lineage entry by pluripotent embryonic stem cells. *PLoS Bio*l. **4**, e121.

Lu, D., Kiriyama, Y., Lee, K.Y. and Giguère, V. (2001). Transcriptional regulation of the estrogen-inducible pS2 breast cancer marker gene by the ERR family of orphan nuclear receptors. *Cancer Res.* **61**, 6755–6761.

Luo, J., Sladek, R., Bader, J.A., Matthyssen, A., Rossant, J., and Giguère, V. (1997). Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388**, 778-782.

Luo, X., Ikeda, Y. and Parker, K.L. (1994). A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation *Cell* **77**, 481-490.

Luo, X., Ikeda, Y. and Parker, K.L. (1995). The cell-specific nuclear receptor steroidogenic factor 1 plays multiple roles in reproductive function *Philos Trans R Soc Lond B Biol Sci* **350**, 279-283.

Louvi, A. and Artavanis-Tsakonas, S. (2006). Notch signalling in vertebrate neural development. *Nat Rev Neurosci*. **7**, 93–102.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K. and Hochedlinger, K. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70.

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. and Evans, R.M. (1995). The nuclear receptor superfamily: the second decade. *Cell* **83,** 841–850.

Mardis, E.R. (2007). ChIP-seq: welcome to the new frontier. *Nat. Methods*. **4**, 613-614.

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. USA* **78**, 7634–7638.

Martin, G.R. and Evans, M.J (1974). The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**, 163–172.

Martin, G.R., and Evans, M.J. (1975a). Differentiation of clonal lines of teratocarcinoma cells: Formation of embryoid bodies *in vitro*. *Proc. Natl Acad. Sci. USA* **72**, 1441–1445.

Martin, G.R. and Evans, M.J. (1975b). Multiple differentiation of clonal teratocarcinoma stem sells following embryoid body formation *in vitro*. *Cell* **6**, 467–474.

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.

Massague, J. (1998). TGF-beta signal transduction. *Annu. Rev. Biochem.***67**, 753–791.

Massague J. (2003). Integration of Smad and MAPK pathways: a link and a linker revisited. *Genes Dev*. **17**, 2993-2997.

Massague, J. (2000). How cells read TGF-beta signals. *Nat. Rev., Mol. Cell Biol*. **1**, 169–178.

Massaro, G.D., Massaro, D. and Chambon, P. (2003) Retinoic acid receptor-□regulates pulmonary alveolus formation in mice after, but not during, perinatal period *Am J Physiol Lung Cell Mol Physiol* **284**, 431-433.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.A. and Ko, Minoru. (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.M. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847.

Matt, N., Ghyselinck, N. B., Wendling, O., Chambon, P. and Mark, M. (2003) Retinoic acid-induced developmental defects are mediated by RARbeta/RXR heterodimers in the pharyngeal endoderm *Development* **130**, 2083-2093.

Mendelsohn, C., Lohnes, D., Decimo, D., Lufkin, T., LeMeur, M., Chambon, P. and Mark, M. (1994). Function of the retinoic acid receptors (RARs) during development (II). Multiple abnormalities at various stages of organogenesis in RAR double mutants *Development* **120**, 2749-2771.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T. and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-116.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S. and Bernstein, B.E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.

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

Mishina, Y. (2003). Function of bone morphogenetic protein signaling during mouse development. *Front Biosci*. **8**, d855– d869.

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., Giguère, 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.

Miura, T., Luo, Y., Khrebtukova, I., Brandenberger, R., Zhou, D., Thies, R.S., Vasicek, T., Young, H., Lebkowski, J., Carpenter, M.K. and Rao M.S. (2004). Monitoring early differentiation events in human embryonic stem cells by massively parallel signature sequencing and expressed sequence tag scan. *Stem Cells Dev.* **13**, 694–715.

Miyazono, K., ten Dijke, P. and Heldin, C.H. (2000). TGF-beta signaling by Smad proteins. *Adv. Immunol*. **75**, 115– 157.

Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E. and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1–15.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P. and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr. Biol.* **15,** 942–947.

Moon, R.T., Bowerman, B., Boutros, M. and Perrimon, N. (2002).The promise and perils of Wnt signaling through beta-catenin. *Science* **296**, 1644-1646.

Morohashi, K. (1997). The ontogenesis of the steroidogenic tissues. *Genes Cells* **2**, 95–106.

Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S. and Parmacek, M.S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev*. **12**, 3579-3590.

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

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S. (2007). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol.* **26**, 101-106.

Nakano, T., Kodama, H. and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098–1101.

Neidhardt, L., Gasca, S., Wertz, K., Obermayr, F., Worpenberg, S., Lehrach, H. and Herrmann, B.G. (2000). Large-scale screen for genes controlling mammalian embryogenesis, using highthroughput gene expression analysis in mouse embryos. *Mech. Dev.* **98**, 77–94.

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.

Ngo, V.N., Davis, R.E., Lamy, L., Yu, X., Zhao, H., Lenz, G., Lam, L.T., Dave, S., Yang, L., Powell, J. and Staudt, L.M. (2006). A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110.

Niakan, K.K., Davis, E.C., Clipsham, R.C., Jiang, M., Dehart, D.B., Sulik, K.K. and McCabe, E.R. (2006). Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Mol. Genet. Metab*. **88**, 261–271.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, 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.

Nishikawa, S.I., Nishikawa, S., Hirashima, M., Matsuyoshi, N. and Kodama, H. (1998). Progressive lineage analysis by cell sorting and culture identifies FLK1+ VE-cadherin+

cells at a diverging point of endothelial and hemopoietic lineages. *Development* **125**, 1747–1757.

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.

Nitta, M., Ku, S., Brown, C., Okamoto, A.Y., and Shan, B. (1999). CPF: an orphan nuclear receptor that regulates liver-specific expression of the human cholesterol 7alpha-hydroxylase gene. *Proc. Natl. Acad. Sci. USA* **96**, 6660–6665.

Niwa, H. (2001). Molecular mechanism to maintain stem cell self-renewal of ES cells. *Cell structure and function* **26**, 137-148.

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., Miyazaki, J. and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or selfrenewal 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* **23**, 917-929.

Nohe, A., Keating, E., Knaus, P. and Petersen, N.O. (2004). Signal transduction of bone morphogenetic protein receptors. *Cell signal* **16**, 291-299.

Nordhoff, V., Hubner, K., Bauer, A., Orlova, I., Malapetsa, A. and Scholer, H.R. (2001). Comparative analysis of human, bovine, and murine Oct4 upstream promoter sequences. *Mamm Genome* **12**, 309-317.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A. and Jenuwein, T. (2001) The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol*. **21**, 4330-4336.

Odorico, J. S., Kaufman, D. S. and Thomson, J. A. (2001). Multilineage differentiation from human embryonic stem cell lines. *Stem Cells* **19**, 193–204.

Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. and Niwa, H. (2006). Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. *Biochem Biophys Res Commun.* **343**, 159–166.

Ogawa, K., Saito, A., Matsui, H., Suzuki, H., Ohtsuka, S., Shimosato, D., Morishita, Y., Watabe, T., Niwa, H. and Miyazono, K. (2007). Activin-Nodal signaling is involved in propagation of mouse embryonic stem cells. *J Cell Sci*. **120,** 55–65.

Okano, M., Lei, H. and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983–1993.

Okita, K., Ichisaka, T. and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317.

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.

Paling, N.R., Wheadon, H., Bone, H.K. and Welham, M.J. (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *J Biol Chem*. **279**, 48063–48070.

Palmieri, S.L., Peter, W., Hess, H. and Schöler, 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.J., Chang, Z.Y., Scholer, H.R. and Pei, D. (2002). Stem cell pluripotency and transcription factor Oct4. *Cell Res.* **12**, 321-329.

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.

Pan, G.J. and Pei, D.Q. (2003). Identification of two distinct transactivation domains in the pluripotency sustaining factor nanog. *Cell Res.* **13**, 499-502.

Pan, G. and Thomson, J.A. (2007). Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res.* **17**, 42-49.

Papaioannou, V.E., Gardner, R.L., McBurney, M.W., Babinet, C., and Evans, M.J. (1978). Participation of cultured teratocarcinoma cells in mouse embryogenesis. *J. Embryol. Exp. Morph.* **44**, 93–104.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L. and Evans, M.J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258**, 70–73.

Pare, J.F., Malenfant, D., Courtemanche, C., Jacob-Wagner, M., Roy, S., Allard D. and Belanger, L. (2004). The fetoprotein transcription factor (FTF) gene is essential to

embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. *J. Biol. Chem.* **279,** 21206–21216.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W. and Daley, G.Q. (2007). Reprgramming of human somatic cells to pluripotency with defined factors. *Nature* **9**, 871-873.

Parker, M.G. (1995). Structure and function of estrogen receptors. Vitam Horm. **15**, 267-287.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. and Helin, K. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* **27,** 3769–3779.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E. and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071.

Pease, A., Solas, D., Sullivan, E, Cronin M.T., Holmes C.P. Fodor S.P. (1994). Light-Generated Oligonucleotide Arrays for Raoid DNA Sequence Analysis. *Proc Natl Acad Sci USA.* **91**, 5022-5026.

Pereira, F.A., Qiu, Y., Zhou, G., Tsai, M.J. and Tsai, S.Y. (1999). The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development *Genes Dev* **13**, 1037-1049.

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.

Pesce, M., Wang, X., Wolgemuth, D.J. and Scholer, H. (1998). Differential expression of the Oct4 transcription factor during mouse germ cell differentiation. *Mech Dev.* **71**, 89-98.

Pettersson, K., Svensson, K., Mattsson, R., Carlsson, B., Ohlsson, R. and Berkenstam, A. (1996). Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. *Mech Dev.* **54**, 211-223.

Pilpel, Y., Sudarsanam, P. and Church, G.M. (2001). Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet.* **29**, 153-159.

Pratt, W.B. and Toft, D.O. (1997). Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr. Rev.* **18**, 306-360.

Qiu, Y., Pereira, F.A., DeMayo, F.J., Lydon, J.P., Tsai, S.Y. and Tsai, M.J. (1997). Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization *Genes Dev* **11**, 1925-1937.

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, M. (2004). Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev. Biol.* **275**, 269– 286.

Rausa, F.M., Galarneau, L., Belanger, L., and Costa, R.H. (1999). The nuclear receptor fetoprotein transcription factor is coexpressed with its target gene HNF-3beta in the developing murine liver intestine and pancreas. *Mech. Dev.* **89**, 185–188.

Resnick, J.L., Bixler, L.S., Cheng, L. and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature*. **359**, 550-551.

Reymond, A., Marigo, V., Yaylaoglu, M.B., Leoni, A., Ucla, C., Scamuffa, N., Caccioppoli, C., Dermitzakis, E.T., Lyle, R., Banfi, S., Eichele, G., Antonarakis, S.E. and Ballabio, A. (2002). Human chromosome 21 gene expression atlas in the mouse. *Nature* **420**, 582–586.

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.

Richardson, R.T., Alekseev, O.M., Grossman, G., Widgren, E.E., Thresher, R., Wagner, E.J., Sullivan, K.D., Marzluff, W.F. and O'Rand, M.G. (2006). Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation. *J. Biol. Chem.* **281,** 21526–21534.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O.L., He, A., Marra, M., Snyder, M. and Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods.* **4**, 651-657.

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.

Rollerova, E and Urbancikova, M. (2000). Intracellular estrogen receptors, their characterization and function (Review). *Endocr Regul.* **34**, 203-218.

Romand, R., Hashino, E., Dolle, P., Vonesch, J. L., Chambon, P. and Ghyselinck, N. B. (2002). The retinoic acid receptors RARalpha and RARgamma are required for inner ear development *Mech Dev* **119**, 213-223.

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.

Rossant, J. and Papaioannou, V.E. (1984). The relationship between embryonic, embryonal carcinoma and embryo-derived stem cells. *Cell differ*. **15**, 155-161.

Rossant, J. and McBurney, M. W. (1982). The developmental potential of a euploid male teratocarcinoma cell line after blastocyst injection. *J. Embryol. Exp. Morph.* **70**, 99–112.

Sadovsky, Y., Crawford, P. A., Woodson, K. G., Polish, J. A., Clements, M. A., Tourtellotte, L. M., Simburger, K. and Milbrandt, J. (1995). Mice deficient in the orphan receptor steroidogenic factor 1 lack adrenal glands and gonads but express P450 side-chain-cleavage enzyme in the placenta and have normal embryonic serum levels of corticosteroids *Proc Natl Acad Sci U S A* **92**, 10939-10943.

Saga, Y., Kobayashi, M., Ohta, H., Murai, N., Nakai, N., Oshima, M. and Taketo, M.M. (1999). Impaired extrapyramidal function caused by the targeted disruption of retinoid X receptor RXRgamma1 isoform *Genes Cells* **4**, 219-228.

Saito, S., Ugai, H., Sawai, K.,Yamamoto,Y.,Minamihashi, A.,Kurosaka, K., Kobayashi, Y., Murata, T., Obata, Y. and Yokoyama, K. (2002). Isolation of embryonic stem-like cells from equine blastocysts and their differentiation *in vitro. FEBS Lett.* **531**, 389–396.

Sanyal, S., Kim, J.Y., Kim, H.J., Takeda, J., Lee, Y.K., Moore, D.D. and Choi, H.S. (2002). Differential regulation of the orphan nuclear receptor SHP gene promoter by orphan nuclear receptor ERR isoforms. *J. Biol. Chem.* **277**, 1739–1748.

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. *Nature Med.* **10**, 55–63.

Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F. and Brivanlou, A.H. (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*. **260**, 404–413.

Schepers, G. E., Teasdale, R. D. and Koopman, P. (2002). Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev. Cell.* **3**, 167–-170.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. and Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* **93**, 10614-10619.

Schmitt, T.M., de Pooter, R.F., Gronski, M.A., Cho, S.K., Ohashi, P.S. and Zúñiga-Pflücker J.C. (2004). Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated *in vitro*. *Nature Immunol.* **5**, 410–417.

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

Serrano, M. A., Gómez, M. C., Lopez, M., Dumas, C. L., Smith, K. E., Leibo, S. P., Dresser, B. L. and Pope, C. E. (2006). Derivation of cat embryonic stem-like cells from *in vitro*-produced blastocysts and their support by intraspecific vs. interspecific feeder cells. *Reprod. Fertil. Dev.* **18**, 210.

Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R. and Gearhart, J.D. (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl. Acad. Sci. USA* **95**, 13726–13731.

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

Shtutman, M., Zhurinsky, J., Simcha, I., Albanese, C., D'Amico, M., Pestell, R. and Ben-Ze'ev, A. (1999). The cyclin D1 gene is a target of the β-catenin/LEF-1 pathway. *Proc. Natl Acad. Sci. USA* **96**, 5522–5527.

Shumacher, A., Faust, C. and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* **383**, 250–253.

Singh, A.M., Hamazaki, T., Hankowski, K.E. and Terada N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem cells* **25**, 2534-2542

Singla, D.K., Schneider, D.J., LeWinter, M.M and Sobel, B.E. (2006). Wnt3a but not wnt11 supports self-renewal of embryonic stem cells. *Biochem Biophys Res Commun*. **345**, 789–795.

Sladek, R., Bader, J.A. and Giguère, V. (1997). The orphan nuclear receptor estrogen-related receptor-α is a transcriptional regulator of the human mediumchain acyl coenzyme A dehydrogenase gene. *Mol. Cell. Biol.* **17**, 5400–5409.

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.

Smit, L., Baas, A., Kuipers, J., Korswagen, H., van de Wetering, M. and Clevers, H. (2004). Wnt activates the Tak1/Nemo-like kinase pathway. *J. Biol. Chem*. **279**, 17232–17240.

Smithies, O. (2005). Many little things: one geneticist's view of complex diseases. *Nature Rev. Genet.* **6**, 419–425.

Solomon, M.J., Larsen, P.L. & 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.

Solter, D. (2006). From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat Rev Genet.* **7**, 319-327.

Solter, D. and Knowles, B. B. (1978). Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc. Natl Acad. Sci. USA* **75**, 5565–5569.

Solter, D., Skreb, N. and Damjanov, I. (1970). Extrauterine growth of mouse egg cylinders results in malignant teratoma. *Nature*, **227**, 503-504.

Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz ,J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O. and Thomson, J.A. (2003). Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci USA* **100**, 13350–13355.

Stevens, L.C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev. Biol.* **21**, 364–382.

Stewart, C.L., Gadi, I. and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. *Dev. Biol.* **161**, 626–628.

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

Stewart, T. A. and Mintz, B. (1981). Successive generations of mice produced from an established culture line of euploid teratocarcinoma cells. *Proc. Natl Acad. Sci. USA* **78**, 6314–6318.

Stewart, T.A. and Mintz, B. (1982). Recurrent germ-line transmission of the teratocarcinoma genome from the METT-1 culture line to progeny *in vivo*. *J. Exp. Zool.* **224**, 465–469.

Sucov, H.M., Dyson, E., Gumeringer, C.L., Price, J., Chien, K.R. and Evans, R.M. (1994). RXR mutant mice establish a genetic basis for vitamin A signaling in heart morphogenesis *Genes Dev* **8**, 1007-1018.

Surani, M.A., Hayashi, K. and Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. *Cell* **128**, 747–762.

Svingen T. and Tonissen K.F. (2006). Hox transcription factors and their elusive mammalian gene targets. *Heredity* **97**, 88-96.

Tada, M., Tada, T., Lefebvre L., Barton, S.C. and Surani, M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J.* **16**, 6510-6520.

Tada, T., Tada, M., Hilton, K., Barton, S.C., Sado, T., Takagi, N. and Surani, M.A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. *Dev. Genes Evol.* **207**, 551–561.

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

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.

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.

Thomas, K.R. and Capecchi, M.R. (1986). Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. *Nature* **324**, 34–38.

Thomas, K.R. and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryoderived stem cells. *Cell* **51**, 503–512.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L. and Melton, D.W. (1989). Germ line transmission and expression of a corrected *HPRT* gene produced by gene targeting in embryonic stem cells. *Cell* **56**, 313–321.

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.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A. and Hearn, J.P. (1995). Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. USA* **92**, 7844–7848.

Thomson, J.A. and Marshall, V.S. (1998). Primate embryonic stem cells. *Curr. Top. Dev. Biol.* **38**, 133–165.

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.

Tremblay, G.B., Bergeron, D. and Giguere, V. (2001). 4-Hydroxytamoxifen is an isoform-specific inhibitor of orphan estrogen-receptor-related (ERR) nuclear receptors beta and gamma. *Endocrinology* **142**, 4572–4575.

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

Tremblay, G.B., Tremblay, A., Labrie, F. and Giguère, V. (1999). Dominant activity of activation function 1 (AF-1) and differential stoichiometric requirements for AF-1 and -2 in the estrogen receptor alpha-beta heterodimeric complex. *Mol Cell Biol.* **19**, 1919-1927.

Tsai, M.J. and O'Malley, B.W. (1994). Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu. Rev. Biochem*. **63**, 451–486.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P. and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811-816.

Van Es, J.H., Barker, N. and Clevers, H. (2003).You Wnt some, you lose some: oncogenes in the Wnt signaling pathway. *Curr. Opin. Genet. Dev*. **13**, 28-33.

Vanacker, J.M., Bonnelye, E., Chopin-Delannoy, S., Delmarre, C., Cavaillès, V. and Laudet, V. (1999) Transcriptional activities of the orphan nuclear receptor ERR alpha (estrogen receptor-related receptor- alpha). *Mol. Endocrinol.* **13**, 764–773.

Vanacker, J.M., Delmarre, C. and Delmarre, C. (1998). Activation of the thyroid receptor-α gene promoter by the orphan nuclear receptor ERRα. *Oncogene* **17**, 2429–2435.

Vanacker, J.M., Pettersson, K., Gustafsson, J.A. and Laudet, V. (1999) Transcriptional targets shared by estrogen receptor-related receptors (ERRs) and estrogen receptor (ER) alpha, but not by ERbeta. *EMBO J.* **18**, 4270–4279.

Velculescu, V.E., Vogelstein, B. and Kinzler, K.W. (2000). Analysing uncharted transcriptomes with SAGE. *Trends Genet*. **16**, 423-425.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* **270**, 484–487.

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.

Voncken, J.W., Roelen, B.A., Roefs, M., de Vries., S, Verhoeven, E., Marino, S., Deschamps, J. and van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A.* **100**, 2468-2473.

Wang, J., Mager, J., Schnedier, E. and Magnuson, T. (2002). The mouse PcG gene eed is required for Hox gene repression and extraembryonic development. *Mamm. Genome* **13**, 493–503.

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, S.H., Tsai, M.S., Chiang, M.F. and Li,H. (2003). A novel NK-type homeobox gene, ENK (early embryo specific NK), preferentially expressed in embryonic stem cells. *Gene Expr. Patterns* **3**, 99–103.

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.

Wendling, O., Ghyselinck, N. B., Chambon, P. and Mark, M. (2001) Roles of retinoic acid receptors in early embryonic morphogenesis and hindbrain patterning *Development* **128**, 2031-2038.

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.

Wichterle, H., Lieberam, I., Porter, J.A. and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385–397.

Wilder, P.J., Kelly, D., Brigman, K., Peterson, C.L., Nowling, T., Gao, Q.S., McComb, R.D., Capecchi, M.R. and Rizzino, A. (1997). Inactivation of the FGF-4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny. *Dev Biol.* **192**, 614-629.

Willert, K., Brown, J.D., Danenberg, E., Duncan, A.W., Weissman, I.L., Reya, T., Yates, J.R. 3rd and Nusse, R. (2003). Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448-452.

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.

Wilson, T.E, Fahrner, T.J. and Milbrandt, J. (1993). The orphan receptors NGFI-B and steroidogenic factor-1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. *Mol. Cell. Biol.* **13**, 5794–5804.

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.

Xia, Z., Patiño, R., Gale, W.L., Maule, A.G. and Densmore, L.D. (1999). Cloning, in vitro expression, and novel phylogenetic classification of a channel catfish estrogen receptor. *Gen Comp Endocrinol.* **113**, 360-368.

Xiang, Y., Zhu, Z., Han, G., Lin, H., Xu, L. and Chen, C.D. (2007). JMJD3 is a histone H3K27 demethylase. *Cell Res*. **17**, 850-857.

Xiang, Y., Zhu, Z., Han, G., Ye, X., Xu, B., Peng, Z., Ma, Y., Yu, Y., Lin, H., Chen, A.P. and Chen, C.D. (2007). JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. *Proc Natl Acad Sci U S A.* **104**, 19226-19231

Xie W., Hong H., Yang N.N., Lin R.J., Simon C.M., Stallcup M.R. and Evans R.M. (1999). Constitutive activation of transcription and binding of coactivator by estrogen-related receptors 1 and 2. *Mol. Endocrinol.* **13**, 2151–2162.

Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P. and Thomson, J.A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol.* **20**, 1261–1264.

Yagi, H., Deguchi, K., Aono, A., Tani, Y., Kishimoto, T. and Komori, T. (1998). Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood* **92**, 108–117.

Yamaguchi, K., Nagai, S., Ninomiya-Tsuji, J., Nishita, M., Tamai, K., Irie, K., Ueno, N., Nishida, E., Shibuya, H. and Matsumoto, K. (1999). XIAP, a cellular member of the inhibitor of apoptosis protein family, links the receptors to TAB1-TAK1 in the BMP signaling pathway. *EMBO J*. **18**, 179– 187.

Yamanaka, S., Li, J., Kania, G., Elliott, S., Wersto, R.P., Van Eyk, J., Wobus, A.M. and Boheler, K.R. (2008). Pluripotency of embryonic stem cells. *Cell Tissue Res*. **331**, 5–22.

Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J. and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell*. **125**, 483-495.

Yang, C., Zhou, D. and Chen, S. (1998) Modulation of aromatase expression in the breast tissue by ERR□-1 orphan receptor. *Cancer Res.* **58**, 5695–5700.

Yang, N., Shigeta, H., Shi, H. and Teng, C.T. (1996) Estrogen-related receptor, hERR1, modulates estrogen receptor-mediated response of human lactoferrin gene promoter.*J. Biol. Chem.* **271**, 5795–5804.

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.

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., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T. and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech. Dev*. **45**, 163-171.

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. USA* **93**, 407–411.

Yoshikawa, T., Piao, Y., Zhong, J., Matoba, R., Carter, M.G., Wang, Y., Goldberg, I. and Ko, M.S. (2006). High-throughput screen for genes predominantly expressed in the ICM of mouse blastocysts by whole mount in situ hybridization. *Gene Expr. Patterns* **6**, 213–224.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505–508.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, II. and Thomson, J.A. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318**, 1917-1920.

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.

Zeng, F. and Schultz, R.M. (2003). Gene expression in mouse oocytes and preimplantation embryos: use of suppression subtractive hybridization to identify oocyte- and embryo-specific genes. *Biol Reprod.* **68**, 31–39.

Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L., Ng, H.H., Lufkin, T., Robson, P. and Lim, B. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of *Pou5f1 Nat Cell Biol.* **8**, 1114-1123.

Zhang, K., Siino, J.S., Jones, P.R., Yau, P.M. and Bradbury, E.M. (2004). A mass spectrometric ‘‘Western blot’’ to evaluate the correlations between histone methylation and histone acetylation. *Proteomics* **4**, 3765–3775.

Zhang, Z. and Teng, C.T. (2000) Estrogen receptor-related receptor-1 interacts with coactivator and constitutively activates the estrogen response elements of the human lactoferrin gene. *J. Biol. Chem.* **275**, 20837–20846.

Zhou, Q., Chipperfield, H., Melton, D.A. and Wong, W.H. (2007). A gene regulatory network in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA.* **104**, 16438-16443.

APPENDICES

Appendix 1. The genes whose expression was reduced in the *Nanog*-depleted ES cells and induced in the *Nanog*-overexpressed ES cells.

AffyID	Gene.Descriptions
1419577_at	RIKEN cDNA A530089I17 gene
1452341_at	enoyl Coenzyme A hydratase, short chain, 1, mitochondrial
1426555_at	serine carboxypeptidase 1
1428075_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 4
1426799_at	RAB8B, member RAS oncogene family
1430985_at	RIKEN cDNA 1810027O10 gene
1418737_at	nudix (nucleoside diphosphate linked moiety X)-type motif 2
1455286_at	BTB (POZ) domain containing 1
1423078_a_at	sterol-C4-methyl oxidase-like
1449254_at	secreted phosphoprotein 1
1448562_at	uridine phosphorylase 1
1449682_s_at	RIKEN cDNA 2410129E14 gene
1433953_at	zinc finger protein 277
1429377_at	RIKEN cDNA 2410004A20 gene
1423911_at	protein phosphatase 2, regulatory subunit B (B56), alpha isoform
1417501_at	F-box only protein 6b
1418649_at	EGL nine homolog 3 (C. elegans)
1434917_at	cordon-bleu
1420619_a_at	amino-terminal enhancer of split
1436926_at	estrogen related receptor, beta
1429278_at	Nucleotide binding protein-like
1421307_at	carbonic anhydrase 13
1426865_a_at	neural cell adhesion molecule 1
1460187_at	secreted frizzled-related sequence protein 1 /// RIKEN cDNA 2210415K03 gene
1448445_at	acid phosphatase 6, lysophosphatidic
1449167_at	erythrocyte protein band 4.1-like 4a
1423702_at	H1 histone family, member 0
1417090_at	reticulocalbin 1
1449590_a_at	muscle and microspikes RAS
1448433_a_at	procollagen C-proteinase enhancer protein
1456242_at	LOC433110
1422912_at	bone morphogenetic protein 4

Appendix 2. The genes which are differentially expressed in both *Nanog*-depleted and *Esrrb*-depleted ES cells.

0610009F02Rik	A830025F02Rik	BC011209	Creb3	Fut8
0610039K22Rik	A930006D11	BC021614	Crtac1	Fzd5
0610039N19Rik	AA409316	BC050188	Csad	Gadd45g
1110001J03Rik	AA517132	BC055811	Csrp1	Gata2
1110019C08Rik	AA517471	BC064011	Ctgf	Gata6
1200009I06Rik	AA606869	Bckdha	Cyba	Gbp2
1300002F13Rik	Aard	Bicd2	Cyp2s1	Gdap5
1300013J15Rik	Abcd4	Birc2	Cyr61	Ggt1
1700007E06Rik	Abhd5	Bmp4	D030041H20Rik	Gja1
1700007J06Rik	Acacb	Bnip2	D030074K08Rik	Gjb3
1700110N18Rik	Acas2l	Btbd11	D130027M04Rik	Gjb5
1810009M01Rik	Acp6	C030003H22Rik	D14Ertd725e	Gli2
2010003I19Rik	Acta1	C230006B20	D17Ertd288e	Gmpr
2010309L07Rik	Actc1	C330005L02Rik	D19Ertd678e	Gng10
2010317E24Rik	Adfp	C530044N13Rik	D19Wsu162e	Gprasp2
2010320M18Rik	Aes	C630016B22Rik	D2Bwg1072e	Grb10
2200002D01Rik	Agtrap	C86987	D8Ertd82e	Gsn
2310006J04Rik	AI467606	Cabc1	D9Ertd278e	Gyg1
2310021P13Rik	AI850995	Cacnb3	Dag1	H19
2310031A18Rik	AI854628	Calca	Dhodh	Hand1
2310066E14Rik	AI987986	Cald1	Dhtkd1	Hck
2310076G13Rik	Aire	Calm1	Dkk1	Hk1
2410003B16Rik	Ak3l1	Capn5	Dlx3	Hoxa1
2410004A20Rik	Akt1	Car2	Dnaja2	Hrasls3
2410006F04Rik	AL022780	Car3	Dnajc6	Hspa1b
2410018C20Rik	Aldoc	Car4	Dp1	Hspa5
2410039E07Rik	Amotl2	Card10	Dtr	Hspb2
2410043F08Rik	Ampd3	Ccdc3	Dysf	Icam1
2410076I21Rik	Anxa1	Cd24a	E2f2	Idb2
2410146L05Rik	Anxa3	Cd97	E2f6	Ifi1
2600010E01Rik	Anxa5	Cdc42ep1	E430026E19Rik	Igf2
2700094F01Rik	Apc	Cdkn1c	Efnb1	Ildr1
2810489O06Rik	Aplp1	Cdkn2b	Egln3	Ilk
3110003A17Rik	App	Cdyl2	Ehd2	Inhbb
3110043J09Rik	Arhgdib	Cebpb	Eif4g3	Irx3
3830417A13Rik	Arhgef16	Cebpd	Elmo3	Itpka
4921513H07Rik	Arhgef5	Celsr1	Elovl1	Jam2
4930444M15Rik	Arl7	Cfl2	Ercc5	Kcnk1
4931431C02Rik	Armcx2	Chc1l	Esrrb	Kctd10
4933426M11Rik	Arpc5	Cited2	Esx1	Kdelr3
4933437F05Rik	Atp1a3	Clcn2	Etv4	Kif1c
5031439A09Rik	AU018574	Cldn4	F3	Kit
5033405K12Rik	AU041783	Cldn6	Fabp3	Klf4
5730453H04Rik	AW047273	Cldn7	Fbxo2	Klf5
5730469M10Rik	AW743107	Cln6	Fgf15	Krt1-18
6030408C04Rik	B130017I01Rik	Clu	Fgf4	Krt1-19
6030432P03Rik	B230104P22Rik	Cnn2	Fgfbp1	Krt2-7
6530411B15Rik	B230317C12Rik	Cobl	Fhl1	Krt2-8
9330159N05Rik	B2m	Commd3	FHOS2	Lad1
9330199F22Rik	B430119L13Rik	Copeb	Flnb	Lcn7
9630008K15Rik	B7h3	Cox7a1	Fndc4	Ldh2
A230098A12Rik	Basp1	Cpn1	Fos	Lefty1
A2m	Bbx	Cpt1a	Fstl1	Lgals1

Lgals3	Pgrmc1	Slc27a2	Vim
Limd1	Phlda1	Slc29a1	Vnn1
Lin28	Pitx2	Slc40a1	Wbp5
Liph	Pknox2	Slc4a2	Wdr1
Lmna	Pla2g10	Slmap	Wnt7b
LOC224833	Plac1	Smarcd3	Wwtr1
Loh11cr2a	Plk2	Sorbs1	Yaf2
Lpgat1	Plk3	Sorl1	Zcwcc2
Lphn2	Plod1	Sparc	Zdhhc9
Lpp	Pls3	Spink3	Zfp296
Lrp2	Pmp22	Spp1	Zwint
Lxn	Podxl	Spry4	Zyx
Ly6a	Pou4f-rs1	Stag3	
Ly6g6d	Ppp1r1a	Stard10	
Ly6g6e	Ppp2cz	Stk17b	
Lzic	Prom1	Stmn1	
Manba	Prss11	Stoml1	
Mapre1	Prss8	Surf5	
Mat2a	Psx1	Tacstd2	
Mbnl3	Psx2	Tagln	
Mfn2	Ptk7	Tax1bp3	
Mitf	Rab15	Tcf15	
Mmd	Rab3il1	Tcl1	
Mnab	Rapgef3	Tdgf1	
Mras	Rasgrp2	Tead3	
Mtmr4	Rbms1	Tesk1	
Myh9	Rbp1	Tex27	
Myl7	Rhob	Tgm2	
Myo1c	Rhou	Thbs1	
Myo1f	Rnf128	Thy1	
NA	Rnf28	Tm4sf2	
Ndrg2	Rpp25	Tm4sf5	
Ngfr	Rtkn	Tm7sf3	
Nid2	Ryk	Tnfsf13	
Nos3	Ryr3	Tnnc1	
Notch2	S100a11	Tnnt1	
Notch4	S100a6	Tp53i5	
Nphs1	Sbsn	Tpm1	
Nr0b1	Sec10l1	Tpm2	
Ntn1	Sema4a	Tpm4	
Nupr1	Sema4c	Trib3	
Oat	Sepn1	Trp53inp1	
Otx2	Serpine1	Trps1	
P2y5	Serpine2	Tsrc1	
Pard6b	Sestd1	Tst	
Pard6g	Sfmbt2	Tubb3	
Parva	Sfn	Tubb6	
Pcolce	Sfrp2	Txnip	
Pde4d	Sgce	Unc5b	
Pdlim2	Sh3glb1	Usp28	
Pea15	Slc11a1	Utf1	
Peg3	Slc16a3	Vax2	
Pfkp	Slc25a24	Vegfc	

Appendix 3. ChIP-qPCR validation of the Esrrb ChIP-sequencing library. The column "Label" represents the labeling number in figure 3.14. The columns "Locus" and "Peak height" respectively represent the Esrrb binding site location and the peak heights detected in the Esrrb ChIP-sequencing dataset. Esrrb ChIP-qPCR was used for validation, and GFP ChIP-qPCR was as a negative control. Experimental triplicates were performed for each Esrrb binding site validation. SD stands for standard deviation.

			Fold enrichment detected by ChIP-qPCR			
Label	**Locus**	**Peak height**	**Esrrb ChIP**	**SD**	**GFP ChIP**	**SD**
1	chr11:117826126-117826153	7	1.15	0.14	1.07	0.51
2	chr2:31126363-31126373	7	3.77	0.14	1.57	0.22
3	chr7:132459189-132459212	7	2.82	0.14	1.94	0.97
4	chr5:7987752-7987785	7	2.21	0.43	5.51	3.57
5	chr5:77993531-77993608	8	0.99	0.23	3.38	0.27
6	chr5:125382632-125382655	8	0.67	0.11	1.06	0.23
7	chr10:76610868-76610907	8	0.68	0.17	1.19	0.08
8	chr8:123710916-123710972	9	1.95	0.29	1.38	0.13
9	chr2:166166316-166166337	9	3.10	0.46	1.33	0.59
10	chr6:28591071-28591097	9	3.19	0.31	1.29	0.63
11	chr19:28626979-28626987	9	5.17	0.15	0.78	0.18
12	chr19:53083200-53083262	10	2.16	0.32	0.57	0.23
13	chr12:19534802-19534806	10	8.18	0.21	1.47	0.30
14	chr10:88883293-88883349	10	5.33	1.83	1.44	1.07
15	chr8:117896827-117896838	10	7.59	0.70	1.15	0.51
16	chr7:72988913-72988963	10	9.10	0.73	0.91	0.34
17	chr5:138158568-138158586	10	3.11	0.31	0.95	0.10
18	chr19:7330163-7330189	11	1.78	0.30	1.00	0.09
19	chr18:81136253-81136263	11	2.94	0.43	1.42	0.21
20	chr15:81973941-81973951	11	6.20	0.91	0.96	0.28
21	chr9:90068275-90068276	11	3.09	0.45	1.06	0.12
22	chr6:83103356-83103372	11	2.75	0.45	1.02	0.46
23	chr14:28928851-28928868	12	3.58	0.53	0.88	0.33
24	chr13:63309615-63309638	12	3.76	0.37	0.53	0.27
25	chr11:115977657-115977663	12	2.50	0.37	0.44	0.20
26	chr7:127796118-127796169	12	2.92	0.57	1.27	0.96
27	chr2:60176177-60176205	12	6.47	0.32	1.50	0.09
28	chr11:69930756-69930758	15	4.53	0.44	0.49	0.08
29	chr9:55355242-55355271	15	7.87	1.35	0.94	0.43
30	chr7:109420334-109420347	15	3.72	0.75	0.77	0.34
31	chr6:117642305-117642306	15	8.59	0.42	0.55	0.05
32	chr2:169170815-169170822	15	5.09	0.30	1.16	0.35
33	chr11:88837312-88837315	20	25.22	2.19	1.04	0.93
34	chr10:75087576-75087578	20	1.32	0.74	0.81	0.21
35	chr8:123175600-123175601	20	9.99	0.54	1.09	0.56

36	chr5:126183366-126183376	20	5.59	0.65	0.95	0.13
37	chr3:101570306-101570306	20	10.09	1.49	1.35	0.16
38	chr16:30661422-30661436	25	3.90	0.12	0.65	0.06
39	chr10:33595082-33595083	25	4.18	0.25	1.82	0.89
40	chr5:129597059-129597059	25	4.47	1.27	1.85	0.14
41	chr4:125382110-125382118	25	7.28	0.12	1.59	1.19
42	chr1:136949008-136949013	25	6.61	0.45	1.18	0.34
43	chr18:7627121-7627136	30	4.55	0.28	0.99	0.11
44	chr15:100128398-100128399	30	4.77	0.25	1.11	0.25
45	chr12:103080111-103080115	30	12.22	1.21	0.84	0.13
46	chr10:22396927-22396927	30	2.91	0.18	0.68	0.49
47	chr2:162765063-162765063	30	16.33	1.27	1.20	0.10
48	chr13:65256633-65256634	32	5.07	0.26	0.81	0.24
49	chr7:81702383-81702383	32	5.98	0.33	0.73	0.08
50	chr11:75009646-75009647	33	6.84	0.44	2.73	0.57
51	chr12:105298125-105298125	34	14.94	0.45	0.82	0.04
52	chr5:106116681-106116684	35	6.50	0.68	0.83	0.38
53	chr10:79801820-79801822	36	7.36	0.47	1.30	0.06
54	chr18:80349663-80349664	37	19.13	2.39	0.48	0.06
55	chr10:66491588-66491591	38	13.08	1.75	2.48	0.34
56	chrX:67690979-67690988	39	22.77	2.38	0.99	0.77
57	chr3:88732281-88732284	40	22.29	0.32	1.50	0.45
58	chr9:114650216-114650220	42	9.06	0.42	1.20	0.16
59	chr10:84935339-84935346	43	12.86	0.57	1.04	0.15
60	chr12:102232475-102232484	44	7.16	0.33	1.80	0.48
61	chr5:73130516-73130516	45	6.56	0.84	0.51	0.04
62	chr11:88487368-88487370	46	14.74	1.24	1.16	0.27
63	chr6:118568131-118568135	47	8.02	1.40	0.76	0.09
64	chr13:52234757-52234757	48	23.64	0.31	0.93	0.36
65	chr2:152390344-152390344	49	8.71	0.28	0.84	0.49
66	chr4:117570108-117570112	50	11.83	1.30	1.69	0.50
67	chr5:113570065-113570067	51	9.70	0.81	1.58	0.36
68	chr7:28678589-28678592	52	13.76	2.27	1.46	0.52
69	chr2:179985759-179985759	53	13.02	2.14	0.72	0.42
70	chr12:72055019-72055020	54	17.35	1.21	1.21	0.44
71	chr5:149051358-149051359	55	36.67	1.01	1.19	0.37
72	chr19:61280300-61280300	56	20.29	0.56	1.28	0.63
73	chr1:193419779-193419783	57	12.68	0.46	1.07	0.70
74	chr4:62942115-62942120	58	29.91	0.72	0.89	0.63
75	chr10:77031270-77031270	59	21.31	0.53	0.67	0.09
76	chr2:92225592-92225596	60	37.98	2.87	1.85	0.44
77	chr11:90046146-90046146	61	22.90	1.60	2.07	0.54
78	chr7:101987191-101987193	62	6.77	1.49	0.63	0.07
79	chr2:31293830-31293831	63	19.14	0.43	0.89	0.23
80	chr5:136386836-136386836	64	28.32	0.33	1.02	0.34
81	chr2:167229476-167229477	65	4.86	0.29	1.40	0.14
82	chr7:83779281-83779285	66	41.60	4.75	1.50	0.13
83	chr15:96161593-96161593	67	39.42	0.52	1.14	0.35

84	chr8:123162021-123162021	68	19.95	0.61	0.83	0.04
85	chr14:23086460-23086468	69	46.13	1.00	1.30	0.11
86	chr10:128140554-128140560	70	16.50	0.50	1.25	0.20
87	chr11:60183027-60183033	71	14.27	5.14	1.08	0.58
88	chr6:54628678-54628678	72	30.82	3.01	1.00	0.22
89	chr5:33853498-33853499	73	17.56	0.56	1.47	0.58
90	chr6:91546030-91546033	74	21.39	0.49	1.37	0.82
91	chr10:59673512-59673514	75	20.90	1.09	1.57	0.66
92	chrX:12618787-12618788	76	56.28	6.75	0.94	0.19
93	chr16:6997220-6997221	77	41.83	0.65	0.57	0.17
94	chr2:164834889-164834892	78	16.85	1.21	0.63	0.31
95	chr12:87391144-87391147	148	133.21	10.29	1.54	0.60
96	chr2:5451939-5451940	245	125.57	3.22	0.96	0.26
97	chr9:20765401-20765401	245	89.59	8.88	1.66	0.31
98	chr8:73576520-73576520	254	317.68	56.71	0.77	0.36
99	chr2:19298400-19298400	259	74.26	7.53	0.89	0.27
100	chr18:35169414-35169414	287	126.45	23.10	0.90	0.37

Appendix 4. The genes co-occupied by Esrrb, Oct4 and Nanog

1110012J17Rik	Clstn2	Gm443	Nmnat2	Slc27a4	Zfp462
1700019D03Rik	Cnot10	Gm672	Nmur2	Slc35b4	Zfp533
2610042L04Rik	Col18a1	Gnas	Nrd1	Slc9a9	Zfp64
4732435N03Rik	Cpxm2	Grinl1a	Nrxn1	Smad7	Zic3
4930555I21Rik	Cradd	Gsh2	Odz2	Smarcad1	Zmat4
4931406I20Rik	Crtac1	Gtdc1	Odz4	Snai1	
4933426M11Rik	Csmd1	Gtf2a1	Osm	Sox2	
4933434M16Rik	D030011O10Rik	Hes1	Otx2	Spnb2	
5033405K12Rik	D10Ertd214e	Hk2	Oxr1	Spry2	
9430067K14Rik	D10Ucla1	Hs6st1	Park2	Spry4	
9630008K15Rik	D230005D02Rik	Hsd17b3	Pcaf	Stard13	
A430065P19Rik	D930015E06Rik	Ifitm7	Pdcl2	Sulf1	
A930017N06Rik	D930036F22Rik	Igf2bp1	Pde4d	Suz12	
A930041G11Rik	Dcbld1	Il6st	Pdgfc	Syk	
Abca4	Dkk1	Immp2l	Phc1	Syt13	
Accn1	Dlgap3	Iqgap1	Pigl	Syt9	
Adam19	Dnahc8	Itpr1	Pou5f1	Tbca	
Adcy5	Dok5	Jam2	Ppm1a	Tbl1x	
Adrb3	Dpp6	Jarid2	Ppp2r2d	Tbx3	
AI790205	Dppa5	Kirrel3	Ppp2r5c	Tcf3	
AK037444	Dtnb	L3mbtl3	Ptch1	Tcfcp2l1	
Akap12	Dusp4	Lmnb1	Pum1	Tcte2	
Akap2	Ebf1	Loxl1	Qk	Tdh	
Akr1b8	Egfl6	Lpp	Rage	Tead1	
Arhgap8	Egln3	Lrig1	Ralgps2	Tfrc	
Arhgef18	Ehmt1	Lrig3	Ranbp17	Tkt	
Arid5b	Elovl6	Lrpprc	Rbms3	Tle3	
Atbf1	Enah	Lrrn2	Rbpms	Tle4	
Atp8a2	Eno1	Ly6c	Rest	Tmem17	
AY395631	Eomes	Lyzl1	Rhpn2	Tnfrsf19	
B130055L09Rik	Epha4	Manba	Rif1	Tnfsf11	
B3gnt7	Ephb1	Mapkap1	Riok1	Trim24	
BC022623	Esrrb	Mbip	Ror2	Trp53bp1	
Bmper	Etv5	Mif	Rutbc2	UBE2H	
Btbd11	Eya1	Msi2h	Rybp	Ubqln4	
Btbd9	Fbxo36	Mtf2	Sall1	Ubxd3	
C80913	Fgfr1	Myo10	Sdk1	Uck2	
Camk1d	Foxn2	Myst2	Sec61g	Upp1	
Cbfa2t1h	Frmd4b	Nanog	Sertad2	Upp2	
Cdh5	Ftl1	Ncam1	Sez6	Vti1a	
Cdyl	Fyb	Ndufs4	Sgk	Wdfy3	
Centg2	Fzd5	Negr1	Sipa1l1	Whsc2	
Ches1	Gabbr1	Nfatc2ip	Six4	Wwox	
Chst3	Gja1	Nfib	Slc16a9	Zbtb24	
Clnk	Glis3	Nkx6-1	Slc23a2	Zfp36l1	