The role of Mek/Erk signalling inhibition and Krüppel-like factor 2 in mouse ground state pluripotency

Yeo, Jia Chi

2015

Yeo, J. C. (2015). The role of Mek/Erk signalling inhibition and Krüppel-like factor 2 in mouse ground state pluripotency. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/62942>

<https://doi.org/10.32657/10356/62942>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**The Role of Mek/Erk Signalling Inhibition
and Krüppel-like Factor 2 in Mouse
Ground State Pluripotency**

YEO JIA CHI

B.Sc (Hons)

School of Biological Sciences

2015

**The Role of Mek/Erk Signalling Inhibition
and Krüppel-like Factor 2 in Mouse
Ground State Pluripotency**

YEO JIA CHI

B.Sc (Hons)

School of Biological Sciences

A thesis submitted to the Nanyang Technological University in
partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2015

Acknowledgements

This thesis would not have been possible for the numerous people who have extended their guidance and help throughout these 4 years. Among these, I would firstly like to thank my PhD supervisors Dr Ng Huck Hui, as well as Dr Su I-Hsin, for their invaluable teaching, advice and support in allowing me to explore and cultivate my scientific interests.

I also like to extend my appreciation to Dr Jiang Jianming and Qi Lihua who have been instrumental in getting me up to speed on this project on ground state pluripotency when I first started my PhD. My laboratory colleagues at the Genome Institute of Singapore also played a big role in helping me throughout my research. These people include Dr Winston Chan, Dr Ng Jia Hui, Dr Lu Xinyi, Dr Daniel Yim, Dr Liang Hongqing, Dr Jonathan Goke, Dr Chuti Laowtammathron, Tan Zi Ying, Kevin Gonzales, Friedrich Sachs, Bobby Tan, Lim Yee Siang, and Yaw Lai Ping.

My research collaborators have also played critical roles in my research with their kind assistance with various experiments. My thesis may not have been completed if it were not for their help. I am grateful to Dr Thomas Lufkin and Dr Petra Kraus for generating the *Klf2-LacZ* transgenic mice, as well as for their help in teaching me the various mouse experimental procedures. I also thank Dr Andrew Tan, as well Kelvin Chong for their help in performing the Proximity Ligation Assay (PLA), which helped linked my data coherently. I would also like to thank my thesis advisory committee members Dr Andrew Tan and Dr Newman Sze for their time and advice throughout my PhD.

I would like to thank my parents and friends for their understanding and support for my pursuit in scientific research. I also thank Rachel Lim for walking with me through these 4 years, and I am greatly indebted to her. Lastly, I want to thank God for his strength and blessings in sustaining me. I also eagerly await the exciting things that God has in store for me.

List of Publications

Peer-Reviewed Articles

1. **Yeo JC**, Jiang J, Tan ZY, Yim GR, Ng JH, Göke J, Kraus P, Liang HQ, Gonzales KA, Chong HC, Tan CP, Lim YS, Tan NS, Lufkin T, Ng HH “Klf2 Is an Essential Factor that Sustains Ground State Pluripotency” *Cell Stem Cell* (2014) Jun 5; 14(6) : 864-72
2. Ng JH, Kumar V, Muratani M, Kraus P, **Yeo JC**, Yaw LP, Xue K, Lufkin T, Prabhakar S, Ng HH "In vitro epigenomic profiling of germ cells reveals germ cell molecular signatures." *Dev Cell* (2013) Feb 11 ; 24(3) : 324-33
3. Percharde M, Laval F, Ng JH, Kumar V, Tomaz RA, Martin N, **Yeo JC**, Gil J, Prabhakar S, Ng HH, Parker MG, Azuara V "Ncoa3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming." *Genes Dev* (2012) Oct 15; 26(20):2286-98

Review Articles

4. **Yeo JC**, Ng HH "The transcriptional regulation of pluripotency." *Cell Res* (2013) Jan ; 23(1) : 20-32
5. **Yeo JC**, Ng HH. “Transcriptomic analysis of pluripotent stem cells: insights into health and disease.” *Genome Med.* (2011) Oct 27; 3(10): 68. doi: 10.1186/gm284.

Table of Contents

List of Figures	vi
List of Tables	ix
List of Abbreviations	x
Work performed by collaborators	xiii
Abstract	xiv
1. Introduction	2
1.1 - Mouse embryonic stem cells (mESCs)	3
1.1.1 – Origins and applications of mESCs	3
1.1.2 – The transcriptional regulatory network of mESCs	4
1.1.3 – External signalling and the integration with the ESC transcriptional regulatory core	11
1.2 - Mouse epiblast stem cell (EpiSCs)	14
1.2.1 – Origins and the pluripotential of EpiSCs	14
1.2.2 – The concept of “naïve” and “primed” pluripotency	16
1.2.3 – Interconversion between naïve and primed pluripotent states	18
1.3 – Ground state pluripotency in mESCs	20
1.3.1 – mESC culture via dual Fgf/Mek/Erk and Gsk3/Tcf3 pathway inhibition	20
1.3.2 – Characteristics of 2i ground state cultured mESCs	22
1.3.3 – 2i ground state mESCs resemble E4.5 epiblast cells of the pre-implantation embryo	25
1.4 – Dissecting the mechanism of ground state pluripotency	27

1.4.1 – Gsk3-inhibition alleviated Tcf3 mediated transcriptional repression of Esrrb	27
1.4.2 – The Klf transcription factors	28
1.4.3 – The role of Klf proteins in mESCs and their association with ground state pluripotency	32
1.5 – Objectives of this study	35
2. Materials and Methods	37
2.1 – Cell culture	38
2.2 – Generation of <i>Klf2-LacZ</i> reporter mice	39
2.3 – Embryo collection and mESC derivation	41
2.4 – Proximity ligation assay (PLA)	41
2.5 – Cell cycle and apoptosis detection	42
2.6 – RNAi knockdown	42
2.7 – RNA extraction, reverse transcription and quantitative real-time PCR	43
2.8 – Microarray analysis	45
2.9 – Southern blot and PCR genotyping	46
2.10 – Transfection, Co-immunoprecipitation and Immunoblotting	48
2.11 – Generation of stable Klf2 expressing <i>Klf2</i> -null mESCs	49
2.12 – Recombinant GST-Klf2 expression and <i>in vitro</i> kinase assay	50
2.13 – Chromatin immunoprecipitation (ChIP) and ChIP-seq analysis	51
3. Results	53
3.1 – Klf2 is negatively regulated by Mek/Erk signalling	54

3.1.1 – Gene expression changes of Klf2, Klf4 and Klf5 during ground state pluripotency	54
3.1.2 – Protein level changes of Klf2, Klf4 and Klf5 during ground state pluripotency	56
3.1.3 – Klf2 protein is negatively regulated by the Mek/Erk pathway	58
3.1.4 – Klf2 protein is proteasomally degraded in mESCs	61
3.2 – Klf2 directly interacts with, and is phosphorylated by, Erk2	63
3.2.1 – Bioinformatics analysis of Klf2 reveals potential Erk phosphorylation motifs	63
3.2.2 – Klf2 is phosphorylated in mESCs	65
3.2.3 – Klf2 physically interacts with Erk2 protein	66
3.2.4 – Erk2 is able to phosphorylate Klf2 <i>in vitro</i>	69
3.3 – Klf2 is essential for mESCs under 2i conditions	70
3.3.1 – Generation of Klf2 knockout mice	70
3.3.2 – Derivation of <i>Klf2</i> -null mESCs using LIF+KSR	74
3.3.3 – Functional validation of <i>Klf2</i> -null mESCs	75
3.3.4 – Loss of Klf2 in mESCs results in cell death under 2i	77
3.3.5 – Klf2 overexpression can restore self-renewal in 2i-cultured <i>Klf2</i> -null mESCs	82
3.3.6 – Klf2 overexpression confers independence from Mek-inhibition during ground state pluripotency	83
3.3.7 – Klf2 is important for the establishment of mESCs under 2i conditions	84
3.4 – <i>Klf2</i> -null mESCs exhibit aberrant PGC gene expression under 2i conditions	86

3.4.1 – <i>Klf2</i> -null mESCs retain a pluripotent signature despite massive apoptosis under 2i conditions	86
3.4.2 – <i>Klf2</i> -null mESCs under 2i exhibit aberrant PGC gene expression	88
3.4.3 – Klf2 binds to the promoters of key PGC genes	90
4. Discussion	94
4.1 – Model of Klf regulation in LIF/Serum and 2i mESC conditions	95
4.2 – Varying potencies of Klf2, Klf4 and Klf5 in mESCs	97
4.3 – Post-translational modification of Klf2 by the Mek/Erk signalling pathway	99
4.4 – The role of Klf2 in pre-implantation embryonic development	101
4.5 – The potential role of Klf2 in mESC metabolic regulation	102
4.6 – The association of KLF2 with naïve hESC pluripotency	107
4.7 – Future perspectives and conclusion	112
5. References	115
6. Appendix	129
A – List of the top 200 upregulated genes in <i>Klf2</i> -null mESCs (relative to <i>Klf2</i> -WT) under LIF/Serum, 2i passage 1, 2, and 3	130
B – List of the top 200 downregulated genes in <i>Klf2</i> -null mESCs (relative to <i>Klf2</i> -WT) under LIF/Serum, 2i passage 1, 2, and 3	138
C – Author reprints of articles by Yeo Jia Chi	146

List of Figures

Fig. 1 – The classical ESC Transcriptional Network	7
Fig. 2 – An alternative view: ESC transcription factors as lineage specifiers	10
Fig. 3 – Development of the mouse epiblast from pre-implantation to post-implantation stage	15
Fig. 4 – Molecular differences between mESCs and EpiSCs	17
Fig. 5 – The maintenance of mESC identity using chemical inhibitors	21
Fig. 6 – The molecular features of ground state pluripotent mESCs	24
Fig. 7– Phylogenetic tree of human KLFs	31
Fig. 8 – The transcriptional regulation of Klf2, Klf4 and Klf5 in mESCs	34
Fig. 9 – Gene targeting of <i>Klf2</i> loci	40
Fig. 10 – Gene expression changes of key mESC genes under LIF/Serum or 2i culture	55
Fig. 11 – Western blot for key mESC factors in LIF/Serum or 2i (P3) conditions	56
Fig. 12 – Increase in Klf2 protein levels is independent of Klf2 gene expression	57
Fig. 13 – Increase in Klf2 protein is dependent upon the Mek-inhibitor PD0325901	59
Fig. 14 – Inhibition of serine/threonine phosphatases in mESCs result in rapid Klf2 protein degradation	60
Fig. 15 – Proteasome inhibition results in elevated Klf2 protein levels	62
Fig. 16 – Bioinformatics prediction of Erk phosphorylation sites on Klf2 protein	64
Fig. 17 – Immunoprecipitation of phosphorylated Klf2 from mESCs	66
Fig. 18 – Co-immunoprecipitation (Co-IP) of Klf2 with Erk2	67
Fig. 19 – Erk2 interacts with Klf2 protein in mESCs	68

Fig. 20 – In-vitro [γ - ³² P]-ATP phosphorylation of recombinant GSK-Klf2 using activated Erk2	69
Fig. 21 – Gene expression analysis of 2i-cultured mESCs upon Klf2, Klf4 or Klf5 RNAi knockdown	71
Fig. 22 – Phenotype of <i>Klf2</i> -null mouse embryos	72-73
Fig. 23 – Derivation of <i>Klf2</i> -null mESCs	75
Fig. 24 – Molecular and functional validation of <i>Klf2</i> -null mESCs	76-77
Fig. 25 – Feeder-free <i>Klf2</i> -null mESCs are phenotypically normal in LIF/Serum	78
Fig. 26 – Feeder-free <i>Klf2</i> -null mESCs cannot be maintained under 2i conditions	80
Fig. 27 – Feeder-free <i>Klf2</i> -null mESCs undergo cell death under 2i conditions	81
Fig. 28 – Overexpression of Klf2 rescues <i>Klf2</i> -null mESC cell death under 2i and confers independence from Mek-inhibition	82
Fig. 29 – Abolition of putative Erk phosphorylation sites on Klf2 confers a self-renewal advantage for <i>Klf2</i> -null mESCs in CHIR99021 culture	84
Fig. 30 – Klf2 is an important factor for the establishment of ground state pluripotency	85
Fig. 31 – <i>Klf2</i> -null mESCs retain a pluripotent signature despite undergoing apoptosis in 2i conditions	87
Fig. 32 – <i>Klf2</i> -null mESCs in 2i exhibit elevated PGC gene expression	88
Fig. 33 – <i>Klf2</i> -null mESCs in 2i fail to downregulate PGC expression	90
Fig. 34 – Klf2 binds the promoters of key PGC genes	92
Fig. 35 – Model of Klf2 protein regulation by active Mek/Erk signalling pathways during conventional LIF/Serum or 2i-induced ground state pluripotency	96
Fig. 36 – Protein homology between Klf2, Klf4 and Klf5	98
Fig. 37 – Diagram of the various protein domains in mouse Klf2 protein	100
Fig. 38 – Diagram of the tricarboxylic acid (TCA) cycle during 2i-culture	103

Fig. 39 – GO analysis of <i>Klf2</i> -null mESC microarray data during LIF/Serum or 2i-culture	104
Fig. 40 – ChIP-seq GO analysis of Klf2 bound genes in LIF/Serum vs. 2i-cultured mESCs	105
Fig. 41 – ChIP-seq GO analysis of Klf2, Klf4 and Klf5 bound genes in LIF/Serum cultured mESCs	106
Fig. 42 – Methods to generate naïve-like hESCs	109
Fig. 43 – Model of Klf2 function during ground state pluripotency	114

List of Tables

Table 1 – The tissue expression, function and pathobiology of Klf family proteins	29-30
Table 2 – List of RNAi sequences	42
Table 3 – List of qRT-PCR primers sequences	44
Table 4 – List of antibodies used	48
Table 5 – List of ChIP-qPCR used	52

List of Abbreviations

2i	Two inhibitor (PD0325901 and CHIR99021)
ALK5	Tgf β receptor 1
AP	Alkaline phosphatase
BMP4	Bone morphogenetic protein 4
CalA	Calyculin A
CH	CHIR99021
ChIP-seq	Chromatin immunoprecipitation with massively parallel DNA sequencing
ChIP-qPCR	Chromatin immunoprecipitation with quantitative PCR
Dnmt	DNA methyl-transferase
DMEM	Dulbecco's modified Eagle medium
EGF	Epidermal growth factor
ESCs	Embryonic stem cells
EpiSCs	Epiblast stem cells
Erk	extracellular signal-regulated kinase
Esrrb	Estrogen-related receptor beta
FBS	Fetal bovine serum
Fgf2/FGF2	Fibroblast growth factor 2
Fgf4	Fibroblast growth factor 4
Fgfr	Fibroblast growth factor receptor
GIS	Genome Institute of Singapore
GO	Gene ontology

Gsk3	Glycogen synthase kinase 3
hESC	Human embryonic stem cell
ICM	Inner cell mass
iPSC	Induced pluripotent stem cell
Klf	Krüppel-like factor
KSR	KnockOut serum
LIF	Leukemia inhibitory factor
LSD1	Lysine (K)-specific demethylase 1A
MEF	Mouse embryonic fibroblast
Mek	Mitogen activated protein kinase kinase
mESC	Mouse embryonic stem cell
miRNA	Micro RNA
Nr0b1	Nuclear receptor subfamily 0, group B, member 1
Nr5a2	Nuclear receptor subfamily 5, group A, member 2
NTU	Nanyang Technological University
Oct4	Octamer-binding transcription factor 4
PcG	Polycomb group
PGC	Primordial germ cell
PD	PD0325901
Pdgfra	Alpha-type platelet-derived growth factor receptor
PLA	Proximity ligation assay
Prdm14	PR Domain-Containing Protein 14
PrE	Primitive endoderm

RNAi	RNA-mediated interference
SCF	Stem cell factor
SEM	Standard error of the mean
shRNA	Short hairpin RNA
SILAC	Stable isotopic labeling of amino-acids in cell culture
Sox2	SRY (sex determining region Y)-box2
Stat3	Signal transducer and activator of transcription 3
Tbx3	T-box protein 3
Tcf3	Transcription factor 3
Tgf β	Transforming growth factor beta
Rex1	Reduced expression 1 (Zinc finger protein 42)
XaXa	Double X-chromosome active
XaXi	Single X-chromosome inactivated

Work performed by collaborators

1. Dr Jiang Jianming (Genome Institute of Singapore)
 - Gene targeting of *Klf2* loci via homologous recombination (**Fig. 9a**)
2. Dr Petra Kraus and Dr Thomas Lufkin (Genome Institute of Singapore)
 - Microinjection of *Klf2*- Het and *Klf2*-null mESCs into host blastocysts (**Materials and Methods – 2.2**) and (**Fig. 24e**).
3. Dr Daniel Yim (Genome Institute of Singapore)
 - Co-IP experiments of Klf2 with Erk2 (**Fig. 17** and **Fig. 18**)
4. Mr Bobby Tan (Genome Institute of Singapore)
 - *Klf2*-null mESC teratoma sectioning and staining (**Fig. 24d**)
5. Dr Kelvin Chong and Dr Andrew Tan (Nanyang Technological University)
 - Proximity ligation assay of Klf2 with Erk2 (**Fig. 19** and **Fig. 24g**)
6. Dr Liang Hongqing (Genome Institute of Singapore)
 - FACS analysis of *Klf2*-null mESC cell cycle and apoptotic status (**Fig. 27a** and **Fig. 27c**)
7. Dr Ng Jia Hui (Genome Institute of Singapore)
 - Klf2 ChIP-qPCR and ChIP-seq in mESCs (**Fig. 34a-b**)
8. Dr Jonathan Goke (Genome Institute of Singapore)
 - Klf2 ChIP-seq mapping and analysis (**Fig. 34a**)

Abstract

The maintenance of undifferentiated mouse embryonic stem cells (mESCs) requires the presence of LIF and serum. Interestingly, by using two chemical molecules to inhibit both the pro-differentiative Fgf/Mek/Erk and Gsk3/Tcf3 pathways in mESCs (dual inhibition or “2i”), a pluripotent “ground state”, resembling the mouse pre-implantation epiblast can be established without the need for LIF and serum. While Gsk3-inhibition is known to alleviate Tcf3-mediated repression of *Esrrb*, the molecular mechanism downstream of Mek/Erk inhibition remains to be identified.

Here, it was uncovered that Erk2 phosphorylates the Krüppel-like factor 2 (Klf2), leading to Klf2 proteasomal degradation. Mek/Erk inhibition during 2i conditions thus serves to halt Klf2 phospho-degradation, leading to Klf2 protein stabilisation and maintenance of ground state pluripotency. Indeed, while *Klf2*-null mESCs are viable under LIF/Serum, they undergo apoptosis during 2i culture. Additionally, it was found that Klf2 overexpression is sufficient to replace Mek-inhibition, allowing for mESC self-renewal under Gsk3-inhibition alone.

Taken together, this study highlights the importance of Klf2 during 2i conditions, and defines the Mek/Erk/Klf2 pathway with the Gsk3/Tcf3/*Esrrb* axis to establish ground state pluripotency.

CHAPTER I

INTRODUCTION

Chapter I: Introduction

The two defining features of embryonic stem cells (ESCs) are their ability to self-renew, as well as give rise to all cell lineages of the organism, including germ cells. These unique characteristics of ESCs to form all adult cell types, termed 'pluripotency', besides enabling the *in vitro* study of early mammalian development, has also greatly facilitated the development of regenerative medicine. Specifically, the differentiation of ESCs could be exploited to generate clinically relevant cells for the replacement of worn out tissue. However, this same remarkable plasticity of ESCs to form various tissues, also poses a significant challenge in efficiently controlling ESCs differentiation into the desired cell type of choice. The harnessing of ESCs for therapeutic purposes therefore necessitates a detailed understanding of the factors and mechanisms that sustain ESC pluripotency, as well as those that specify cell lineage commitment.

1.1 – Mouse embryonic stem cells (mESCs)

1.1.1 – Origins and applications of mESCs

mESCs were first isolated from pre-implantation blastocysts by two independent groups in 1981 (Evans and Kaufman, 1981; Martin, 1981). The origins of these pluripotent mESCs were later traced to the transient epiblast compartment of the blastocyst inner cell mass (ICM) (Brook and Gardner, 1997). Alternatively, mESC lines could also be generated from the 8-cell stage embryos (Delhaise *et al.*, 1996), morula stage embryos (Eistetter, 1989), and interestingly from individual blastomeres of the developing embryos (Chung *et al.*, 2006; Wakayama *et al.*, 2007). Importantly, when reintroduced into host blastocysts, these pluripotent mESCs were able to integrate with the host ICM and contribute extensively to the formation of the adult mouse, resulting in the formation of germ-line chimeric mice (Bradley *et al.*, 1984).

When coupled with genomic alteration strategies, the germ-line transmitting ability of mESCs has made the targeted manipulation of the mouse genome possible (Gossler *et al.*, 1986; Robertson *et al.*, 1986). This scientific breakthrough in mESCs and genome editing has greatly facilitated the engineering of various transgenic mice. In the process, transgenic mouse models have revolutionized our ability to study gene function and regulation in mammalian biology (Reviewed by Capecchi, 2005).

1.1.2 – The transcriptional regulatory network of mESCs

The transcriptome is described as the collection of all expressed RNA molecules within a cell at a particular state (Wang *et al.*, 2009). As such, understanding the mESC transcriptional circuitry is an important step towards dissecting the genetic regulatory mechanism of pluripotency and lineage commitment. In ESCs, the pluripotent state is maintained by the core transcription factor trio of Oct4, Sox2 and Nanog (Reviewed by Young, 2011).

Oct4 (encoded by the *Pou5f1* gene) is expressed exclusively within the mouse blastomeres, the pluripotent epiblast cells, as well as in primordial germ cells (PGCs) (Rosner *et al.*, 1990; Scholer *et al.*, 1990). Importantly, Oct4 is critical for the mouse embryonic epiblast development as *Pou5f1*-null ICM differentiates into trophectoderm lineages (Nichols *et al.*, 1998). Similarly, Oct4 is important for mESC pluripotency, and as such, the levels of Oct4 in mESCs are tightly regulated. For instance, a 50% decrease in Oct4 gene expression would result in the aberrant differentiation of mESCs into trophectoderm. Conversely, overexpression of Oct4 by 50% will promote mESC differentiation into mesodermal and endodermal lineages (Niwa *et al.*, 2000).

Sox2, which possesses the highly conserved high-mobility group box DNA-binding domain, is similarly expressed within the ICM, the epiblast, and the extraembryonic ectoderm of E6.5 mouse embryos (Avilion *et al.*, 2003). Similar to Oct4 knockout embryos, the absence of Sox2 in blastocysts results in the failure to form a pluripotent ICM (Avilion *et al.*, 2003). In addition, *Sox2*-null mESCs will differentiate primarily into trophoctoderm lineages (Masui *et al.*, 2007). As Oct4 and Sox2 are known to synergise and regulate various ESC genes (Nakatake *et al.*, 2006; Nishimoto *et al.*, 1999; Rodda *et al.*, 2005; Tokuzawa *et al.*, 2003; Yuan *et al.*, 1995), it is thus not surprising that Oct4 overexpression is sufficient to rescue the trophoctoderm differentiation phenotype of *Sox2*-null mESCs.

Nanog, the third core ESC transcription factor, was initially discovered through a screen for pluripotency factors that replaced leukaemia inhibitor factor (LIF) (Chambers *et al.*, 2003; Mitsui *et al.*, 2003). While *Nanog*^{-/-} mESCs cannot be directly established from *Nanog*-null embryos due to the absence of a pluripotent ICM (Silva *et al.*, 2009), *Nanog*^{-/-} mESCs could nevertheless still be established via *Nanog* gene knockout in wildtype mESCs (Chambers *et al.*, 2007; Mitsui *et al.*, 2003). Therefore, it is hypothesized that although Nanog is important for the acquisition of pluripotency, Nanog is dispensable once pluripotency is achieved (Chambers *et al.*, 2007).

Given the important roles of these core transcription factors in ESCs, subsequent chromatin immunoprecipitation (ChIP) studies sought to map the genomic binding sites of these transcription factors in both mouse and human ESCs (hESCs) (Boyer *et al.*, 2005; Loh *et al.*, 2006). Together, these studies found that the core ESC transcription factors are able to sustain pluripotency through: 1) the activation of pluripotency-associated factors while simultaneously suppressing lineage-specific genes; and 2) through the stable autoregulatory Oct4, Sox2 and Nanog network to sustain their expression (Young, 2011) (**Fig. 1**). This model may suggest why ESCs can sustain self-renewal and pluripotency, while still remaining poised to differentiate when the signal to shutdown the core ESC transcription network is given.

Figure 1. The classical ESC Transcriptional Network. The core ESC transcription factors Oct4, Sox2 and Nanog form an autoregulatory circuit by regulating their own expression, and that of each other. Together, these three factors maintain an overall ESC transcriptional profile by activating (a) ESC-related transcription factors; (b) signalling transcription factors; (c) chromatin modifiers; and (d) ESC-associated microRNAs (miRNAs). On the other hand, Oct4, Sox2 and Nanog, possibly acting together with Polycomb group proteins (PcG), suppress lineage-specific genes to inhibit ESC differentiation. Figure is obtained from Young., 2011.

In recent years, there has been an alternative model proposed to explain how pluripotency is maintained in ESCs. Unlike the prevailing model in which pluripotency factors serve to promote self-renewal and inhibit differentiation (Young, 2011); this alternative concept suggests that pluripotency factors are actually lineage specifiers that direct or inhibit differentiation into certain tissue lineages (**Fig. 2a**) (Reviewed by Loh and Lim, 2011; Loh et al., 2015). Indeed, this idea is not completely foreign. For instance, it is known that the 50% overexpression of Oct4 can cause mESC differentiation into mesoderm and endoderm lineages (Niwa et al., 2000). Likewise, Sox2 overexpression reportedly induces ectoderm differentiation in mESCs (Kopp et al., 2008); while Nanog overexpression can enhance primitive streak differentiation, as well as inhibit hESC differentiation into ectoderm (Teo et al., 2011; Vallier et al., 2009; Wang et al., 2012).

As such, due to the multiple expression of various ESC “lineage” factors, ESCs hence have the potential to differentiate into any tissue lineage (and thus maintaining pluripotency). More importantly, this new model suggests that the pluripotent state is actually the delicate balance of competing differentiation signals in a “transcription factor stalemate”; whereby no single lineage dominates (**Fig. 2**) (Loh and Lim, 2011; Loh et al., 2015; Shu and Deng, 2013).

Experimentally, this novel idea has been partially validated in two studies that utilized lineage specifiers to reprogram somatic cells into induced pluripotent stem cells (iPSCs) (Montserrat et al., 2013; Shu et al., 2013). In mice, Shu *et al.* for instance have found that Gata3, and other mesendoderm factors, could replace for

Oct4 in reprogramming (Shu et al., 2013). This was presumably done by the counteracting of Sox2-induced ectodermal genes. Indeed, RNAi depletion of the master ectoderm-related *Dlx3* gene could mimic the overexpression of *Oct4* or *Gata3*. Likewise, Shu *et al.* also found that the inhibitory effect of Sox2 on mesendodermal differentiation could be replaced by the ectoderm specifier *GMNN* (Shu et al., 2013).

Similarly in human iPSC generation, Montserrat *et al.* found that GATA3 could replace for exogenous OCT4, and that the ectodermal specifier *ZNF521* could replace for SOX2 during reprogramming (Montserrat et al., 2013). Interestingly, it was observed that knockdown of *GATA3* in human iPSCs resulted in cell death. This suggests that GATA3 could have previously unidentified roles in the maintenance of human ESC pluripotency (Montserrat et al., 2013).

Despite the remarkable ability of lineage specifiers to induce pluripotency (Montserrat et al., 2013; Shu et al., 2013), there are several questions that remain to be addressed. Firstly, while the balance of various lineage-specifiers can explain the ability of pluripotent ESCs to differentiate into all tissue lineages; it does not explain how a differentiated somatic cell is able to revert into pluripotency (i.e. iPSCs reprogramming). Secondly, it is also unclear why only some, but not all lineage factors, could replace for the classical Yamanaka reprogramming factors (Shu et al., 2013). This leads to the third question: could it be possible that lineage-factors have previously unknown functions/effects on signaling or epigenetic pathways that may directly or inadvertently reactivate the classical pluripotency regulators?

While this “precarious-balance” or “see-saw” model may present a novel perspective for ESC transcriptional regulation (**Fig. 2**) (Loh and Lim, 2011; Loh et al., 2015; Montserrat et al., 2013; Shu et al., 2013), it is still early days to determine if this is indeed a valid model to explain how pluripotency is maintained.

Figure 2. An alternative view: ESC transcription factors as lineage specifiers. a) In this alternative view of ESC transcriptional regulation, a variety of transcription factors drive differentiation towards a specific lineage, while repressing those of others. Green arrow indicates lineage induction, while the red arrow indicates lineage repression. Overall, this simultaneous expression and repression in ESCs leads to a “transcription factor stalemate” or “precarious balance”, in which no tissue lineage dominates and a pluripotent ESC state is established (Loh et al., 2015). Figure is obtained from the review by Loh et al., 2015. **b)** Another model for the interpretation of ESC transcription factors as lineage specifiers. In this “see-saw” model, the pluripotent state is achieved by the balancing of two opposing lineage differentiation signals. ME: Mesendoderm; ECT: Ectoderm. Figure is obtained from the review by Shu and Deng, (Shu and Deng, 2013)

1.1.3 – External signalling and the integration with the ESC transcriptional regulatory core

Traditionally, mESCs were cultured and kept undifferentiated on a layer of mitotically inactivated feeder cells in serum-supplemented media (Evans and Kaufman, 1981; Martin, 1981). However, given the undefined nature of feeder cells and serum, it was a significant challenge to identify and study the various extrinsic signalling factors and cellular pathways that control mESC pluripotency and differentiation.

Subsequent work over the next two decades identified the IL-6 family cytokine LIF as the key active component produced by feeder cells which sustains mESC growth (Smith *et al.*, 1988; Williams *et al.*, 1988), and that the BMP4 molecule could substitute for the use of serum in mESC culture (Ying *et al.*, 2003). More importantly, the combinatorial usage of LIF with BMP4 in a defined feeder and serum-free culture delineated the pathways that were responsible for mESC self-renewal and pluripotency (Ying *et al.*, 2003).

Together, LIF and BMP4 were found to exert their physiological effect via the respective phosphorylation of downstream transcription factors Stat3 (Niwa *et al.*, 1998) and Smad1 (Ying *et al.*, 2003) in mESCs. More importantly, genome-wide transcription factor mapping studies via ChIP with massively-parallel sequencing (ChIP-seq), revealed Stat3 and Smad1 colocalisation with Oct4, Sox2 and Nanog binding sites (Chen *et al.*, 2008). This study by Chen *et al.* thus demonstrates how

these two key mESC signalling pathways are integrated into the core transcriptional network to sustain pluripotency.

In contrast to the self-renewing effect of LIF and BMP4, the autocrine production of fibroblast growth factor 4 (Fgf4) in mESCs (Ma *et al.*, 1992; Rathjen *et al.*, 1999), acting via the Mek/Erk signalling pathway, is a known inducer of mESC lineage commitment (Kunath *et al.*, 2007; Stavridis *et al.*, 2007). This pro-differentiation Fgf/Mek/Erk signalling effect can be further observed through the impaired mESC differentiation when Fgf/Mek/Erk signalling components were either chemically or genetically disrupted (Burdon *et al.*, 1999; Chen *et al.*, 2006; Cheng *et al.*, 1998; Kunath *et al.*, 2007; Stavridis *et al.*, 2007).

Although it is clear that active Mek/Erk signalling primes mESCs for differentiation (Kunath *et al.*, 2007), the cellular mechanism by which this occurs is still not well understood. Through studies in other cell types, Erk is known to regulate gene expression via various mechanisms such as transcription factor phosphorylation (Johnson and Lapadat, 2002; Ramos, 2008), chromatin remodelling complex phosphorylation (Simone *et al.*, 2004), as well as direct DNA binding as a transcriptional repressor (Hu *et al.*, 2009). Therefore, it may be possible that Erk could adopt similar regulatory mechanisms towards the priming of mESCs differentiation. Indeed, recent data found that active Erk signalling was responsible for lineage priming in mESCs through the establishment of poised RNA polymerase II transcription complexes at developmental genes (Tee *et al.*, 2014). However, it is

possible that other forms of Erk-mediated cellular regulation in mESCs remains to be uncovered.

Lastly, mESCs are also responsive towards Wnt signalling, and Wnt stimulation in mESCs is necessary to prevent their differentiation into the developmentally advanced mouse epiblast stem cell (EpiSCs) (ten Berge *et al.*, 2011). During active canonical Wnt signalling, β -catenin is stabilized, allowing for the association of β -catenin with the TCF/LEF family for gene activation (Sokol, 2011). In mESCs, Tcf3 is the most abundant Tcf protein (Pereira *et al.*, 2006), and Tcf3 genomic binding sites are known to be integrated with the core Oct4, Sox2 and Nanog ESC transcriptional circuitry (Cole *et al.*, 2008). Because Tcf3 functions primarily as a pluripotency transcriptional repressor (Pereira *et al.*, 2006; Tam *et al.*, 2008; Yi *et al.*, 2008), *Tcf3*-null mESCs is known to exhibit impaired differentiation (Guo *et al.*, 2011; Pereira *et al.*, 2006). Unfortunately, how exactly does β -catenin control Tcf3 gene regulation is still unclear (Sokol, 2011). Therefore, further studies are clearly necessary to better delineate the Wnt-signalling pathway in mESCs.

1.2 – Mouse epiblast stem cells (EpiSCs)

1.2.1 – Origins and the pluripotential of EpiSCs

After implantation, the mouse epiblast will transform into a single-cell epithelial layer located within the egg cylinder (**Fig. 3**). Depending on its location, the epiblast cells will experience different lineage specification inductive signals from the adjacent tissue (Beddington and Robertson, 1999). For female embryos at this post-implantation stage, the egg cylinder epiblast cells will undergo random X-chromosome inactivation to form a XaXi status (Heard, 2004). Notably, while the post-implantation epiblast has been specified for differentiation, transplant studies between egg cylinders suggests that this lineage fate has not been fixed (Gardner and Beddington, 1988).

Interestingly, a LIF-independent pluripotent stem cell could also be isolated from the post-implantation mouse epiblast (Brons *et al.*, 2007; Tesar *et al.*, 2007). Termed as “epiblast stem cells” (EpiSCs) due to their tissue of origin, these self-renewing EpiSCs exhibit the ability to differentiate into all three embryonic germ layers in teratomas (Brons *et al.*, 2007; Tesar *et al.*, 2007), as well as germ cells *in vitro* (Hayashi and Surani, 2009). Despite this pluripotent nature, EpiSCs have limited developmental potential, and are considered inefficient in forming mouse chimeras (Brons *et al.*, 2007; Guo *et al.*, 2009; Tesar *et al.*, 2007).

This poor chimerism may however be overcome by the forced expression of *E-cadherin* in EpiSCs to allow for their integration and co-development with the mouse

embryo (Ohtsuka et al., 2012). While *E-cadherin* is important for mESC pluripotency and is able to replace Oct4 during iPSC reprogramming (Redmer et al., 2011), EpiSCs expressing *E-cadherin* do not exhibit obvious signs of reprogramming or conversion into a mESC-like state (Ohtsuka et al., 2012). As such, the exact mechanism in which *E-cadherin* allows EpiSCs to contribute towards mouse chimerism remains unclear.

Figure 3. Development of the mouse epiblast from pre-implantation to post-implantation stage. mESCs are derived from the epiblast of pre-implantation blastocysts, while EpiSCs are isolated from the epiblast of post-implantation embryos. Figure is obtained from the review by Blair *et al.*, 2011.

1.2.2 – The concept of “naïve” and “primed” pluripotency

Despite expressing the mESC pluripotency factors Oct4, Sox2 and Nanog, EpiSCs nevertheless differ greatly from mESCs with regards to their physical features, overall transcriptome, epigenetic status, and requirement for exogenous factors to maintain a stem cell identity (**Fig. 4**) (Bao *et al.*, 2009; Brons *et al.*, 2007; Hayashi *et al.*, 2008; Sun *et al.*, 2012; Tesar *et al.*, 2007).

The most notable of these differences is that EpiSCs require exogenous Activin and Fgf signalling for self-renewal; unlike mESCs which need LIF and BMP4 (Reviewed by Hanna *et al.*, 2010b). It is also known that EpiSCs exhibit reduced levels of pluripotency genes such as *Rex1*, *Stella* and the *Klfs*, with a higher expression of lineage differentiation markers (Tesar *et al.*, 2007). Lastly, while female mESCs exist in the pre X-inactivation state with two active X-chromosomes (XaXa), EpiSCs have already undergone X-inactivation (XaXi) resembling a more developmentally advanced stage (**Fig. 4**) (Hanna *et al.*, 2010b).

These data thus indicate that EpiSCs are more developmentally advanced than mESCs, and are in a “primed” pluripotent state. On the contrary, pluripotent mESCs which resembles the pre-implantation mouse epiblast are considered to be developmentally “naïve” (Nichols and Smith, 2009).

Characteristic	Naive Mouse Pluripotent State	Primed Mouse Pluripotent State
Alternative names	Preimplantation inner cell mass-like; mouse ESC-like	Post-implantation epiblast-like; mouse epiblast stem cell-like
Stem cell types	Embryonic stem cells (ESCs); embryonic germ cells (EGCs); spermatogonial germ stem cell (maGSC); induced pluripotent stem cell (iPSC)	Epiblast stem cell (EpiSCs)
Developmental Potential		
Embryonic body formation	Yes	Yes
Teratoma	Yes	Yes
Blastocyst chimera	Yes	No
Susceptibility for primordial germ cell specification	Low	High
Growth properties		
Single-cell clonogenicity	High	Low
Ability to grow independent of feeders and/or in defined medium	Yes	Yes
Doubling time in vitro	10–14 hr	14–16 hr
Morphology	Domed	Flattened
Regulation by exogenous signaling pathways		
Positive regulators of the state	LIF/Stat3 BMP4 WNT IGF	TGF- β Activin FGF2 ERK1/2 WNT IGF
Negative regulators of the state	TGF- β Activin FGF2 ERK1/2	BMP4
Gene and marker expression signatures		
SSEA1, alkaline phosphatase	+	+
TRA1-60, TRA1-81, SSEA3/4	—	—
Oct4, Sox2	+	+
Nanog, Klf2, Klf4, Rex1, Stella	High ++	Low/absent +/-
Lineage specification markers (e.g., FGF5, Blimp1, Cer1)	Absent	Positive with heterogeneous expression pattern
MHC class I	Nearly absent	Present
Oscillatory gene patterns	Nanog, Rex1	Blimp1
Epigenetic state		
Oct4 enhancer activity	Distal element	Proximal element
XX status	XaXa	XaXi

Figure 4. Molecular differences between mESCs and EpiSCs. Figure is obtained from the review by Hanna *et al.*, 2010b.

On a separate note, hESCs which are derived from human pre-implantation blastocysts (Thomson *et al.*, 1998), interestingly resemble primed EpiSCs rather than their naive mESCs developmental counterpart (Brons *et al.*, 2007; Tesar *et al.*, 2007). This is because hESCs, like EpiSCs require FGF2 and Activin A stimulation to maintain pluripotency (Vallier *et al.*, 2005), and do not require LIF signaling

(Daheron *et al.*, 2004). This finding indicates that during the hESC derivation process, the cells could have progressed developmentally to resemble the primed EpiSC-like state; highlighting developmental differences between human and rodent ESCs (Reviewed by De Los Angeles *et al.*, 2012).

1.2.3 – Inter-conversion between naive and primed pluripotent states

Because mESCs are derived from a developmentally earlier timepoint (pre-implantation stage), as compared to EpiSCs (post-implantation stage), the transition of mESC into an EpiSC would correspond to a normal differentiation step along the developmental program. Indeed, EpiSCs can be easily obtained through mESC differentiation via Fgf2 and Activin A stimulation (Guo *et al.*, 2009), or through Fgf2 and LIF/Stat3 inhibition (Greber *et al.*, 2010). Similarly, using EpiSC culture techniques, it is also possible to derive EpiSCs from mouse pre-implantation blastocysts (Najm *et al.*, 2011). This indicates that the signalling environment plays a strong influence towards the formation of naive and primed pluripotent states.

In contrast, the reversal of primed EpiSCs into naive mESCs occurs at an extremely low rate simply through the switching into LIF/Serum culture media (1 in less than 1×10^6 cells) (Guo *et al.*, 2009; Hanna *et al.*, 2009). This conversion rate however may be drastically improved through the exogenous expression of pluripotency factors Klf2, Klf4, Nanog, Nr5a2 or Esrrb in EpiSCs (Festuccia *et al.*, 2012; Guo *et al.*, 2009; Hall *et al.*, 2009; Hanna *et al.*, 2009; Silva *et al.*, 2009). A chemically defined

approach to convert primed to naive-like mouse pluripotency is also possible via the simultaneous inhibition of LSD1, ALK5, MEK, FGFR and GSK3 (Zhou *et al.*, 2010)

Interestingly, the EpiSC into mESC transition can be further shortened into 3-4 days through the synergistic overexpression of the Prdm14 and Klf2 transcription factors (Gillich *et al.*, 2012). This is in contrast to the usual process of 8-10 days using single factors. By itself, the primordial germ cell factor Prdm14 has no significant impact of the reversion of primed to naive pluripotency. However, it was found through transcriptomic analysis that Prdm14 functions to prepare EpiSC conversion through the simultaneous repression of lineage genes, as well as the activation of early epiblast genes. In addition, Prdm14 is also able to enhance Klf2 recruitment onto mESC genes loci such as *Nr5a2* and *Oct4* to rapidly drive the EpiSC into mESC conversion (Gillich *et al.*, 2012).

1.3 – Ground-state pluripotency and mESCs

1.3.1 – mESC culture via dual Fgf/Mek/Erk and Gsk3/Tcf3 pathway inhibition

As discussed in Section 1.1.3 – “External signalling and the integration with the ESC transcriptional regulatory core”, the culture of mESCs conventionally requires the presence of LIF and serum/BMP4 for proper self-renewal and maintenance of pluripotency. However, even during LIF/Serum culture, the autocrine pro-differentiation Fgf4/Mek/Erk signalling in mESCs remains active (Ying *et al.*, 2003; Ying *et al.*, 2008). This suggests that LIF/Serum serves to sustain mESC pluripotency downstream of the Fgf/Mek/Erk signalling pathway to halt mESC differentiation (**Fig. 5a**).

Following up with this knowledge, Ying *et al.* surprisingly found that this requirement for exogenous LIF/Serum in mESCs could be dispensable if small molecular kinase inhibitors were used (**Fig. 5b**) (Ying *et al.*, 2008). This dual “2i” inhibitor treatment regime, consisting of PD0325901 and CHIR99021 to respectively target the Mek and Gsk3 kinases, serve to block the pro-differentiative effects of Fgf4 autocrine stimulation (Kunath *et al.*, 2007), as well as inhibit the transcriptional repression of pluripotent genes by Tcf3 (Wray *et al.*, 2011). Most importantly, 2i culture allowed for the isolation of germ-line transmitting mESCs from previously recalcitrant CBA, MF1 and NOD mouse strains (Hanna *et al.*, 2009; Ying *et al.*, 2008), as well as the cross-species germline transmitting rat ESCs (Buehr *et al.*, 2008; Li *et al.*, 2008).

Figure 5. The maintenance of mESC identity using chemical inhibitors. (a) In conventional mESC culture, LIF and serum/BMP4 serve to block mESC differentiation downstream of the pro-differentiation Fgf4/Mek/Erk signalling pathway. (b) mESCs can be maintained without exogenous LIF and serum/BMP4 using a triple inhibitor cocktail comprising FGF-inhibitor (SU5402), Mek-inhibitor (PD184352) with Gsk3-inhibitor (CHIR99021) (“3i” culture). Similarly, mESCs can also be cultured in a double inhibitor mixture consisting of a more potent Mek-inhibitor (PD0325901) with Gsk3-inhibitor (CHIR99021) (“2i culture”). Figure adapted from Ying *et al.*, 2008.

1.3.2 – Characteristics of 2i ground state cultured mESCs

During the culture of mESCs under LIF/Serum conditions, the expression of pluripotency genes such as *Nanog*, *Rex1* and *Stella* are heterogeneous among cells (Chambers *et al.*, 2007; Hayashi *et al.*, 2008; Toyooka *et al.*, 2008). Even after FACS-purification to enrich for either a homogenously high or homogeneously low expressing cell population, these sorted mESCs would again re-establish a heterogeneous *Nanog*, *Rex1* and *Stella* gene expression. This finding indicates that gene expression in LIF/Serum mESCs exists in a dynamic fluctuating state.

Interestingly upon the transfer into 2i culture, *Nanog* and *Rex1* levels in mESCs become homogeneously-high (**Fig. 6a**) (Wray *et al.*, 2010). Since *Nanog*-low cells are easily prone to differentiation, and that *Rex1*-low mESCs are unable to contribute towards chimera formation (Chambers *et al.*, 2007; Toyooka *et al.*, 2008), the ability to establish a *Nanog/Rex1*-high mESC state indicates that mESC pluripotency has been stabilised under 2i conditions (Wray *et al.*, 2010). Interestingly, the *Nanog* gene loci in mESCs switch from a monoallelic into a biallelic expression during 2i culture, suggesting a plausible explanation for the homogenously high levels of *Nanog* (Miyazari and Torres-Padilla, 2012). Because of the unique features between LIF/Serum and 2i-cultured mESCs, mESCs under 2i conditions are said to be in a novel and distinct pluripotent “ground state” (Ying *et al.*, 2008).

Through the detailed transcriptomic and epigenomic analysis of 2i-cultured mESCs, Marks *et al.* found that in addition to the homogenous levels of pluripotency factors,

the transient expression of lineage-specific genes in LIF/Serum mESCs were repressed under 2i (Marks *et al.*, 2012). In addition, evidence of a stable mESC pluripotent state can be further seen through the reduced levels of poised bivalent domains, and the increased occurrences of RNA Pol II transcriptional pausing at lineage-specific genes during 2i ground state mESC culture (**Fig. 6b**) (Marks *et al.*, 2012).

At the DNA methylation level, a globally hypomethylated state in mESCs is established during 2i culture (Ficz *et al.*, 2013; Habibi *et al.*, 2013; Leitch *et al.*, 2013). While the exact details of this DNA hypomethylation process are unclear, it is postulated that DNA demethylation could have resulted from the downregulation of *Dnmt3a/b/l* genes under 2i (Leitch *et al.*, 2013). Likewise, 2i could also enhance Tet1/2 activity to drive active DNA demethylation (Ficz *et al.*, 2013; Habibi *et al.*, 2013). However, the most interesting feature of these 2i ground state mESCs is that their DNA hypomethylation pattern resembles the hypomethylated state of the *in vivo* E3.5 ICMs or E9.5 migratory PGCs (Ficz *et al.*, 2013; Habibi *et al.*, 2013). In contrast to 2i, bisulphite-sequencing of LIF/Serum cultured mESCs revealed the close association with the DNA methylation state of post-implantation E6.5 embryos (Habibi *et al.*, 2013).

Figure 6. The molecular features of ground state pluripotent mESCs. (a) Using a *Nanog*-GFP and *Rex1*-GFP reporter mESC cultured under LIF/Serum, it is observed a range of GFP-hi and GFP-lo cell populations exists. This reflects the heterogeneous nature of mESC pluripotency in LIF/Serum. However, when the same *Nanog*-GFP and *Rex1*-GFP reporter mESCs were cultured under 2i ground state conditions, all the cells would exhibit a homogeneously high level of pluripotency gene expression. Figure adapted from Wray *et al.*, 2010. **(b)** 2i ground state mESCs exist in a stable pluripotent state, marked with reduced lineage-specific gene priming, reduced levels of developmentally “poised” bivalent domains, as well as increased levels of RNA Pol II transcriptional pausing at lineage-specific genes. Figure adapted from Marks *et al.*, 2012.

1.3.3 – 2i ground state mESCs resemble E4.5 epiblast cells of the pre-implantation embryo

While mESCs have been traditionally isolated from the ICM of pre-implantation embryos, the origins of these pluripotent cells have been controversial. Specifically, it is being debated whether mESCs are a cell-culture induced adaptation with no *in vivo* developmental counterpart. The origins of mESCs were further complicated by the knowledge that a developmentally primed EpiSC can be converted into a naive mESC state through the modification of external culture environment (an extended LIF/Serum culture) (Bao *et al.*, 2009), and that a rare transient population of mESC can exist in a 2-cell embryo like state (Macfarlan *et al.*, 2012).

To experimentally address this issue, Boroviak *et al.* adopted a single-cell transcriptional analysis of early mouse embryos and found that 2i or 2i+LIF mESCs most closely resembled cells of the *in vivo* E4.5 epiblast (Boroviak *et al.*, 2014). In contrast, mESCs cultured in LIF/Serum correlated poorly with the transcriptional profile of E4.5 epiblast cells, consistent with the reported heterogeneous nature of LIF/Serum mESCs (Marks *et al.*, 2012).

Going further, if indeed 2i ground state mESCs most resemble the E4.5 epiblast, each single E4.5 epiblast cell should theoretically give rise to a single mESC line. To this end, using 2i+LIF conditions and a *Pdgfra-GFP* reporter to distinguish primitive endoderm cells (PrE), Boroviak *et al.* found that mESC colonies could only be isolated from epiblast cells, but not from PrE cells. Furthermore, the number of

mESC colonies derived per embryo was in close approximation with the expected number of E4.5 epiblast cells that exist within the *in vivo* ICM. This finding also indicates that each epiblast cell has the potential to become a mESC under 2i+LIF (Boroviak *et al.*, 2014).

Taken together, the work by Boroviak *et al.* has demonstrated that 2i or 2i+LIF ground state mESCs resemble the pluripotent E4.5 mouse epiblast cells. In contrast, LIF/Serum mESCs which exhibit heterogeneous marker expression do not associate closely with cells of the E4.5 mouse epiblast (Boroviak *et al.*, 2014).

1.4 – Dissecting the mechanism of ground state pluripotency

1.4.1 – Gsk3-inhibition alleviates Tcf3 mediated transcriptional repression of *Esrrb*

Due to the unique molecular and features of mESCs under 2i culture, the next step was to elucidate the cellular pathways by which Mek-inhibition through PD0325901, and Gsk3-inhibition by CHIR99021, can establish ground state pluripotency.

Tcf3 has been previously shown to repress the mESC pluripotency network, and that Gsk3-inhibition serves to promote pluripotency through the alleviation of this Tcf3-mediated repression (Wray *et al.*, 2011). However, whether Tcf3 functions as a general repressor of all pluripotency genes, or whether Tcf3 targets a specific subset of genes was not clear. To address this question, Martello *et al.* through the correlation of Tcf3 binding sites with transcriptional data, found that several pluripotency genes such as *Esrrb*, *Klf2*, *Nanog*, *Nr0b1* and *Tcfcp2l* may serve as potential Gsk3/Tcf3 repressed targets.

Validation of these candidates however only indicated *Esrrb* to be the genuine functional target of Gsk3-inhibition. Specifically, it was observed that *Esrrb*-null mESCs cannot self-renew properly under 2i, or under CHIR99021 conditions, without the support of exogenous LIF. More importantly however, the authors demonstrated that *Esrrb* overexpression was sufficient to substitute for Gsk3-inhibition to sustain mESC self-renewal under 1i (Mek-inhibitor only). Together, the data by Martello *et al.* clearly demonstrates *Esrrb* as the critical downstream target for

Gsk3/Tcf3 repression, and that the addition of CHIR99021 was necessary to alleviate the repression of *Esrrb* (Martello *et al.*, 2012). The addition of the Gsk3-inhibitor CHIR99021 thus alleviates Gsk3/Tcf3 repression of the *Esrrb* gene.

Because *Esrrb* is a potent pluripotency factor capable of replacing Klf4 in iPSC reprogramming (Feng *et al.*, 2009), and substituting for Nanog function in mESCs (Festuccia *et al.*, 2012), it is hence not surprising that *Esrrb* plays a key role in sustaining ground state pluripotency.

1.4.2 – The Klf transcription factors

Krüppel-like factors (Klfs) are a family of DNA-binding zinc finger containing transcription factors that are involved in a multitude of biological processes. Named after the *Drosophila melanogaster* Krüppel protein due to its DNA sequence homology, Klfs are known to regulate cellular proliferation, cell differentiation, migration, animal development and even pluripotency (Reviewed by McConnell and Yang, 2010; Tetreault *et al.*, 2013). Given this diverse function, many Klf protein members have been demonstrated to be key regulators of various mouse and human developmental pathways and diseases (**Table 1**). Currently, at least 17 validated Klf family proteins have been identified in mammals, which can be generally grouped into three classes based on their structural homology and effect on transcriptional regulation (**Fig. 7**) (Tetreault *et al.*, 2013).

Name	Expression pattern	Tissue function/ Disease association	Gene knockout phenotype
Klf1	Expression is found within erythroid cells and in the fetal liver. Also detected in the adult bone marrow and spleen	Involved in erythropoiesis and megakaryocyte differentiation	Embryonic lethal at E14.5 due to defective hematopoiesis β -thalassemia
Klf2	Mainly found in the lungs, and vascular endothelial cells. Also found in T-lymphocytes, erythroid cells, as well as in white adipose tissue	Important regulator of blood vessel dynamics, and lung development. Klf2 is important for T-cell survival/migration, and for the inhibition of adipogenesis. Also an induced pluripotent stem cell (iPSCs) reprogramming factor.	Embryonic lethality between E12.5 - E14.5 arising from defective blood vessel integrity; impaired lung and T-lymphocyte development; and defective adipocyte differentiation
Klf3	Widely expressed in various tissues. High expression in erythroid cells	Inhibitor of adipocyte differentiation	Reduced formation of white adipose tissue
Klf4	Found in gut, skin and lung epithelial cells; endothelial cells; early precursor B cells	Important for the normal homeostasis of intestinal epithelial cells; Klf4 functions as a tumor suppressor during colon cancer. Conversely, Klf4 exhibits oncogenic properties in breast cancer and squamous cell carcinoma. Klf4 is a promoter of adipogenesis, and is also an iPSCs reprogramming factor.	Perinatal lethal just after birth due to dehydration arising from defective skin barrier formation. Knockout mice also have abnormal colonic goblet cell differentiation.
Klf5	Highly enriched in the gut epithelia. Also found in the epidermis, vascular smooth muscle and white adipose tissue.	Important for intestinal homeostasis; cardiac remodeling; promoter of adipogenesis; generation of iPSCs. Dysregulated Klf5 expression is found in colon, breast, prostate cancer and other solid tumors.	Embryonic lethal at E6.5 due to placental defects. Klf5 heterozygotes and Klf5 tissue-specific deletion demonstrate a reduced susceptibility to stress-induced cardiac hypertrophy. Klf5 heterozygotes also exhibit impaired formation of white adipose tissue, and impaired skeletal growth. Klf5 heterozygous deletion drastically reduces the incidence of intestinal tumor formation.
Klf6	Strong expression within the embryonic placenta and developing hindgut, heart, lung and kidney. Klf6 is also found in the adult heart, liver, lung, kidney, intestine, and endothelial cells	Promotes vascular remodeling and stimulates adipogenesis. Klf6 functions as a tumor suppressor in prostate cancer. As such, the silencing/mutation of Klf6 is reported in a variety of cancers	Embryonic lethal at E12.5 due to failed erythropoiesis and yolk sac vascularization
Klf7	High levels detected within the brain, spinal cord, the developing central and peripheral nervous system	Inhibitor of adipocyte differentiation. Single-nucleotide polymorphism (SNPs) of the Klf7 gene is reportedly linked to type 2 diabetes	Neonatal lethality by P3 due to severe neurological defects arising from poor neurite outgrowth and axonal misprojection
Klf8	Low but ubiquitous expression	High expression of Klf8 in ovarian and other cancers	Knockout mice are viable but with a reduced lifespan. The cause is yet to be determined.

Klf9	Broad expression; high levels of Klf9 is found in the developing brain, thymus, epithelia and gut/bladder smooth muscles	Loss/reduced levels of Klf9 results in endocrine-responsive cancers such as endometrial cancer	Knockout mice exhibit normal lifespans, but with behavioral abnormalities. Female mice demonstrate impaired uterine development and defective embryo implantation. Klf9 knockout mice also have shorter small intestinal villi
Klf10	Broad expression in the adult pancreas, kidney, lung, brain, liver, heart and testis	Involved in T-cell differentiation and activation. Klf10 is also important for bone development. Klf10 is proposed to be tumor suppressor in breast cancer	Knockout mice have normal lifespan, but Klf10-null female mice have osteopenia, Cardiac hypertrophy is observed in males. Mice also have defects in the structure and healing of tendons
Klf11	Expression in the erythroid cells of the fetal liver. Ubiquitous expression in the adult	Klf11 is a tumor suppressor in pancreatic cancer. Klf11 gene variants which arise due to polymorphisms may be linked with early-onset type 2 diabetes	Knockout mice exhibit normal development, lifespan and fertility
Klf12	Expression found within the developing brain. Klf12 is also detected in the developing/adult kidney	Elevated Klf12 levels results in oncogenic activity. Implicated in breast and gastric cancer	-
Klf13	Broad expression, with high levels in the developing heart, brain, thymus, epidermis, gut and bladder epithelia	Important for cardiac development, as well as B and T lymphocyte formation	Klf13 knockout mice exhibit enlargement of the spleen, with anomalous erythroblast differentiation. Knockout mice also have abnormal T and B lymphocyte formation.
Klf14	Ubiquitous expression. Expressed only from the maternal allele.	Neighbouring gene regions of the paternal Klf14 allele is associated with basal cell carcinoma. Klf14 variation at the gene loci is also implicated in metabolic diseases.	Mouse are viable, but phenotype is not known
Klf15	Low expression in the developing heart. High expression in the adult kidney, liver, pancreas, heart, skeletal muscle, lung and ovary	Important regulator of gluconeogenesis; prevents cardiac hypertrophy and fibrosis	Knockout mice are viable with normal fertility. Mice have increased susceptibility to stress-induced cardiac hypertrophy; hypoglycemic phenotype arising from defective amino acid metabolism
Klf16	High levels found in the developing brain. Low Klf16 detected in the thymus, duodenum, kidney, liver, heart, bladder and lung	-	-
Klf17	Expressed in spermatids and oocytes	Tumor suppressor function to inhibit metastasis and EMT	-
Klf18	Predicted Klf protein with unknown biological expression or function	-	-

Table 1. The tissue expression, function and pathobiology of Klf family proteins. Data obtained, adapted and updated from the review by (McConnell and Yang, 2010; Tetreault et al., 2013). New information added for *Klf8* (Funnell et al., 2013), *Klf14* (de Assuncao et al., 2014; Parker-Katirae et al., 2007; Small et al., 2011), *Klf17* (Gumireddy et al., 2009), *Klf18* (Pei and Grishin, 2013).

Figure 7. Phylogenetic tree of human KLFs. Group 1 KLF proteins serve as transcriptional repressors via their interaction with carboxy-terminal binding proteins (CtBP). Group 2 KLF proteins function primarily as transcription activators. Group 3 KLF protein are also transcriptional repressors acting with the Sin3A transcriptional corepressor. Based on the phylogenetic analysis, KLF15 and KLF17 are the most distant from the known KLF family. Figure obtained from the review by Tetrault *et al.* 2013.

1.4.3 – The role of Klf proteins in mESCs and their association with ground state pluripotency

We have previously shown that the three Krüppel-like factors, Klf2, Klf4 and Klf5 play functionally redundant roles in LIF/Serum mESCs to sustain pluripotency (Jiang *et al.*, 2008). It was found that only a triple *Klf2/Klf4/Klf5* knockdown could induce mESC differentiation and that this phenotype may be rescued through the expression of a single Klf2, Klf4 or Klf5 (Jiang *et al.*, 2008). Furthermore through the ChIP mapping using a custom genomic-tiling microarray (covering 402 genes and 4 gene clusters), Jiang *et al.* found that there was significant overlap in Klf2, Klf4, Klf5 binding sites. This thus indicates that all three Klf proteins may potentially work together to regulate a common set of downstream target genes (Jiang *et al.*, 2008).

Further evidence for a triple Klf protein redundancy in mESCs is partially supported by single *Klf2* or *Klf5* gene knockout studies in mESCs. For instance, *Klf2*^{-/-} embryos are embryonic lethal between E12.5-E14.5 due to intra-embryonic haemorrhaging (Kuo *et al.*, 1997; Wani *et al.*, 1998). However, *Klf2*-null mESCs were still able to self-renew under LIF/Serum conditions with growth rates similar to wildtype mESCs (Wu *et al.*, 2005). In addition, *Klf2* knockout mESCs were capable of generating mouse chimeras, indicating that they are still pluripotent despite the absence of *Klf2* (Pei *et al.*, 2011).

Klf5 knockout embryos are also embryonic lethal due to implantation failures arising from trophectoderm defects (Ema *et al.*, 2008). *Klf5*^{-/-} mESCs could nevertheless still

be generated from *Klf5*-Het mESCs through the disruption of the single remaining wildtype *Klf5* allele. These *Klf5*^{-/-} mESCs while prone to spontaneous differentiation, were still capable of embryoid body differentiation and teratoma formation; indicating that *Klf5* knockout mESCs were still pluripotent (Ema *et al.*, 2008). Notably, Ema *et al.* found that Klf4 overexpression could help rescue the spontaneous differentiation phenotype of *Klf5*^{-/-} mESCs. This indicates a possible functional redundancy between the two transcription factors (Ema *et al.*, 2008).

Currently, only the *Klf4* knockout phenotype in mESCs is not known. However, *Klf4*^{-/-} embryos can develop past the pre-implantation ICM stage, and will only die shortly after birth due to dehydration arising from a defective skin barrier (Katz *et al.*, 2002; Segre *et al.*, 1999). This suggests that at least for the *in vivo* pluripotent ICM stage, the absence of Klf4 has no overt effect on embryonic tissue differentiation.

Despite this reported functional redundancy, the transcriptional pathways which control *Klf2*, *Klf4* and *Klf5* expression are differentially regulated. Specifically, it was discovered that *Klf4* and *Klf5* expression was induced by the LIF/Stat3 pathway (Hall *et al.*, 2009; Niwa *et al.*, 2009), and that *Klf2* was LIF-independent and under Oct4 transcriptional regulation (**Fig. 8**) (Hall *et al.*, 2009). Hall *et al.* further discovered that Klf2 was the most potent among the three Klf proteins in sustaining mESC self-renewal without LIF stimulation, and that Klf2 was more efficient than Klf4/Klf5 in reprogramming EpiSCs into a mESC-like state (Hall *et al.*, 2009).

Figure 8. The transcriptional regulation of Klf2, Klf4 and Klf5 in mESCs. During conventional LIF/Serum culture, the triple Klf2, Klf4 and Klf5 all contribute towards the maintenance of mESC pluripotency. However, the activation of these three Klf proteins is under different pathways. *Klf4* and *Klf5* were found to be mainly under the influence of LIF-signalling. On the other hand, *Klf2* expression is under the direct transcriptional regulation of Oct4. Figure obtained from Hall *et al.* 2009.

Because *Stat3*-null mESCs cannot be propagated under serum, but were otherwise normal under 2i conditions (Ying *et al.*, 2008), the most striking result from Hall *et al.* was that ectopic *Klf2* expression was able to sustain *Stat3*-null mESC self-renewal akin to that of 2i. This rescue however was not possible using *Klf4* or *Klf5*, hence hinting that there could be some association of Klf2 with ground state pluripotency (Hall *et al.*, 2009).

Indeed, this association of 2i ground state pluripotency with Klf2 is further suggested from another study by Greber *et al.* When EpiSCs are treated with the Mek-inhibitor PD0325901, an increase in *Klf2* gene expression is observed (Greber *et al.*, 2010). Conversely, when mESCs were treated with FGF2, *Klf2* gene expression was found to be suppressed. Mek-inhibition by PD0325901 thus appears to alleviate *Klf2* repression by the Fgf/Mek/Erk pathway.

1.5 – Objectives of this study

Although 2i culture has been successful in the derivation of germline-competent ESCs from recalcitrant mouse strains and also from the rat, the exact mechanism by which ground state pluripotency is established remains to be fully understood. Specifically, while Gsk3-inhibition has been linked to the transcriptional derepression of *Esrrb*, the pathway downstream of Mek-inhibition by PD0325901 has not been identified.

Because *Klf2* overexpression, like 2i culture, is able to sustain *Stat3*-null mESC self-renewal (Hall *et al.*, 2009), and Mek-inhibition is able to upregulate *Klf2* expression (Greber *et al.*, 2010); it is hypothesized that *Klf2* regulation may be downstream of Mek/Erk signaling. This thesis thus serves to investigate how the Mek/Erk signalling pathway regulates *Klf2* in mESCs. In addition, this study seeks to understand the role of *Klf2* during ground state pluripotency.

These objectives include:

- a) Determine the effects of LIF/Serum and 2i media on the transcription and protein levels of *Klf2*, *Klf4* and *Klf5*
- b) Identify the mechanisms by which the Fgf/Mek/Erk signalling pathway is antagonistic towards *Klf2*
- c) Test and examine the effects of *Klf2* depletion on mESC self-renewal and pluripotency during LIF/Serum or 2i culture

- d) Test and examine the consequences of Klf2 depletion towards the derivation of mESCs from mouse blastocysts using LIF/Serum or 2i.
- e) Elucidate the downstream targets of Klf2 in ground state mESCs and how they promote ground state pluripotency

Together, this study will provide an integrated model as to how ground state pluripotency is achieved via the dual Fgf/Mek/Erk and Gsk3/Tcf3 signalling pathway inhibition. More importantly, the understanding of how ground state pluripotency is established would offer an essential roadmap towards the capturing ground state ESCs from other mammalian species, including humans.

CHAPTER II

MATERIALS AND

METHODS

Chapter II: Materials and Methods

2.1 – Cell Culture

Feeder-free E14 mESCs were cultured on 0.1% gelatin-coated plates at 37°C with 5% CO₂ in Dulbecco's modified Eagle medium (DMEM; GIBCO), supplemented with 15% heat-inactivated ESC-certified fetal bovine serum (FBS; GIBCO), 0.055 mM β -mercaptoethanol (GIBCO), 2 mM L-glutamine, 0.1 mM MEM nonessential amino acid, 20 μ g/ml gentamicin (GIBCO) and 1,000 units/ml of LIF (Chemicon). E14 mESCs were passaged at a ratio of 1:6 every two days. Klf2 transgenic mESC lines were cultured on mitomycin-C treated CD1 MEF feeder cells in similar mESC cell culture media, and passaged every two days at a 1:6 ratio. For chemically-defined 2i conditions, mESCs were either cultured on CD1 MEF feeder-cells, or on 0.1% coated gelatin plates, using serum free N2B27 media supplemented with Mek inhibitor PD0325901 (1 μ M) and Gsk3 β inhibitor CHIR99021 (3 μ M) (Stemgent) (Ying *et al.*, 2008). 2i-cultured mESCs are passaged every three days using Accutase (Chemicon) at a ratio of 1:6.

CD1 MEFs were isolated from E13.5 CD1 mouse embryos by dissociation in 0.25% trypsin at 37 °C for 10 min and cultured in 10% FBS-DMEM with 20 μ g/ml of gentamicin (GIBCO). 293T cells were cultured in 10% FBS-DMEM containing 20 μ g/ml of gentamicin (GIBCO).

2.2 – Generation of *Klf2-LacZ* reporter mice

The procedure in generating *Klf2-LacZ* reporter mice is as summarized (**Fig. 9**). V6.4 (C57BL/6J x 129/SvJae) mESCs were electroporated with a linearised targeting vector consisting of *FRT-En2SA-IRES-LacZ-loxP-βactinP-Neo-FRT-loxP* (KOMP Repository Clone ID: DPGS00163_A_G03). Electroporated cells were then plated onto mitomycin-C treated CD1 feeder cells under G418 selection. After 10 days, G418-resistant colonies were then isolated and individually expanded. Genomic DNA is then extracted from these clones for PCR and southern blot genotyping. Two independent heterozygous *Klf2^{LacZ-Neo/+}* mESCs lines (Line #3 and #7), were then microinjected into 2-8 cell stage C56BL/6J embryos (Kraus *et al.*, 2010) to generate germ-line transmitting chimeric mice (Line ID #3 only. Homologous gene targeting of the *Klf2* allele was performed by Dr Jiang Jianming from the Genome Institute of Singapore (GIS). Microinjection of heterozygous *Klf2^{LacZ-Neo/+}* was done by Dr Petra Kraus and Dr Thomas Lufkin from GIS.

The resultant *Klf2^{LacZ-NeoZ/+}* mice were then bred and maintained on a mixed C57BL/6 and 129 genetic background. All animal procedures were approved by the Singapore A*STAR Biopolis Biological Resource Center (BRC) Institutional Animal Care and Use Committee (IACUC).

Figure 9. Gene targeting of *Klf2* loci. (a) Wildtype *Klf2* exons 2 and 3 are replaced with an *IRES-LacZ-Neomycin* cassette via homologous recombination. This creates a *Klf2*-null locus, with more than 90% of *Klf2* protein sequence being deleted. Gene targeting performed by Dr Jiang Jianming (GIS) (b) For Southern blot genotyping, genomic DNA is digested with *AflIII* restriction enzymes and hybridized with a probe specific for a region within the 3'-homology arm. Wildtype *Klf2* allele will produce a 6.7 kb band, while the *Klf2*-null allele will result in a 12.2 kb DNA fragment. The *Klf2*-null Southern blot genotyping process was designed and performed by Yeo Jia Chi (c) PCR genotyping will give rise to a 527bp and 721bp DNA fragment corresponding to a wildtype or *Klf2*-null allele. The *Klf2*-null PCR genotyping process was designed and performed by Yeo Jia Chi. Figure 9 is based on Fig. S2a-c from Yeo *et al.*, 2014.

2.3 – Embryo Collection and mESC derivation

All E3.5 blastocysts were flushed from mouse uteri using M2 media (Sigma), and cultured individually in 24-well plates containing CD1 feeder-cells. For derivation of mESCs under LIF conditions, blastocysts were grown in knockout DMEM (KO-DMEM; GIBCO) supplemented with 20% knockout serum (KSR; GIBCO) and 1,000 units/ml of LIF (Chemicon) as previously described (Bryja *et al.*, 2006). As for the derivation of mESCs under 2i conditions, blastocysts were cultured with serum free N2B27 media with PD0325901 (1 μ M) and Gsk3 inhibitor CHIR99021 (3 μ M) for 5-8 days as previously described (Ying *et al.*, 2008). Following attachment and subsequent culture, the blastocyst outgrowths were isolated, trypsinised, washed in PBS, and re-plated onto fresh feeders under LIF/KSR or 2i media. After an additional 4-8 days of culture, mESC colonies would start developing from the dissociated blastocyst outgrowths.

2.4 – Proximity Ligation Assay (PLA)

To probe for Klf2 and Erk2 protein interactions, mouse monoclonal anti-Erk2 antibody (clone D2, sc-1647) (1:100 concentration; Santa Cruz) was paired with rabbit polyclonal anti-Klf2 antibody (1:100 concentration; purified with Protein G column) (Jiang *et al.*, 2008). All PLA experiments were performed by Kelvin Chong and Dr Andrew Tan from Nanyang Technological University (NTU) according to manufacturer instructions (Olink Biosciences, Sweden).

2.5 – Cell cycle and Apoptosis Detection

Klf2-WT and *Klf2*-null mESCs were trypsinised, collected, and washed with PBS. To detect for cell cycle profile, the mESCs were stained with 2 µg/ml Hoechst 33342 at 37°C for 40mins. To detect for early and late apoptotic mESCs, 1×10^5 cells were mixed and incubated for 15 min in the dark at 25°C with 10 µg/ml Annexin V FITC and 5 µg/ml of propidium iodide (Annexin V Apoptosis Detection Kit; Santa Cruz). Stained cells were FACS analyzed using a BD LSRII (BD Biosciences) with constant parameters set for all samples. Appropriate compensation was also performed to reduce the spillover of FITC onto the propidium iodide channel during the detection of apoptotic cells. All experiments were performed by Dr Liang Hongqing from GIS.

2.6 – RNAi knockdown

Transfection of E14 mESCs with shRNAs against *Klf2*, *Klf4* and *Klf5* (**Table 2**) was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 1-2 µg of plasmids were transfected into mESCs in 6-well plates, with puromycin (Sigma) selection introduced 8 h post-transfection at 1 µg/ml concentration. Drug selection was maintained for 2-4 days prior to RNA extraction and real-time PCR analysis.

Gene	RNAi target sequence 1	RNAi target sequence 2
<i>Klf2</i>	GGCAAGACCTACACCAAGA	GTGAGAAGCCTTATCATTG
<i>Klf4</i>	GAGGAACTCTCTCACATGA	GACCACCTTGCCTTACACA
<i>Klf5</i>	GACCATGCGTAACACAGAT	GTACCAGCTGTTGAATACA

Table 2. List of RNAi sequences

2.7 – RNA extraction, reverse transcription and quantitative real-time PCR

Total RNA was extracted from cells using TRIzol reagent (Invitrogen) and purified using conventional isopropanol precipitation techniques. Prior to reverse transcription (RT), DNA contamination was eliminated by DNase I treatment (Ambion) treatment at 37°C for 30 min. DNase I inactivation was done by heating at 75°C for 10 min. RT was subsequently performed using SuperScript II Kit (Invitrogen) according to the manufacturer's instructions. Quantitative real-time PCR (qRT-PCR) were performed on the ABI PRISM 7900 Sequence Detection System, using the KAPA SYBR FAST q-PCR mix (KAPA Biosystems). The sequence of qRT-PCR primers used in this study is listed in **Table 3**.

Gene	Primer Sequence (Forward)	Primer Sequence (Reverse)
<i>Klf2</i>	ACCAAGAGCTCGCACCTAAA	GTGGCACTGAAAGGGTCTGT
<i>Klf4</i>	CTGAACAGCAGGGACTGTCA	GTGTGGGTGGCTGTTCTTTT
<i>Klf5</i>	ACGTACACCATGCCAAGTCA	GTGGGAGAGTTGGCGAATTA
<i>Oct4</i>	CACGAGTGGAAGCAACTCA	AGATGGTGGTCTGGCTGAAC
<i>Nanog</i>	AAGTACCTCAGCCTCCAGCA	GTGCTGAGCCCTTCTGAATC
<i>Sox2</i>	CGAACTGGAGAAGGGGAGAG	AAGCGTTAATTTGGATGGGA
<i>Rex1 (Zfp42)</i>	CTCAAGCCGGGTGCAAGAA	CGGATATGGGTGCGCAAGTT
<i>Esrrb</i>	AAGTGGGGATGCTGAAGGA	TTCACAGAGAGTGGTCAGGG
<i>Nr5a2</i>	AGAAAAGCTGAGCGCATTTG	GCAGCTTCATTTGGTCATCA
<i>Nr0b1</i>	CCAGGCCATCAAGAGTTTCT	CCCTCAATGTATTTACGCA
<i>Prdm14</i>	TGGACCCTTTTCGAGGTAGAGT	AAAATATCTGCCCTTGGTGCT
<i>Tbx3</i>	GACCATGGAACCCGAAGAAG	ACCATCCACCGAGAGTTGTG
<i>Tfcp2l1</i>	AGGTGCTGACCTCCTGAAGA	GTTTTGCTCCAGCTCCTGAC
<i>c-Kit</i>	GATCTGCTCTGCGTCCTGTT	TGTCGCCAGCTTCAACTATT
<i>Prdm1</i>	ACCAAGGAACCTGCTTTTCA	GGCATTCTTGGGAAGTGTGT
<i>Dnd1</i>	GTCAACGGGCAAAGGAAGTA	AACAAGTAGGCTGGCACCAT
<i>Tcfap2c</i>	TCCTTGAGGGAGAAGTTGGA	AGACATAGGCGAAGTCCCGT
<i>Ifitm1</i>	CCACAATCAACATGCCTGAG	CGTATCACCCACCATCTTCC
<i>Dppa3</i>	CCAATGAAGGACCCTGAAACT	AACAAAGTGCGGACCCTTCT
<i>β-actin</i>	AGCCATGTACGTAGCCATCC	GCTGTGGTGGTGAAGCTGTA

Table 3. List of qRT-PCR primer sequences.

2.8 – Microarray analysis

RNA was extracted from feeder-free *Klf2*-WT and *Klf2*-null mESCs maintained in LIF/Serum culture, or during 2i passage 1, 2, or 3, using Trizol (Invitrogen). After initial isopropanol precipitation, total RNA samples were treated with DNase I (Ambion) for 30mins at 37°C to remove genomic DNA contamination. Following which, the DNase-treated RNA was then re-purified and eluted using the PureLink RNA Mini Kit (Invitrogen). 500 ng of this purified total RNA was used for cDNA amplification and biotin-tagging by the Illumina Total Prep RNA amplification kit. Amplified biotin-tagged cDNA was then used for hybridization onto Illumina MouseWG-6 Expression Bead Chip v2.0 according to the manufacturer's instructions. Hybridized microarray chips were finally labeled with Cy3, and scanned using an Illumina Bead Scanner. Raw data were then translated using Bead Studio, and analyzed using the "Significance analysis of Microarrays" (SAM) software to determine gene expression fold-changes. All microarray data are publically available at the Array Express database (<http://www.ebi.ac.uk/arrayexpress>) under the accession number E-MTAB-2366.

2.9 – Southern blot and PCR genotyping

Genomic DNA (gDNA) from mESCs or mouse tail biopsies extracted and precipitated using conventional phenol-chloroform techniques, following overnight Proteinase K digestion at 55°C. For Southern blot, 10 µg of gDNA was digested overnight with AflIII restriction enzyme (New England Biolabs) at 37°C and resolved on a 0.8% agarose gel. The DNA gel was then denatured using 0.5 M NaOH, and transferred onto a positively-charged nylon membrane (Amersham Hybond N+). The transferred membrane was then pre-hybridized 42°C with 0.2 mg/ml sonicated salmon sperm DNA in DIG Easy Hyb buffer (Roche) for at least 4 hrs, and then hybridized at 42°C in DIG Easy Hyb buffer with 20 ng/ml of Digoxigenin-labelled probe (Roche) overnight.

After hybridization, the membranes were washed once at room temperature with 2×SSC, and washed twice with 0.5×SSC + 0.1% SDS at 53°C for 15 mins each. The DNA membrane was then blocked with 1×Blocking reagent (Roche) for 30 mins at room temperature, and then incubated with anti-DIG-AP antibody (Roche) (1:10000 concentration in 1× Blocking reagent) for 30 mins at room temperature. The membrane was then washed once with 1× MAB at room temperature for 5 mins, and washed twice at room temperature with 1×MAB + 0.3% Tween 20 for 15 mins each. Finally, the membrane was equilibrated briefly in detection buffer (100 mM Tris-HCl, pH 9.5; 100 mM NaCl); incubated with CSPD (Roche) at 37°C for 10 mins, before DNA band visualization on X-ray film (Amersham).

The DIG labeled probe was generated via the PCR-amplification of *Klf2-LacZ* gene targeting construct (KOMP Clone ID: DPGS00163_A_G03) using the DIG DNA labeling kit (Roche) with forward primer (5'GAAATGCCTTTGAGTTTACTGTCCCCAACG 3') and the reverse primer (5'GGTGGGGCAAATTATGGCTCAAAGTAGCA 3').

For PCR genotyping, multiplex PCR was performed using the following three primers: a) Common forward primer *Klf2*-WT/*Klf2-LacZ* (5' ACAGCACACACAGTCCCTGC 3'); b) Reverse primer *Klf2*-WT (5' GGCTCCGGGTAGTAGAAGGC 3'); and c) Reverse primer *Klf2-LacZ* (5' TCTCCCCTTCAGTCTTCCTG 3'). The *Klf2*-WT allele is identified by a 527bp DNA band, and the *Klf2* knockout allele will result in a 721bp DNA band when resolved on agarose gel.

2.10 – Transfection, Co-immunoprecipitation and Immunoblotting

HEK293T cells were transfected with 200 ng to 1 µg of *3HA-Klf2* or *3Flag-Erk2* plasmids using Lipofectamine 2000 according to the manufacturer instructions (Invitrogen). Immunoprecipitation (IP) and immunoblotting (IB) were conducted as described previously in (Yim et al., 2015) using cell lysis buffer constituted by 20 mM HEPES (pH 7.5), 137 mM NaCl, 1% Triton X-100, 10% Glycerol, 1.5 mM MgCl₂, 1 mM EGTA, 0.2 mM sodium orthovanadate (Na₃VO₄) and supplemented with protease inhibitor cocktail. Western blotting of proteins was done following conventional procedures. All IP experiments were performed by Dr Daniel Yim from GIS.

The sources and concentrations of antibodies are as listed in **Table 4**.

Antibody	Concentration	Source
α-Klf2	1:3000	Jiang <i>et al.</i> , 2008
α-Klf4	1:3000	
α-Klf5	1:3000	
α-Oct4	1:2000	Santa Cruz (sc-8628)
α-Nanog	1:2000	Cosmobio (REC-RCAB0002P-F)
α-HA	1:1000	Santa Cruz (sc-7392)
α-Erk2	1:2000	Santa Cruz (sc-1647)
α-βactin	1:5000	Santa Cruz (sc-47778)
α-phosphoserine/threonine	1:1000	ECM Biosciences (PP2551)

Table 4. List of Antibodies used

2.11 – Generation of stable *Klf2* expressing *Klf2*-null mESCs

To generate *Klf2* expressing *Klf2*-null mESCs, a *piggybac* transposase system was employed which allows the stable genomic integration of exogenous DNA sequences (Wang *et al.*, 2008). To this end, 2 µg of *piggybac* transposon plasmids (pCyl50) containing 3HA-tagged wildtype *Klf2*, 3HA-tagged serine/threonine *Klf2* mutants, or GFP control were transfected into 1×10^6 *Klf2*-null mESCs. Simultaneously, 2µg of transposase plasmid (pCyl43) were also co-transfected with the pCyl50 plasmids to enable genomic integration. After 12 hrs post-transfection, the LIF/Serum cell culture media was replaced with 2i media and cultured for another 48 hrs.

Due to the presence of a hygromycin resistance cassette, transfected *Klf2*-null mESCs were passaged with hygromycin selection (100µg/ml) under 2i or 1i (CHIR99021 only) post 48hrs transfection. Brightfield images were then taken every 2 days to track *Klf2*-null mESC cell survival. Cell colony quantification was done using ImageJ using a automated colony counter with settings “Size (pixel²)” of 200-infinity and “Circularity” of 0.00-1.00.

2.12 – Recombinant GST-Klf2 expression and *in vitro* kinase assay

Full length *Klf2* sequence was cloned into pET42b vectors (Novagen) for the production of N-terminal GST-tagged and C-terminal 6xHis-tagged Klf2 protein. For the production of GST proteins, empty pET42b vectors were used. Recombinant Klf2 protein or GST protein is then produced using IPTG-induced BL21 bacteria, and purified using GST-sepharose (GE Healthcare Life Sciences) and Ni-NTA-sepharose purification (Qiagen).

In vitro kinase assay of GST-Klf2 or GST proteins was performed using commercial activated Erk2 protein (Sigma). Briefly, the kinase reactions were carried at 30°C for 30mins using 100ng active Erk2 with 5µg of recombinant GST-Klf2, GST or MBP in kinase reaction buffer (50mM HEPES, 50mM NaCl, 10mM MgCl₂, 5mM NaF, 0.05% Triton X-100, 0.05mM DTT, 50mM unlabelled ATP, and 10 µCi of [γ -³²P]-ATP). Reactions were inactivated by addition of 2 x SDS sample buffer and heated to 95°C for 10 mins. Following *in vitro* phosphorylation, protein samples were then run on a 10% SDS-PAGE gel and dried onto a piece of filter paper. The dried SDS-PAGE gel was then exposed onto a phosphoimager screen before image capture using Typhoon scanner (Amersham). Myelin basic protein (MBP) was purchased from Sigma and used to verify a positive kinase phosphorylation reaction.

2.13 – Chromatin immunoprecipitation (ChIP) and ChIP-seq analysis

Chromatin immunoprecipitation was performed as previously described (Karwacki-Neisius *et al.*, 2013). Briefly, cells were fixed with 1% formaldehyde at room temperature for 10min, and inactivated by the addition of glycine to a final concentration of 200mM. Cell lysates were sonicated and chromatin extracts were immunoprecipitated overnight at 4°C using Protein G Dynal Magnetic Beads that have been coupled with antibodies against Klf2 or Klf4 (rabbit antiserum raised against Klf2 or Klf4) (Jiang *et al.*, 2008). ChIP-Seq library was prepared using the NEB Next® ChIP-Seq Library kit (NEB Biolabs) according to the manufacturer's instructions, and sequenced using the HiSeq 2000 system (Illumina). All ChIP-seq and ChIP-qPCR experiments were performed by Dr Ng Jia Hui from GIS.

For all ChIP-qPCR experiments, DNA was phenol-chloroform extracted from reverse cross-linked samples and quantitative PCR analyses were performed in real-time using the ABI Vii ATM 7 Real-Time PCR System (Life Technologies) and SYBR Green Master Mix. ChIP-qPCR primers are as listed in **Table 5**.

For Klf2 ChIP-seq, reads were mapped against the mm9 reference genome using Bowtie (0.12.8) (Langmead *et al.*, 2009). Peak calling was done with MACS 1.4.2 (Zhang *et al.*, 2008). Peaks from 2i and LIF/Serum were then combined into one set, removing overlapping peaks. Peaks that were detected in the Input control were also removed. This set was then divided into three equal sized sets according to the number of Klf2 ChIP-Seq reads that mapped to the individual loci. Bioinformatics

analysis of Klf2 ChIP-seq was done by Dr Jonathan Goke from GIS and Yeo Jia Chi.

Analysis of Klf4 and Klf5 ChIP-seq was performed by Yeo Jia Chi.

The Klf2 ChIP-seq data is publically available at the Array Express database (<http://www.ebi.ac.uk/arrayexpress>) under the access number E-MTAB-2365.

Gene	Primer Sequence (Forward)	Primer Sequence (Reverse)
<i>c-Kit</i> (ChIP)	TTTCTCTGTCTCTCCCGTTTCAGG	TGTGAACTCCTCAGATCCTTCGTG
<i>Prdm1</i> (ChIP)	GAGCAGTACTGACGCGCACCT	GTAGTCAGTCGCTCGCTCACTCG
<i>Tcfap2c</i> (ChIP)	CCGCCTCTTAAGGGAGTGAGA	CCTCGTTCAACAACCCTCCTC

Table 5. List of ChIP-qPCR primers used

CHAPTER III

RESULTS

Chapter III: Results

3.1 – Klf2 is negatively regulated by Mek/Erk signalling

3.1.1 – Gene expression changes of *Klf2*, *Klf4* and *Klf5* during ground state pluripotency

Because *Klf2* is hypothesized to be potentially regulated by the Mek/Erk signalling pathway, it would be interesting to examine the effect of ground state pluripotency on the various Klf proteins. As such, the gene expression changes of 2i-cultured or LIF/Serum cultured feeder-free E14 mESCs were examined.

While the mRNA levels of key pluripotency genes *Oct4* and *Nanog* were comparable between the two mESC culture conditions (**Fig. 10a**), it was observed that 2i-cultured mESCs had low *Klf4* and *Klf5* gene expression unlike LIF/Serum mESCs (**Fig. 10b**). This reduction in *Klf4* and *Klf5* mRNA levels during ground state pluripotency is expected due to the absence of LIF-stimulation to induce *Klf4* and *Klf5* transcription (Hall *et al.*, 2009; Niwa *et al.*, 2009). As the three Klf proteins are capable of replacing for each other (Jiang *et al.*, 2008), the low *Klf4/Klf5* transcript levels present in 2i-cultured mESCs (**Fig. 10b**) should therefore result in a strong upregulation of *Klf2* expression. However, only a marginal 2-fold gain of *Klf2* expression in 2i-cultured mESCs was observed (**Fig. 10b**).

Figure 10. Gene expression changes of key mESC genes under LIF/Serum or 2i culture. (a) *Oct4* and *Nanog* mRNA changes in LIF/Serum or 2i passage 3 (P3). **(b)** *Klf* mRNA changes in mESCs in LIF/Serum or 2i (P3). All data presented are the averages of biological triplicates $\pm$ standard error of the means (SEM). Figure 10 is based on Figure 1a-b from Yeo *et al.*, 2014.

3.1.2 – Protein level changes of Klf2, Klf4 and Klf5 during ground state pluripotency

To examine the correlation between gene expression changes and protein levels, whole cell lysates from LIF/Serum and 2i-cultured E14 mESCs were isolated and probed with previously generated Klf2, Klf4 and Klf5 antibodies (Jiang *et al.*, 2008).

Similar to the transcriptional data (**Fig. 10**), Klf4 and Klf5 protein levels were found to be reduced in 2i-cultured mESCs (**Fig. 11**). However, unlike the modest gain in *Klf2* transcript levels, it was observed that Klf2 protein amounts were strikingly increased during 2i culture (**Fig. 11**). Because this large Klf2 protein spike was unlikely to have arisen from a small 2-fold gain of *Klf2* expression, it was speculated that Klf2 protein in mESCs could be post-translationally regulated.

Figure 11. Western blot for key mESC factors in LIF/Serum or 2i (P3) conditions. Arrow demarcates the Klf4 protein band. The lower Klf4 band is non-specific. Figure 11 is based on Figure 1c of Yeo *et al.*, 2014.

Evidence for a post-translational modification of Klf2 can also be inferred from the rapid gain of Klf2 protein within 30 mins of switching from LIF/Serum into 2i mESC culture (**Fig. 12a**). Moreover, *Klf2* transcript levels during this same 30 min period remained almost constant; arguing that post-translational processes may be responsible for the elevated Klf2 protein levels (**Fig. 12b**).

No changes to Klf4 and Klf5 protein or mRNA levels were observed during this 30 min duration of 2i-culture (**Fig. 12a-b**). This suggests that the decrease in Klf4 and Klf5 protein levels observed during ground state pluripotency were likely occurring through gene transcription (**Fig. 10b and 11**).

Figure 12. Increase in Klf2 protein levels is independent of *Klf2* gene expression. (a) Western blot of Klf protein levels in LIF/Serum and after 30mins of 2i culture. (b) Gene expression changes of *Klf* transcripts following LIF/Serum and after 30mins of 2i culture. Data is the average of biological triplicates $\pm$ SEM. Figure 12 is based on Figure 1d-e from Yeo *et al.*, 2014.

3.1.3 – Klf2 protein is negatively regulated by the Mek/Erk pathway

2i ground state pluripotency is maintained by the two small molecule inhibitors, PD0325901 and CHIR99021, which targets the Mek and Gsk3 proteins respectively. As Mek (acting via Erk) and Gsk3 are known to phosphorylate numerous substrates (Doble and Woodgett, 2003; Roberts and Der, 2007), it is necessary to delineate whether it was Mek-inhibition, Gsk3-inhibition, or both inhibitors that were responsible for the elevated Klf2 protein levels under 2i culture.

To address this question, LIF/Serum cultured mESCs were treated with PD0325901 only, CHIR99021 only, or both PD0325901 + CHIR99021 for 2h before protein extraction and western blot. Interestingly, it was found that the addition of PD0325901 alone was sufficient to induce the elevation of Klf2 protein levels similar to that found in 2i-culture (**Fig. 13**). Conversely, Gsk3-inhibition via CHIR99021 had no effect on Klf2 protein levels (**Fig. 13**). Similar to results present in **Fig. 12a**, Klf4 and Klf5 protein levels remain constant throughout the course of Mek or Gsk3-inhibition.

Figure 13. Increase in Klf2 protein is dependent upon the Mek-inhibitor PD0325901. Western blot for Klf proteins in LIF/Serum mESCs that have been treated with DMSO, Mek-inhibitor PD0325901 (PD), Gsk3-inhibitor CHIR99021 (CH), or PD0325901 + CHIR99021 (PD/CH) for 2hrs. Figure 13 is based on Figure 1f of Yeo *et al.*, 2014.

As Mek-inhibition by PD0325901 results in an increase of Klf2 protein levels (**Fig. 13**), it is of interest to examine the effect of hyperphosphorylation on Klf2. Using Calyculin A (CalA) to inhibit serine/threonine protein phosphatases, it was found that compared with control DMSO-treatment, CalA-treated mESCs rapidly lost Klf2 protein within 30 mins (**Fig. 14**). On the contrary, CalA treatment did not result in the appreciable loss of Klf4 or Klf5 protein (**Fig. 14**). This observed loss of Klf2 protein following CalA-treatment suggests that any putative phosphorylation on Klf2 by Mek/Erk signalling may lead to Klf2 protein degradation.

Figure 14. Inhibition of serine/threonine phosphatases in mESCs results in rapid Klf2 protein degradation. Western blot for Klf proteins in mESCs following a time course treatment with the serine/threonine phosphatase inhibitor Calyculin A. Figure 14 is based on Figure 1g of Yeo *et al.*, 2014.

3.1.4 – Klf2 protein is proteasomally degraded in mESCs

Klf2 protein has been previously reported to be ubiquitinated and proteasomally degraded (Wang *et al.*, 2013; Xie *et al.*, 2011; Zhang *et al.*, 2004). As such, the observed rapid loss of Klf2 following CalA-treatment may suggest a similar mechanism of proteasomal degradation of Klf2 protein in mESCs (**Fig. 14**). Inhibition of proteasome activity is thus expected to result in the accumulation of Klf2 protein.

To test this hypothesis, mESCs were treated with the proteasome inhibitor MG132. Indeed, it was found that Klf2 protein levels in LIF/Serum cultured mESCs were elevated after 1h of MG132 treatment (**Fig. 15**). Notably, this increase in Klf2 protein following proteasome inhibition (**Fig. 15**) mimicked the spike in Klf2 protein levels resulting from 2i culture (**Fig. 11**). Concurrently, no changes to Klf4 and Klf5 protein levels were observed during MG132 treatment (**Fig. 15**). This suggests that the regulation of Klf4 and Klf5 protein in 2i ground state mESCs previously observed in **Fig. 10b and 11**, is again, most likely at the transcriptional level.

Figure 15. Proteasome inhibition results in elevated Klf2 protein levels. Western blot of Klf protein levels in mESCs following 1hr treatment with various concentrations of the proteasome inhibitor MG132. Figure 15 is based on Figure 1h of Yeo *et al.*, 2014.

3.2 – Klf2 directly interacts with, and is phosphorylated by, Erk2

3.2.1 –Bioinformatics analysis of Klf2 reveals potential Erk phosphorylation motifs

The current data thus far indicates a potential post-translational regulation of Klf2 by the Mek/Erk pathway. Because Erk lies downstream of Mek-signalling (Roberts and Der, 2007), and that Erk targets the serine or threonine residues of the consensus P-X-**(S/T)**-P sequence (Lewis *et al.*, 1998), three potential Erk phosphorylation motifs at the T171, S175 and S247 position were identified on Klf2 (**Fig. 16a**).

Alignment of Klf2 protein sequences from different species revealed the conservation of these potential Erk phosphorylation sites among mammals; and to a certain extent with birds and amphibians (**Fig. 16b**). This degree of evolutionary conservation suggests that these potential Erk-phosphorylation sites may be functionally important, and that Erk may phosphorylate Klf2.

a**Mouse Klf2 protein sequence (NM_032478.2)**

```

1  MALSEPI LPSFATFASPCRERGLQERWPRNEPEAGGTTEDLNNVLD F ILSMGLDGLGAENP
61  PEPPPQPPPPAFYYPEPGAPPPYSIPAASLGTELLRPDLDPQGPALHGRFLLAPPGRLV
121 KAEPPEVDGGGYGCAPGLAHGPRGLKLEGAPGATGACMRGPAGRPPPP PDTPPLSP DGPL
181 RIPASGPRNPFPFFPGPGSPFGGPGPALHYGPAPGAFGLFEDAAAAAALGLAPPATRG
241 LLTPSSP LELLEAKPKRGRRSWPRKRAATHCTSYTNCGKTYTKSSHLKAHLRTHTEKPE
301 YHCNWECCGWKFARSDELTRHYRKHTGHRPFQCHLCDRAFSRSDHLALHMKRHM

```

Red – Putative Erk phosphosite
Black – Putative Gsk3 phosphosite

b

Figure 16. Bioinformatics prediction of Erk phosphorylation sites on Klf2 protein. (a) Full-length mouse Klf2 protein sequence with predicted putative Erk and Gsk3 phosphorylation sites based on consensus motifs. **(b)** Alignment of candidate Klf2 phosphorylation sites reveals sequence conservation among species. Red and green coloured bars represent putative Erk or Gsk3 phosphorylation sites respectively. Figure 16a is based on Figure 2b of Yeo *et al.*, 2014. Figure 16b is unpublished data by Yeo Jia Chi.

3.2.2 – Klf2 is phosphorylated in mESCs

In agreement with the predicted Erk phosphorylation sites, data mining of published mESCs phosphoproteomic studies revealed phospho-peptides at the identical T171, S175 and S247 residues on Klf2 (Li *et al.*, 2011; Pines *et al.*, 2011). This evidence further supports our claim that Klf2 protein may be phosphorylated by Erk. In addition to T171, S175 and S247, the proteomics study by Li *et al.* and Pines *et al.* also indicated phosphorylation at T243 and S246 (**Fig. 16a**). Interestingly, T243 is a Gsk3 phosphorylation site and has been reported to be important for FBW7 mediated Klf2 ubiquitinylation and degradation in endothelial cells (Wang *et al.*, 2013). The purpose of Klf2 T243 phosphorylation in mESC pluripotency is still unknown, and would serve as an interesting topic of investigation for future studies.

To validate whether Klf2 is phosphorylated in mESCs, immunoprecipitated (IP) Klf2 from LIF/Serum cultured mESC lysates were probed using α -phosphoserine/threonine antibody. Importantly, serine/threonine phosphorylation from IP-Klf2 could be detected, but not from the negative control IgG pulldown samples (**Fig. 17**). Furthermore, while comparable levels of IP-Klf2 were obtained for non-treated, PD0325901 treated or CalA-treated mESC lysates, a marked loss of phosphorylated Klf2 was only observed following PD0325901 Mek-inhibition (**Fig. 17**). Therefore, the above results demonstrate that phosphorylated Klf2 protein exists in mESCs, and that the presence of phospho-Klf2 is largely regulated by the Mek/Erk pathway. However, given the technical challenges of generating a site-specific

phospho-antibody to accurately identify the Klf2 phosphorylation sites, the actual Klf2 phospho-residue(s) in mESCs remains to be elucidated.

Figure 17. Immunoprecipitation of phosphorylated Klf2 from mESCs. Immunoprecipitated Klf2 proteins were isolated from non-treated, 2hr PD0325901 (1 μ M) (PD03), or from 30min CalA treated mESCs. Detection of potential phosphoserine or phosphothreonine residues on Klf2 was then performed using a pan-phosphoserine/threonine antibody. IP experiments were done by Dr Daniel Yim (GIS). Figure 17 is based on Figure 2a of Yeo *et al.*, 2014.

3.2.3 – Klf2 physically interacts with Erk2 protein

As mESCs predominantly express Erk2 (Kunath *et al.*, 2007), it is most likely that Klf2 protein interacts with Erk2 in mESCs. As an initial trial, 3HA-tagged Klf2 were co-expressed with 3Flag-tagged Erk2 in HEK293T cells before IP using α -HA antibodies. Indeed, it was observed that the IP of exogenous 3HA-Klf2 could result in the co-immunoprecipitation (co-IP) of transfected 3Flag-Erk2 or endogenous Erk2 in HEK293T (**Fig. 18a-b**). Likewise, the reciprocal co-IP of transfected 3HA-Klf2 using

endogenous Erk2 in HEK293T cells is also possible (**Fig. 18c**). Together, these data strongly demonstrate a Klf2-Erk2 protein-protein interaction.

Figure 18. Co-immunoprecipitation (Co-IP) of Klf2 with Erk2. (a) Co-IP of exogenous 3Flag-Erk2 using 3HA-Klf2 in HEK293T cells. (b) Co-IP and detection of endogenous Erk2 using 3HA-Klf2 in HEK293T cells. Arrow demarcates the Erk2 protein band. (c) Co-IP of 3HA-Klf2 using endogenous Erk2 in HEK293T cells. WCL = whole cell lysates. IP experiments were done by Dr Daniel Yim (GIS). Figure 18a-b is based on Figure S1a-b of Yeo *et al.*, 2014. Figure 18c is based on Figure 2c of Yeo *et al.*, 2014.

To further study this Erk2-Klf2 protein-protein interaction in mESCs, an *in situ* detection method termed “Proximity Ligation Assay” (PLA) was employed to detect for endogenous Klf2-Erk2 association (Soderberg *et al.*, 2006). In brief, PLA exploits the close association of two proteins (<40 nm distance) to link their cognate antibodies and generate a fluorescence signal. Through PLA, each protein-protein interaction would result in a single fluorescence spot which can be visualized and quantified using confocal microscopy.

Following PLA in mESCs using α -Erk2 and α -Klf2, numerous PLA-spots in mESCs were observed. These signal spots were not detected when either antibody was removed (**Fig. 19**). This data demonstrates that the observed Klf2-Erk2 PLA-signals are specific, and indicative of Klf2 and Erk2 protein co-localization in mESCs.

Figure 19. Erk2 interacts with Klf2 protein in mESCs. PLA and confocal imaging of mESCs probed with both α -Erk2 and α -Klf2 antibodies. Negative control PLA experiments were done in mESCs using a single α -Erk2 or single α -Klf2 antibody. Cell nuclei were stained with DAPI, and the scale bar represents 10 μ m. PLA experiments were performed by Kelvin Chong and Dr Andrew Tan from NTU. Figure 19 is based on Figure 2d of Yeo *et al.*, 2014.

3.2.4 –Erk2 is able to phosphorylate Klf2 *in vitro*

To demonstrate that the interaction of Klf2 with Erk2 is direct, and that Klf2 can be phosphorylated by Erk2, an *in vitro* kinase assay using [γ - 32 P]-ATP was next performed. Indeed, it was found that activated recombinant Erk2 was able to phosphorylate GST-tagged recombinant Klf2 protein, as well as the Myelin Basic Protein (MBP) (**Fig. 20**). Negative control GST-only proteins however were not phosphorylated by activated Erk2 (**Fig. 20**). In light of the previous data which supports a Klf2-Erk2 interaction (**Fig. 18 and 19**), the ability of active Erk2 to phosphorylate recombinant GST-Klf2 demonstrates a direct and functional Klf2-Erk2 interaction.

Figure 20. In-vitro [γ - 32 P]-ATP phosphorylation of recombinant GST-Klf2 using activated Erk2. Figure 20 is based on Figure 2e of Yeo *et al.*, 2014.

3.3 – Klf2 is essential for mESCs under 2i conditions

3.3.1 –Generation of *Klf2* knockout mice

As 2i-cultured mESCs express elevated levels of Klf2 protein with reduced levels of Klf4 and Klf5 protein (**Fig. 11**), it was hypothesized that Klf2, rather than Klf4 or Klf5, is important in mESCs during ground state pluripotency.

To test this hypothesis, a shRNA depletion of Klf2, Klf4 or Klf5 in mESCs using previously established RNAi constructs was performed (Jiang *et al.*, 2008). However, possibly due to the short transient duration of the RNAi system (2-5 days), only a modest decrease in Oct4, Rex1 and c-Myc transcription following Klf2 depletion in 2i culture was observed (**Fig. 21a**). The overall mESC gene expression signature however remained fairly constant among the three Klf protein knockdowns (**Fig. 21b**).

In order to definitively test the importance of Klf2 towards ground state pluripotency, *Klf2* heterozygous knockout mice were generated via the homologous recombination of exons 2 and 3 in the *Klf2* coding sequence with a *IRES-LacZ* and a *Neomycin*-selection cassette (**Fig. 9**) (refer to “Materials and Methods”). This removal of *Klf2* exons 2/3 represents over 90% loss in Klf2 protein sequence to generate a *Klf2*-null allele. Subsequently, single-allele targeted *Klf2* heterozygous (*Klf2*-Het) mESCs were used to generate germ-line transmitting mouse chimeras, and the resultant *Klf2*-Het transgenic mice kept for breeding.

Figure 21. Gene expression analysis of 2i-cultured mESCs upon Klf2, Klf4 or Klf5 RNAi knockdown. (a) Gene expression of *Oct4*, *Sox2* and *c-Myc* mRNA relative to control *Luciferase* knockdown. (b) Gene expression of pluripotency markers *Nanog*, *Sox2*, *Esrrb*, *Nr0b1*, *Tbx3* and *Nr5a2* relative to control *Luciferase* knockdown. All data represents the average of biological triplicates $\pm$ SEM. RNA samples were obtained from mESCs 4-days post RNAi transfection. Figure 21 is unpublished work by Yeo Jia Chi.

According to previously published work, *Klf2*^{-/-} mouse embryos are embryonic lethal, and die between E12.5 to E14.5 due to intra-embryonic haemorrhaging (Kuo *et al.*, 1997; Wani *et al.*, 1998). In agreement with the earlier studies, our generated *Klf2*-null mice were also embryonic lethal due to intra-embryonic haemorrhaging (**Fig. 22a**), with no live *Klf2*-null mice arising from *Klf2*-Het crosses (**Fig. 22b**). Furthermore, as *Klf2* is expressed in the embryonic vasculature during development (Kuo *et al.*, 1997), a strong LacZ signal could also be detected in the blood vessels of *Klf2*-null embryos (**Fig. 22c-d**). Taken together, our data corroborates with known literature and demonstrates the specificity and phenotype arising from our *Klf2* gene targeting.

Figure 22. Phenotype of *Klf2*-null mouse embryos. (a) Brightfield images of E16.5 *Klf2*-wildtype and *Klf2*-null mouse embryos. (b) Summary of mouse number and their respective genotypes arising through heterozygous *Klf2*^{+/-} × *Klf2*^{+/-} crosses. (c) β-galactosidase staining of *Klf2*-wildtype and *Klf2*-null E12.5 mouse embryos. (d) Close-up β-galactosidase staining of *Klf2*-null E12.5 mouse embryo. White arrows denote the major blood vessels which are stained with LacZ. Figure 22a-c is based on Figure S2d-f of Yeo *et al.*, 2014. Figure 22d is unpublished work by Yeo Jia Chi.

3.3.2 – Derivation of *Klf2*-null mESCs using LIF + KSR

As *Klf2*-null embryos are embryonic lethal (**Fig. 22a-b**) (Kuo *et al.*, 1997; Wani *et al.*, 1998), *Klf2*-null blastocysts for mESC derivation could only be obtained through *Klf2*-Het parent matings. In our hands, the traditional method of LIF with fetal bovine serum (LIF/Serum) culture worked poorly in deriving mESCs from blastocysts. We therefore adopted a more efficient mESC-derivation technique as described by Bryja *et al.*, (Bryja *et al.*, 2006). In this system, blastocysts are mainly cultured in LIF with KnockOut Serum (LIF/KSR) conditions for optimal mESC growth (**Fig. 23a**). However during passaging, the trypsinised blastocyst outgrowths would be cultured with LIF/Serum for 24hrs of post-trypsinization recovery before reverting back into LIF/KSR (**Fig. 23a**) (Bryja *et al.*, 2006). The process would then be repeated till a stable mESC line is established.

Using this LIF/KSR method, a total of 19 mESC lines were established from 61 blastocysts. Among these mESC lines, 5 of them were wildtype (WT), 11 were heterozygous (Het), and 3 were *Klf2*-null (**Fig. 23b**). Of note, the genotypes of the LIF/KSR derived mESCs approximately followed the expected Mendelian ratio of 1:2:1, corresponding with a Het × Het mating scheme. Indeed, chi-square test revealed a p-value of 0.639, indicating that our *Klf2* mESC genotypes did not differ significantly from the expected Mendelian ratio. Taken together, the results demonstrate that *Klf2*-null blastocysts under LIF/KSR conditions are viable and had comparable efficacies with *Klf2*-WT or *Klf2*-Het blastocysts in forming mESCs.

Figure 23. Derivation of *Klf2*-null mESCs (a) Schematic of the mESC derivation process from E3.5 blastocysts using an alternating LIF/KSR (denoted “SR”) and LIF/Serum (denoted “FCS”) approach. Figure was obtained from Bryja *et al.*, 2006. (b) Summary of genotyped mESC lines obtain via LIF/KSR from *Klf2*^{+/-} × *Klf2*^{+/-} crosses. χ^2 -value and p-value of observed genotyped mESC lines are 0.895 and 0.639 respectively. The result is not significant at $p \leq 0.05$. Figure 23b was based on Figure S2g of Yeo *et al.*, 2014.

3.3.3 – Functional validation of *Klf2*-null mESCs

The generated *Klf2*-null mESCs were next validated and found to stain positively for alkaline phosphatase (**Fig. 24a**). With the exception for *Klf2*, these *Klf2*-null mESCs were also observed to possess the expression of key ESC genes and proteins (**Fig. 24b-c**). In addition, the *Klf2*-null mESCs were able to form teratomas comprising tissue of all three embryonic germ layers (**Fig. 24d**); as well as contribute towards chimeric mouse formation (**Fig. 24e-f**). Although *Klf2*-null mESCs could not give

rise to germ-line transmitting chimeras, the cumulative data was sufficient to allow us to conclude that *Klf2*-null mESCs are pluripotent.

PLA of *Klf2*-null mESCs using α -Klf2 and α -Erk2 antibodies was also negative (**Fig. 24g**), further supporting our earlier conclusion that a specific Klf2-Erk2 protein-protein interaction exists in mESCs.

Figure 24. Molecular and functional validation of *Klf2*-null mESCs (a) Alkaline phosphatase staining of *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum on feeders. Scale bar = 100 μ m. (b) Reverse transcription PCR of key mESC genes from *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum on feeders. (c) Western blot for *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum on feeders. (d) Teratoma sections generated from *Klf2*-null mESCs. Experiments done by Bobby Tan (GIS) (e) Mouse chimera generated from *Klf2*-null mESCs. Microinjection performed by Dr Petra Kraus and Dr Thomas Lufkin (GIS) (f) Tail DNA PCR genotyping of non-chimeric (NC) and *Klf2*-null chimeric mouse. (g) PLA confocal images of α -Erk2/ α -Klf2 interactions in *Klf2*-WT and *Klf2*-null mESCs. Scale bar = 10 μ m. PLA experiments were performed by Kelvin Chong and Dr Andrew Tan (NTU). Figures 24a-e were obtained from Figures 3a-e of Yeo *et al.*, 2014. Figures 24f-g were obtained from Figures S2i-j of Yeo *et al.*, 2014.

3.3.4 – Loss of Klf2 in mESCs result in cell death under 2i

Because mouse embryonic fibroblast feeder-cells are known to produce LIF, the growth of *Klf2*-null mESCs on feeders will complicate the analysis of signalling

pathways. Therefore, *Klf2*-WT and *Klf2*-null mESCs were cultured and adapted onto feeder-free 0.1% gelatin coated surfaces.

It was found that both feeder-free *Klf2*-WT and *Klf2*-null mESC lines could be cultured for more than 15 passages under LIF/Serum conditions without any loss of AP signal (**Fig. 25a**). The overall ESC gene expression signature of feeder-free *Klf2*-null mESCs under LIF/Serum culture was also retained in comparison to *Klf2*-WT mESCs (**Fig. 25b**). These data collectively indicate that *Klf2*-null mESCs phenotypically resemble *Klf2*-WT mESCs during LIF/Serum culture conditions.

Figure 25. Feeder free *Klf2*-null mESCs are phenotypically normal in LIF/Serum (a) Alkaline phosphatase staining of feeder-free *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum. Scale bar = 100μm. **(b)** Gene expression profiles of LIF/Serum cultured feeder-free *Klf2*-WT and *Klf2*-null mESCs. Data represents the average of biological triplicates ± SEM. Figures 25a-b were obtained from Figures 3f-g of Yeo *et al.*, 2014.

However under 2i conditions, it was found that feeder-free *Klf2*-null mESCs could not be cultured for more than 3-4 passages (**Fig. 26a-b**). This collapse of *Klf2*-null mESCs under 2i-culture could however be rescued with the addition of exogenous LIF (**Fig. 26a-b**). It is believed that LIF stimulation serves to activate *Klf4* and *Klf5* expression (Hall *et al.*, 2009), and compensate for the loss of *Klf2*.

The failure to maintain *Klf2*-null mESCs was not due to cell proliferation defects as the cell cycle profile of 2i-cultured *Klf2*-WT and *Klf2*-null mESCs did not differ from each other (**Fig. 27a**). Furthermore, the addition of epidermal growth factor (EGF), a known mESC mitogen (Heo *et al.*, 2006), was unable to restore the propagation defect of 2i-cultured *Klf2*-null mESCs (**Fig. 27b**).

Rather, because 2i-cultured *Klf2*-null mESCs exhibit increased populations of Annexin V positive cells, as well as propidium iodide positive cells, it is proposed that *Klf2*-null mESCs undergo cell death during ground state conditions (**Fig. 27c**).

Figure 26. Feeder free *Klf2*-null mESCs cannot be maintained under 2i conditions. (a) Brightfield images of feeder-free passage 4 *Klf2*-WT and *Klf2*-null mESCs under LIF/Serum, 2i, or 2i+LIF. Scale bar = 200 μ m. (b) Cumulative cell count of feeder-free LIF/Serum, 2i, or 2i+LIF cultured *Klf2*-WT and *Klf2*-null mESCs. Data points represent the average of biological triplicates $\pm$ SEM. Figures 26a-b were obtained from Figures 3h-I of Yeo *et al.*, 2014.

Figure 27. Feeder free *Klf2*-null mESCs undergo cell death under 2i conditions. (a) Cell cycle profile analysis of *Klf2*-WT and *Klf2*-null mESCs cultured under LIF/Serum (L+F), 2i (2i), or 2i+LIF (2i+L) for various passages. (b) Bright-field images of *Klf2*-null mESCs cultured for three passages under LIF/Serum, 2i, 2i+EGF, or 2i+LIF conditions. Scale bar = 200 μ m. (c) Annexin V and propidium iodide FACS analysis of *Klf2*-WT and *Klf2*-null mESCs cultured in 2i for increasing passages. Figure 27c was obtained from Figure 3j of Yeo *et al.*, 2014, and performed by Dr Liang Hongqing (GIS). Figure 27a is unpublished work by Dr Liang Hongqing (GIS), and Figure 27b is unpublished work by Yeo Jia Chi.

3.3.5 – *Klf2* overexpression can restore self-renewal in 2i-cultured *Klf2*-null mESCs

To validate whether the non-viability of *Klf2*-null mESCs during 2i culture is arising specifically from *Klf2* deletion, a rescue experiment via *Klf2* overexpression was performed in *Klf2*-null mESCs. To this end, a *piggybac* transposon system was employed (Wang *et al.*, 2008) to stably introduce 3HA-tagged wild-type *Klf2* transgenes into *Klf2*-null mESCs. Indeed, it was found that *Klf2* overexpression was able to prevent the cell death of *Klf2*-null mESCs during 2i culture for at least 10 passages (**Fig. 28a**). The overexpression of negative control GFP however failed to rescue the cell death of *Klf2*-null mESCs (**Fig. 28a**). Taken together, these data indicate that our observed cell-death phenotype of *Klf2*-null mESCs under 2i was specifically due to the removal of Klf2 protein, and not due to non-specific cell effects.

Figure 28. Overexpression of *Klf2* rescues *Klf2*-null mESC cell death under 2i and confers independence from Mek-inhibition. (a) Brightfield image of *Klf2*-null mESC transfected with *piggybac* (PB) vectors containing GFP or wildtype 3HA-Klf2. Transfected *Klf2*-null mESCs were cultured under 2i for the indicated number of passages. Scale bar = 200µm. (b) Brightfield image of PB-GFP or PB-3HA-Klf2 transfected *Klf2*-null mESCs cultured under CHIR99021 (CH) only for the indicated number of passages. Scale bar = 200µm. Figures 28a-b were obtained from Figures 3k-l of Yeo *et al.*, 2014.

3.3.6 – *Klf2* overexpression confers independence from Mek-inhibition during ground state pluripotency

It has been demonstrated that *Klf2* is the downstream target of active Mek/Erk signalling, and that *Klf2* is necessary for mESCs maintenance in ground state conditions. It is thus expected that *Klf2* overexpression should mimic the effects of Mek-inhibitor PD0325901, and allow for the continuous propagation under “1i” or Gsk3 inhibition alone (CHIR99021 only).

Again using the *piggybac* overexpression system (Wang *et al.*, 2008), it was found that the stable introduction of 3HA-tagged wildtype *Klf2* was sufficient to replace Mek-inhibitor and sustain *Klf2*-null mESC self-renewal for at least 8 passages (**Fig. 28b**). This is in contrast to GFP transfected *Klf2*-null mESCs which died within 2 passages (**Fig. 28b**).

Considering that Erk2 phosphorylates *Klf2*, and that *Klf2* overexpression can replace Mek-inhibitor PD0325901, it would be interesting to study whether abolishing putative Erk-phosphorylation sites on *Klf2* could enhance mESC self-renewal under 1i conditions. *Klf2* constructs bearing various serine/threonine → alanine mutations were then generated to disrupt potential Erk phosphorylation at these sites. On its own, single point mutations of Erk target residues did not confer much advantage over wildtype *Klf2* (data not shown). However, through the overexpression of combinatorial double or quadruple point mutations, it was observed that phospho-deficient *Klf2* had a distinct advantage over wildtype *Klf2* in 1i culture of *Klf2*-null mESC colony forming assays (**Fig. 29a-b**).

Figure 29. Abolition of putative Erk phosphorylation sites on *Klf2* confers a self-renewal advantage for *Klf2*-null mESC in CHIR99021 culture. (a) Cell colony count of *Klf2*-null mESCs stably transfected with *piggybac* *Klf2* containing various serine/threonine point-mutations. Cells were cultured for three passages under CHIR99021. Cell colony numbers were then quantified using ImageJ. Data obtained is the average of biological triplicates $\pm$ SEM. **(b)** Brightfield images of representative *Klf2*-null mESCs stably transfected with various *PB-Klf2* containing various serine/threonine point mutations. Cells were cultured for three passages under CHIR99021 alone. Scale bar = 200 μ m. Figures 29a-b were obtained from Figures S2k-l of Yeo *et al.*, 2014.

3.3.7 – *Klf2* is important for the establishment of mESCs under 2i conditions

With the failure of *Klf2*-null mESCs to be maintained under 2i ground state conditions (**Fig. 26a, 27c**), it was thus interesting to examine if *Klf2*-null mESC lines could be derived directly from blastocysts using 2i media. The 2i culture system has thus far been a superior method in the derivation of mESCs, and has been successfully used to derive mESCs from recalcitrant CBA, MF1 and NOD mouse

strains (Hanna *et al.*, 2009; Ying *et al.*, 2008), as well as the cross-species rat ESCs (Buehr *et al.*, 2008; Li *et al.*, 2008).

Through the crossing of *Klf2*-Het parents, a total of 85 *Klf2* blastocysts were harvested from 14 separate matings. From these blastocysts, a total of 41 mESCs were established under 2i + feeder conditions. After genotyping, it was interestingly observed that wildtype and heterozygous *Klf2* mESCs were established at a 1:1 sub-mendelian ratio, and that only a single *Klf2*-null mESC was obtained under 2i (**Fig. 30**). More importantly, this abnormally skewed 2i-derived *Klf2* mESC genotype frequency deviates considerably from the expected 1:2:1 genotype ratio from Het × Het matings (χ^2 value = 17.634, p-value < 0.001) (**Fig. 23b**). Therefore, the inefficiency in generating *Klf2*-null and *Klf2*-Het mESCs under 2i emphasizes the necessity of *Klf2* during ground state pluripotency.

Genotype	No. of ESCs
<i>Klf2</i> ^{+/+}	20
<i>Klf2</i> ^{+/-}	20
<i>Klf2</i> ^{-/-}	1

Figure 30. *Klf2* is an important factor for the establishment of ground state pluripotency. Summary of genotyped mESC lines obtain via 2i from *Klf2*^{+/-} × *Klf2*^{+/-} crosses. χ^2 value = 17.634, p<0.001. The obtained *Klf2*-null mESC genotype under 2i conditions is significantly different from the expected 1:2:1 Mendelian ratio at p<0.001. Figure 30 is obtained from Figure 3m of Yeo *et al.*, 2014.

3.4 – *Klf2*-null mESCs exhibit aberrant PGC gene expression under 2i conditions

3.4.1 – *Klf2*-null mESCs retain a pluripotent signature despite massive apoptosis under 2i conditions

Despite the massive apoptosis under passage 3 of 2i conditions, quantitative real-time PCR revealed that *Klf2*-null mESCs had similar expression levels of key pluripotency genes such as *Oct4* and *Sox2*; as compared to *Klf2*-WT mESCs (**Fig. 31a**). Likewise, expression of other important pluripotency factors such as *Klf4*, *Klf5*, *Nanog* and *Rex1* were retained in 2i-cultured *Klf2*-null mESCs (**Fig. 31a-b**). However, it was observed that *Klf4*, *Klf5* and *Nanog* expression were higher in 2i-cultured *Klf2*-null mESCs (**Fig. 31a-b**); and that *Rex1* expression was slightly reduced in ground state *Klf2*-null mESCs (**Fig. 31b**). Due to the cell death phenotype and limited amounts of protein available from 2i-cultured *Klf2*-null mESCs, it is difficult to establish if there is any enhanced *Klf4* or *Klf5* activity to compensate for the loss of *Klf2* during ground state conditions.

Indeed, subsequent microarray of 2i cultured *Klf2*-null mESCs at passage 3 did not reveal gene ontology (GO) enrichment for terms relating to the loss of pluripotency or lineage differentiation (**Fig. 31c**) (See **Appendix A-B** for the list of top 200 gene changes). Rather, among the upregulated and downregulated genes in *Klf2*-null mESCs under 2i passage 3, we found an enrichment of GO-terms relating to cellular and metabolic processes (**Fig. 31c**). Not surprisingly, due to the apoptotic phenotype of *Klf2*-null mESCs at passage 3 of 2i culture, GO-terms such as “death” and “cell death” were found among the list of upregulated genes. Taken together, in spite of the

massive apoptosis, the results suggest that *Klf2*-null mESCs still retained a pluripotent gene expression signature under 2i culture.

Figure 31. *Klf2*-null mESCs retain a pluripotent signature despite undergoing apoptosis in 2i conditions. (a) *Klf* gene expression profile of *Klf2*-WT and *Klf2*-null mESCs under 2i passage 3. Data represents the average of biological triplicates $\pm$ SEM. (b) Gene expression profile of *Oct4*, *Sox2*, *Nanog* and *Rex1* in *Klf2*-WT and *Klf2*-Null mESCs under 2i passage 3. Data represents the average of biological triplicates $\pm$ SEM. (** p-value $\leq$ 0.001; * p-value $\leq$ 0.05; ND – no difference) (c) GO analysis of upregulated and downregulated genes in *Klf2*-null mESCs in 2i passage 3. Figures 31a-c were obtained from Figures S3a-c of Yeo *et al.*, 2014.

3.4.2 – *Klf2*-null mESCs under 2i exhibit aberrant PGC gene expression

While the global transcriptomic changes of *Klf2*-null mESCs under 2i passage did not reveal any loss of mESC pluripotency or lineage differentiation (**Fig. 31**), a closer inspection of the list of gene alterations revealed elevated levels of several PGC related genes such as *Dppa3*, *c-Kit*, *Tcfap2c*, *Dnd1*, *Ifitm1* and *Prdm1* in *Klf2*-null mESCs over *Klf2*-WT mESCs (**Fig. 32a-b**) (Kurimoto *et al.*, 2008; Magnusdottir and Surani, 2014). Due to the previous synergistic association of *Klf2* with the PGC factor *Prdm14* to rapidly convert EpiSCs into naïve pluripotent cells (Gillich *et al.*, 2012), and the association of *Klf2* with PGC development (Kurimoto *et al.*, 2008), it was hence of particular interest to further examine this *Klf2*-PGC relationship during ground state pluripotency.

Figure 32. *Klf2*-null mESCs in 2i exhibit elevated PGC gene expression. (a) Microarray heat-map of PGC genes levels in *Klf2*-WT and *Klf2*-null mESCs in 2i culture passage 3 (P3). Red and green boxes denote increased or decreased gene expression respectively. (b) Gene expression analysis of various PGC genes in *Klf2*-WT and *Klf2*-null mESCs in 2i (P3). Data is the average of biological triplicates $\pm$ SEM. Figures 32a-b were obtained from Figures S3d-e of Yeo *et al.*, 2014.

It is known that *Klf2*-null mESCs at 2i passage 3 have already started to undergo apoptosis (**Fig. 27c**). As such, it would have been difficult to determine at this late state (2i passage 3) if the observed PGC gene transcriptional changes in **Fig. 32** were a direct consequence of *Klf2* ablation; or whether it was an artefact of mESC undergoing apoptosis.

To address this issue, a separate time-course microarray was performed on *Klf2*-null mESCs in LIF/Serum, 2i passage 1, and 2i passage 2 (**Fig. 33a**). These earlier cell culture stages were prior to the major onset of *Klf2*-null mESC cell death (**Fig. 26a and 27c**); and as such, allowed us to directly examine the PGC transcriptional changes.

Indeed, time-course microarray confirmed that *Klf2*-null mESCs under 2i conditions exhibited elevated PGC gene expression; and that these changes occurred as early as the initial 2i passage 1 (**Fig. 33a**). However, unlike *Klf2*-WT mESCs which gradually downregulate PGC gene expression relative to the starting LIF/Serum PGC transcript levels (**Fig. 33b; blue line**), it was found that *Klf2*-null mESCs continued to sustain PGC transcription (**Fig. 33b; red line**).

This defective PGC gene suppression in *Klf2*-null mESCs under 2i conditions led us to speculate that *Klf2* may be playing a role in the proper silencing of PGC gene expression in mESCs.

Figure 33. *Klf2*-null mESCs in 2i fail to downregulate PGC expression. (a) Timecourse microarray heat-map of PGC genes levels in feeder-free *Klf2*-WT and *Klf2*-null mESCs in 2i culture passage 1 (P1), passage 2 (P2). Red and green boxes denote increased or decreased gene expression respectively. (b) Relative gene expression of PGC genes in feeder-free *Klf2*-WT and *Klf2*-null mESCs in various 2i passages. Data is the average of biological triplicates $\pm$ SEM. Figures 33a-b were obtained from Figures 4a-b of Yeo *et al.*, 2014.

3.4.3 – *Klf2* binds to the promoters of key PGC genes

By analyzing the *Klf2* ChIP-seq data from LIF/Serum or 2i-cultured mESCs data, *Klf2* binding could be detected at the promoters of key PGC factors such as *Prdm1*, *Tcfap2c*, and *c-Kit* (**Fig. 34a**). This binding of *Klf2* at PGC promoters was also validated using ChIP-qPCR (**Fig. 34b**), indicating that *Klf2* could be involved in the regulation of PGC genes.

While the level of Klf2 binding at PGC genes was similar in both LIF/Serum and 2i conditions (**Fig. 34a**), it was observed that Klf4 binding at these same genes was decreased only under 2i conditions (**Fig. 34c**). This meant that during LIF/Serum conditions, both Klf2 and Klf4 could bind at PGC genes. However during ground state conditions, due to the absence of LIF-stimulation and reduced Klf4 protein levels (**Fig. 11**), only Klf2 is left binding at PGC gene promoters.

Therefore in the context of 2i-cultured *Klf2*-null mESCs, the further loss of *Klf2* from gene knockout would result in the failure of Klf-mediated PGC gene regulation. We hypothesize that due to the extremely restrictive nature of 2i ground state culture, any *Klf2*-null mESCs which may be initiating PGC differentiation would not have been able to survive. This stringency for pluripotent mESCs during 2i culture may thus explain our observed *Klf2*-null mESC cell death phenotype in **Fig. 27c**.

Because 2i-cultured *Klf2*-null mESCs can be rescued by the addition of exogenous LIF (**Fig. 26a**), it is believed that LIF will induce the expression *Klf4* to compensate for the absence of Klf2 protein. While it has not been experimentally validated, it is hypothesized that LIF-induced Klf4 would then be able to bind and suppress aberrant PGC gene expression in *Klf2*-null mESCs.

Figure 34. Klf2 binds the promoters of key PGC genes. (a) Klf2 ChIP-seq peaks at the promoters of key PGC genes in mESCs during LIF/Serum or 2i culture passage 3 (P3). Klf2 ChIP-seq tracks were mapped by Dr Jonathan Goke (GIS) (b) Klf2 ChIP-qPCR validation of Klf2 binding at key PGC genes in mESCs under LIF/Serum or 2i (P3). Data is the average of biological triplicates $\pm$ SEM (c) Klf4 ChIP-qPCR at key PGC genes in mESCs under LIF/Serum or 2i (P3). Data is the average of biological triplicates $\pm$ SEM. ChIP-qPCR experiments were performed by Dr Ng Jia Hui (GIS). Figure 34a was obtained from Figure 4c of Yeo *et al.*, 2014. Figure 34c was obtained from Figure S4f of Yeo *et al.*, 2014. Figures 34b is unpublished work by Dr Ng Jia Hui (GIS).

While the binding of Klf2 at PGC genes suggests a possible role of Klf2 in PGC transcriptional regulation during 2i ground state pluripotency, the mechanism of this hypothesized Klf2 repression of PGC genes is still unclear. It should also be highlighted that while Klf2 is considered a PGC-associated factor (Kurimoto *et al.*, 2008), mESCs and PGCs are two different cell types, each with their own unique transcriptional regulatory circuitry. It would hence be premature to assume that the repressive activity of Klf2 on PGC gene transcription in mESCs is similar to the function of Klf2 during PGC development.

CHAPTER IV

DISCUSSION

Chapter IV: Discussion

4.1 – Model of Klf regulation in LIF/Serum and 2i mESC conditions

In this thesis, Klf2 is found to be an important transcription factor towards the establishment and sustenance of mESCs during 2i ground state pluripotency. Importantly, it was found that Klf2 is negatively regulated at the post-translational level by Mek/Erk signalling. As such, the inhibition of the Mek/Erk pathway will result in the stabilisation of Klf2 protein, which in turn is able to sustain mESC self-renewal in the absence of exogenous LIF to induce Klf4/Klf5 expression. In the light of the research data, a model of Klf protein regulation under conventional LIF/Serum or during 2i ground state pluripotency is proposed (**Fig. 35**).

During LIF/Serum conditions, both the LIF-signalling and the pro-differentiation Fgf/Mek/Erk signalling pathways are active. While Fgf-signalling is antagonistic towards the stability of Klf2 protein, not all of the Klf2 protein would be degraded. As a result, the Klf2, Klf4 and Klf5 proteins in LIF/Serum cultured mESCs will exhibit a triple-redundancy effect as previously reported (Jiang *et al.*, 2008). However under 2i conditions, due to the absence of LIF stimulation, the levels of Klf4 and Klf5 are insufficient to support mESC self-renewal. Therefore, the inhibition of Mek/Erk signalling is required to halt Klf2 phospho-dependent protein degradation. Elevated Klf2 protein levels under 2i would then be able to sustain mESC self-renewal (**Fig. 35**).

Figure 35. Model of Klf2 protein regulation by active Mek/Erk signalling pathways during conventional LIF/Serum or 2i-induced ground state pluripotency. Figure 35 is obtained from Figure 4d of Yeo *et al.*, 2014.

4.2 – Varying potencies of Klf2, Klf4 and Klf5 in mESCs

Although a functional redundancy of Klf2, Klf4 and Klf5 has been previously reported in mESCs (Jiang *et al.*, 2008), not all of these Klf proteins exhibit equal potency.

For instance, *Klf2* overexpression was most effective in supporting LIF-independent and Stat3^{-/-} mESC growth under serum culture conditions; and that Klf2 could revert developmentally “primed” EpiSCs into mESCs at a two-fold higher rate than Klf4 (Hall *et al.*, 2009). Similarly, Klf2 is a more efficient reprogramming factor than Klf1, Klf4 or Klf5 during the conversion of fibroblasts into iPSCs (Feng *et al.*, 2009; Nakagawa *et al.*, 2008). Furthermore, it was found that only Klf2, but not Klf4 or Klf5, had a synergistic effect with the PGC factor Prdm14 to rapidly drive to conversion of primed EpiSCs into naïve mESCs (Gillich *et al.*, 2012). Therefore, it appears that among the three Klf proteins, Klf2 is the strongest in promoting a mESC identity, with Klf4 showing intermediate and Klf5 having the least activity.

With the similarities in the Klf2, Klf4 and Klf5 consensus DNA-binding motif (Jiang *et al.*, 2008), and the greater than 80% protein sequence homology within the DNA-binding domains of Klf2, Klf4, and Klf5 (Hall *et al.*, 2009), the exact mechanism for these functional differences among the three Klf proteins however is still not clear.

As the non DNA-binding region of a separate Klf3 were found to be important for its correct target specificity (Burdach et al., 2014), a plausible explanation behind the varying potencies of Klf2, Klf4 and Klf5 may lie in their non DNA-binding regions. In terms of their full protein sequence, Klf2 and Klf4 protein shares 46% similarity, while Klf5 being the most divergent, shares only 29% and 24% homology with Klf2 and Klf4 respectively (**Fig. 36**) (Hall *et al.*, 2009).

Due to the above differences, it is hypothesized that the non DNA-binding regions of Klf2, Klf4, and Klf5 may each have their unique and distinctive set of protein partners which could influence their biological activity. As there has not been any prior studies done to confirm this prediction, it would be of interest to perform a proteomic screen for the protein interacting partners of Klf2, Klf4 and Klf5 in mESCs.

Protein comparison		Conservation (%)	
		Total protein	Zinc finger domains (ZFDs)*
Klf2 (354 aa)	Klf4 (483 aa)	46%	90%
Klf2 (354 aa)	Klf5 (446 aa)	29%	80%
Klf4 (483 aa)	Klf5 (446 aa)	26%	81%

Figure 36. Protein homology between Klf2, Klf4 and Klf5. Table summarizing the percentage alignment for pairwise comparison over the full-length protein, or the zinc finger domains. (*) Alignment of the terminal 82/83 residues that comprise the three zinc finger domains. The figure and data is obtained from Hall *et al.*, 2009.

4.3 – Post-translational modification of Klf2 by the Mek/Erk signalling pathway

Regulation of Klf transcriptional activity can occur via post-translational modifications. Klf4 and Klf5 for instance are known to be acetylated, phosphorylated, ubiquitinated and sumoylated (McConnell and Yang, 2010). Such post-translational modifications will either alter the Klf protein half-life, or affect its protein-protein interaction with other transcriptional regulators to bring about the desired gene regulatory effect (McConnell and Yang, 2010).

With regards to Klf2, it is known that Klf2 can be ubiquitinated and proteasomally degraded by the E3-ligases, WWP1 (Zhang *et al.*, 2004) and Smurf 1 (Xie *et al.*, 2011). It is also known that Klf2 is phosphorylated in mESCs (Li *et al.*, 2011; Pines *et al.*, 2011), and that Gsk3 phosphorylation of Klf2 in endothelial cells could specifically result in FBW7-mediated Klf2 degradation (Wang *et al.*, 2013).

Based on the biochemical data in mESCs, Klf2 protein has been for the first time directly linked with the Mek/Erk signalling pathway (**Fig. 13**). Specifically, it was found that Klf2 is phosphorylated by Erk2 (**Fig. 17-20**), and the phospho-Klf2 is degraded in the proteasome (**Fig. 15**). Whether Klf2 protein in mESCs is ubiquitinated by WWP1, Smurf1, or SCF(FBW7) as previously reported (Wang *et al.*, 2013; Xie *et al.*, 2011; Zhang *et al.*, 2004), or if Klf2 degradation is dependent upon other yet-to-be identified factors is still not clear.

Through motif analysis and mining of mass spectrometry data, candidate Mek/Erk phosphorylation sites on Klf2 at the T171, S175, S246 and S247 residue (**Fig. 16**).

Indeed, determining the functional relevance of these phosphorylation sites may reveal the mechanism in which Klf2 supports mESC pluripotency.

Figure 37. Diagram of the various protein domains in mouse Klf2 protein. Numbers denote their respective amino acid positions. Unpublished work by Yeo Jia Chi.

Klf2 can be divided into three functional regions – a) the activation domain (residues 1-110); b) the inhibitory domain (residues 111-267); and c) the DNA-binding zinc-finger domain (residues 268-354) (**Fig. 37**) (Conkright *et al.*, 2001). Therefore, our predicted putative T171, S175, S246 and S247 phospho-residues are located within the Klf2 inhibitory domain, and could play suggest an autoregulatory role, or function as a binding site for other co-regulatory proteins. Indeed, such co-binding with Klf2 has been previously reported for the E3 ubiquitin ligase WWP1 and Smurf1 (Xie *et al.*, 2011; Zhang *et al.*, 2004) and well as the p300/CBP transcriptional co-activators to modulate NF- κ B expression (SenBanerjee *et al.*, 2004).

Mass spectrometry data also indicated Klf2 phosphorylation at the T243 position (**Fig. 16a**) (Li *et al.*, 2011), which when combined with the S247phosphosite, corresponds with the (S/T)-X-X-X-(S/T)-P consensus sequence of GSK3 (Doble and Woodgett, 2003). Although Gsk3-inhibition by CHIR99021 did not induce Klf2

protein stabilisation unlike Mek-inhibition (**Fig. 13**), it is possible that the T243 site could have other functions for regulating Klf2 activity.

4.4 – The role of Klf2 in pre-implantation embryonic development

2i ground state mESCs have been proposed to resemble the *in vivo* pre-implantation epiblast cells (Boroviak *et al.*, 2014; Nichols *et al.*, 2009; Nichols and Smith, 2009), and recent experimental evidence via transcriptional analysis of ground state mESCs with epiblast cells has supported this idea (Boroviak *et al.*, 2014). However, because *Klf2*-null ICMs are functional, and are able to develop till the E12.5-E14.5 stage (Kuo *et al.*, 1997; Wani *et al.*, 1998); our failure to efficiently derive *Klf2*-null mESCs from *Klf2* knockout blastocysts under 2i appears to produce a discrepancy (**Fig. 30**). Specifically, if Klf2 were important for ground state pluripotency as proposed by this thesis, and that if ground state mESCs resemble the E4.5 epiblast cells, *Klf2*-null embryos should be defective in forming a pluripotent ICM.

To account for these discrepancies, there are two possible explanations. First, although mESCs are derived from the pre-implantation ICM, the actual developmental duration for this ICM stage is very transient. As such, any potential defects which arise from the absence of Klf2 may not manifest within this timeframe. Indeed, this idea of a delayed action may be true as it required two passages of 2i culture (approximately 5-6 days) before *Klf2*-null mESCs start to undergo apoptosis (**Fig. 26, 27c**). In addition, the compensatory effect of Klf4 and Klf5, which is

expressed in high levels in the E3.5 and E4.5 ICM (Jiang *et al.*, 2008), may also compensate for the loss of Klf2 during this duration.

Secondly, the developing mouse blastocysts is known to produce LIF (Nichols *et al.*, 1996), which may also serve to rescue the loss of Klf2 *in vivo* through the activation of Klf4 and Klf5. This hypothesis is supported by our observation that LIF addition into 2i-cultured Klf2-null mESCs was able to rescue the cell-death phenotype (**Fig. 26, 27c**).

4.5 – The potential role of Klf2 in mESC metabolic regulation

During ground state pluripotency, RNA-seq has found that mESCs will upregulate the expression of metabolic genes (Marks *et al.*, 2012). Furthermore during 2i conditions, it was recently observed that the entry of exogenous glutamate into the mitochondrial tricarboxylic acid (TCA) cycle in 2i-cultured mESCs is repressed (Carey *et al.*, 2014).

Therefore in order to maintain sufficient glutamate for its cellular activity, ground state mESCs would increase the production of glucose-derived α -ketoglutarate (α KG) via the TCA cycle. Some of these α KG would then be rerouted for *de novo* glutamate synthesis (**Fig. 38**) (Carey *et al.*, 2014). However as excess α KG is being accumulated during 2i-culture, the activity of α KG-dependent dioxygenase such as the Jumonji C (JmjC)-domain-containing histone demethylases and Tet family of DNA demethylases would consequently be enhanced (Kaelin, 2011). Overall, this phenomenon is reported to favour the demethylation of repressive DNA and histone

marks, and contributes to the regulation of pluripotency-associated gene expression (Carey et al., 2014).

Figure 38. Diagram of the tricarboxylic acid (TCA) cycle during 2i-culture. During ground state pluripotency, the pathway to convert exogenous glutamine into α KG is suppressed. Instead, elevated levels of α KG are produced from glucose via the TCA cycle. Due to this metabolic reprogramming, the high levels of α KG will in turn enhance the activity of α KG-dependent enzymes to influence chromatin structure and gene expression. Figure adapted and modified from Carey *et al.*, 2014.

Although *Klf2*-null mESCs have aberrant upregulation of PGC genes during 2i conditions (**Fig. 33a**), time-course microarray data also found that *Klf2*-null mESCs exhibit large changes to their metabolic gene expression (**Fig. 39**). This apparent metabolic dysregulation during the absence of *Klf2* lead us to speculate that in addition to preventing PGC differentiation, *Klf2* could also be implicated in mESC metabolism. To determine if *Klf2* could potentially bind and regulate metabolic genes, *Klf2* ChIP-seq data from LIF/Serum and 2i-cultured mESCs (Yeo et al., 2014), were mapped and overlapped with each other.

Figure 39. GO analysis of *Klf2*-null mESC microarray data during LIF/Serum or 2i-culture. Microarray data from *Klf2*-WT and *Klf2*-null mESCs under LIF/Serum, 2i passage 1, or 2i passage 2 were analysed to identify for enriched GO terms. The top 5 GO processes (ranked by p-values) are listed. Unpublished work by Yeo Jia Chi.

Among the *Klf2*-bound genes, a total of 9281 of them were common *Klf2*-bound genes in both LIF/Serum and 2i-culture (**Fig. 40**). Separately, 125 *Klf2*-bound genes were LIF/Serum specific (~1.3% of all *Klf2* LIF/Serum bound genes); and 1816 genes were 2i-culture specific (~14.8% of all *Klf2* 2i bound genes) (**Fig. 40**). Importantly, GO analysis of all *Klf2*-bound genes under different cell culture conditions revealed a strong enrichment of terms relating both embryonic development and metabolic processes (**Fig. 40**).

Because Klf2 is known to be involved in the regulation of mESC pluripotency and other embryonic developmental processes (Hall et al., 2009; Jiang et al., 2008; McConnell and Yang, 2010), the discovery of Klf2 binding at development genes was not unexpected. However, the presence of Klf2 at a majority of metabolic genes (**Fig. 40**) strongly implicates that Klf2 may also play additional functions in mESC metabolic regulation.

Figure 40. ChIP-seq GO analysis of Klf2 bound genes in LIF/Serum vs. 2i-cultured mESCs. The top 10 GO terms (ranked by p-values) are listed. Klf2 ChIP-seq data for LIF/Serum and 2i-cultured mESCs was obtained from this study (Yeo et al., 2014). Unpublished work by Yeo Jia Chi.

Interestingly, this regulation of metabolic and developmental genes is a common feature of all three Klf2, Klf4 and Klf5 proteins in mESCs under LIF/Serum conditions. Of the 8691 commonly bound genes by Klf2, Klf4 and Klf5, GO analysis revealed the presence of various metabolic and organismal developmental processes (**Fig. 41: red box**); suggesting that Klf2, Klf4 and Klf5 may regulate them in a common manner during LIF/Serum culture. Probing the mESC metabolome following gene knockdown/knockout of *Klf2*, *Klf4* and *Klf5* may reveal new insights into the potential pathways involved.

Of note, among the 8691 common Klf-bound genes, this comprised 92.1% of all Klf2-bound genes; 70.7% of all Klf4-bound genes; and 60.8% of all Klf5-bound genes. This data meant that a substantial proportion of Klf4 and Klf5 binding were not shared with Klf2 (**Fig. 41**). Therefore, it is possible that Klf4 and Klf5 may have other functions in mESC biology.

It should however be noted that although Klf2 ChIP-seq for 2i-cultured mESCs is available (Yeo et al., 2014), Klf4/Klf5 ChIP-seq in ground state mESCs has not yet been performed. As such, the analysis of Klf2, Klf4 and Klf5 genomic binding sites should only be interpreted in the context of conventional LIF/Serum culture.

Figure 41. ChIP-seq GO analysis of Klf2, Klf4 and Klf5 bound genes in LIF/Serum cultured mESCs. The top 5 GO terms (ranked by p-values) are listed. The red box denotes the GO terms of the commonly bound Klf2, Klf4 and Klf5 genes. Klf2 ChIP-seq data was obtained from this study (Yeo et al., 2014), while the Klf4/Klf5 ChIP-seq data were obtained from Aksoy *et al.* (Aksoy et al., 2014). Unpublished work by Yeo Jia Chi.

4.6 – The association of KLF2 with naïve hESC pluripotency

Due to the developmentally “primed” nature of hESCs, which resemble mouse EpiSC state rather than mESCs (Brons et al., 2007; Tesar et al., 2007); there is great interest for ESC and developmental biologists to examine the possibility of a common pluripotent ground state in mammals (Pera, 2014).

While there has been several early attempts to generate hESC or human iPSCs (hiPSCs) with naïve mESC pluripotency features (Buecker et al., 2010; Hanna et al., 2010a; Lengner et al., 2010; Li et al., 2009; Wang et al., 2011; Xu et al., 2010); these “naïve-like” hESCs were either reliant on sustained exogenous transgenes for cell

culture maintenance, or have been inadequately characterised for naïve pluripotency features (Reviewed by Yeo and Ng, 2013).

Given the above mentioned experimental weaknesses, there has recently been a string of independent publications describing an improved method of generating naïve-like hESCs using minimal or no exogenous transgene support (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014; Valamehr et al., 2014; Ware et al., 2014). By utilizing a combination of small molecules and growth factors, these 6 studies were able to create their own unique naïve-like hESC state with slightly varying molecular and cellular features (**Fig. 42**).

	Gafni <i>et al.</i> (2013)	Chan <i>et al.</i> (2013)	Valamehr <i>et al.</i> (2014)	Ware <i>et al.</i> (2014)	Theunissen <i>et al.</i> (2014)	Takashima <i>et al.</i> (2014)
Inhibitors	MEKi (PD0325901)	MEKi (PD0325901)	MEKi (PD0325901)	MEKi (PD0325901)	MEKi (PD0325901)	MEKi (PD0325901)
	GSK3i (CHIR99021)	GSKi (BIO)	GSKi (CHIR99021)	GSKi (CHIR99021)	GSKi (CHIR99021)	GSKi (CHIR99021)
	JNKi					
	P38i (SP600125/SB203580)					
	PKCi (Gö6983)					PKCi (Gö6983)
	ROCKi (Y-27632)		ROCKi (Thiazovivin)		ROCKi (Y-27632)	
		BMPi (Dorsomorphin)				
					BRAFi (SB590885)	
					SRCi (WH-4-023)	
Growth factors	bFGF		bFGF	bFGF		
	TGFβ				Activin	
	hLIF	hLIF	hLIF	hLIF	hLIF	hLIF
Basal media	Albumax + N2 or 20% KSR	mTesr1 (bFGF+TGFβ)	20% KSR	20% KSR	N2 + B27	N2 + B27
Naïve transcripts expressed	<i>NANOG</i> , <i>DUSP</i> family	<i>NANOG</i> , <i>DPPA3</i> , <i>KLF4</i> and <i>TBX3</i>	<i>KLF2</i> , <i>KLF4</i> , <i>DPPA3</i> , <i>TBX3</i> , <i>GDF3</i> and <i>ESRRB</i>	<i>miR-581b</i> , <i>miR-520b</i> and <i>miR-250c</i>	<i>NANOG</i> , <i>KLF2</i> , <i>KLF4</i> , <i>REX1</i> , <i>DPPA3</i>	<i>NANOG</i> , <i>KLF4</i> , <i>TFCP2L1</i>
Epigenetic status (Histones)	Reduction of H3K27me3 at developmental genes	Reduction of H3K27me3 at human epiblast-specific genes	Not examined	Reduction of H3K27me3 at developmental genes	Reduction of H3K27me3 at bivalent genes	Not examined
Epigenetic status (DNA methylation)	Downregulation of DNMT3A, DNMT3B, DNMT3L Maintenance of DNMT1 and TET enzymes	Downregulation of DNMT3L	Not examined	Not examined	Not examined	Global DNA hypomethylation. Downregulation of DNMT3B. Upregulation of DNMT3L.
X-chromosome status	XaXa	Not tested	Unclear. Increase in X-chromosome gene expression but <i>XIST</i> still present	XaXa	XaXi	XaXa
Aerobic respiration	Not tested	Not tested	Not tested	Yes	Not tested	Yes
Teratoma formation	Yes	Yes	Yes	Yes	Yes	Yes
Mouse-human chimerism	Chimerism with mice till E10.5	Not tested	Not tested	Not tested	No chimeras obtained	Integration with mature mouse ICM

Figure 42. Methods to generate naïve-like hESCs. The molecular and cellular features of these naïve-like hESCs are as summarized from (Chan *et al.*, 2013; Gafni *et al.*, 2013; Takashima *et al.*, 2014; Theunissen *et al.*, 2014; Valamehr *et al.*, 2014; Ware *et al.*, 2014). Unpublished work by Yeo Jia Chi.

Despite these reported divergences (**Fig. 42**), transcriptomic analysis found that these various naïve-like hESCs share common gene modules relating to RNA processing, ribosome biogenesis, and mitochondrial metabolism (Huang et al., 2014). Indeed, increased mitochondrial activity was described in several naïve-like hESCs (Takashima et al., 2014; Ware et al., 2014). Taken together, these data suggests that unlike the highly glycolytic primed EpiSC/hESCs, naïve human pluripotency may entail a switch into oxidative respiration similar to that found in naïve mESCs (Zhou et al., 2012b).

Because Klf2 appears to be regulating mESC metabolism (**Fig. 39-40**), and that different naïve-like hESCs all share a common metabolic gene expression module (Huang et al., 2014); it is plausible that human KLF2 may also be involved in naïve hESC pluripotency. Further hints for the role of KLF2 in naïve human pluripotency may be further inferred from two additional observations.

Firstly, among the different methodologies, a common component for the derivation of naïve hESCs is the Mek-inhibitor PD0325901 (**Fig. 42**). Assuming that the phospho-degradation of Klf2 by Mek/Erk signaling is conserved between mice and humans; this unanimous inclusion of PD0325901 could indicate that hESCs may be reliant on the stabilization of human KLF2 protein to sustain naïve pluripotency.

Secondly, while naïve-like hESC could be achieved through a variety of means (**Fig. 42**), it was found that only hESCs that were derived through transient *KLF2/NANOG* expression; or chemical/growth-factor treatment to mimic *KLF2/NANOG*-induced *OCT4* distal enhancer activation; had the closest resemblance to the human pre-

implantation ICM transcriptome (Huang et al., 2014; Takashima et al., 2014; Theunissen et al., 2014). Although the process by which *KLF2/NANOG* establishes naïve hESCs is still unclear, their combined ability to stringently recreate a naïve pluripotent state, akin to the human ICM, suggests an important role for these transcription factors.

Taken together, while naïve pluripotency appears to be achievable in hESCs, more work is clearly needed to delineate the transcriptional and signaling requirements responsible for maintaining this state. Given the 1) possible metabolic regulatory role of Klf2 in mESCs, the 2) robust reversion of naïve human pluripotency using *KLF2/NANOG*, and 3) the common use of Mek-inhibitor to create a naïve hESC state; these three points collectively argue for an important role for KLF2 in the regulation of naïve human pluripotency.

4.7 – Future Perspectives and Conclusion

Moving forward, although Klf2 has been demonstrated to be an essential factor for ground state pluripotency, further research is clearly needed to determine how Klf2 functions at the protein level to promote ground state pluripotency.

Firstly, as Klf2 has been implicated with mESC metabolism (**Fig. 39-40**), it would be of great value to perform a liquid chromatography/tandem mass spectrometry (LC-MS/MS) (Zhou et al., 2012a) to profile the metabolome of *Klf2*-null mESCs as they transit from LIF/Serum to ground state pluripotency. Indeed, identifying metabolite increases or decreases may indicate any potential metabolic pathways which may be regulated by Klf2.

Secondly, the identification of Klf2 protein-binding partners is useful to understand how Klf2 is able to sustain mESC self-renewal. Specifically, quantitative mass spectrometric methods such as SILAC can be employed to compare between LIF/Serum or 2i-cultured mESCs to observe how the different signaling pathways in mESCs could affect Klf2 protein interactions (Blagoev et al., 2003). Likewise, quantitative proteomics may also be employed to identify Klf2 post-translational modifications during LIF/Serum or 2i-culture that could modulate Klf2 activity (Choudhary and Mann, 2010).

Thirdly, as naïve hESCs differ slightly from naïve mESCs due to the absence of *ESRRB* expression; the naïve human pluripotency is “more fragile” than that of mESCs (Takashima et al., 2014). As such, even though KLF2 and KLF4 are both

expressed in naïve hESCs, the single disruption of *KLF4* is sufficient to impair hESC colony growth (Takashima et al., 2014). This finding indicates that unlike the triple redundancy effect of *Klf2*, *Klf4* and *Klf5* in mESCs (Jiang et al., 2008), human *KLF2* and *KLF4* proteins may not share overlapping functions in naïve hESCs. Given the potential cross-species differences between mouse and humans, it would hence be interesting to delineate the transcriptional and signaling requirements of human KLF proteins in naïve hESC pluripotency.

Lastly, it is known that FGF-signalling inhibition in early mouse embryos results in the blockade of primitive endoderm differentiation and the formation of an all *Nanog*-positive epiblast ICM (Nichols *et al.*, 2009; Yamanaka *et al.*, 2010). As such, with *Klf2* being the downstream target of Mek/Erk signalling, it would thus be interesting to study how *Klf2* deletion could affect this “salt and pepper” arrangement and segregation of *Nanog* and *Gata6*-positive ICM cells (Chazaud *et al.*, 2006). In addition, the ease of CRISPR/Cas9 to simultaneously disrupt multiple genes (Wei *et al.*, 2013) would allow us to examine the effects of *Klf2*, *Klf4* and *Klf5* functional redundancy and compensation during ICM development.

In conclusion, this study sought to uncover how Mek-inhibition is important for ground state pluripotency. In this regard, it was found that *Klf2* is negatively regulated by active Mek/Erk signalling, and that Mek-inhibition by PD0325901 serves to halt *Klf2* protein degradation in mESCs. Importantly, it was found that *Klf2* is a critical factor for the establishment and maintenance of 2i ground state pluripotency. As such, together with the knowledge that Gsk3-inhibition serves to

relieve Tcf3-mediated *Esrrb* repression (Martello *et al.*, 2012), our new data on Klf2 presents a complete picture into how the simultaneous inhibition of two pro-differentiation pathways in mESCs can create the pluripotent ground state (**Fig. 43**) (Yeo *et al.*, 2014).

Figure 43. Model of Klf2 function during ground state pluripotency. Figure obtained from the graphical abstract of Yeo *et al.*, 2014.

CHAPTER V

REFERENCES

Chapter V: References

- Aksoy, I., Giudice, V., Delahaye, E., Wianny, F., Aubry, M., Mure, M., Chen, J., Jauch, R., Bogu, G.K., Nolden, T., *et al.* (2014). Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nat Commun* 5, 3719.
- 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.
- Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* 461, 1292-1295.
- Beddington, R.S., and Robertson, E.J. (1999). Axis development and early asymmetry in mammals. *Cell* 96, 195-209.
- Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., and Mann, M. (2003). A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* 21, 315-318.
- Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol* 16, 516-528.
- 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., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947-956.
- 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.
- Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., *et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, 191-195.
- Brook, F.A., and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A* 94, 5709-5712.
- Bryja, V., Bonilla, S., Cajanek, L., Parish, C.L., Schwartz, C.M., Luo, Y., Rao, M.S., and Arenas, E. (2006). An efficient method for the derivation of mouse embryonic stem cells. *Stem Cells* 24, 844-849.
- Buecker, C., Chen, H.H., Polo, J.M., Daheron, L., Bu, L., Barakat, T.S., Okwieka, P., Porter, A., Gribnau, J., Hochedlinger, K., *et al.* (2010). A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell* 6, 535-546.
- Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q.L., and Smith, A. (2008). Capture of authentic embryonic stem cells from rat blastocysts. *Cell* 135, 1287-1298.

- Burdach, J., Funnell, A.P., Mak, K.S., Artuz, C.M., Wienert, B., Lim, W.F., Tan, L.Y., Pearson, R.C., and Crossley, M. (2014). Regions outside the DNA-binding domain are critical for proper in vivo specificity of an archetypal zinc finger transcription factor. *Nucleic Acids Res* 42, 276-289.
- Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol* 210, 30-43.
- Capecchi, M.R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* 6, 507-512.
- Carey, B.W., Finley, L.W., Cross, J.R., Allis, C.D., and Thompson, C.B. (2014). Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature*.
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113, 643-655.
- Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* 450, 1230-1234.
- Chan, Y.S., Goke, J., Ng, J.H., Lu, X., Gonzales, K.A., Tan, C.P., Tng, W.Q., Hong, Z.Z., Lim, Y.S., and Ng, H.H. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* 13, 663-675.
- Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell* 10, 615-624.
- Chen, S., Do, J.T., Zhang, Q., Yao, S., Yan, F., Peters, E.C., Scholer, H.R., Schultz, P.G., and Ding, S. (2006). Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A* 103, 17266-17271.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106-1117.
- Cheng, A.M., Saxton, T.M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C.G., Cardiff, R.D., Cross, J.C., Muller, W.J., *et al.* (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* 95, 793-803.
- Choudhary, C., and Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* 11, 427-439.
- Chung, Y., Klimanskaya, I., Becker, S., Marh, J., Lu, S.J., Johnson, J., Meisner, L., and Lanza, R. (2006). Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. *Nature* 439, 216-219.
- Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* 22, 746-755.

- Conkright, M.D., Wani, M.A., and Lingrel, J.B. (2001). Lung Kruppel-like factor contains an autoinhibitory domain that regulates its transcriptional activation by binding WWP1, an E3 ubiquitin ligase. *J Biol Chem* 276, 29299-29306.
- Daheron, L., Opitz, S.L., Zaehres, H., Lensch, M.W., Andrews, P.W., Itskovitz-Eldor, J., and Daley, G.Q. (2004). LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* 22, 770-778.
- de Assuncao, T.M., Lomberg, G., Cao, S., Yaqoob, U., Mathison, A., Simonetto, D.A., Huebert, R.C., Urrutia, R.A., and Shah, V.H. (2014). New Role for Kruppel-like Factor 14 as a Transcriptional Activator Involved in the Generation of Signaling Lipids. *J Biol Chem* 289, 15798-15809.
- De Los Angeles, A., Loh, Y.H., Tesar, P.J., and Daley, G.Q. (2012). Accessing naive human pluripotency. *Curr Opin Genet Dev* 22, 272-282.
- Delhaise, F., Bralio, V., Schuurbiens, N., and Dessy, F. (1996). Establishment of an embryonic stem cell line from 8-cell stage mouse embryos. *Eur J Morphol* 34, 237-243.
- Doble, B.W., and Woodgett, J.R. (2003). GSK-3: tricks of the trade for a multi-tasking kinase. *J Cell Sci* 116, 1175-1186.
- Eistetter, H.R. (1989). Pluripotent Embryonal Stem Cell Lines Can Be Established from Disaggregated Mouse Morulae. *Development, Growth & Differentiation* 31, 275-282.
- Ema, M., Mori, D., Niwa, H., Hasegawa, Y., Yamanaka, Y., Hitoshi, S., Mimura, J., Kawabe, Y., Hosoya, T., Morita, M., *et al.* (2008). Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. *Cell Stem Cell* 3, 555-567.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-156.
- Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J., *et al.* (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* 11, 197-203.
- Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. *Cell Stem Cell* 11, 477-490.
- Ficiz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., *et al.* (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell* 13, 351-359.
- Funnell, A.P., Mak, K.S., Twine, N.A., Pelka, G.J., Norton, L.J., Radziewicz, T., Power, M., Wilkins, M.R., Bell-Anderson, K.S., Fraser, S.T., *et al.* (2013). Generation of mice deficient in both KLF3/BKLF and KLF8 reveals a genetic

- interaction and a role for these factors in embryonic globin gene silencing. *Mol Cell Biol* 33, 2976-2987.
- Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., *et al.* (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature* 504, 282-286.
 - Gardner, R.L., and Beddington, R.S. (1988). Multi-lineage 'stem' cells in the mammalian embryo. *J Cell Sci Suppl* 10, 11-27.
 - Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W., Pasque, V., Magnusdottir, E., and Surani, M.A. (2012). Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* 10, 425-439.
 - Gossler, A., Doetschman, T., Korn, R., Serfling, E., and Kemler, R. (1986). Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc Natl Acad Sci U S A* 83, 9065-9069.
 - Greber, B., Wu, G., Bernemann, C., Joo, J.Y., Han, D.W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., *et al.* (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* 6, 215-226.
 - Gumireddy, K., Li, A., Gimotty, P.A., Klein-Szanto, A.J., Showe, L.C., Katsaros, D., Coukos, G., Zhang, L., and Huang, Q. (2009). KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. *Nat Cell Biol* 11, 1297-1304.
 - Guo, G., Huang, Y., Humphreys, P., Wang, X., and Smith, A. (2011). A PiggyBac-based recessive screening method to identify pluripotency regulators. *PLoS One* 6, e18189.
 - Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* 136, 1063-1069.
 - Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., *et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* 13, 360-369.
 - Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., *et al.* (2009). Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* 5, 597-609.
 - Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010a). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A* 107, 9222-9227.
 - Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J., *et al.* (2009).

- Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* 4, 513-524.
- Hanna, J.H., Saha, K., and Jaenisch, R. (2010b). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* 143, 508-525.
 - Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 3, 391-401.
 - Hayashi, K., and Surani, M.A. (2009). Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development* 136, 3549-3556.
 - Heard, E. (2004). Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* 16, 247-255.
 - Heo, J.S., Lee, Y.J., and Han, H.J. (2006). EGF stimulates proliferation of mouse embryonic stem cells: involvement of Ca²⁺ influx and p44/42 MAPKs. *Am J Physiol Cell Physiol* 290, C123-133.
 - Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H.S., Woodard, C., Wang, H., Jeong, J.S., *et al.* (2009). Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* 139, 610-622.
 - Huang, K., Maruyama, T., and Fan, G. (2014). The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell* 15, 410-415.
 - Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* 10, 353-360.
 - Johnson, G.L., and Lapadat, R. (2002). Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 298, 1911-1912.
 - Kaelin, W.G., Jr. (2011). Cancer and altered metabolism: potential importance of hypoxia-inducible factor and 2-oxoglutarate-dependent dioxygenases. *Cold Spring Harb Symp Quant Biol* 76, 335-345.
 - Karwacki-Neisius, V., Goke, J., Osorno, R., Halbritter, F., Ng, J.H., Weisse, A.Y., Wong, F.C., Gagliardi, A., Mullin, N.P., Festuccia, N., *et al.* (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. *Cell Stem Cell* 12, 531-545.
 - Katz, J.P., Perreault, N., Goldstein, B.G., Lee, C.S., Labosky, P.A., Yang, V.W., and Kaestner, K.H. (2002). The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development* 129, 2619-2628.
 - Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* 26, 903-911.

- Kraus, P., Leong, G., Tan, V., Xing, X., Goh, J.W., Yap, S.P., and Lufkin, T. (2010). A more cost effective and rapid high percentage germ-line transmitting chimeric mouse generation procedure via microinjection of 2-cell, 4-cell, and 8-cell embryos with ES and iPS cells. *Genesis* 48, 394-399.
- Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* 134, 2895-2902.
- Kuo, C.T., Veselits, M.L., Barton, K.P., Lu, M.M., Clendenin, C., and Leiden, J.M. (1997). The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. *Genes Dev* 11, 2996-3006.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* 22, 1617-1635.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., *et al.* (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* 20, 311-316.
- Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., *et al.* (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* 141, 872-883.
- Lewis, T.S., Shapiro, P.S., and Ahn, N.G. (1998). Signal transduction through MAP kinase cascades. *Adv Cancer Res* 74, 49-139.
- Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.L., *et al.* (2008). Germline competent embryonic stem cells derived from rat blastocysts. *Cell* 135, 1299-1310.
- Li, Q.R., Xing, X.B., Chen, T.T., Li, R.X., Dai, J., Sheng, Q.H., Xin, S.M., Zhu, L.L., Jin, Y., Pei, G., *et al.* (2011). Large scale phosphoproteome profiles comprehensive features of mouse embryonic stem cells. *Mol Cell Proteomics* 10, M110 001750.
- Li, W., Wei, W., Zhu, S., Zhu, J., Shi, Y., Lin, T., Hao, E., Hayek, A., Deng, H., and Ding, S. (2009). Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell Stem Cell* 4, 16-19.
- Loh, K.M., and Lim, B. (2011). A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* 8, 363-369.
- Loh, K.M., Lim, B., and Ang, L.T. (2015). Ex uno plures: molecular designs for embryonic pluripotency. *Physiol Rev* 95, 245-295.

- Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38, 431-440.
- Ma, Y.G., Rosfjord, E., Huebert, C., Wilder, P., Tiesman, J., Kelly, D., and Rizzino, A. (1992). Transcriptional regulation of the murine k-FGF gene in embryonic cell lines. *Dev Biol* 154, 45-54.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57-63.
- Magnusdottir, E., and Surani, M.A. (2014). How to make a primordial germ cell. *Development* 141, 245-252.
- Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., *et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590-604.
- Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Gottgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* 11, 491-504.
- Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78, 7634-7638.
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., *et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* 9, 625-635.
- McConnell, B.B., and Yang, V.W. (2010). Mammalian Kruppel-like factors in health and diseases. *Physiol Rev* 90, 1337-1381.
- 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.
- Miyanari, Y., and Torres-Padilla, M.E. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* 483, 470-473.
- Montserrat, N., Nivet, E., Sancho-Martinez, I., Hishida, T., Kumar, S., Miquel, L., Cortina, C., Hishida, Y., Xia, Y., Esteban, C.R., *et al.* (2013). Reprogramming of human fibroblasts to pluripotency with lineage specifiers. *Cell Stem Cell* 13, 341-350.
- Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D., and Tesar, P.J. (2011). Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell* 8, 318-325.

- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 26, 101-106.
- Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M.S., *et al.* (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol* 26, 7772-7782.
- Nichols, J., Davidson, D., Taga, T., Yoshida, K., Chambers, I., and Smith, A. (1996). Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis. *Mech Dev* 57, 123-131.
- Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* 136, 3215-3222.
- Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* 4, 487-492.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379-391.
- Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol* 19, 5453-5465.
- 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 self-renewal of ES cells. *Nat Genet* 24, 372-376.
- Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* 460, 118-122.
- Ohtsuka, S., Nishikawa-Torikai, S., and Niwa, H. (2012). E-cadherin promotes incorporation of mouse epiblast stem cells into normal development. *PLoS One* 7, e45220.
- Parker-Katiraei, L., Carson, A.R., Yamada, T., Arnaud, P., Feil, R., Abu-Amero, S.N., Moore, G.E., Kaneda, M., Perry, G.H., Stone, A.C., *et al.* (2007). Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet* 3, e65.
- Pei, J., and Grishin, N.V. (2013). A new family of predicted Kruppel-like factor genes and pseudogenes in placental mammals. *PLoS One* 8, e81109.
- Pei, L., Leblanc, M., Barish, G., Atkins, A., Nofsinger, R., Whyte, J., Gold, D., He, M., Kawamura, K., Li, H.R., *et al.* (2011). Thyroid hormone receptor

- repression is linked to type I pneumocyte-associated respiratory distress syndrome. *Nat Med* 17, 1466-1472.
- Pera, M.F. (2014). In search of naivety. *Cell Stem Cell* 15, 543-545.
 - 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.
 - Pines, A., Kelstrup, C.D., Vrouwe, M.G., Puigvert, J.C., Typas, D., Misovic, B., de Groot, A., von Stechow, L., van de Water, B., Danen, E.H., *et al.* (2011). Global phosphoproteome profiling reveals unanticipated networks responsive to cisplatin treatment of embryonic stem cells. *Mol Cell Biol* 31, 4964-4977.
 - Ramos, J.W. (2008). The regulation of extracellular signal-regulated kinase (ERK) in mammalian cells. *Int J Biochem Cell Biol* 40, 2707-2719.
 - Rathjen, J., Lake, J.A., Bettess, M.D., Washington, J.M., Chapman, G., and Rathjen, P.D. (1999). Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci* 112 (Pt 5), 601-612.
 - Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W., and Besser, D. (2011). E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep* 12, 720-726.
 - Roberts, P.J., and Der, C.J. (2007). Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* 26, 3291-3310.
 - Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotent cells by retroviral vector. *Nature* 323, 445-448.
 - 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.
 - 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.
 - Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J* 9, 2185-2195.
 - Segre, J.A., Bauer, C., and Fuchs, E. (1999). Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* 22, 356-360.
 - SenBanerjee, S., Lin, Z., Atkins, G.B., Greif, D.M., Rao, R.M., Kumar, A., Feinberg, M.W., Chen, Z., Simon, D.I., Luscinskas, F.W., *et al.* (2004). KLF2 Is a novel transcriptional regulator of endothelial proinflammatory activation. *J Exp Med* 199, 1305-1315.

- Shu, J., and Deng, H. (2013). Lineage specifiers: new players in the induction of pluripotency. *Genomics Proteomics Bioinformatics* 11, 259-263.
- Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., *et al.* (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* 153, 963-975.
- Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138, 722-737.
- Simone, C., Forcales, S.V., Hill, D.A., Imbalzano, A.N., Latella, L., and Puri, P.L. (2004). p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet* 36, 738-743.
- Small, K.S., Hedman, A.K., Grundberg, E., Nica, A.C., Thorleifsson, G., Kong, A., Thorsteindottir, U., Shin, S.Y., Richards, H.B., Soranzo, N., *et al.* (2011). Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* 43, 561-564.
- 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.
- Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., *et al.* (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods* 3, 995-1000.
- Sokol, S.Y. (2011). Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development* 138, 4341-4350.
- Stavridis, M.P., Lunn, J.S., Collins, B.J., and Storey, K.G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development* 134, 2889-2894.
- Sun, B., Ito, M., Mendjan, S., Ito, Y., Brons, I.G., Murrell, A., Vallier, L., Ferguson-Smith, A.C., and Pedersen, R.A. (2012). Status of genomic imprinting in epigenetically distinct pluripotent stem cells. *Stem Cells* 30, 161-168.
- Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficuz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., *et al.* (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* 158, 1254-1269.
- Tam, W.L., Lim, C.Y., Han, J., Zhang, J., Ang, Y.S., Ng, H.H., Yang, H., and Lim, B. (2008). T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* 26, 2019-2031.
- Tee, W.W., Shen, S.S., Oksuz, O., Narendra, V., and Reinberg, D. (2014). Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. *Cell* 156, 678-690.

- ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., Siu, R.K., and Nusse, R. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat Cell Biol* 13, 1070-1075.
- Teo, A.K., Arnold, S.J., Trotter, M.W., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev* 25, 238-250.
- Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196-199.
- Tetreault, M.P., Yang, Y., and Katz, J.P. (2013). Kruppel-like factors in cancer. *Nat Rev Cancer* 13, 701-713.
- Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., *et al.* (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 15, 471-487.
- 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.
- 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.
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* 135, 909-918.
- Valamehr, B., Robinson, M., Abujarour, R., Rezner, B., Vranceanu, F., Le, T., Medcalf, A., Lee, T.T., Fitch, M., Robbins, D., *et al.* (2014). Platform for Induction and Maintenance of Transgene-free hiPSCs Resembling Ground State Pluripotent Stem Cells. *Stem Cell Reports* 2, 366-381.
- Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* 118, 4495-4509.
- Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P., *et al.* (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* 136, 1339-1349.
- Wakayama, S., Hikichi, T., Suetsugu, R., Sakaide, Y., Bui, H.T., Mizutani, E., and Wakayama, T. (2007). Efficient establishment of mouse embryonic stem cell lines from single blastomeres and polar bodies. *Stem Cells* 25, 986-993.
- Wang, R., Wang, Y., Liu, N., Ren, C., Jiang, C., Zhang, K., Yu, S., Chen, Y., Tang, H., Deng, Q., *et al.* (2013). FBW7 regulates endothelial functions by targeting KLF2 for ubiquitination and degradation. *Cell Res* 23, 803-819.

- Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* *105*, 9290-9295.
- Wang, W., Yang, J., Liu, H., Lu, D., Chen, X., Zenonos, Z., Campos, L.S., Rad, R., Guo, G., Zhang, S., *et al.* (2011). Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc Natl Acad Sci U S A* *108*, 18283-18288.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* *10*, 57-63.
- Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* *10*, 440-454.
- Wani, M.A., Means, R.T., Jr., and Lingrel, J.B. (1998). Loss of LKLF function results in embryonic lethality in mice. *Transgenic Res* *7*, 229-238.
- Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., *et al.* (2014). Derivation of naive human embryonic stem cells. *Proc Natl Acad Sci U S A* *111*, 4484-4489.
- Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* *13*, 36-47.
- 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.
- Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol* *13*, 838-845.
- Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. *Biochem Soc Trans* *38*, 1027-1032.
- Wu, J., Srinivasan, S.V., Neumann, J.C., and Lingrel, J.B. (2005). The KLF2 transcription factor does not affect the formation of preadipocytes but inhibits their differentiation into adipocytes. *Biochemistry* *44*, 11098-11105.
- Xie, P., Tang, Y., Shen, S., Wang, Y., Xing, G., Yin, Y., He, F., and Zhang, L. (2011). Smurf1 ubiquitin ligase targets Kruppel-like factor KLF2 for ubiquitination and degradation in human lung cancer H1299 cells. *Biochem Biophys Res Commun* *407*, 254-259.
- Xu, Y., Zhu, X., Hahm, H.S., Wei, W., Hao, E., Hayek, A., and Ding, S. (2010). Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc Natl Acad Sci U S A* *107*, 8129-8134.

- Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* 137, 715-724.
- Yeo, J.C., Jiang, J., Tan, Z.Y., Yim, G.R., Ng, J.H., Goke, J., Kraus, P., Liang, H., Gonzales, K.A., Chong, H.C., *et al.* (2014). Klf2 is an essential factor that sustains ground state pluripotency. *Cell Stem Cell* 14, 864-872.
- Yeo, J.C., and Ng, H.H. (2013). The transcriptional regulation of pluripotency. *Cell Res* 23, 20-32.
- Yi, F., Pereira, L., and Merrill, B.J. (2008). Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* 26, 1951-1960.
- Yim, D.G., Ghosh, S., Guy, G.R., and Virshup, D.M. (2015). Casein kinase 1 regulates Sprouty2 in FGF-ERK signaling. *Oncogene* 34, 474-484.
- 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.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519-523.
- Young, R.A. (2011). Control of the embryonic stem cell state. *Cell* 144, 940-954.
- 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.
- Zhang, X., Srinivasan, S.V., and Lingrel, J.B. (2004). WWP1-dependent ubiquitination and degradation of the lung Kruppel-like factor, KLF2. *Biochem Biophys Res Commun* 316, 139-148.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.
- Zhou, B., Xiao, J.F., Tuli, L., and Ransom, H.W. (2012a). LC-MS-based metabolomics. *Mol Biosyst* 8, 470-481.
- Zhou, H., Li, W., Zhu, S., Joo, J.Y., Do, J.T., Xiong, W., Kim, J.B., Zhang, K., Scholer, H.R., and Ding, S. (2010). Conversion of mouse epiblast stem cells to an earlier pluripotency state by small molecules. *J Biol Chem* 285, 29676-29680.
- Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., *et al.* (2012b). HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J* 31, 2103-2116.

CHAPTER VI

APPENDIX

Chapter VI: Appendix

A. List of the top 200 upregulated genes in *Klf2*-null mESCs (relative to *Klf2*-WT) under LIF/Serum, 2i passage 1, 2 and 3.

No.	Klf2-null mESC LIF/Serum		Klf2-null mESC 2i Passage 1		Klf2-null mESC 2i Passage 2		Klf2-null mESC 2i Passage 3	
	Gene ID	Fold Change	Gene ID	Fold Change	Gene ID	Fold Change	Gene ID	Fold Change
1	H19	53.990	LOC666403	18.051	Rhox9	31.844	LOC666403	28.428
2	H19	18.549	Plk1	14.206	LOC666403	15.157	RHOX9	22.402
3	LOC666403	18.095	Rhox9	13.429	Plk1	13.859	PLK1	18.912
4	Plk1	17.525	H2-BI	9.759	H2-BI	13.128	RHOX6	14.136
5	Igf2	10.830	Dppa3	8.248	Rhox6	9.094	DDX3Y	11.042
6	Rhox9	8.996	Eif2s3y	7.214	Cotl1	7.319	H2-BL	10.870
7	2900060B14Rik	8.554	Cpsf4l	6.443	Eif2s3y	7.209	LPL	9.906
8	Ly6a	7.317	6720469N11Rik	5.659	Rnf213	6.906	H19	8.183
9	Car2	5.961	Acot1	5.405	S100a11	6.678	MRAS	7.612
10	Robo4	5.232	Lgals1	5.322	H19	6.637	SH3TC2	6.349
11	Ccrn4l	5.137	Rnf213	5.141	Cpsf4l	6.510	2810423A18RIK	6.134
12	Rhox5	4.579	Ddx3y	5.110	Gjb3	5.876	RNF213	6.068
13	Aurkc	4.496	Ifitm1	4.936	Dppa3	5.795	S100A11	5.996
14	LOC666185	4.011	Acot1	4.800	Mras	5.791	RBM8A	5.722
15	Taf15	3.989	Cotl1	4.796	Ifitm1	5.789	ACOT1	5.477
16	Fgfr2	3.896	Zdhhc13	4.568	Ddx3y	5.613	DND1	5.132
17	Mep1b	3.824	Gjb3	4.216	6720469N11Rik	5.513	EIF2S3Y	4.855
18	Ccnd2	3.731	Ap3b2	4.155	Cotl1	5.496	9330175B01RIK	4.605

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

19	LOC380655	3.714	Fgfr2	4.089	Acot1	4.795	DAB2	4.435
20	Ddx3y	3.700	Mbp	4.087	Tes	4.641	MRAS	4.373
21	Cdkn1c	3.667	Irx3	4.011	Mfge8	4.546	D030040K20RIK	4.336
22	Ctgf	3.619	2810423A18Rik	4.007	Dnd1	4.479	AMFR	4.319
23	Prmt3	3.598	S100a11	3.951	Tes	4.395	MMP15	4.296
24	Rhox6	3.567	Car12	3.940	Mbp	4.391	GSTO1	4.200
25	Plac1	3.464	H19	3.761	Ifitm3	4.315	KNR1	4.090
26	Stard8	3.458	Rhox6	3.724	Lpl	4.302	PCSK9	4.033
27	2810423A18Rik	3.446	Zdhhc13	3.713	2810423A18Rik	4.297	3830408P04RIK	4.009
28	Eif2s3y	3.426	Cotl1	3.643	Mbp	4.253	CD9	3.984
29	Serpina3m	3.416	Klk1b22	3.626	Ap3b2	4.201	LY6A	3.953
30	Smoc1	3.384	S100a1	3.602	Ifitm1	4.140	AURKC	3.845
31	Gadd45g	3.351	Zfp521	3.447	Car12	4.137	6720469N11RIK	3.828
32	Dok2	3.310	Tes	3.345	Phlda2	4.114	LOC100046120	3.762
33	LOC381284	3.273	Zdhhc13	3.328	2810003C17Rik	4.051	ACSL5	3.743
34	S100a11	3.254	Plcg2	3.311	2900060B14Rik	4.041	ADAMTS9	3.727
35	Sct	3.170	3830408P04Rik	3.287	Acot1	3.972	CMTM7	3.698
36	Car4	3.168	Tes	3.286	Cyba	3.970	HTATIP2	3.671
37	Ldoc1	3.154	Tcfap2c	3.284	Aurkc	3.959	SERPINB6C	3.670
38	Nipa1	3.134	Fhl1	3.219	Phf19	3.952	LOC100045542	3.665
39	Dppa3	3.098	Ccdc5	3.195	Sox7	3.942	2410078J06RIK	3.660
40	Zdhhc13	3.092	Amfr	3.167	Lgals1	3.882	DAB2	3.642
41	Hist1h4i	3.062	2810003C17Rik	3.136	Plcg2	3.729	HTATIP2	3.640
42	Zdhhc13	3.038	D030040K20Rik	3.130	Acsl5	3.692	RBM8A	3.635
43	Zdhhc13	3.025	Fhl1	3.129	3830408P04Rik	3.643	STARD8	3.602
44	Stard8	3.003	Tcfap2c	3.114	Hexa	3.592	LY6G6E	3.585

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

45	5730543M03Rik	2.993	Ifitm1	3.093	H19	3.584	CYB5	3.546
46	LOC100045342	2.971	9330175B01Rik	3.052	Rbm8a	3.560	PHLDA2	3.490
47	Klk8	2.934	Sox7	3.045	Fgfr2	3.524	PRMT3	3.488
48	Nucb1	2.924	Hexa	3.031	Cmtm7	3.500	CYBA	3.468
49	BC017612	2.922	D11Ert636e	3.000	Zdhhc13	3.493	ARC	3.458
50	2410004A20Rik	2.919	Tcfap2c	2.949	Ifitm1	3.486	CMTM7	3.457
51	Naprt1	2.911	Prmt3	2.929	Pdlim4	3.474	PLCG2	3.449
52	Hist1h4a	2.902	Extl2	2.922	D030040K20Rik	3.469	RAB11FIP5	3.421
53	Ndp52	2.898	6330404F12Rik	2.888	Zdhhc13	3.441	1110032E23RIK	3.408
54	C130035G06Rik	2.893	EG630499	2.876	Gata2	3.430	EXTL2	3.381
55	Mbnl1	2.879	Mbp	2.845	Mfge8	3.398	AHNAK	3.337
56	Amfr	2.866	Phlda2	2.843	Cmtm7	3.349	LOC100041569	3.327
57	Cyp7b1	2.856	Dnd1	2.843	Cd9	3.344	LPCAT3	3.295
58	LOC100046586	2.853	Acsl5	2.840	LOC224532	3.329	ZDHHC13	3.268
59	6720469N11Rik	2.830	Eno1	2.831	Ccnd3	3.303	COTL1	3.242
60	Grwd1	2.820	Phf19	2.818	Dab2	3.277	NIPA1	3.228
61	Bmp4	2.807	Cyb5	2.797	Dner	3.230	RAB11FIP5	3.224
62	Akt1s1	2.797	Rps6kl1	2.770	Klk1b22	3.223	2900060B14RIK	3.210
63	Mep1b	2.790	Tdrd12	2.757	Rps6kl1	3.213	SLC39A4	3.189
64	Ccnd2	2.758	Lrrc15	2.730	9330175B01Rik	3.109	CIDEA	3.188
65	Zscan4c	2.748	Spnb3	2.699	Spnb3	3.064	2900011O08RIK	3.177
66	Samd9l	2.738	Smoc1	2.687	Tcfap2c	3.051	C630002L24RIK	3.168
67	LOC381844	2.725	Copz2	2.661	Rbm47	3.028	PHF19	3.165
68	LOC100046049	2.723	LOC224532	2.656	Sh3tc2	3.001	HERC2	3.164
69	Tlcd1	2.701	Aurkc	2.612	LOC100046120	3.000	STARD8	3.148
70	3830408P04Rik	2.701	Ets1	2.590	Fhl1	2.986	9130213B05RIK	3.147

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

71	Psmb10	2.677	Rbm8a	2.580	Ccdc5	2.964	B230311B06RIK	3.142
72	Lxn	2.673	Copz2	2.569	EG630499	2.956	COTL1	3.134
73	Cyb5	2.664	Peg3	2.553	Stard8	2.939	SOX7	3.132
74	Tcstv3	2.636	Nipa1	2.543	Tmsb10	2.926	ZFP521	3.078
75	Cyfip1	2.623	2900060B14Rik	2.531	Prmt3	2.915	CCRN4L	3.076
76	Ccdc5	2.620	Slc39a4	2.530	Amfr	2.913	LOC100045343	3.073
77	Cmtm4	2.617	Herc2	2.527	Mmp15	2.910	GLRX	3.056
78	Krt8	2.609	Cyba	2.516	Slc39a4	2.893	A130010C12RIK	3.024
79	Ndp52	2.596	2410008K03Rik	2.500	Tcfap2c	2.846	TGM1	3.009
80	Slc40a1	2.590	LOC100040525	2.494	Vgll4	2.839	FHOD1	2.981
81	Rnd3	2.579	Zfp473	2.469	S100a1	2.826	LOC380655	2.975
82	Zfp473	2.575	Gadd45a	2.463	Nuak1	2.814	FBP2	2.970
83	Retsat	2.573	Elmo1	2.453	Bahd1	2.789	TMEM117	2.946
84	Hist1h4m	2.571	Lxn	2.453	Dab2	2.788	B4GALNT4	2.940
85	Chd9	2.542	Jmjd5	2.452	Cyb5	2.784	FMR1NB	2.937
86	Hist1h3e	2.537	Tmem143	2.451	8430410A17Rik	2.774	CLN6	2.926
87	Psmb10	2.525	Pvrl2	2.447	Punc	2.771	MME	2.920
88	ENSMUSG00000074075	2.521	6330562C20Rik	2.443	Gna14	2.767	CPSF4L	2.915
89	Itpr2	2.514	Lpl	2.442	Extl2	2.741	PCOLCE	2.897
90	Ube2cbp	2.510	Lpcat3	2.425	Fhl1	2.729	DHRS7	2.883
91	Htatip2	2.502	H2-T23	2.424	Stard8	2.727	GPRC5B	2.880
92	Rbm8a	2.486	Kctd15	2.402	Zdhhc13	2.726	H19	2.879
93	Chd9	2.471	LOC100046120	2.399	Tcfap2c	2.723	2410078J06RIK	2.872
94	BC061212	2.469	Gata2	2.392	Npepl1	2.718	ZFP473	2.865
95	LOC383368	2.465	C330023J09Rik	2.388	LOC676420	2.717	ZDHHHC13	2.856
96	Raet1c	2.464	H2-T23	2.382	D11Erttd636e	2.711	FGFR2	2.854

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

97	Hist1h4j	2.455	Dab2	2.381	Gadd45a	2.707	PPP1R1A	2.849
98	Esam	2.442	Mfge8	2.370	Htra1	2.706	NUDT6	2.837
99	Satb1	2.439	Slc15a2	2.363	Zfp473	2.705	KLF4	2.834
100	Tle6	2.427	Rnf103	2.354	Lrrc34	2.677	CAR12	2.821
101	Sh3tc2	2.426	Stard8	2.346	H2-T23	2.676	TMEM82	2.813
102	Jmjd5	2.423	Lrrc34	2.342	Atmin	2.671	NPC1	2.809
103	A430089I19Rik	2.414	Mbnl1	2.325	Pga5	2.671	2610009E16RIK	2.796
104	Pml	2.412	Sh3tc2	2.316	Gsto1	2.669	DNER	2.767
105	Tmem143	2.412	Nudt19	2.284	Kcnc4	2.668	IFITM1	2.760
106	Smpd13b	2.399	Arpc1b	2.280	Ccrn4l	2.668	TEAD4	2.758
107	Zfp473	2.398	Npepl1	2.278	Npc1	2.641	EPHB1	2.754
108	LOC100039706	2.397	Grwd1	2.277	H2-T23	2.638	MFGE8	2.742
109	Mep1b	2.393	C030027H14Rik	2.262	C130021I20	2.616	MBP	2.737
110	2010004N17Rik	2.393	Npc1	2.258	Arpc1b	2.606	VWF	2.718
111	Nudt6	2.391	Stard8	2.255	Herc2	2.596	BC031853	2.707
112	Nckap1l	2.383	Ifitm3	2.251	Lrpap1	2.592	WFDC2	2.706
113	Gja1	2.377	Car12	2.249	Mme	2.590	ZFP473	2.706
114	Cklf	2.375	Nucb1	2.246	Agpat9	2.578	GRWD1	2.691
115	D14Ert449e	2.372	C630002L24Rik	2.226	Htra1	2.558	PON2	2.688
116	Nomo1	2.368	Mras	2.221	Mras	2.558	AP3B2	2.686
117	2700024D06Rik	2.353	Cln6	2.220	LOC100045542	2.557	GPT2	2.685
118	Tatdn2	2.348	Dhdh	2.219	Aard	2.544	LOC100046883	2.680
119	Wnt7b	2.347	LOC100046883	2.219	Htatip2	2.540	DENND1C	2.677
120	Herc2	2.344	Gadd45a	2.219	2310014H01Rik	2.539	9330186A19RIK	2.663
121	Mmp15	2.336	2010004N17Rik	2.202	Ets1	2.538	KRT7	2.662
122	Adamts9	2.333	Bcap29	2.184	Slc15a2	2.537	C230070D10RIK	2.652

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

123	Eef1d	2.329	Ddit4	2.184	Arpc1b	2.528	SMOC1	2.651
124	Vgll3	2.325	Tmem143	2.183	Htati2	2.510	AHNAK	2.651
125	Ahcy	2.316	Zfp473	2.183	Peg3	2.506	ENSMUSG00000068790	2.644
126	Dus4l	2.312	LOC100039706	2.181	Tes	2.484	FBXO2	2.643
127	Ciapi1	2.309	Tes	2.174	Fbxl12	2.479	TMEM51	2.640
128	Gata2	2.295	Cyfp1	2.168	Aldh16a1	2.468	AKT1S1	2.636
129	Dhdh	2.294	Bahd1	2.167	Zfp473	2.464	FHOD1	2.633
130	Tmem143	2.282	Galnt11	2.164	Fhod1	2.457	ALDH16A1	2.627
131	H2-BI	2.281	Raet1c	2.159	Prss8	2.455	XLR4A	2.611
132	Irx3	2.271	Ccrn4l	2.152	Cln6	2.454	RPS6KL1	2.610
133	Elmo1	2.268	Rhbdf1	2.147	Tmem150	2.450	CYFIP1	2.607
134	4930502E18Rik	2.265	Slc6a13	2.144	Sphk2	2.447	CYP7B1	2.594
135	Lonp2	2.265	Plrg1	2.131	Mcart6	2.441	S100A13	2.592
136	Extl2	2.265	D14Ert449e	2.128	Smoc1	2.429	NUCB1	2.591
137	Pih1d1	2.259	Pdlim4	2.118	Zfp521	2.427	TMSB10	2.590
138	Hist1h3f	2.253	Rps3a	2.117	C630002L24Rik	2.421	NUCB1	2.587
139	LOC546233	2.251	Dab2	2.110	Rdm1	2.407	GNA14	2.578
140	Cited2	2.238	Mfge8	2.102	C030027H14Rik	2.399	ETS2	2.577
141	Rnf213	2.237	Punc	2.099	Bcl9l	2.393	9830143E02RIK	2.572
142	B230311B06Rik	2.229	Tmem82	2.095	Nudt6	2.393	EMP1	2.569
143	Atmin	2.226	Heatr5a	2.093	Pon2	2.387	RNPEP	2.564
144	S100a6	2.223	Smox	2.092	Nucb1	2.373	TES	2.559
145	Hist1h3d	2.222	Rbm8a	2.090	Gm22	2.373	2310044H10RIK	2.559
146	Klk1b22	2.218	Gstz1	2.084	Tmem82	2.371	HEXA	2.557
147	Csnk2a2	2.215	Aatk	2.081	Grwd1	2.369	FRMD6	2.543
148	Mgat4b	2.210	Serpinh1	2.078	Rhbdf1	2.366	FHL1	2.539

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

149	Cmtm4	2.204	Mogat2	2.075	Ttc28	2.360	AGL	2.529
150	2410008K03Rik	2.197	3110078M01Rik	2.075	Car12	2.359	PRSS8	2.520
151	Nucb1	2.196	Stat3	2.074	Jmjd5	2.357	STAT3	2.519
152	Rps13	2.194	Slit2	2.071	Npepl1	2.354	TMEM143	2.515
153	Pygb	2.193	Sat2	2.069	Nipa1	2.343	CCDC5	2.514
154	Pigt	2.185	Etfb	2.063	Irx3	2.335	AATK	2.505
155	9530027K23Rik	2.181	P2rx4	2.052	Aatk	2.333	EIF2AK2	2.496
156	Htatip2	2.172	Gng12	2.042	Atmin	2.329	TULP2	2.493
157	Lgals1	2.166	2210410E06Rik	2.038	Fhod1	2.319	2810417H13RIK	2.490
158	Nr1h2	2.155	Tmem150	2.038	Tmem143	2.307	VGLL4	2.487
159	BC002216	2.151	Arl5a	2.036	Gstz1	2.307	2810003C17RIK	2.482
160	Id3	2.148	S100a13	2.035	Eno1	2.302	FXYD6	2.471
161	EG329521	2.147	Smox	2.030	Gja4	2.302	PYGB	2.467
162	LOC381996	2.147	Sct	2.029	Irf1	2.296	FAM102A	2.453
163	9130213B05Rik	2.146	Seh1l	2.026	Lpcat3	2.291	BC046404	2.448
164	Pnkp	2.141	Gylt1b	2.026	Sphk2	2.290	S100A1	2.441
165	4933421H10Rik	2.134	Lrpap1	2.023	Emb	2.286	NUP210	2.440
166	Ciapi1	2.132	Nr1h2	2.021	Sod2	2.282	PRKCZ	2.435
167	1110057K04Rik	2.131	Chchd8	2.018	Pla1a	2.279	FHL1	2.435
168	LOC100041230	2.130	LOC209372	2.016	S100a13	2.277	PGA5	2.435
169	Plcg2	2.124	Dner	2.016	Tdrd12	2.274	TMEM143	2.434
170	AU018091	2.123	Cd9	2.012	1700061G19Rik	2.269	UBE2E2	2.424
171	Tmem92	2.121	Htatip2	2.012	Tbc1d17	2.268	RNPEP	2.423
172	Pyroxd1	2.117	Lass2	2.011	Rnf103	2.264	AP3B2	2.421
173	Arl2bp	2.117	Wfdc2	2.010	C330023J09Rik	2.262	FCGRT	2.417
174	D030040K20Rik	2.109	Rin1	2.009	Rap2c	2.261	SCT	2.415

Appendix A - Top 200 upregulated genes in *Klf2*-null mESCs

175	Tlr2	2.097	Hist1h1c	2.006	Rnpep	2.261	VRK1	2.409
176	Arpc4	2.093	Fbxl12	2.005	4933407C03Rik	2.256	EEF1D	2.404
177	Ap2b1	2.092	Ap3b2	2.005	Cklf	2.253	DUT	2.404
178	BC052688	2.092	Htatip2	2.003	GlrX	2.251	IL4I1	2.396
179	Cklf	2.085	Chchd4	2.001	Sox7	2.248	BC038881	2.393
180	LOC209372	2.084	4931417G12Rik	1.984	Phlda1	2.236	4933402E13RIK	2.393
181	Cog8	2.084	Tbc1d17	1.982	Rnpep	2.232	TMEM150	2.390
182	Anxa2	2.083	Nudt6	1.979	Stat3	2.219	XBP1	2.385
183	Retsat	2.083	C130021I20	1.979	Gadd45a	2.207	NUDT19	2.368
184	LOC100044390	2.079	Pex2	1.977	4933402E13Rik	2.201	9530027K23RIK	2.367
185	Nutf2	2.077	Hist1h4i	1.973	LOC100040525	2.200	PDLIM4	2.366
186	Hook2	2.075	Rap2c	1.970	Akt1s1	2.197	3632451O06RIK	2.364
187	Fhod1	2.074	Acot2	1.969	Ly6g6e	2.193	NR1H2	2.356
188	Emp1	2.072	Rps13	1.968	Serpib6c	2.189	SPHK2	2.353
189	Gstm2	2.072	LOC676420	1.960	Ephb1	2.187	BCL9L	2.352
190	LOC100040854	2.070	Ttc28	1.959	Nomo1	2.184	RBM47	2.340
191	Smyd5	2.070	B230311B06Rik	1.959	Rbm8a	2.175	MFGE8	2.336
192	LOC245128	2.070	Rec8	1.957	Plekhg5	2.173	IFITM1	2.332
193	Csnk2a2	2.065	Lym5	1.957	Irf1	2.167	GPC3	2.331
194	Etfb	2.055	Lass2	1.956	Nr1h2	2.165	RNF103	2.329
195	Junb	2.053	Sphk2	1.955	Hist1h1c	2.163	NPEPL1	2.329
196	LOC212386	2.053	Whsc1	1.953	Bcap29	2.161	RDM1	2.327
197	Tbc1d17	2.052	Mmp15	1.951	Pvrl2	2.159	JMJD5	2.326
198	C630002L24Rik	2.051	Tia1	1.950	InsI3	2.157	C1GALT1	2.325
199	Bax	2.049	Bcap29	1.949	BC031853	2.152	KLK1B22	2.315
200	4930504E06Rik	2.042	Abca3	1.947	Jam2	2.148	MAPK8IP2	2.314

B. List of the top 200 downregulated genes in *Klf2*-null mESCs (relative to *Klf2*-WT) under LIF/Serum, 2i passage 1, 2 and 3

No.	Klf2-null mESC LIF/Serum		Klf2-null mESC 2i Passage 1		Klf2-null mESC 2i Passage 2		Klf2-null mESC 2i Passage 3	
	Gene ID	Fold Change	Gene ID	Fold Change	Gene ID	Fold Change	Gene ID	Fold Change
1	Klf2	0.065	Klf2	0.033	Klf2	0.051	KLF2	0.014
2	Col4a1	0.149	Grb10	0.110	Grb10	0.122	ENOX1	0.165
3	EG433229	0.169	Dpp4	0.160	Dpp4	0.160	LEFTY1	0.188
4	1110017D15Rik	0.176	Col4a1	0.184	Enox1	0.186	DPP4	0.199
5	Dido1	0.201	LOC675899	0.218	Tgfb1	0.193	ADAM23	0.200
6	4930455C21Rik	0.211	1700123J19Rik	0.228	1110017D15Rik	0.206	CCNO	0.218
7	LOC226017	0.220	Npm3-ps1	0.244	Neurog3	0.212	DNMT3L	0.228
8	Erdr1	0.238	Enox1	0.246	LOC675899	0.215	SNURF	0.237
9	Snurf	0.245	LOC386405	0.249	Ccno	0.218	FER1L3	0.239
10	LOC386405	0.252	Erdr1	0.267	Dnmt3l	0.230	F2RL1	0.240
11	Ide	0.261	Rusc2	0.267	Glpr2	0.245	DNMT3L	0.240
12	LOC641366	0.262	Dnmt3l	0.277	Dnmt3l	0.249	LOC100040479	0.242
13	Got1l1	0.270	Fkbp9	0.282	Trim2	0.255	SERPINE2	0.249
14	Snurf	0.279	Snurf	0.284	Col4a1	0.255	LOC675899	0.249
15	Rusc2	0.286	Pard6g	0.286	Gtl2	0.257	LOC381844	0.250
16	LOC100045280	0.292	4930455C21Rik	0.287	Glpr2	0.260	EPHX1	0.256
17	Wdfy1	0.303	Slc1a3	0.289	Lefty1	0.261	COL4A1	0.261
18	Klhl26	0.311	Manba	0.293	A230050P20Rik	0.264	LEFTY2	0.263
19	Ppic	0.319	Dnmt3l	0.294	Tcf15	0.264	DUSP6	0.263
20	Erdr1	0.321	Fez1	0.298	Sox4	0.269	TRIM2	0.270

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

21	Nuak2	0.333	1110017D15Rik	0.299	F2rl1	0.269	MID1	0.273
22	Npm3-ps1	0.337	Dusp6	0.301	Dnmt3b	0.269	1110017D15RIK	0.273
23	Fxyd6	0.341	LOC641366	0.301	Pard6g	0.270	GM428	0.282
24	Sfrp1	0.342	Dido1	0.304	Fez1	0.271	LOC641366	0.283
25	Rev1	0.344	Tgfb1	0.305	Igsf9	0.274	A230050P20RIK	0.285
26	Calb2	0.346	Pla2g1b	0.305	Plekha4	0.276	LOC676461	0.290
27	4930583H14Rik	0.350	Zscan5b	0.309	1700019N12Rik	0.278	4930455C21RIK	0.293
28	LOC675899	0.351	Fer1l3	0.313	Snurf	0.279	GALNT10	0.293
29	Crabp1	0.354	Plekha4	0.313	Ccnd1	0.280	CDKN2A	0.296
30	Thy1	0.356	Lrmp	0.319	Slc1a3	0.282	IDE	0.297
31	Rusc2	0.356	Ctsc	0.320	Serpine2	0.286	SPIRE2	0.302
32	Ephx1	0.357	Tmem40	0.322	Dido1	0.286	F2RL1	0.304
33	1700123J19Rik	0.359	Asah3l	0.326	LOC386405	0.295	TMEM40	0.307
34	Sox21	0.360	Snurf	0.327	Zfp330	0.300	FEZ1	0.313
35	LOC675228	0.368	Nuak2	0.330	Calca	0.300	PQLC1	0.315
36	Grb10	0.370	Sh3gl2	0.331	Spire2	0.300	BC080695	0.315
37	1110001J03Rik	0.373	Spire2	0.335	Gm428	0.301	NEUROG3	0.316
38	Pgpep1	0.376	Trim2	0.336	Crmp1	0.303	AQP9	0.317
39	Zfp330	0.377	Tcf15	0.337	Rarg	0.304	BNC2	0.320
40	Egr1	0.381	Rusc2	0.339	Agfg2	0.307	TEX19.1	0.322
41	Ccnt2	0.382	Phlda3	0.344	Arhgef18	0.309	PLEKHA4	0.324
42	Trp53inp1	0.386	LOC100040479	0.346	Pla2g1b	0.310	LOC100045019	0.326
43	H3f3a	0.390	Mycl1	0.346	Rusc2	0.311	ERCC5	0.327
44	scl0001118.1_0	0.391	Rtn1	0.347	BC057371	0.311	GTL2	0.329
45	Rps24	0.393	Galnt10	0.348	Ccnd1	0.313	SNURF	0.329
46	LOC100045280	0.395	Igsf9	0.350	F2rl1	0.313	DIDO1	0.329

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

47	1810032O08Rik	0.396	F2rl1	0.353	Erdr1	0.318	1700019N12RIK	0.330
48	LOC381283	0.398	LOC100045019	0.358	Gadd45g	0.320	ERDR1	0.331
49	6820406G21Rik	0.399	Manba	0.360	Ptp4a3	0.322	PARD6G	0.332
50	LOC626152	0.401	A230050P20Rik	0.361	BC057371	0.324	4930583H14RIK	0.332
51	Enah	0.401	Tex13	0.364	Adam23	0.325	MYLPF	0.338
52	Tmem118	0.402	Rarg	0.367	Snurf	0.326	ZSCAN4C	0.341
53	Phlda3	0.402	Rtn1	0.369	St5	0.328	DUSP4	0.345
54	Ampd3	0.403	Zfp330	0.370	Ide	0.328	WDFY1	0.348
55	C230070D10Rik	0.404	F2rl1	0.374	Jak1	0.329	GJA1	0.349
56	LOC674912	0.405	Fez1	0.376	Asah3l	0.331	ERDR1	0.352
57	Tmem40	0.405	Calb2	0.377	Mycl1	0.333	GLIPR2	0.353
58	Ndufa13	0.405	Ei24	0.377	scl0002540.1_6	0.339	SNHG11	0.354
59	Repin1	0.409	Sema4a	0.380	Rusc2	0.340	UBD	0.355
60	1110001N06Rik	0.410	Xlr4a	0.381	LOC641366	0.344	ZFP330	0.356
61	Sccpdh	0.412	Serpine2	0.382	Ephx1	0.344	C130035G06RIK	0.356
62	Pqlc1	0.414	Ccnd1	0.382	Manba	0.347	BC057079	0.357
63	Hspb8	0.414	Ddc	0.385	Pqlc1	0.349	RARG	0.357
64	Hnrnp1	0.416	Ccnd1	0.386	LOC100040479	0.350	ZSCAN5B	0.358
65	LOC100045019	0.416	Slc6a15	0.387	Galnt10	0.351	GSTA3	0.360
66	Tgfb1	0.419	Pfkm	0.389	Armxc2	0.352	LOC329984	0.364
67	Plac8	0.422	Glipr2	0.389	Snhg11	0.352	GLIPR2	0.368
68	Serpine2	0.423	Mcam	0.390	1700123J19Rik	0.355	LOC100040854	0.370
69	Ddit4l	0.427	2700038C09Rik	0.392	St5	0.357	TGFBI	0.372
70	Mid1	0.429	Glipr2	0.394	Ephx2	0.362	STAG3	0.373
71	Ugcg	0.429	Clk1	0.394	LOC100045019	0.365	SNHG11	0.375
72	Whrn	0.430	Pqlc1	0.395	LOC381844	0.365	FEZ1	0.375

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

73	Prnp	0.431	A130092J06Rik	0.397	Snhg11	0.365	SLC1A3	0.378
74	Hnrnp1	0.431	Mylpf	0.398	Manba	0.369	A630082K20RIK	0.387
75	Mcam	0.431	Adam23	0.400	Raet1b	0.372	EPHX2	0.389
76	Hnrnp1	0.431	BC057371	0.400	4930583H14Rik	0.372	D630023F18RIK	0.389
77	Fbln1	0.433	Stag3	0.401	Tcf15	0.373	PTP4A3	0.389
78	Npm3	0.434	Dnmt3b	0.401	LOC100040854	0.374	MID1IP1	0.397
79	Snx10	0.437	St5	0.401	Gja1	0.374	CPM	0.401
80	A630097D09Rik	0.437	Ephx1	0.402	A130092J06Rik	0.374	BC057371	0.402
81	LOC674611	0.438	Sfrp1	0.404	4930455C21Rik	0.374	PITX2	0.403
82	2610319H10Rik	0.439	Mid1ip1	0.405	Dnajb6	0.378	F830002E14RIK	0.404
83	Usp37	0.440	Mif4gd	0.406	Ercc5	0.378	LOC381727	0.405
84	Sh3gl2	0.440	Pfkm	0.407	Lrmp	0.379	TCF15	0.406
85	Stag3	0.444	Pla2g10	0.410	Znf512b	0.380	1700029P11RIK	0.407
86	Slc19a2	0.445	Susd4	0.410	Pfkip	0.380	TMEM118	0.407
87	Camta1	0.445	Jak1	0.411	Khdrbs1	0.380	LOC100045198	0.408
88	Ndufa5	0.446	Slc9a3r2	0.412	Zscan4c	0.380	LOC386085	0.408
89	Sdhc	0.448	Ptp4a3	0.413	Sema4a	0.382	2310016C08RIK	0.408
90	C330034C07Rik	0.448	Calca	0.414	Wdfy1	0.383	1810032O08RIK	0.409
91	Dclk2	0.449	Ccnd1	0.415	Dusp6	0.384	ZNF512B	0.410
92	Igf2bp3	0.449	Rbpms	0.415	Igf2bp2	0.384	HNRNP1	0.413
93	Ppic	0.450	Gtl2	0.417	LOC329984	0.387	KRTDAP	0.413
94	9830141K10Rik	0.451	Sox4	0.418	Slc6a15	0.387	MANBA	0.415
95	Brp17	0.451	1300013J15Rik	0.418	Enah	0.388	ODZ4	0.416
96	Clgn	0.452	Faah	0.419	Ppic	0.389	HDC	0.417
97	Tex13	0.453	St5	0.421	Nuak2	0.391	BST2	0.417
98	Ercc5	0.454	A630082K20Rik	0.422	Tnpo3	0.394	LOC674912	0.417

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

99	5730406M06Rik	0.454	Tcf15	0.422	Hrb	0.394	ZP3	0.418
100	Prph	0.455	Evc2	0.423	Chst8	0.394	1810032O08RIK	0.418
101	Exoc4	0.455	Klf9	0.425	BC057079	0.395	BC057371	0.420
102	Klhdc10	0.456	LOC674912	0.425	Tmem40	0.398	ZYX	0.421
103	Rhox10	0.456	Ccng1	0.426	Tmem118	0.399	SLC9A3R2	0.423
104	Unc13b	0.456	Hsd17b7	0.426	Bnc2	0.399	PDLIM3	0.423
105	Ccng1	0.458	Klhl26	0.427	Sh3gl2	0.400	ASAH3L	0.424
106	Bach2	0.458	Piwil2	0.429	Rasgrp1	0.400	MANBA	0.424
107	Dnaic1	0.460	Hmgn3	0.429	Pycr2	0.401	IGSF9	0.425
108	EG626367	0.460	LOC674611	0.430	Rbpms	0.402	RUSC2	0.425
109	Atp13a1	0.461	D630023F18Rik	0.430	Klf9	0.403	DNMT3B	0.428
110	Dpm3	0.461	Dnajb6	0.430	Zscan5b	0.403	1300013J15RIK	0.430
111	Igf2bp3	0.461	Otx2	0.433	Lefty2	0.407	RUSC2	0.431
112	E130112E08Rik	0.461	Ercc5	0.433	Zdhhc21	0.407	PYCR2	0.431
113	LOC269515	0.463	Agfg2	0.436	A630082K20Rik	0.408	MEST	0.433
114	1110001N06Rik	0.463	LOC381904	0.437	Pfkm	0.408	C330021F23RIK	0.433
115	Mtf1	0.463	1500011K16Rik	0.438	Upp1	0.408	ENSMUSG00000074075	0.433
116	R3hdm1	0.463	Igf2bp3	0.438	Igf2bp3	0.410	SALL3	0.434
117	5730406M06Rik	0.464	BC032203	0.439	C130035G06Rik	0.411	REV1	0.437
118	Gars	0.464	AW555464	0.439	Mest	0.411	FBLN1	0.437
119	Ythdf1	0.464	Hrb	0.440	Igfbp2	0.413	MYLPF	0.437
120	1190003J15Rik	0.466	Enah	0.440	Mid1	0.413	BC061212	0.438
121	Ints7	0.467	Trp53inp1	0.441	Rnf130	0.414	GPR23	0.438
122	Mrps33	0.468	Mid1	0.441	Pdzk1	0.414	LRMP	0.439
123	Fanca	0.468	A530057A03Rik	0.442	Fbln1	0.416	LYPLA1	0.439
124	Trim32	0.470	Znf512b	0.443	LOC381904	0.416	4933403G14RIK	0.442

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

125	Casp2	0.470	Mia1	0.445	Evc2	0.416	CDKN2A	0.445
126	Adam23	0.470	Zfp36l1	0.445	Ccnd1	0.419	LOC633016	0.446
127	Mcm5	0.471	Atp2a3	0.445	Npm3-ps1	0.419	RBPM5	0.447
128	Ugcg	0.472	Ei24	0.449	Fkbp9	0.419	LOC229810	0.451
129	2900019M05Rik	0.472	Igfbp2	0.450	Hmgb2l1	0.421	SNURF	0.452
130	Myeov2	0.473	Neurog3	0.451	Acp6	0.421	ZFP292	0.453
131	Pgpep1	0.474	Arl8a	0.452	Ercc5	0.423	OSBPL3	0.453
132	LOC633016	0.474	Crtap	0.452	Calb2	0.423	INPP5D	0.454
133	Fzd5	0.475	Snhg11	0.452	Fer1l3	0.423	CTGF	0.454
134	Sumo1	0.475	Arhgef18	0.453	Rtn1	0.424	A430089I19RIK	0.454
135	Bbc3	0.476	Acp6	0.454	Hmgn3	0.424	MREG	0.455
136	Glipr2	0.476	Ccno	0.454	Tmc7	0.424	BDH1	0.455
137	Slc38a5	0.477	Igf2bp3	0.456	Nfkb1a	0.426	ACP6	0.456
138	Fn1	0.477	Ccng1	0.456	Ei24	0.427	HNRNPH1	0.456
139	Gli2	0.478	Cat	0.457	Hmgn3	0.427	SOX4	0.457
140	Nr5a2	0.478	Tmsb4x	0.458	Pdzk1	0.427	ODZ4	0.458
141	Flywch2	0.478	1700019N12Rik	0.458	2700038C09Rik	0.428	SCL0002540.1_6	0.459
142	Rtn1	0.480	Def6	0.459	Camk2b	0.429	APOE	0.459
143	A230057G18Rik	0.481	Cobl	0.459	Phlda3	0.430	PHLDA3	0.460
144	Zdhhc21	0.481	BC057371	0.460	1300013J15Rik	0.433	WNT8A	0.461
145	Tcea3	0.481	Ppic	0.463	Zyx	0.434	PIWIL2	0.462
146	Ezh2	0.482	G3bp2	0.463	Mid1ip1	0.434	D10BWG1379E	0.462
147	Clgn	0.483	Isyna1	0.463	Fez1	0.435	EXOC4	0.462
148	2410003K15Rik	0.483	Slc19a2	0.464	Dpysl2	0.436	SORD	0.464
149	Nxt2	0.484	1810032O08Rik	0.465	Prph	0.437	PLAC8	0.464
150	Psmc8	0.484	Zfp57	0.466	Npm3	0.437	TMEM92	0.465

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

151	Grb10	0.485	Crmp1	0.466	Foxp1	0.438	PRNP	0.465
152	Fez1	0.485	Cd97	0.468	Got1l1	0.440	PLA2G1B	0.465
153	Tnpo3	0.485	LOC100045403	0.468	Mylpf	0.441	PICALM	0.468
154	Gstk1	0.486	Pfkip	0.469	D630023F18Rik	0.441	2700038C09RIK	0.469
155	Wfdc10	0.486	Gars	0.469	Kbtbd2	0.444	RNF130	0.469
156	Mobp	0.487	Mgst3	0.469	1500011K16Rik	0.444	MGMT	0.469
157	Txnip	0.488	Cpn1	0.470	Hsd17b7	0.446	VMN2R-PS14	0.469
158	Gtl2	0.488	Cd97	0.470	E130306D19Rik	0.447	EPHA2	0.470
159	Snurf	0.490	Rbm28	0.471	LOC380927	0.447	LOC381398	0.472
160	Osbpl3	0.490	Gstk1	0.472	6330414G02Rik	0.448	KIF5A	0.472
161	Mcf2l	0.491	Sap30	0.473	Inpp5d	0.449	GADD45G	0.474
162	Mkrn1	0.491	Tnpo3	0.473	Rbm28	0.450	CDS1	0.475
163	Stau2	0.492	Inpp5d	0.474	Tex13	0.451	L1TD1	0.475
164	Stard4	0.494	Ccnc	0.474	Pfkm	0.451	KLF9	0.475
165	Kbtbd2	0.494	Odz4	0.475	Gars	0.451	LOC329750	0.475
166	Cdc20	0.494	Zfp212	0.475	Calca	0.452	CTSC	0.475
167	D130072O21Rik	0.495	2310003C23Rik	0.476	LOC212386	0.452	IGF2BP3	0.476
168	Svop	0.495	Cdc20	0.477	Papola	0.453	ADAP1	0.476
169	2210408F11Rik	0.495	Kcnu1	0.477	Zfp292	0.454	KHDRBS1	0.476
170	Hrb	0.495	Calca	0.478	Ddc	0.455	MKRN1	0.477
171	Pdia4	0.496	Mgmt	0.479	Klhl26	0.455	LOC665290	0.478
172	Ccng1	0.496	1110001J03Rik	0.480	Ripk4	0.456	D1PAS1	0.478
173	Lancl1	0.497	Tuba3a	0.482	C330021F23Rik	0.457	TPM2	0.479
174	Hmgn3	0.497	Tmc7	0.482	Armxc1	0.457	CPHX	0.480
175	4933403G14Rik	0.499	Snhg11	0.483	Pla2g10	0.457	AF067061	0.480
176	6330403E01Rik	0.499	Pacsin1	0.483	Smad2	0.458	DDX25	0.481

Appendix B - Top 200 downregulated genes in *Klf2*-null mESCs

177	1810032O08Rik	0.499	Ethe1	0.484	Khdrbs1	0.460	HMGN3	0.481
178	Fer1l3	0.499	LOC381302	0.484	Cpn1	0.460	RBM28	0.481
179	1700019D03Rik	0.500	C330021F23Rik	0.485	Rlf	0.460	KHDRBS1	0.481
180	Dnajc5	0.501	5033414K04Rik	0.485	Plac8	0.464	SSBP1	0.481
181	Hgsnat	0.501	Sdhc	0.486	Mreg	0.464	B930014J03RIK	0.482
182	Arl8a	0.502	Tmem118	0.486	Tmc6	0.464	ZIC2	0.482
183	LOC381889	0.502	Nucks1	0.487	Upp1	0.465	AW555464	0.483
184	Inpp5d	0.502	Ptpn14	0.488	Gpr125	0.465	SESN2	0.483
185	C77080	0.503	Zfhx3	0.489	Pla2g12a	0.465	LUC7L2	0.484
186	Zfp292	0.503	Hmgn3	0.489	Mest	0.466	2810026P18RIK	0.486
187	BC026590	0.503	Ampd3	0.490	Sdhc	0.467	FAM115C	0.488
188	C030025P15Rik	0.503	Tex19.1	0.490	Epha2	0.468	CLPS	0.488
189	2610019E17Rik	0.505	Ssbp4	0.490	Lypla1	0.469	IRF2BP2	0.489
190	Ndufb3	0.505	LOC100045280	0.490	LOC208080	0.469	AU018829	0.489
191	A630082K20Rik	0.506	Rab7l1	0.490	2210408F11Rik	0.469	CTNNAL1	0.490
192	Srrm2	0.506	Wdfy1	0.491	Rtn1	0.470	CCNC	0.490
193	Flnc	0.506	Casp2	0.491	Clk1	0.472	BB287469	0.490
194	Sec61b	0.507	9830141K10Rik	0.492	Prnp	0.473	ERCC5	0.490
195	Mex3a	0.507	Ddit4l	0.492	Mgst3	0.473	DYRK3	0.491
196	Sfrs14	0.508	Rev1	0.493	LOC100047486	0.473	HNRNPH1	0.491
197	D630036F01Rik	0.508	Fbln1	0.493	Sall3	0.473	LOC546233	0.491
198	1700007E06Rik	0.508	Epha2	0.494	Trim32	0.473	PDK1	0.491
199	Hnrpd1	0.509	Zfp330	0.494	Nupr1	0.474	RTEL1	0.492
200	Sumo1	0.510	Igfbp2	0.495	Upp1	0.474	E030026I10RIK	0.492

Klf2 Is an Essential Factor that Sustains Ground State Pluripotency

Jia-Chi Yeo,^{1,2} Jianming Jiang,¹ Zi-Ying Tan,^{1,3} Guo-Rong Yim,¹ Jia-Hui Ng,¹ Jonathan Göke,¹ Petra Kraus,^{1,6} Hongqing Liang,¹ Kevin Andrew Uy Gonzales,^{1,4} Han-Chung Chong,² Cheng-Peow Tan,¹ Yee-Siang Lim,¹ Nguan-Soon Tan,^{2,7} Thomas Lufkin,^{1,6} and Huck-Hui Ng^{1,2,3,4,5,*}

¹Genome Institute of Singapore, 60 Biopolis Street, #02-01 Genome Building, Singapore 138672, Singapore

²School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

³Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

⁴Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 117456, Singapore

⁵Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Block MD6, Centre for Translational Medicine, 14 Medical Drive #14-01T, Singapore 117599, Singapore

⁶Department of Biology, Clarkson University, Potsdam, NY 13699, USA

⁷Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos, A*STAR, Singapore 138673, Singapore

*Correspondence: nghh@gis.a-star.edu.sg

<http://dx.doi.org/10.1016/j.stem.2014.04.015>

SUMMARY

The maintenance of mouse embryonic stem cells (mESCs) requires LIF and serum. However, a pluripotent “ground state,” bearing resemblance to preimplantation mouse epiblasts, can be established through dual inhibition (2i) of both prodifferentiation Mek/Erk and Gsk3/Tcf3 pathways. While Gsk3 inhibition has been attributed to the transcriptional derepression of *Esrrb*, the molecular mechanism mediated by Mek inhibition remains unclear. In this study, we show that Krüppel-like factor 2 (Klf2) is phosphorylated by Erk2 and that phospho-Klf2 is proteasomally degraded. Mek inhibition hence prevents Klf2 protein phosphodegradation to sustain pluripotency. Indeed, while *Klf2*-null mESCs can survive under LIF/Serum, they are not viable under 2i, demonstrating that Klf2 is essential for ground state pluripotency. Importantly, we also show that ectopic Klf2 expression can replace Mek inhibition in mESCs, allowing the culture of *Klf2*-null mESCs under Gsk3 inhibition alone. Collectively, our study defines the Mek/Erk/Klf2 axis that cooperates with the Gsk3/Tcf3/*Esrrb* pathway in mediating ground state pluripotency.

INTRODUCTION

The self-renewing pluripotent state was first captured in mouse embryonic stem cells (mESCs) over 30 years ago (Evans and Kaufman, 1981; Martin, 1981). The standard culture of mESCs requires the presence of Leukemia inhibitory factor (LIF) and serum to maintain a self-renewing and pluripotent state (Nichols and Smith, 2011). However, under LIF/Serum culture conditions, the prodifferentiation Fibroblast growth factor/Mitogen-activated protein kinase kinase/Extracellular signal-regulated kinase (Fgf/Mek/Erk) pathway, which initiates lineage commitment

(Kunath et al., 2007; Stavridis et al., 2007), and the repressive Gsk3/Tcf3 pathway, which induces mESCs differentiation into epiblast stem cells (EpiSCs), remain active (ten Berge et al., 2011; Wray et al., 2011). Interestingly, by using two small molecules, PD0325901 and CHIR99021 (denoted as 2i), to inhibit Mek and Gsk3 signaling, it was found that mESCs could be kept pluripotent in the absence of LIF/Serum stimulation (Ying et al., 2008). Importantly, the use of 2i in place of LIF/Serum is considered a major advancement for pluripotent stem cell biology because it enables the successful derivation of ESCs from previously refractory mouse strains (Blair et al., 2011).

mESCs cultured under 2i exhibit several unique features, including uniformly high expression of several pluripotency genes (Wray et al., 2010), reduced expression of lineage-specific genes with a lower occurrence of bivalent domains (Marks et al., 2012), and the possession of a genome-wide DNA hypomethylated state similar to the mouse inner cell mass (ICM) (Bagci and Fisher, 2013). Therefore, mESCs under 2i are proposed to exist in a distinct “ground state” of pluripotency resembling the ICM of preimplantation blastocysts (Nichols et al., 2009; Ying et al., 2008). Given these similar features, investigating how 2i ground state pluripotency operates should reveal new insights into the molecular design of pluripotency and preimplantation embryogenesis.

How exactly then does the dual inhibition of Fgf/Mek/Erk and Gsk3/Tcf3 signaling pathways enable ground state pluripotency? Analysis of Tcf3 genomic binding sites through chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) reveals that Gsk3 signaling inhibition alleviates Tcf3-mediated transcriptional repression of *Esrrb* (Martello et al., 2012). Notably, because *Esrrb* overexpression could replace CHIR99021 under 2i conditions, this finding demonstrates the importance of *Esrrb*, which is downstream of Gsk3/Tcf3 inhibition in ground state pluripotency.

In contrast to the Gsk3/Tcf3 pathway, the exact mechanism by which Mek/Erk signaling inhibition affects ground state pluripotency is less clear. Here, by analyzing the change in protein expression of Krüppel-like factor 2 (Klf2), Klf4, and Klf5 under LIF/Serum and 2i, we found that Mek/Erk stimulation leads to the phosphodependent degradation of Klf2 protein.

Figure 1. Active Mek/Erk Signaling Negatively Regulates Klf2 Protein in a Posttranslational Manner

(A) Oct4 and Nanog mRNA changes in mESCs under LIF/Serum or 2i. Data are the averages of biological triplicates $\pm$ standard error of the mean (SEM).

(B) Klf mRNA changes in mESCs under LIF/Serum or 2i. Data are the averages of biological triplicates $\pm$ SEM.

(C) Western blot of key mESC factors in mESCs under LIF/Serum or 2i. Arrow demarcates the Klf4 protein band.

(D) Klf protein level changes in mESCs after 30 min of 2i exposure.

(E) Klf gene expression changes in mESCs after 30 min of 2i exposure. Data are the averages of biological triplicates $\pm$ SEM.

(F) Effect of 2 hr treatment using PD0325901 (PD), CHIR99021 (CH), or PD0325901 + CHIR99021 (PD/CH) upon Klf protein levels in LIF/Serum mESCs.

(G) Time course of Klf protein changes in mESCs treated with 100 nM CalA.

(H) Klf protein changes in LIF/Serum mESCs treated with various concentrations of MG132 for 1 hr.

Consequently, the addition of Mek inhibitor PD0325901 halts Mek/Erk-induced Klf2 protein degradation and sustains elevated Klf2 protein levels. Importantly, we also demonstrate that Klf2 is an essential factor during ground state conditions as *Klf2*-null mESCs undergo cell death under 2i. Together with the report that Gsk3 inhibition acts to alleviate Tcf3-mediated repression of *Esrrb* (Martello et al., 2012), our study reveals that Mek/Erk inhibition sustains mESCs through Klf2 protein stabilization, thus completing the model for how 2i supports ground state pluripotency.

RESULTS

Klf2 Is Negatively Regulated by Mek/Erk Signaling

We have previously shown that mESCs cultured under LIF/Serum conditions require the collective presence of Klf2, Klf4, and Klf5 to sustain self-renewal (Jiang et al., 2008). Therefore, we were interested to know if this triple redundancy exists during ground state pluripotency. At the transcriptional level, we found that LIF/Serum and 2i-cultured mESCs exhibit comparable Oct4 and Nanog expression (Figure 1A). In addition, 2i-cultured mESCs expressed low levels of Klf4 and Klf5 and

showed a marginal 2-fold upregulation of Klf2 expression (Figure 1B). The change in Klf4 and Klf5 expression is expected because LIF is not present under 2i conditions to induce Klf4 and Klf5 expression (Hall et al., 2009; Niwa et al., 2009). However, 2i appears to have a modest positive effect on Klf2 expression compared to LIF/Serum conditions (Greber et al., 2010).

We next probed into the protein level changes of these three Klf proteins in ground state mESCs and found that, correlating with transcriptional data, Klf4 and Klf5 protein levels were reduced under 2i (Figure 1C). However, unlike the marginal upregulation of *Klf2* transcript (Figure 1B), we found a striking increase in Klf2 protein following 2i culture (Figure 1C). Because this increase in Klf2 protein levels under 2i is unlikely to be fully accounted for by the 2-fold upregulation of Klf2 mRNA (Figure 1B), we hence hypothesized that Klf2 protein may be posttranslationally regulated. In support of this hypothesis, we found the increase of Klf2 protein under 2i to be rapid and detectable within 30 min of switching from LIF/Serum (Figure 1D). Importantly, because Klf2 transcript was only increased slightly within this timeframe despite the substantial increase of Klf2 protein levels (Figure 1E), our data strongly implicate a posttranslational regulation of Klf2 under 2i. We did not however observe any changes to Klf4 and Klf5 protein or mRNA during this 30 min period (Figures 1D and 1E), suggesting that the decrease of Klf4 and Klf5 during ground state pluripotency (Figure 1C) is regulated at the transcriptional level.

2i culture contains the inhibitors PD0325901 and CHIR99021, which block the activity of Mek and Gsk3, respectively. Therefore,

to determine which of these inhibitors is responsible for the increase in Klf2 protein, we added the inhibitors alone or in combination to mESCs cultured in LIF/Serum for 2 hr before protein extraction. We found that the Mek inhibitor PD0325901 alone was sufficient to elevate the Klf2 protein level to that observed in 2i (Figure 1F). Gsk3 inhibition with CHIR99021, however, did not result in any increase in Klf2 protein level (Figure 1F). Similar to the data presented in Figures 1D and 1E, Klf4 and Klf5 protein levels are refractory toward either Mek or Gsk3 signaling inhibition in this experiment (Figure 1F).

Because Klf2 protein appears to be negatively regulated by Mek signaling, we next examined the effect of hyperphosphorylation on Klf proteins. Using Calyculin A (CalA) to inhibit serine/threonine protein phosphatases, we found a rapid loss of Klf2 protein within 30 min, but we did not find this loss for Klf4 and Klf5 protein (Figure 1G). This data suggest that any putative phosphorylation on Klf2 by Mek signaling could lead to the degradation of Klf2 protein. Indeed, blockade of proteasome activity by MG132 in LIF/Serum cultured mESCs was found to halt Klf2 protein degradation, leading to elevated Klf2 protein levels similar to that resulting from treatment with Mek inhibitor (Figure 1H).

Taken together, these results demonstrate that active Mek signaling destabilizes Klf2 protein. By inhibiting Mek with PD0325901, this phosphodependent degradation of Klf2 is relieved, resulting in Klf2 protein accumulation under ground state conditions.

Klf2 Directly Interacts with and Is Phosphorylated by Erk2

To establish that Klf2 is phosphorylated, we immunoprecipitated (IP) Klf2 from mESC lysates and found the presence of phosphoserine/threonine-specific residues on Klf2 (Figure 2A). Furthermore, while comparable levels of Klf2 IP were obtained among the different mESC samples, we could only observe a strong loss of phospho-Klf2 signal following PD0325901 treatment (Figure 2A). This demonstrates that Klf2 serine/threonine phosphorylation is regulated via a Mek signaling mechanism.

Because the downstream targets of Mek are the serine/threonine kinases Erk1/2 (Roberts and Der, 2007), and because previous phosphoproteomic studies in mESCs have identified phosphoresidues at several Erk consensus phosphorylation P-X-S/T-P motifs in Klf2 protein (Li et al., 2011; Pines et al., 2011) (Figure 2B), we hence hypothesize that Klf2 may be phosphorylated by Erk.

Given that mESCs predominantly express Erk2 (Kunath et al., 2007), we investigated a potential interaction between Klf2 and Erk2. As an initial trial using HEK293T cells, we found that Klf2 and Erk2 were indeed able to bind with each other. Specifically, we observed that exogenous HA-tagged Klf2 protein could be coimmunoprecipitated (co-IP) with endogenous Erk2 (Figure 2C). Likewise, the reciprocal pulldown of exogenous HA-tagged Klf2 could result in the co-IP of coexpressed Flag-tagged Erk2 or endogenous Erk2 (Figures S1A and S1B available online).

To further validate the Klf2-Erk2 interaction in mESCs, we employed the proximity ligation assay (PLA) (Söderberg et al., 2006) to detect for in situ endogenous Klf2-Erk2 association. In principle, PLA functions by exploiting the close proximity of protein-protein interactions to link their respective cognate antibodies and generate a fluorescence signal. Following PLA,

we found numerous Klf2-Erk2 binding signals in mESCs, but not when either α -Klf2 or α -Erk2 antibodies were absent (Figure 2D). This indicates that the observed Klf2-Erk2 PLA signals are specific and that the Klf2 and Erk2 protein colocalize with each other.

Lastly, we performed an in vitro kinase assay using [γ -³²P]-ATP and found that activated recombinant Erk2 could specifically phosphorylate GST-tagged recombinant Klf2, but not the negative control GST protein (Figure 2E). Therefore, in the context of our previous data, which support a Klf2-Erk2 interaction (Figures 2A–2D), the phosphorylation of recombinant GST-Klf2 by active Erk2 clearly demonstrates a direct and functional Klf2-Erk2 relationship.

Klf2 Is Critical for mESCs Cultured under Ground State Conditions

Because mESCs cultured under 2i express a high level of Klf2 with a reduced level of Klf4 and Klf5 protein (Figure 1C), we hypothesized that Klf2, rather than Klf4 or Klf5, is important for ground state pluripotency. To examine the role of Klf2 during ground state conditions, we sought to generate *Klf2* knockout mESCs. This was done by replacing a genomic segment covering exon 2 and 3 of the *Klf2* gene with an *IRES-LacZ* reporter and *Neo*-selection cassette (Figures S2A–S2C). This targeting strategy generates a *Klf2*-null allele. Subsequently, single-allele targeted *Klf2*-heterozygous (*Klf2*-Het) mESCs were used to generate germ-line-transmitting chimeras, and the eventual *Klf2*-Het transgenic mice were used for breeding.

In agreement with previous studies (Kuo et al., 1997; Wani et al., 1998), our *Klf2* knockout mice were all embryonic lethal due to intraembryonic hemorrhaging (Figures S2D and S2E). Additionally, because Klf2 is expressed during the development of the embryonic vasculature (Kuo et al., 1997), we also found strong LacZ signal in the blood vessels of *Klf2*-null embryos (Figure S2F). Therefore, our corroborating in vivo data with the literature validates the specificity and phenotype of our *Klf2* gene targeting.

Because *Klf2*-null mice are embryonic lethal, *Klf2*-null blastocysts have to be isolated via the mating of *Klf2*-Het parents. Using LIF with knockout serum (KSR), we successfully generated 19 mESC lines. Among these obtained were 5 *Klf2* wild-type (*Klf2*-WT), 11 *Klf2*-Het and 3 *Klf2*-null mESC lines (Figures S2G and S2H). This observed genotype ratio is consistent with the expected 1:2:1 Mendelian ratio following a *Klf2*-Het mating scheme (Figures S2G and S2H). Importantly, our derived *Klf2*-null mESCs were found to be pluripotent: they exhibited positive alkaline phosphatase (AP) staining, expression of key mESC genes and proteins except Klf2, an ability to form teratomas, and a contribution toward mouse chimeras (Figures 3A–3E and Figure S2I).

Because feeder cells secrete factors that interfere with signaling pathway studies, we adapted the *Klf2*-null mESCs onto gelatin surfaces. These feeder-free *Klf2*-null mESCs were AP positive and retained an overall mESC gene expression signature (Figures 3F and 3G). PLAs of these *Klf2*-null mESCs using antibodies against Erk2 and Klf2 were negative (Figure S2J), demonstrating that our previously observed Klf2-Erk2 PLA signals did not arise from nonspecific interactions (Figure 2D). Importantly, although these feeder-free *Klf2*-null mESCs could be cultured under LIF/Serum, we found that *Klf2*-null mESCs cannot be maintained under 2i for more than three or four

Figure 2. Klf2 Directly Interacts with and Is Phosphorylated by Erk2

(A) Detection of phospho-serine/threonine residues from immunoprecipitated Klf2 in nontreated, 2 hr PD0325901 treated (1 μM) (PD03), or 30 min of CaIA (100 nM) treated mESCs.

(B) Motif analysis of mouse Klf2 protein sequence. Potential Erk or Gsk3 phosphorylation sites are marked in yellow or green, respectively.

(C) Co-IP of transfected HA-tagged Klf2 using endogenous Erk2 in 293T cells. WC, whole cell lysate.

(D) PLA and confocal imaging of LIF/Serum mESCs probed with both α-Klf2 and α-Erk2 antibodies. Negative control PLA experiments were done using single α-Klf2 or α-Erk2 antibody only. Cell nuclei are stained with DAPI. Scale bar represents 10 μm.

(E) In vitro phosphorylation of recombinant GST-Klf2 by active Erk2 using [γ-32P]-ATP.

See also Figure S1.

passages (Figures 3H and 3I). This nonviability of *Klf2*-null mESCs under 2i however can be rescued with exogenous LIF (Figures 3H and 3I). Because these *Klf2*-null mESCs exhibited elevated Annexin V and propidium iodide staining following culture under 2i, we propose that *Klf2*-null mESCs undergo cell death under ground state condition (Figure 3J).

To rescue the cell death phenotype under 2i, we stably overexpressed wild-type HA-tagged Klf2 in our *Klf2*-null mESCs using *piggybac* vectors (Wang et al., 2008). Indeed, we found that ectopic Klf2 expression was able to prevent *Klf2*-null mESC death, allowing the continuous culture of *Klf2*-null mESCs under 2i for up to 10 passages (Figure 3K). In addition, we also found that Klf2 overexpression was sufficient to replace PD0325901, enabling the culture of *Klf2*-null mESCs in CHIR99021 alone (1i) (Figure 3L).

Because Klf2 protein is negatively regulated by Erk phosphorylation, we next generated various Klf2 serine/threonine to alanine point mutations at predicted Erk phosphorylation sites. This was to examine if Klf2 phosphomutants could confer any positive growth effects on *Klf2*-null mESCs under 1i. Because we observed that the single-site mutants showed modest or weak effect, we next proceeded to generate double and quadruple Klf2 serine/threonine mutants. Notably, these Klf2 double and quadruple mutants in 1i culture were found to promote a 30% and 60% increase, respectively, in *Klf2*-null mESC colony numbers (Figures S2K and S2L). Therefore, taken together, our data demonstrate the essentiality of Klf2 protein downstream of Mek/Erk signaling pathway inhibition.

Figure 3. Klf2-null mESCs Fail to Self-Renew in Ground State Conditions

(A) AP staining of *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum on feeder. Scale bar represents 100 μ m.
 (B) Reverse transcription PCR of key mESC genes from *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum on feeder.
 (C) Western blot for *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum on feeder. Arrow demarcates the Klf4 protein band.
 (D) Teratoma sections generated from feeder-grown LIF/Serum *Klf2*-null mESCs.
 (E) Generation of *Klf2*-null chimeric mouse using feeder-grown LIF/Serum *Klf2*-null mESCs.

(legend continued on next page)

Given this necessity for Klf2 for 2i-cultured mESCs, we examined if the absence of Klf2 is detrimental toward establishing mESCs from blastocysts under 2i conditions. Starting from 85 blastocysts obtained through *Klf2*-Het crosses, we eventually derived 41 mESC lines under 2i with feeder conditions. Among these, we genotyped 20 *Klf2*-WT, 20 *Klf2*-Het, and a single *Klf2*-null mESC line (Figure 3M). This abnormally skewed *Klf2* genotype frequency using the 2i derivation method far deviates from the 1:2:1 expected ratio from a standard Het × Het mating. Because 2i culture has been extremely effective in deriving mESCs from recalcitrant mouse strains (Blair et al., 2011), our severe inefficiency in generating *Klf2*-null and *Klf2*-Het mESCs under 2i emphasizes again this critical necessity for Klf2 in ground state pluripotency.

Klf2-null mESCs under 2i Exhibit Aberrant Primordial-Germ-Cell-Associated Gene Expression

Although 2i-cultured *Klf2*-null mESCs undergo cell death around 2i passage 3 (Figures 3I and 3J), expression levels of key pluripotency genes were still fairly consistent albeit with elevated Klf4, Klf5, and Nanog levels (Figures S3A and S3B). To determine the cause of *Klf2*-null mESC death under 2i, we performed microarray gene expression profiling of *Klf2*-WT and *Klf2*-null mESCs at 2i passage 3. Although gene ontology (GO) analysis of transcriptional changes between *Klf2*-WT and *Klf2*-null mESCs only revealed enrichment of terms related to metabolic and cellular processes (Figure S3C), closer inspection into the list of gene changes revealed elevated levels of primordial germ cell (PGC) genes such as *Dppa3*, *c-Kit*, *Tcfap2c*, *Dnd1*, *Ifitm1*, and *Prdm1* (Figures S3D and S3E). Because Klf2 is known to synergize with the PGC factor Prdm14 to convert EpiSCs into naive pluripotency (Gillich et al., 2012), and because Klf2 expression is associated with PGC development (Kurimoto et al., 2008), we next examined the relationship of Klf2 with PGC gene regulation during ground state pluripotency.

In *Klf2*-null mESCs, we observed that elevated PGC gene expression could be detectable during the early transition phase from LIF/Serum into 2i passage 1 (Figure 4A). However, unlike 2i-cultured *Klf2*-WT mESCs, which gradually downregulate PGC gene expression relative to the starting LIF/Serum PGC transcript levels (Figure 4B), we noticed that *Klf2*-null mESCs maintained a sustained PGC gene expression despite continuous 2i culture (Figure 4B). This lead us to hypothesize that without Klf2, ground state mESCs could be potentially defective in their ability to properly suppress PGC gene expression.

To examine whether Klf2 may be involved in regulating these PGC genes, we next conducted Klf2 ChIP-seq for LIF/Serum and 2i cultured mESCs. Indeed, Klf2 binding was detected at

the promoters of key PGC genes such as *Prdm1*, *Tcfap2c*, and *c-Kit* in both LIF/Serum and 2i conditions (Figure 4C). This suggests that Klf2 may play a possible regulatory role at these PGC genes. Notably, because we observed a concurrent decrease in Klf4 binding at *Prdm1*, *Tcfap2c*, and *c-Kit* promoters following 2i culture (Figure S3F), it is probable that it is Klf2 rather than Klf4 that plays a role to suppress PGC gene expression under 2i.

Hence in the context of *Klf2*-null mESCs, the loss of Klf2 without the additional support from LIF-induced Klf4 results in aberrant PGC gene transcription and potential differentiation into PGC-like lineages. However, due to the restrictive nature of 2i culture for the pluripotent ground state, any differentiating *Klf2*-null mESCs would not be able to survive. This may account for our observed cell death phenotype of *Klf2*-null mESCs under 2i (Figure 3J).

DISCUSSION

In light of our findings, we propose a model for the regulation of Klf2 by Mek/Erk signaling under conventional LIF/Serum condition or 2i-induced ground state (Figure 4D). Under LIF/Serum condition, both the Fgf/Mek/Erk and LIF signaling pathways in mESCs are active. Although active Fgf-signaling is antagonistic toward Klf2 protein stability, not all Klf2 protein is degraded. As such, some Klf2 protein may continue to function with the LIF-induced Klf4/Klf5 to exert a triple-redundancy effect in mESCs (Jiang et al., 2008). However, during 2i-induced ground state, because of the absence of LIF, the level of Klf4/Klf5 protein is inadequate to sustain mESCs. Under this condition, inhibition of Mek signaling is necessary to prevent Mek/Erk phosphodependent protein degradation of Klf2, which leads to an elevated Klf2 protein level to sustain the self-renewal of mESCs.

Despite the reported functional redundancy of Klf2/Klf4/Klf5 in LIF/Serum mESCs, not all Klf proteins exhibit equal potency. During somatic cell reprogramming, Klf2 is a more efficient reprogramming factor than Klf1, Klf4, or Klf5 (Feng et al., 2009; Nakagawa et al., 2008). Similarly, Prdm14, when synergized with Klf2, could rapidly drive EpiSCs into mESCs conversion, but not so for Klf4 or Klf5 (Gillich et al., 2012). It therefore appears that among the Klf proteins in mESCs, Klf2 exhibits the strongest effect in promoting mESC pluripotency.

Because *Klf2*-null mESCs cannot properly downregulate PGC gene expression and would eventually die under 2i culture, the binding of Klf2 at key PGC genes suggests a possible function of Klf2 to properly regulate PGC expression during ground state pluripotency. However, the mechanism of this hypothesized Klf2 repression at PGC gene loci is unclear, and may possibly involve recruiting of Polycomb group proteins for gene repression. We

(F) AP staining of feeder-free adapted *Klf2*-WT and *Klf2*-null in LIF/Serum. Scale bar represents 100 μ m.

(G) LIF/Serum gene expression profiling of feeder-free *Klf2*-WT and *Klf2*-null mESCs. Data are the averages of biological triplicates $\pm$ SEM.

(H) Bright-field image of feeder-free *Klf2*-WT and *Klf2*-null mESCs at passage 4 under LIF/Serum, 2i, or 2i+LIF. Scale bar represents 200 μ m.

(I) Cumulative cell count of feeder-free *Klf2*-WT and *Klf2*-null mESCs under LIF/Serum, 2i, or 2i+LIF. Data are the averages of biological triplicates $\pm$ SEM.

(J) Annexin V and propidium iodide FACS analysis of *Klf2*-WT and *Klf2*-null mESCs cultured in 2i from passage 1, 2, and 3.

(K) Bright-field image of *Klf2*-null mESCs transfected with *piggybac* (PB) vectors containing GFP or wildtype 3HA-tagged Klf2. mESCs were cultured under 2i for the indicated number of passages. Scale bar represents 200 μ m.

(L) Bright-field image of *Klf2*-null mESCs transfected with *PB-GFP* or wild-type *PB-3HA-Klf2* plasmids cultured under CHIR99021 only (CH) for the indicated number of passages. Scale bar represents 200 μ m.

(M) Table summarizing the genotypes of 2i/feeder-derived Klf2 mESC lines isolated from blastocysts obtained through *Klf2*-Het parent crosses.

See also Figure S2.

Figure 4. *Klf2*-null mESCs Exhibit Aberrant PGC Lineage Priming under 2i

(A) Microarray heat-map of PGC genes of feeder-free *Klf2*-WT and *Klf2*-null mESCs cultured in LIF/Serum, 2i passage 1 (P1), and 2i passage 2 (P2). Red and green boxes denote increased or decreased gene expression, respectively.

(B) Relative gene expression changes of PGC genes in feeder-free *Klf2*-WT and *Klf2*-null mESCs in LIF/Serum, 2i (P1), and 2i (P2). Data are the averages of biological triplicates $\pm$ SEM.

(C) Klf2 ChIP-seq peaks at key PGC promoters in mESCs cultured under LIF/Serum or 2i passage 3 (P3).

(D) Model of Klf2 protein regulation by active Mek/Erk signaling pathway under conventional LIF/Serum culture condition or 2i-induced ground state. See also Figure S3.

also wish to highlight that while Klf2 is generally considered to be a PGC-associated factor (Kurimoto et al., 2008), mESCs and PGCs are two separate cell types, each with their unique developmental potential and transcriptional networks. It would be premature to conclude whether the role of Klf2 to suppress PGC genes during ground state pluripotency is similar to how Klf2 affects in vivo PGC development.

While 2i ground state mESCs are said to resemble the in vivo preimplantation ICM, our inability to culture or efficiently derive Klf2-null mESCs under 2i appears to produce a discrepancy. This is because Klf2-null embryos can develop past the ICM stage until E12.5–E14.5 (Kuo et al., 1997; Wani et al., 1998) (Figures S2D and S2F). To address this issue, we propose two explanations. First, while pluripotent mESCs are derived from the ICM, the actual duration for which the ICM exists in vivo is transient. Any defects arising from the loss of Klf2 may not be able to manifest within this timeframe. Second, the developing mouse blastocyst is known to express LIF (Nichols et al., 1996), and this endogenous LIF may compensate for the loss of Klf2 in vivo.

To conclude, we have uncovered Klf2 to be a critical factor for mESC ground state pluripotency, and that the absence of Klf2 is detrimental for mESC self-renewal under 2i conditions. Importantly, our results reveal how Mek inhibition, acting to stabilize Klf2 protein, is essential to establish and sustain mESC ground state pluripotency. Together with the understanding that Gsk3 inhibition relieves Tcf3-mediated *Esrrb* repression (Martello et al., 2012), our finding provides an integrated model of how dual inhibition of two major prodifferentiation pathways in mESCs via transcriptional and protein level regulation can maintain a pluripotent ground state (Figure 4D).

EXPERIMENTAL PROCEDURES

Cell Culture

Feeder-free E14 mESCs were cultured on 0.1% gelatin-coated plates at 37°C with 5% CO₂ in Dulbecco's modified Eagle medium (DMEM; GIBCO), supplemented with 15% heat-inactivated ESC-certified fetal bovine serum (FBS; GIBCO), 0.055 mM β-mercaptoethanol (GIBCO), 2 mM L-glutamine, 0.1 mM MEM nonessential amino acid, 20 μg/ml gentamicin (GIBCO), and 1,000 units/ml of LIF (Chemicon). mESCs were passaged at a ratio of 1:6 every 2 days. Klf2-transgenic mESC lines were cultured on 0.1% gelatin or mitomycin-C treated CD1 MEF feeder cells in similar mESC cell culture media and passaged at a 1:6 ratio every 2 days.

For chemically defined 2i conditions, mESCs were cultured either on CD1 MEF feeder cells or on 0.1% coated gelatin plates using serum-free N2B27 media supplemented with Mek inhibitor PD0325901 (1 μM) and Gsk3β inhibitor CHIR99021 (3 μM) (Stemgent) as previously described (Ying et al., 2008). 2i mESCs were passaged with Accutase (Chemicon) at a 1:6 ratio every 2 days.

CD1 MEFs were isolated from E13.5 CD1 mouse embryos by dissociation in 0.25% trypsin at 37°C for 10 min and cultured in 10% FBS-DMEM, supplemented with 2 mM L-glutamine, 0.055 mM β-mercaptoethanol (GIBCO), and 20 μg/ml of gentamicin (GIBCO). HEK293T cells were cultured in similar 10% FBS-DMEM media.

Generation of Klf2-LacZ Reporter Mice

The scheme to target the *Klf2* locus is as summarized in Figure S2A. V6.4 mESCs were electroporated with a linearized targeting vector consisting of *FRT-En2SA-IRES-LacZ-loxP-βactinP-Neo-FRT-loxP* (KOMP Repository Clone ID: DPGS00163_A_G03) and plated onto mitomycin-C treated CD1 feeder cells under G418 selection. After 10 days, G418-resistant colonies were then isolated and expanded and genomic DNA was extracted for genotyping by PCR and Southern blot. Two independent heterozygous *Klf2^{LacZ-Neo/+}* mESC lines were then microinjected into two- to eight-cell

stage C56BL/6J embryos (Kraus et al., 2010) to generate germ-line-transmitting chimeric mice. The resultant *Klf2^{LacZ-Neo/+}* mice were then bred and maintained on a mixed C57BL/6 and Sv129 genetic background.

Embryo Collection and mESC Derivation

All E3.5 blastocysts were flushed from mouse uteri using M2 media (Sigma) and cultured individually in 24-well plates containing CD1 feeder cells. For derivation of mESCs under LIF conditions, blastocysts were cultured as previously described (Bryja et al., 2006). For 2i derivation of mESCs, isolated blastocysts were cultured in N2B27 media supplemented with 2i as described (Ying et al., 2008).

PLA

PLA experiments were performed in accordance with the manufacturer's protocol (Olink Biosciences, Sweden) using antibody pairs to probe for protein interactions of interest. Mouse monoclonal anti-Erk2 (clone D2, sc-1647) (1:100; Santa Cruz) was paired with rabbit polyclonal anti-Klf2 (1:100; purified with protein G column). Following the primary antibody incubation, a pair of secondary proximity probes—Duolink in situ PLA probe anti-rabbit plus with Duolink in situ PLA probe anti-mouse minus (Olink Biosciences, Sweden)—were diluted at a 1:5 ratio and mixed in antibody diluent buffer. Sections were then incubated for 60 min at 37°C in a preheated humidified chamber. Duolink In Situ Detection Reagents Orange kit (Olink Biosciences, Sweden) was used for the subsequent assay. The sections were washed twice for 5 min with Duolink wash buffer A (10 mM Tris-HCl [pH 7.4]; 150 mM NaCl, and 0.05% Tween 20) and then incubated 30 min with prepared hybridization ligation mix from the kit at 37°C in a preheated humidified chamber. Next, the sections were washed twice with Duolink wash buffer A and incubated in the dark with prepared amplification-detection mix from the kit for 100 min at 37°C in the same chamber. After incubation, sections were washed twice for 10 min with Duolink wash buffer B (200 mM Tris-HCl [pH 7.4] and 100 mM NaCl) and 0.01× Duolink wash buffer B for 2 min.

Sections were counterstained and mounted in Vectorshield mounting media with DAPI (Vector Laboratories, USA). Negative controls were carried out in the absence of the primary antibody. PLA images were taken using the LSM 710 confocal microscope (Carl Zeiss, Germany) with the Plan-Apochromat 63×/1.4 oil differential interference contrast objective. Analyses were performed with the ZEN 2012 Light Edition software (Carl Zeiss, Germany).

Cell Apoptosis Detection

Cells were collected by trypsinization and washed with PBS. To detect early and late apoptotic cells, Annexin V Apoptosis detection kit (Santa Cruz) was used. Briefly, 1×10^5 cells were mixed and incubated for 15 min at room temperature in the dark with 10 μg/ml Annexin V FITC and 5 μg/ml propidium iodide. Stained cells were analyzed by BD LSRII (BD Biosciences) and parameters were set consistent for all samples. For apoptosis detection, compensation was set to eliminate the spillover of the FITC channel to the propidium iodide channel.

Microarray Analysis

RNA was extracted from feeder-free Klf2-WT and Klf2-null mESCs maintained in LIF/Serum culture, or during 2i passage 1, 2, or 3, using Trizol (Invitrogen). RNA samples were then purified using standard isopropanol precipitation techniques before DNase1 treatment. RNA was then purified again using the PureLink RNA Mini Kit (Invitrogen), and 500 ng of total RNA was used for cDNA amplification and biotin-tagging by the Illumina TotalPre RNA amplification kit. Amplified biotin-tagged cDNA was then used for hybridization onto Illumina MouseWG-6 Expression BeadChip v2.0 according to the manufacturer's instructions. Hybridized and Cy3-labeled microarray chips were then scanned and the results were processed using BeadStudio and Significance analysis of Microarrays (SAM).

ACCESSION NUMBERS

ChIP-seq and microarray data are available in the Array Express database (<http://www.ebi.ac.uk/arrayexpress>) under accession numbers E-MTAB-2365 and E-MTAB-2366, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes three figures, three tables, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.stem.2014.04.015>.

ACKNOWLEDGMENTS

We are grateful to the Biomedical Research Council (BMRC) and Agency for Science, Technology and Research (A*STAR) for funding. J.Y., Z.T., and K.G. are supported by the NTU Research, SINGA, and NGS Scholarships, respectively. J.G. is supported by a fellowship within the Postdoctoral Programme of the German Academic Exchange Service (DAAD). We thank W. Yue and G. Feng for their assistance in autoradiographic experiments. We also thank members of the Ng laboratory for their support and comments on this manuscript.

Received: October 14, 2013

Revised: March 14, 2014

Accepted: April 17, 2014

Published: June 5, 2014

REFERENCES

- Bagci, H., and Fisher, A.G. (2013). DNA demethylation in pluripotency and reprogramming: the role of tet proteins and cell division. *Cell Stem Cell* **13**, 265–269.
- Blair, K., Wray, J., and Smith, A. (2011). The liberation of embryonic stem cells. *PLoS Genet.* **7**, e1002019.
- Bryja, V., Bonilla, S., Cajánek, L., Parish, C.L., Schwartz, C.M., Luo, Y., Rao, M.S., and Arenas, E. (2006). An efficient method for the derivation of mouse embryonic stem cells. *Stem Cells* **24**, 844–849.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotent cells from mouse embryos. *Nature* **292**, 154–156.
- Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J., et al. (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat. Cell Biol.* **11**, 197–203.
- Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W., Pasque, V., Magnúsdóttir, E., and Surani, M.A. (2012). Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* **10**, 425–439.
- Greber, B., Wu, G., Bernemann, C., Joo, J.Y., Han, D.W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., and Schöler, H.R. (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* **6**, 215–226.
- Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* **5**, 597–609.
- Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* **10**, 353–360.
- Kraus, P., Leong, G., Tan, V., Xing, X., Goh, J.W., Yap, S.P., and Lufkin, T. (2010). A more cost effective and rapid high percentage germ-line transmitting chimeric mouse generation procedure via microinjection of 2-cell, 4-cell, and 8-cell embryos with ES and iPS cells. *Genesis* **48**, 394–399.
- Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895–2902.
- Kuo, C.T., Veselits, M.L., Barton, K.P., Lu, M.M., Clendenin, C., and Leiden, J.M. (1997). The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. *Genes Dev.* **11**, 2996–3006.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22**, 1617–1635.
- Li, Q.R., Xing, X.B., Chen, T.T., Li, R.X., Dai, J., Sheng, Q.H., Xin, S.M., Zhu, L.L., Jin, Y., Pei, G., et al. (2011). Large scale phosphoproteome profiles comprehensive features of mouse embryonic stem cells. *Mol. Cell Proteomics* **10**, M110 001750.
- Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604.
- Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491–504.
- 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.
- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochizuki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**, 101–106.
- Nichols, J., and Smith, A. (2011). The origin and identity of embryonic stem cells. *Development* **138**, 3–8.
- Nichols, J., Davidson, D., Taga, T., Yoshida, K., Chambers, I., and Smith, A. (1996). Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis. *Mech. Dev.* **57**, 123–131.
- Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215–3222.
- Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118–122.
- Pines, A., Kelstrup, C.D., Vrouwe, M.G., Puigvert, J.C., Typas, D., Misovic, B., de Groot, A., von Stechow, L., van de Water, B., Danen, E.H., et al. (2011). Global phosphoproteome profiling reveals unanticipated networks responsive to cisplatin treatment of embryonic stem cells. *Mol. Cell Biol.* **31**, 4964–4977.
- Roberts, P.J., and Der, C.J. (2007). Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310.
- Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., and Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000.
- Stavridis, M.P., Lunn, J.S., Collins, B.J., and Storey, K.G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development* **134**, 2889–2894.
- ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., Siu, R.K., and Nusse, R. (2011). Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.* **13**, 1070–1075.
- Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **105**, 9290–9295.
- Wani, M.A., Means, R.T., Jr., and Lingrel, J.B. (1998). Loss of LKLF function results in embryonic lethality in mice. *Transgenic Res.* **7**, 229–238.
- Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. *Biochem. Soc. Trans.* **38**, 1027–1032.
- Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.* **13**, 838–845.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523.

Ncoa3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming

Michelle Percharde,¹ Fabrice Laval,¹ Jia-Hui Ng,² Vibhor Kumar,³ Rute A. Tomaz,¹ Nadine Martin,⁴ Jia-Chi Yeo,² Jesús Gil,⁴ Shyam Prabhakar,³ Huck-Hui Ng,² Malcolm G. Parker,¹ and Véronique Azuara^{1,5}

¹Institute of Reproductive and Developmental Biology, Faculty of Medicine, Imperial College, London W12 0NN, United Kingdom; ²Gene Regulation Laboratory, ³Transcriptional Regulation Laboratory, Genome Institute of Singapore, Singapore 138672, Republic of Singapore; ⁴MRC-Clinical Sciences Centre, Faculty of Medicine, Imperial College, London W12 0NN, United Kingdom

Embryonic stem cell (ESC) pluripotency depends on a well-characterized gene regulatory network centered on Oct4, Sox2, and Nanog. In contrast, little is known about the identity of the key coregulators and the mechanisms by which they may potentiate transcription in ESCs. Alongside core transcription factors, the orphan nuclear receptor Esrrb (estrogen-related receptor β) is vital for the maintenance of ESC identity and furthermore is uniquely associated with the basal transcription machinery. Here, we show that Ncoa3, an essential coactivator, is required to mediate Esrrb function in ESCs. Ncoa3 interacts with Esrrb via its ligand-binding domain and bridges Esrrb to RNA polymerase II complexes. Functionally, Ncoa3 is critical for both the induction and maintenance of pluripotency. Through chromatin immunoprecipitation (ChIP) sequencing and microarray experiments, we further demonstrate that Ncoa3 shares overlapping gene regulatory functions with Esrrb and cooperates genome-wide with the Oct4–Sox2–Nanog circuitry at active enhancers to up-regulate genes involved in self-renewal and pluripotency. We propose an integrated model of transcriptional and coactivator control, mediated by Ncoa3, for the maintenance of ESC self-renewal and somatic cell reprogramming.

[**Keywords:** Ncoa3; Esrrb; embryonic stem cells; coactivator; pluripotency; reprogramming; SRC-3]

Supplemental material is available for this article.

Received May 2, 2012; revised version accepted August 27, 2012.

Embryonic stem cells (ESCs) are derived from the inner cell mass of early embryos and can be propagated for extended periods of time in culture while retaining their pluripotency; i.e., the capacity for multilineage specification and differentiation (Smith 2001). Coordinating with essential components of the LIF and BMP signaling pathways (Niwa et al. 1998; Matsuda et al. 1999; Ying et al. 2003), intrinsic transcription factors are crucial for specifying the undifferentiated state of mouse ESCs. Oct4 is essential for establishing and maintaining pluripotency (Nichols et al. 1998; Niwa et al. 2000; Boiani et al. 2005) and interacts with Sox2 (Avilion et al. 2003; Chew et al. 2005) in ESCs, where both factors cotarget multiple genes through cooperative binding at Oct–Sox sequence elements (Boyer et al. 2005; Loh et al. 2006). Moreover, Oct4 and Sox2, along with Klf4 and cMyc, can mediate repro-

gramming of mouse embryonic fibroblasts (MEFs), generating ESC-like induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka 2006). A third well-characterized core pluripotency factor is Nanog, which is notably capable of promoting ESC self-renewal even in the absence of LIF (Chambers et al. 2003; Mitsui et al. 2003).

Recent work has begun to identify new components of the ESC core transcriptional network. Along with Oct4, Sox2, and Nanog (OSN), these factors participate in auto- and cross-regulation to activate each other's expression as well as downstream self-renewal regulators while simultaneously repressing genes that promote differentiation (Chambers and Tomlinson 2009). Perturbation of these factors collapses the self-renewal circuitry and triggers specific or mixed-lineage differentiation (Ivanova et al. 2006; Loh et al. 2006). Whole-genome mapping of DNA-binding sites further revealed that the OSN triad colocalizes with a variable yet overlapping set of pluripotency-associated transcription factors at many promoters and enhancers (Chen et al. 2008). Clusters with a relatively high number of factors correlate with expression of nearby genes

⁵Corresponding author

E-mail v.azuara@imperial.ac.uk

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.195545.112>.

(Kim et al. 2008). It remains less clear, however, how this well-established network functionally interacts with basal transcription machinery complexes to activate transcription and, most importantly, which factors mediate such interactions. While the list of ESC-specific transcription factors is rapidly expanding, relatively little is known about the identity of essential coregulators (Fong et al. 2011) and the mechanisms by which they may potentiate transcription.

To gain greater insight into how transcription factors function mechanistically in ESCs, we focused here on estrogen-related receptor β (Esrrb). This orphan nuclear receptor binds many target sites that are co-occupied by OSN and Klf4 in ESCs and up-regulates the expression of these factors themselves (Chen et al. 2008). Levels of *Esrrb* decline rapidly upon differentiation (Xie et al. 2009), and its depletion results in a loss of ESC characteristics (Ivanova et al. 2006; Loh et al. 2006). Overexpression of *Esrrb* inhibits lineage commitment toward the meso- and neuro-ectodermal paths and allows robust self-renewal in the absence of LIF (Zhang et al. 2008). In addition, Esrrb can substitute for Klf4 in somatic cell reprogramming (Feng et al. 2009), further highlighting a pivotal role in both maintaining and inducing pluripotency. In somatic cells, the activity of estrogen-related receptors (ERRs) is known to be regulated through the recruitment of coactivators to the AF-2 portion of their ligand-binding domain (LBD) (Giguere 2008); for example, PGC-1 in metabolic tissues (Lin et al. 2005; Charest-Marcotte et al. 2010). These coactivators in turn act as scaffolds for other protein complexes required to bring about transcription (Rosenfeld et al. 2006). Although well studied within the context of adult somatic systems, little is known about the mechanisms by which Esrrb function is conferred in ESCs.

In this study, we show that Esrrb transcriptional and self-renewal activity is absolutely dependent on protein-protein interactions mediated via its LBD/AF-2 domain. Despite this, we found that the coactivators *PGC-1 α* and *PGC-1 β* commonly required for ERR activity are not expressed in ESCs and instead identify Nco3 (also known as SRC-3/AIB1) as an essential coactivator of Esrrb. We show that *Nco3* expression positively correlates with the undifferentiated ESC state and itself is required for both the induction and maintenance of pluripotency. Nco3 binds to Esrrb specifically via its AF-2 region in ESCs and thus is corecruited to *Esrrb*, *Klf4*, *Nanog*, and *Sox2* enhancer regions as a critical step to trigger Esrrb-dependent gene activation. Through chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq) and transcriptome analysis in *Nco3* knockdown ESCs, we further establish that Nco3 and Esrrb have overlapping gene regulatory functions and cooperate genome-wide with the Oct4–Sox2–Nanog circuitry at active enhancers to up-regulate genes involved in self-renewal and pluripotency. Finally, we demonstrate mechanistically that Nco3's presence links Esrrb to RNA polymerase II (RNAPol2) complexes, identifying a crucial role for the Nco3–Esrrb partnership in the ESC core transcriptional network and shedding new light on how pluripotency transcription factors may be bridged to the general transcription machinery to activate transcription.

Results

The LBD and AF-2 regions are essential for Esrrb transcriptional activity in ESCs

How the assembly of transcription factor clusters leads to active transcription and which essential mediating components contribute to this process are yet to be fully investigated in ESCs. Combining mutagenesis and luciferase reporter assays, we here asked whether Esrrb activity is conferred in ESCs through specific protein recruitment. Flag-tagged Esrrb mutants lacking either the entire LBD (Δ LBD) or N-terminal AF-1 (Δ AF-1) domain—regions key for mediating nuclear receptor protein-protein interactions—were generated (Supplemental Fig. S1A), and their expression was verified alongside wild-type Esrrb in ESCs (Supplemental Fig. S1B). To test these deletion mutants, we initially took advantage of previously described constructs that carry luciferase downstream from a minimal *Oct4* promoter, together with ERR response element (ERRE)-containing regulatory fragments found at the *Esrrb*, *Klf4*, or *Sox2* loci (Fig. 1A; Feng et al. 2009), and an additional Esrrb-dependent *Nanog*-Luc reporter (Rodda et al. 2005; van den Berg et al. 2008). These Enhancer-Luc vectors all show a loss of luciferase activity following knockdown of endogenous *Esrrb* in ESCs, an activity readily regained by coexpressing RNAi-resistant wild-type *Esrrb* cDNA (Supplemental Fig. S1C). As opposed to wild-type control, we found that the Δ LBD mutant failed to restore transcription from *Esrrb*, *Klf4*, *Nanog*, and *Sox2* reporters. In contrast, deleting the AF-1 domain had no such suppressive effect on Esrrb activity (Supplemental Fig. S1C). Unlike other nuclear receptors, ERR activity is not thought to be ligand-dependent but nevertheless involves the recruitment of essential coactivators to the AF-2 region, a domain of 12 helices forming a unique protein-binding surface within the LBD (Wurtz et al. 1996). Accordingly, we repeated these experiments using Esrrb point mutants in either helix 3 (AF-2 mut1) or helix 12 (AF-2 mut2)—two analogous mutations previously found to inactivate the AF-2 region of the estrogen receptor (Danielian et al. 1992; Henttu et al. 1997)—along with a DNA-binding domain (DBD) mutant unable to bind DNA generated as a control (Fig. 1B). Despite equal expression to wild-type Esrrb in ESCs (Fig. 1C), all three point mutants were similarly unable to rescue the luciferase activity of *Esrrb*, *Klf4*, *Nanog*, and *Sox2* reporters (Fig. 1D), unequivocally demonstrating that Esrrb in ESCs requires an active LBD/AF-2 domain to trigger transcription.

AF-2 region inactivation results in a loss of self-renewal in ESCs

Overexpressing *Esrrb* is sufficient to allow robust self-renewal in ESCs, negating the need for LIF to prevent differentiation (Zhang et al. 2008). To link the AF-2 dependence of Esrrb transcriptional activity with its self-renewal function, we next generated stable ESC lines that constitutively express Flag-tagged wild-type, AF-2 mut1, or AF-2 mut2 Esrrb (Fig. 1E). Cells were plated at low density and cultured for 5 d in the absence of LIF, then

Figure 1. The AF-2 region is essential for *Esrrb* activity in ESCs. (A) Schematic of *Esrrb*, *Klf4*, and *Sox2* reporters used in luciferase assays. Enhancer fragments containing ERREs from the indicated genes are inserted downstream from luciferase (Feng et al. 2009). An additional *Nanog*-Luc reporter was also used (Rodda et al. 2005). (B) Diagram indicating Flag-tagged wild-type (Wt) and *Esrrb* point mutants. (DBD mut) C120G/C123G; (AF-2 mut1) K259A; (AF-2 mut2) L424A/F425A. (C) Expression of Flag-tagged wild-type *Esrrb* or point mutants in ESCs. (D) Luciferase assays for *Esrrb*-bound enhancers. ESCs were transfected with the indicated reporter plus either control (shScrambled) or shRNA to endogenous *Esrrb* (sh*Esrrb*) and with or without RNAi-resistant *Esrrb* rescue constructs as depicted in B. Data are the mean $\pm$ SEM of three transfections. Three experiments were performed with similar results. (E) Levels of CAG-Flag-*Esrrb* construct expression in ESCs following stable transfection with wild-type, AF-2 mut1, or AF-2 mut2 *Esrrb* forms. (F,G) Ability of wild-type or *Esrrb* point mutants to confer LIF independence in stably transfected ESCs. Each cell line was plated at low density in normal ESC medium and, the following day, cultured in medium minus LIF for 5 d. Colonies were fixed and stained for AP and scored as undifferentiated, mixed, or differentiated. Data are the mean $\pm$ SEM of three independent experiments. Student's *t*-test was used to compare numbers of undifferentiated colonies in wild-type and either AF-2 mutant ESCs or empty vector control; (*) *P*-value < 0.05.

stained for alkaline phosphatase (AP) to score the number of undifferentiated, mixed, and differentiated colonies formed under these conditions (Supplemental Fig. S1D). Strikingly, and in contrast to overexpressing wild-type *Esrrb* cells, we observed a near-complete absence of compact, AP-positive colonies in both AF-2 mut1 and AF-2 mut2 ESC cultures (Fig. 1F,G), indicating a loss in the ability of *Esrrb* point mutants to confer LIF independence. To support this, we also examined the effect of treating ESCs with an ERR-specific AF-2 antagonist, diethylstilbestrol (DES). As previously reported, this compound interacts with all three ERR isoforms to block coactivator binding and thereby prevent their transcriptional activity (Supplemental Fig. S1E; Tremblay et al. 2001; Greschik et al. 2004). We demonstrated that DES treatment consistently inhibited transcription from *Esrrb*-dependent reporters in ESCs (Supplemental Fig. S1F) and significantly impaired the self-renewal ability of these cells in a dose-dependent fashion (Supplemental Fig. S1G,H). Taken together, these

findings reveal the importance of the AF-2 region to potentiate *Esrrb* function and imply a crucial role for a cognate *Esrrb* coactivator in ESCs.

Ncoa3 is highly expressed in undifferentiated ESCs

In somatic cells, the best-characterized ERR coactivators belong to the PGC-1 family. PGC-1 α and PGC-1 β both potentiate ERR activity in various adult cell types to activate genes involved in processes such as cellular energy metabolism and homeostasis (Huss et al. 2002; Kamei et al. 2003; Charest-Marcotte et al. 2010). Most recently, a third member of the PGC-1 family, PRC, was identified as a regulator of NRF-1 and mitochondrial function (Vercauteren et al. 2009). Among all three PGC-1 members, only *PRC* was expressed in ESCs (Supplemental Fig. S2A), and thus its role as a putative *Esrrb* coactivator was further investigated. Knockdown of *PRC* in ESCs did not, however, compromise *Esrrb* transcriptional activity or ESC

self-renewal, as seen upon depletion of *Esrrb* itself (Supplemental Fig. S2B–E). We next turned to the SRC/p160 family that comprises Ncoa1, Ncoa2, and Ncoa3, three alternative coactivator proteins also capable of interacting with the ERR family in mammalian cell models (Hong et al. 1999; Zhang and Teng 2000; van den Berg et al. 2010). *Ncoa3* notably emerged as an interesting candidate, being highly expressed in self-renewing ESC cultures, in contrast to *Ncoa1* and *Ncoa2* (Fig. 2A), and promptly down-regulated alongside *Esrrb* and other pluripotency factors during differentiation, as assessed upon addition of retinoic acid and/or LIF withdrawal (Fig. 2B,C). A sharp decline in *Ncoa3* transcript levels was also confirmed upon embryoid body (EB)-mediated differentiation, further pointing to a close association between *Ncoa3* expression and the undifferentiated ESC state (Fig. 2D).

Ncoa3 acts as an *Esrrb* coactivator in pluripotent ESCs

To establish a functional link between Ncoa3 and *Esrrb*, we first checked whether these two protein factors could

indeed interact. Coimmunoprecipitation (co-IP) experiments were performed in COS-1 cells coexpressing Ncoa3 and Flag-*Esrrb*, readily demonstrating an Ncoa3–*Esrrb* interaction (Fig. 2E). As expected, this interaction was mediated via the *Esrrb* AF-2 region, as AF-2 mutants failed to bind to Ncoa3 (Supplemental Fig. S2F). Co-IPs were next performed using ESC extracts and, importantly, confirmed that endogenous Ncoa3 and *Esrrb* proteins interact in ESCs (Supplemental Fig. S2G). Moreover, Ncoa3 binding to *Esrrb* was abrogated in the presence of DES (Fig. 2F), a result that reflects the decreased self-renewal ability of DES-treated ESCs (Supplemental Fig. S1). To test the requirement for Ncoa3 on *Esrrb*-dependent enhancers, ESCs were cotransfected with a pool of siRNAs targeting *Ncoa3* together with *Esrrb*, *Klf4*, *Nanog*, and *Sox2* luciferase vectors. *Ncoa3* loss closely phenocopied the suppressive effect of *Esrrb* depletion itself (Fig. 2G; Supplemental Fig. S2H). Furthermore, restoring *Esrrb* expression in depleted ESCs failed to rescue luciferase reporter activity in the absence of *Ncoa3*

Figure 2. Ncoa3 acts as *Esrrb* coactivator in ESCs. (A) Expression profiling of *Ncoa1*, *Ncoa2*, and *Ncoa3* transcripts in ESCs as assessed by qRT-PCR and normalized to two housekeeping genes. Expression levels in immortalized brown adipose tissue (IMBAT) cells are shown as a control. Data are the mean $\pm$ SEM of three biological replicates. (B) Immunofluorescence staining of ESCs maintained in proliferative conditions or induced to differentiate via addition of 1 μ M retinoic acid (atRA) and removal of LIF for 96 h. Bars, 50 μ m. (C) Western blotting showing Ncoa3 down-regulation along with *Esrrb*, *Nanog*, and *Oct4* upon LIF withdrawal. Tubulin is shown as a loading control. (D) Relative expression levels of *Ncoa3* and other pluripotency factors during EB-mediated differentiation. Data are normalized to housekeeping genes and expressed relative to EB day 0 as the mean $\pm$ SEM of three independent experiments. (E) Co-IP experiments performed in COS-1 cells cotransfected with Flag-*Esrrb* and Ncoa3. (F) Interaction of endogenous *Esrrb* and Ncoa3 proteins in ESC extracts following 24 h of treatment with ethanol or 12.5 μ M DES. At least two experiments were performed with similar results. (G) Loss of ESC-specific luciferase reporter activity 48 h after depletion of endogenous *Esrrb* or *Ncoa3* in ESCs. Data are the mean $\pm$ SEM of three transfections. Three experiments were performed with similar results.

(Supplemental Fig. S2I), indicating that Nco3 is indeed essential for the activation of these ESC-specific enhancers by Esrrb. We next demonstrated that Nco3 is recruited by Esrrb to genomic fragments containing ERREs using DNA pull-down experiments with biotinylated wild-type or ERRE mutated probes incubated together with Nco3 and Flag-Esrrb (Fig. 3A). These proteins both readily purified with *Esrrb*, *Klf4*, *Sox2*, and *Nanog* DNA fragments. Furthermore, point mutations in ERRE motifs that abrogated Esrrb binding also reduced Nco3 detection (Fig. 3A), confirming that Nco3 associates via Esrrb at ERREs in the vicinity of target genes. Similarly, Nco3 and Esrrb were corecruited to endogenous *Esrrb*, *Klf4*, *Nanog*, and *Sox2* target genes in ESCs, as assessed by ChIP and quantitative PCR (qPCR) (Fig. 3B). Nco3 recruitment was also verified here to be Esrrb-dependent, as transient depletion of Esrrb protein reduced Nco3 enrichment at these sites (Fig. 3C). Importantly, both Nco3 and Oct4 protein levels remained unaltered at this time point (Fig. 3D), excluding the possibility that the loss of Nco3 binding is due to ESC differentiation. Given the co-occupancy of many ERRE-containing enhancers with the core pluripotency factors (Chen et al. 2008), we next examined the Oct4 and Nanog dependency of Nco3–Esrrb recruitment. As previously reported for the *Nanog* locus (van den Berg et al. 2008), removing Oct4 protein prevented the detection of Esrrb at all four ESC candidate loci, and this was strictly mirrored by a loss of Nco3 recruitment (Supple-

mental Fig. S3A,B), a trend also observed to a lesser extent upon conditional *Nanog* deletion (Supplemental Fig. S3C,D). However, while Nco3 and Esrrb readily interact (Fig. 2; Supplemental Fig. S2), we could find no evidence for an interaction between Nco3 and Oct4/Nanog themselves (Supplemental Fig. S3E–H), thus highlighting the specificity of the Nco3–Esrrb association at these sites. In line with this, DNA pull-down assays further demonstrated that although Oct4 and Sox2 bind to an Oct–Sox site also contained within the *Nanog* probe, they were unable alone to recruit Nco3 (Fig. 3E). Collectively, these results identify Nco3 as an important Esrrb coactivator in ESCs and point toward a critical role for the Nco3–Esrrb partnership in regulating pluripotency.

Nco3 contributes to the maintenance of pluripotency in ESCs

To explore the functional requirement for Nco3 in ESCs, we next knocked down *Nco3* using two independent puromycin-selectable shRNA vectors. The extent of the knockdown was validated by qRT–PCR (Fig. 4A) and Western blot (Fig. 4B) and, strikingly, resulted in a clear and rapid loss of ESC characteristics. By day 4 post-selection, colonies displayed a flattened or spread morphology and stained negative for AP (Fig. 4C), and no stable *Nco3* knockdown ESC clones could be established. *Nco3* depletion was promptly accompanied by a decrease

Figure 3. Nco3 is recruited via Esrrb to ERREs at target loci. (A) DNA pull-down assays using previously described wild-type (Wt) or ERRE mutated sequences from *Esrrb*, *Klf4*, *Sox2* (Feng et al. 2009), or *Nanog* (van den Berg et al. 2008). Biotinylated probes (40–50 base pairs [bp]) were incubated with extracts from COS-1 cells transfected with Flag-Esrrb and Nco3 and recovered on streptavidin beads, and DNA-associated proteins were visualized by Western blotting. (B) Esrrb and Nco3 enrichment levels at *Esrrb*, *Klf4*, *Nanog*, and *Sox2* ERREs and an intergenic (Inter.) control region in ESCs as assessed by ChIP and qPCR and expressed relative to input. Data are the mean $\pm$ SEM of three biological replicates. (C) Nco3 and Esrrb enrichment levels at candidate loci following *Esrrb* depletion in ESCs, expressed relative to input. Enrichment for shGFP transfected cells is set at 100% in each individual experiment. Data are the mean $\pm$ SD of at least two independent experiments. Dotted lines in B and C indicate background enrichment by control IgGs (Santa Cruz Biotechnology). (D) Western blotting showing specific Esrrb protein depletion 48 h after transfection with shEsrrb. Note that Nco3 and Oct4 levels are unchanged at this time. (E) DNA pull-down assays using *Nanog* wild-type or ERRE mutated probe, which also contains a neighboring Oct–Sox site (van den Berg et al. 2008), and cell extracts from COS-1 transfected with Nco3, Oct4, and Sox2.

Figure 4. Ncoa3 plays critical roles in the maintenance of ESC self-renewal. (A,B) Specific knockdown of *Ncoa3* in ESCs. ESCs were transfected with two independent puromycin-selectable shRNA vectors targeting *Ncoa3*, and, 48 h after transfection, *Ncoa3* expression was quantified by qRT-PCR (A) and Western blotting (B). (C) AP staining following *Ncoa3* knockdown. ESCs were transfected with shScrambled or shNcoa3 vectors, and, 24 h later, puromycin was added. Selection was maintained for 4 d, then cells were fixed and stained for AP. (D,E) Quantitative gene expression analysis of cells 4 d post-selection for *Ncoa3* depletion. Data are the mean $\pm$ SD of at least two independent experiments. (F) Generation of stable ESC lines overexpressing combinations of *Esrrb* and *Ncoa3*. E14Tg2A (E14) ESCs were first transfected with linearized CAG-Control or CAG-Esrrb vectors and selected with puromycin. Pooled clones for each stable cell line were then transfected with linearized control vector, pCDNA4/TO-GFP, or pCDNA4/TO-Ncoa3, and overexpressing cells were selected using Zeocin. (G) ESC lines generated in F were plated at low density and cultured for 5 d without LIF. Colonies were fixed and stained for AP; shown are three independent experiments. The statistical significance of the difference in undifferentiated colonies between the indicated cell lines was calculated using Student's *t*-test; (*) *P*-value < 0.05, (**) *P*-value < 0.01. (H) Representative colony morphology for each cell line from G.

in self-renewal gene expression (Fig. 4D) and concomitant derepression of lineage-associated markers (Fig. 4E), supporting an effect of Ncoa3 on differentiation. Conversely, we asked whether, through the potentiation of Esrrb activity, elevated levels of Ncoa3 could enhance resistance to differentiation in ESCs. For this, we established a series of stable ESC lines that overexpress either *Esrrb* (ESC^{Esrrb}), *Ncoa3* (ESC^{Ncoa3}), or both factors together (ESC^{Esrrb+Ncoa3}) alongside control cells (ESC^{control}) through two sequential rounds of transfection and selection, as depicted in Figure 4F. To compare the behavior of these lines, cells were plated at low density and cultured minus LIF for 5 d as previously performed (Fig. 1). The ESC^{Esrrb+Ncoa3} line showed a significant increase in the number of undifferentiated colonies relative to ESC^{Esrrb} cells (Fig. 4G), demonstrating the additive effect conferred by overexpressing *Esrrb* and *Ncoa3* together. Interestingly, a consistent

increase in the number of undifferentiated and mixed colonies was also observed in ESC^{Ncoa3} as compared with control cells (Fig. 4G,H), suggesting that overexpressing *Ncoa3* alone could support the self-renewal capacity of ESCs. This was confirmed in ESCs stably expressing *Ncoa3* using an alternative (pPyCAGIP) episomal vector (Supplemental Fig. S4A; Chambers et al. 2003). These cells similarly exhibited lower levels of differentiation when cultured in the absence of LIF as well as under self-renewing conditions—an effect associated with increased expression of Esrrb targets *Esrrb* and *Klf4* (Supplemental Fig. S4B; Feng et al. 2009). Following several passages, however, *Ncoa3*-overexpressing ESCs cultured minus LIF eventually differentiated due to the down-regulation of endogenous *Esrrb* (data not shown). Furthermore, these cells were capable of EB-mediated differentiation (Supplemental Fig. S4C). Collectively, these results show that

Ncoa3 plays critical roles in maintaining an intact ESC state and a self-renewal phenotype but, consistent with its role as a coactivator, is insufficient alone to maintain *Esrrb* levels and block differentiation.

Ncoa3 is required for somatic cell reprogramming

Interestingly, *Ncoa3* was found to be up-regulated during iPSC generation alongside *Esrrb* and *Nanog* (Supplemental Fig. S5A; Mikkelsen et al. 2008), suggesting that *Ncoa3* may also be important for the induction of pluripotency by reprogramming. Detection of high levels of *Ncoa3* protein was confirmed in two independently derived iPSC clones (Fig. 5A) from Oct4GFP MEFs. These MEFs carry a GFP transgene downstream from the *Oct4* distal enhancer, which becomes selectively activated during successful somatic cell reprogramming (Fig. 5B; Supplemental Fig. S5B; Yeom et al. 1996). To test the functional importance of *Ncoa3* in this process, we specifically depleted *Ncoa3* in Oct4GFP MEFs by RNAi (Fig. 5C; Supplemental Fig. S5C) prior to transduction with Oct4, Sox2, Klf4, and cMyc (OSKM) reprogramming retroviruses (Supplemental Fig. S5D). Both scrambled and *Ncoa3* knockdown MEFs showed no difference in cell proliferation (Supplemental Fig. S5E) or apoptosis (Supplemental Fig. S5F) at the time of reprogramming and infected equally highly with retroviruses (Fig. 5D). In contrast, *Ncoa3* depletion induced a

drastic reduction in reprogramming efficiency, as judged by the significantly lower number of GFP-positive (Fig. 5E–F) and AP-positive (Fig. 5G) colonies generated. We next asked whether *Ncoa3* knockdown might inhibit reprogramming simply by inducing MEF senescence, a known impediment of reprogramming. For this, experiments were repeated in MEFs lacking *p53*, which allows enhanced speed and efficiency of reprogramming without cMyc (Banito et al. 2009; Kawamura et al. 2009; Marion et al. 2009). *p53*^{−/−} MEFs transduced with OSK, however, showed similarly striking reductions in the number of AP-positive colonies following *Ncoa3* depletion (Fig. 5H), indicating that the observed reprogramming defect was not due to p53-mediated senescence. Taken together, these findings demonstrate that as well as being critical for the maintenance of ESCs, *Ncoa3* is furthermore required for the induction of pluripotency upon reprogramming.

Ncoa3 participates genome-wide with Esrrb and OSN at active ESC-specific enhancers

To gain greater understanding of how *Ncoa3* functions in ESCs, *Ncoa3* DNA-binding sites were next mapped by ChIP-seq. We first verified that *Ncoa3* (this study) and *Esrrb* (Chen et al. 2008) data sets showed strong overlap at *Esrrb*, *Klf4*, *Nanog*, and *Sox2* target genes and local binding profiles (Supplemental Fig. S6A). Focusing on ESC-

Figure 5. Essential requirement for *Ncoa3* in somatic cell reprogramming. (A) *Ncoa3* protein is up-regulated in fully reprogrammed iPSC lines. Western blotting depicts two independently generated iPSC clones from two different Oct4GFP genetic backgrounds (clone 1: 129/Ola; clone 2: MF1/B6) (Yeom et al. 1996) in comparison with uninfected proliferative MEFs. (B) Immunofluorescence confirming *Ncoa3* and GFP expression in iPSC clone 1. Bars, 50 μ m. (C) Efficient *Ncoa3* knockdown in MEFs analyzed 3 d after infection with lentiviral particles expressing shScrambled or shNcoa3 shRNA. Data are the mean $\pm$ SEM of three independent experiments. (D) Equal retroviral mCherry infection of scrambled or *Ncoa3* knockdown MEFs. (E) Typical GFP-positive colonies arising from transduction of shScrambled-treated Oct4GFP MEFs with OSKM reprogramming factors. Bars, 100 μ m. (F) Inhibition of reprogramming following depletion of *Ncoa3*, analyzed by quantifying GFP-positive colonies 12–14 d after infection with OSKM vectors. The number of colonies produced from shScrambled-infected MEFs is set as 100% for each individual experiment. Three experiments were performed using different batches of MEFs and viruses. (**) *P*-value < 0.01, (***) *P*-value < 0.001, Student's *t*-test. (G) AP staining of iPSC colonies 14 d after infection with OSKM reprogramming factors. (H) AP staining of iPSC colonies generated in *p53*^{−/−} shScrambled or shNcoa3 MEFs using OSK reprogramming factors.

relevant loci, we next compared all Ncoa3–Oct4–Sox2 cobound sites with Esrrb–Oct4–Sox2 and found significant overlaps between these groups (P -value = 1.11×10^{-16}) (Fig. 6A). In contrast, the insulator-binding transcription factor CTCF did not show any significant overlap (P -value = 0.84). Using a de novo motif discovery

algorithm (Heng et al. 2010), we identified, as expected, a canonical ERRE motif among significantly overrepresented binding sequences found at Ncoa3 peaks (Fig. 6B; Supplemental Fig. S7). Oct–Sox elements were also detected at these sites, reflecting the frequent colocalization of Esrrb with Oct4 and Sox2. GREAT gene ontology

Figure 6. Ncoa3 is an integrated component of the ESC transcriptional network. (A) Significant overlap between Esrrb and Ncoa3 peaks within 200 bp at Oct4–Sox2-bound sites. The intersection with CTCF is shown for comparison. (B) MEME de novo motif discovery analysis of top Ncoa3-bound sequences. (C) Spatial heat map depicting binding peak intensity of indicated factors, centered on the peak of maximum Ncoa3 binding for each site. The two-dimensional matrix was provided as input for the K-mean clustering program in Matlab to determine the classes of each position. (D) Venn diagram showing the overlap of genes that are significantly regulated by Ncoa3 and Esrrb, as determined by microarray following *Ncoa3* (this study) and *Esrrb* (Feng et al. 2009) knockdown in ESCs (fold change >1.5). The Fisher's exact test P -value for the correlation of fold change is indicated. See also Supplemental Table 5. (E) Heat map of class I genes bound by Ncoa3 and Esrrb. First, all of the genes regulated by Esrrb within 100 kb of the class I region were taken, then their fold change upon *Esrrb* (left) or *Ncoa3* (right) knockdown was plotted according to *Esrrb* knockdown. Shades of green and red indicate down-regulation and up-regulation, respectively. Note that the majority of genes is down-regulated. (F) Co-IP of Esrrb in ESCs treated for 24 h with ethanol or 12.5 μ M DES, immunoblotted for the indicated proteins. (G) Co-IP in ESCs showing the interaction of endogenous Ncoa3 and RNApol2. Two representative blots are shown. (H) Co-IP of Esrrb immunoblotted for RNApol2 in ESCs following 48 h of transfection with siControl or siNcoa3 oligos. The square brackets in F–H indicate bands corresponding to RNApol2. All co-IPs were performed at least twice with similar results.

(GO) analysis at Nco3–Esrrb cobound sites revealed significant enrichment for developmental processes and embryonic phenotypes (Supplemental Fig. S6B). In addition, abnormal fertility and organ size were also enriched, agreeing with published mouse knockout models for *Nco3* and/or *Esrrb* (Luo et al. 1997; Mitsunaga et al. 2004; Chen et al. 2010). Taken together, these data suggest a high degree of similarity between Nco3 and Esrrb transcriptional networks in ESCs. To look more widely at the different types of transcription factor clusters associated with Nco3 peaks, we next used spatial heat map analysis (Fig. 6C). This distinguished three main classes of Nco3 targets based on their co-occupancies by Esrrb and/or OSN as well as Nr5a2, p300, H3K27ac, and H3K4me1. Here we confirmed that ~40% of all Nco3 peaks colocalized with Esrrb at active enhancers, as denoted by the presence of p300, H3K27ac, and H3K4me1 (Fig. 6C; Heintzman et al. 2007; Creighton et al. 2010). Sites bound by both Esrrb/Nco3 and OSN (class Ib) represented key self-renewal regulators that are actively transcribed in ESCs, such as *Esrrb*, *Klf2/4/5*, *Zfp42*, *Tcfcp2l1*, *Oct4*, and *Sox2*, and included GO processes involving stem cells and blastocyst development (Supplemental Table S1). In contrast, sites containing Esrrb and Nco3 but with reduced binding for OSN (class Ia) (Fig. 6C) were uniquely associated with developmental phenotypes involving germ cell integrity and fertility (Supplemental Table S2), including *Prdm14*, a recently described factor critical for germ cell lineage (Yamaji et al. 2008; Gillich et al. 2012). To closely investigate transcriptional regulation by Nco3 and Esrrb, microarray analysis was performed on ESCs following *Nco3* knockdown and compared with *Esrrb* knockdown at the same time point (Feng et al. 2009). Notably, 62.5% of the misregulated genes in *Esrrb*-depleted ESCs were also misregulated in *Nco3* knockdown ESCs (Fig. 6D), most often in the same direction (Fig. 6E), showing that gene regulation by Nco3 and Esrrb is significantly correlated (P -value = 3.92×10^{-49}). Among them, many target genes with (class Ib) or without (class Ia) high OSN enrichment were significantly down-regulated upon *Nco3* depletion, as validated by qRT-PCR (Supplemental Fig. S8A), confirming that Nco3 is strictly required for Esrrb-dependent gene activation. Interestingly, at some sites, Nco3 also colocalized with Nr5a2—another orphan nuclear receptor previously involved in stem cells and reprogramming (Gu et al. 2005; Heng et al. 2010)—together with (class I) and without (class II) Esrrb and/or OSN (Fig. 6C). GO analysis of class II genes, however, consisted of targets associated with extra-embryonic development (Supplemental Table S3), indicating that although Nr5a2 could potentially recruit Nco3 to a set of genes, these targets may be less relevant in ESCs. Finally, the last class identified (class III) (Supplemental Table S4) included Nco3-bound sites that were not co-occupied by Esrrb, Nr5a2, or OSN (Fig. 6C), suggesting that Nco3 may also be recruited by other factors. While marks of active enhancers—p300, H3K27ac, and H3K4me1—were absent at the majority of these sites (class IIIa), a small subset (class IIIb) colocalized with H3K4me1 alone (Fig. 6C), potentially delineating inactive/poised Nco3-bound

enhancers in ESCs (Creighton et al. 2010). Interestingly, microarray analysis of class IIIb revealed that out of the genes that are misregulated upon *Nco3* knockdown, all were up-regulated (Supplemental Fig. S8B).

Nco3 bridges Esrrb to the general transcription machinery

Recently, Esrrb, among other pluripotency factors, was identified as being uniquely associated with RNAPol2 complexes (van den Berg et al. 2010), suggesting a critical function for Esrrb, distinct from OSN, in activating transcription. Thus, we investigated the importance of the Esrrb AF-2 domain and Nco3 for this function. We first demonstrated by co-IP that RNAPol2 does indeed interact with endogenous Esrrb in ESCs (Fig. 6F, top left). Strikingly, however, upon short treatment with DES, we observed a reduction of the Esrrb–RNAPol2 interaction (Fig. 6F, top left). This appeared to be specifically associated with loss of Nco3 binding (Fig. 6F, top right), as neither the interaction between Esrrb and Oct4 (van den Berg et al. 2008) nor the efficiency of anti-Esrrb immunoprecipitation itself was affected (Fig. 6F, bottom). These results suggested that Nco3 could mediate the interaction between Esrrb and RNAPol2. To test this further, we performed co-IPs between RNAPol2 and Nco3 itself and demonstrated an interaction between these proteins in ESCs (Fig. 6G). Finally, to confirm that Nco3 is required for the association between Esrrb and RNAPol2, co-IPs were repeated in cells treated with siControl or siNco3 oligos to deplete endogenous Nco3 protein. Transfection with siNco3 specifically reduced the interaction between endogenous Esrrb and RNAPol2 (Fig. 6H; Supplemental Fig. S8C–E). Taken together, these findings reveal Nco3 as a key limiting factor that bridges Esrrb to the general transcription machinery in ESCs.

Discussion

Here, we present the first detailed characterization of a nuclear receptor coactivator in the maintenance of ESC pluripotency. Although it is not a tissue-specific factor, we observed that elevated Nco3 expression is a defining property and essential requirement of ESCs and iPSCs. Numerous transcription factors have been well characterized in this context. In contrast, there exist few matching studies for critical coactivators. Our work notably complements recent findings by Fong et al. (2011) identifying a novel Oct4/Sox2 coactivator, the XPC nucleotide excision repair complex, which is essential to activate *Nanog* and support the ESC state while also safeguarding genome integrity.

In this study, we provide significant data to suggest that Nco3 action in ESCs is primarily mediated through Esrrb function. We show that Nco3 binds to Esrrb specifically via its AF-2 domain and that Nco3 recruitment is strictly required to potentiate Esrrb activity in ESCs. AF-2 domain inactivation in ESCs or depletion of *Nco3* itself both suppress transcription from Esrrb-dependent gene reporters and result in a loss of ESC self-renewal. Conversely,

overexpressing *Ncoa3* in ESCs enhances Esrrb-mediated resistance to differentiation but is not sufficient alone to block differentiation, in agreement with *Ncoa3* acting as an Esrrb coactivator and not a transcription factor. Through genome-wide ChIP-seq analysis, we demonstrate the presence of *Ncoa3* at a large number of active Esrrb-bound enhancers that are shared with the OSN triad and enriched for p300, H3K27ac, and H3K4me1 (Heintzman et al. 2007; Creighton et al. 2010). GO analysis at this set of targets shows enrichment for key self-renewal regulators whose expression is significantly reduced in *Ncoa3* knockdown ESCs, as determined by microarray and qRT-PCR, identifying *Ncoa3* as an integrated component of the ESC transcriptional network.

A distinct group of *Ncoa3*- and Esrrb-bound loci showing reduced OSN occupancies is exclusively associated with germ cell and fertility-related phenotypes. This corroborates with a primordial germ cell (PGC)-specific defect observed *in vivo* in the absence of *Esrrb* (Mitsunaga et al. 2004) and an infertility phenotype in *Ncoa3* knock-out mice (Xu et al. 2000). It furthermore suggests that Esrrb and *Ncoa3* might also act in synergy during PGC reprogramming. Consistently, we found that germ cell-relevant targets such as *Prdm14* are significantly down-regulated upon *Ncoa3* knockdown (Supplemental Fig. S8A), supporting a putative role for the *Ncoa3*–Esrrb partnership in this process. Aside from Esrrb, it is still possible that other nuclear receptors or transcription factors may also recruit *Ncoa3* to some genes. Nr5a2, for example, maps in ESCs to many Esrrb-, *Ncoa3*-, and OSN-bound targets (Heng et al. 2010; this study) and could potentially interact with *Ncoa3* via its LBD, although this remains to be formally demonstrated. In contrast, we could find no evidence to support a direct role for OSN in recruiting *Ncoa3* at these sites, although these core factors most likely stabilize the Esrrb–*Ncoa3* complex through protein–protein interactions with Esrrb at multitranscription factor clusters (Supplemental Fig. S3; van den Berg et al. 2008).

Critically, we shed new light on how the *Ncoa3*–Esrrb self-renewal function is conferred in ESCs. A recent study reported the Oct4 interactome of 166 proteins, including many transcription factors and chromatin-modifying complexes not previously known to associate with the ESC network (van den Berg et al. 2010). Among these Oct4 partners, Esrrb was uniquely found to be associated with the basal transcription machinery. An association between *Ncoa3* and Esrrb was also observed there. Our mechanistic work on *Ncoa3* thus elucidates the functional importance of this interaction and further highlights the relevance of Oct4's broad range of interaction partners. Here, we confirm that endogenous Esrrb and RNAPol2 are physically associated in ESCs and demonstrate that *Ncoa3* mediates this association. We propose that *Ncoa3* functionally bridges Esrrb and RNAPol2 at loci co-occupied by Oct4 as well as Sox2, Nanog, and additional pluripotency factors (see Fig. 7), highlighting a distinct role for the *Ncoa3*–Esrrb partnership to the core ESC transcription factors. *Ncoa3* is also established in this study as being required for efficient somatic cell

Figure 7. Proposed model. Esrrb interacts with Nanog and Oct4/Sox2 at ESC enhancers and recruits *Ncoa3* via its AF-2 region. *Ncoa3* interacts with RNAPol2 and bridges Esrrb at enhancers to the general transcriptional machinery. These factors would thus act, together with other factors such as chromatin proteins and Mediator (Kagey et al. 2010), to bring about transcription of ESC-associated genes.

reprogramming—a property shared with other Oct4/Esrrb partners, including core members of Trithorax/MLL complexes (Ang et al. 2011). These chromatin-modifying complexes are essential for open chromatin and pluripotency and may therefore open the path to the reprogramming process by facilitating Oct4 occupancy at genomic loci encoding self-renewal genes. Esrrb–*Ncoa3*–RNAPol2 bridging would in turn perpetuate strong transcriptional activation at Oct4/Esrrb-bound sites, allowing for the full establishment of iPSCs or the maintenance of ESC self-renewal.

In summary, our work elucidates a previously unrecognized functional link between the core ESC transcriptional network and a nuclear receptor coactivator and, importantly, provides novel mechanistic insights into the role of *Ncoa3*–Esrrb in sustaining both ESC self-renewal and reprogramming. It would be of interest in future studies to elucidate whether coactivator recruitment is indeed a general requirement for many ESC pluripotency factors. Moreover, acting as “sensors,” these cofactors could also be capable of integrating stimuli from multiple signaling pathways, thus adding another layer of control to the complex network governing pluripotency and differentiation.

Materials and methods

Cell culture and RNAi experiments

Mouse E14Tg2A (E14) ESCs were cultured on 0.1% gelatin as previously described (Azuara et al. 2006). COS-1 and 293T were grown in DMEM supplemented with 10% FCS, penicillin/streptomycin, and L-glutamine. MEFs and iPSCs were cultured as previously described (Feng et al. 2009). RNAi experiments followed by puromycin selection were performed using pLKO.1 or pSuper vectors. Transfection of ESCs was performed with Lipofectamine 2000 according to the manufacturer's protocols. See Supplemental Table S6 for shRNA sequences.

Antibodies

The primary antibodies used in this study for immunofluorescence, Western blotting, and ChIP were as follows: anti-Esrrb (R&D Systems, H6705 and H6707), anti-Nanog (Cosmo,

RCAB0002P), anti-Ncoa3 (Santa Cruz Biotechnology, M-397; BD Biosciences, #611105), anti-Oct4 (Santa Cruz Biotechnology, C-10), anti-Sox2 (Abcam, ab97959), anti-Flag (Sigma, M2), anti-RNApol2 (Santa Cruz Biotechnology, N-20), anti-Gapdh (Millipore), anti-Tubulin (Sigma), anti-cleaved Caspase-3 (Cell Signaling, #9664), and anti-BrdU (Abcam, ab6326).

Co-IP and ChIP

For co-IP experiments performed with ectopically expressed proteins, COS-1 cells in 10-cm dishes were transfected with 3–5 μ g of each expression vector using Fugene HD (Roche), according to the manufacturer's guidelines. Cells were collected 48 h later and lysed in RIPA buffer. Eight hundred micrograms of protein extract was precleared with protein A Sepharose beads (Sigma, P3391) for 2 h at 4°C, then incubated overnight at 4°C with the indicated antibodies. Protein A beads were added for 2 h, then washed with RIPA and TSE buffer (2mM EDTA, 20 mM Tris at pH 8, 150 mM NaCl) three times and loaded on 7% acrylamide gels. Detection of endogenous interactions was performed using 1- μ g extracts similarly prepared from untransfected E14 ESCs. Anti-Ncoa3 (#611105; mouse monoclonal) was used to coimmunoprecipitate Ncoa3 expressed in COS-1 cells when examining interactions with Nanog (rabbit polyclonal); all other co-IP, immunofluorescence, and ChIP experiments were performed with M-397. ChIP experiments were carried out with chromatin prepared from E14 cells as previously described (Stock et al. 2007) using 8–10 μ g of primary antibody and 600 μ g of precleared chromatin per immunoprecipitation. Immunoprecipitated DNA fragments were purified and analyzed via SYBR Green qPCR.

DNA pull-down assays

COS-1 cells were transfected with Ncoa3 together with Flag-Esrrb or Oct4 and Sox2 expression vectors. Forty-eight hours after transfection, cells were lysed in HKMG buffer (10 mM HEPES at pH 7.9, 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 0.5% NP-40, 1 mM DTT, protease inhibitors [Complete EDTA-free; Roche]) and lysates precleared for 1 h with pre-equilibrated streptavidin-coupled Dynabeads (Invitrogen, M-280). Pairs of complementary DNA oligonucleotides with the sense oligonucleotide biotinylated at the 5' end (Sigma) were annealed and incubated for 2 h at 4°C with precleared cell lysates. DNA-bound proteins were collected by incubation for 1 h at 4°C with streptavidin-coupled Dynabeads (Invitrogen), washed four times in HKMG buffer, and separated by SDS-PAGE. Ncoa3, Flag-Esrrb, Oct4, and Sox2 were detected by Western blot. See Supplemental Table S6 for probe sequences.

Luciferase assays

ESCs were plated on gelatin-coated 96-well plates at a density of 2×10^4 cells per well and, 5 h later, transfected with the indicated luciferase reporters and expression constructs using Lipofectamine 2000. For knockdown experiments, either 30 ng of shRNA vector or 2–3 pmol of siRNA was cotransfected per well. pCmv.Renilla or pCDNA4/TO.eGFP was used to normalize luciferase values in each well. Firefly and Renilla luciferase activities were measured 48 h post-transfection using the Steady-Lite Plus kit (PerkinElmer).

Retroviral production

For reprogramming experiments, PlatE cells were used to produce pMXs-based retroviruses containing the cDNA for *Pou5f1*/

Oct4, *Sox2*, *Klf4*, or *cMyc*, as previously described (Feng et al. 2009). PlatE-generated mCherry retrovirus was used to monitor MEF infection efficiency in each experiment.

iPSC generation

Embryonic day 13.5 (E13.5) Oct4-GFP MEFs were derived from embryos produced via the intercross between wild-type MF1 females and homozygous Oct4-GFP B6 males (Yeom et al. 1996). For iPSC induction, 2×10^4 Oct4-GFP MEFs were seeded per well of a 24-well plate and, 5 h later, infections were performed using equal amounts of each indicated retrovirus in the presence of 8 μ g/mL polybrene. Medium was exchanged for MEF medium 24 h post-infection, and, 48 h after infection, MEFs were collected and reseeded onto inactivated feeders in iPSC medium. Emerging iPSC colonies were monitored until day 12–14 when cells were harvested or individual colonies were picked for further analysis.

ChIP-seq and microarray analysis

ChIP-seq and microarray analysis were performed on mouse ESCs as previously described (Chen et al. 2008; Feng et al. 2009). Detection of Ncoa3 sequenced peaks was carried out using MACS with a *P*-value cut-off of 1.0×10^{-5} .

Accession numbers

ChIP-seq and microarray data are accessible at the Gene Expression Omnibus database under accession number GSE40193.

Detailed Materials and Methods, antibodies, and sequences are available in the Supplemental Material.

Acknowledgments

We are grateful to Paul Robson, Vincent Cavaillès, and Mark Christian for providing plasmids, and to Eleni Daniel and Anthony Okolo for IMBAT cDNA. Thanks to Rachel Amoroux for E13.5 Oct4GFP embryos; Sadaf Khan for p53^{-/-} MEFs; Matias Autio for iMEFs; Hitoshi Niwa and Ian Chambers for the ZHBTc4 and RCN β HB ESC lines, respectively; and Ana Banito and Lai-Ping Yaw for their technical assistance. Thanks also to Mark Christian, Helle Jorgensen, and Tristan Rodriguez, as well as members of the Epigenetics and Development Group and Molecular Endocrinology Group for discussions and/or critical reading of the manuscript. This work was supported by the BBSRC, Genesis Research Trust, FCT, MRC, Genome Institute of Singapore, and Imperial College London.

References

- Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, et al. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**: 183–197.
- Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**: 126–140.
- Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, et al. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**: 532–538.
- Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V, Walsh M, et al. 2009. Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev* **23**: 2134–2139.

- Boiani M, Gentile L, Gambles VV, Cavaleri F, Redi CA, Scholer HR. 2005. Variable reprogramming of the pluripotent stem cell marker Oct4 in mouse clones: Distinct developmental potentials in different culture environments. *Stem Cells* **23**: 1089–1104.
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947–956.
- Chambers I, Tomlinson SR. 2009. The transcriptional foundation of pluripotency. *Development* **136**: 2311–2322.
- Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**: 643–655.
- Charest-Marcotte A, Dufour CR, Wilson BJ, Tremblay AM, Eichner LJ, Arlow DH, Mootha VK, Giguere V. 2010. The homeobox protein Prox1 is a negative modulator of ER α /PGC-1 α bioenergetic functions. *Genes Dev* **24**: 537–542.
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106–1117.
- Chen X, Liu Z, Xu J. 2010. The cooperative function of nuclear receptor coactivator 1 (NCOA1) and NCOA3 in placental development and embryo survival. *Mol Endocrinol* **24**: 1917–1934.
- Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P, et al. 2005. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* **25**: 6031–6046.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**: 21931–21936.
- Danielian PS, White R, Lees JA, Parker MG. 1992. Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *EMBO J* **11**: 1025–1033.
- Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J, et al. 2009. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* **11**: 197–203.
- Fong YW, Inouye C, Yamaguchi T, Cattoglio C, Grubisic I, Tjian R. 2011. A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell* **147**: 120–131.
- Giguere V. 2008. Transcriptional control of energy homeostasis by the estrogen-related receptors. *Endocr Rev* **29**: 677–696.
- Gillich A, Bao S, Grabole N, Hayashi K, Trotter MW, Pasque V, Magnusdottir E, Surani MA. 2012. Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* **10**: 425–439.
- Greschik H, Flaig R, Renaud JP, Moras D. 2004. Structural basis for the deactivation of the estrogen-related receptor γ by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity. *J Biol Chem* **279**: 33639–33646.
- Gu P, Goodwin B, Chung AC, Xu X, Wheeler DA, Price RR, Galardi C, Peng L, Latour AM, Koller BH, et al. 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.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318.
- Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH, Orlov YL, Huss M, Yang L, Lufkin T, et al. 2010. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* **6**: 167–174.
- Henttu PM, Kalkhoven E, Parker MG. 1997. AF-2 activity and recruitment of steroid receptor coactivator 1 to the estrogen receptor depend on a lysine residue conserved in nuclear receptors. *Mol Cell Biol* **17**: 1832–1839.
- Hong H, Yang L, Stallcup MR. 1999. Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. *J Biol Chem* **274**: 22618–22626.
- Huss JM, Kopp RP, Kelly DP. 2002. Peroxisome proliferator-activated receptor coactivator-1 α (PGC-1 α) coactivates the cardiac-enriched nuclear receptors estrogen-related receptor- α and - γ . Identification of novel leucine-rich interaction motif within PGC-1 α . *J Biol Chem* **277**: 40265–40274.
- Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**: 533–538.
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430–435.
- Kamei Y, Ohizumi H, Fujitani Y, Nemoto T, Tanaka T, Takahashi N, Kawada T, Miyoshi M, Ezaki O, Kakizuka A. 2003. PPAR γ coactivator 1 β /ERR ligand 1 is an ERR protein ligand, whose expression induces a high-energy expenditure and antagonizes obesity. *Proc Natl Acad Sci* **100**: 12378–12383.
- Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Izpisua Belmonte JC. 2009. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460**: 1140–1144.
- Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**: 1049–1061.
- Lin J, Handschin C, Spiegelman BM. 2005. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab* **1**: 361–370.
- Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**: 431–440.
- Luo J, Sladek R, Bader JA, Matthysen A, Rossant J, Giguere V. 1997. Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR- β . *Nature* **388**: 778–782.
- Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, Fernandez-Capetillo O, Serrano M, Blasco MA. 2009. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**: 1149–1153.
- Matsuda T, Nakamura T, Nakao K, Arai T, Katsuki M, Heike T, Yokota T. 1999. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J* **18**: 4261–4269.
- Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**: 49–55.
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**: 631–642.
- Mitsunaga K, Araki K, Mizusaki H, Morohashi K, Haruna K, Nakagata N, Giguere V, Yamamura K, Abe K. 2004. Loss of PGC-specific expression of the orphan nuclear receptor ERR- β

- results in reduction of germ cell number in mouse embryos. *Mech Dev* **121**: 237–246.
- Nichols J, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**: 379–391.
- Niwa H, Burdon T, Chambers I, 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, Smith AG. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**: 372–376.
- Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**: 24731–24737.
- Rosenfeld MG, Lunyak VV, Glass CK. 2006. Sensors and signals: A coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev* **20**: 1405–1428.
- Smith AG. 2001. Embryo-derived stem cells: Of mice and men. *Annu Rev Cell Dev Biol* **17**: 435–462.
- Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG, Pombo A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**: 1428–1435.
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663–676.
- Tremblay GB, Kunath T, Bergeron D, Lapointe L, Champigny C, Bader JA, Rossant J, Giguere V. 2001. Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERR β . *Genes Dev* **15**: 833–838.
- van den Berg DL, Zhang W, Yates A, Engelen E, Takacs K, Bezstarosti K, Demmers J, Chambers I, Poot RA. 2008. Estrogen-related receptor β interacts with Oct4 to positively regulate Nanog gene expression. *Mol Cell Biol* **28**: 5986–5995.
- van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**: 369–381.
- Vercauteren K, Gleyzer N, Scarpulla RC. 2009. Short hairpin RNA-mediated silencing of PRC (PGC-1-related coactivator) results in a severe respiratory chain deficiency associated with the proliferation of aberrant mitochondria. *J Biol Chem* **284**: 2307–2319.
- Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, Moras D, Gronemeyer H. 1996. A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* **3**: 206.
- Xie CQ, Jeong Y, Fu M, Bookout AL, Garcia-Barrio MT, Sun T, Kim BH, Xie Y, Root S, Zhang J, et al. 2009. Expression profiling of nuclear receptors in human and mouse embryonic stem cells. *Mol Endocrinol* **23**: 724–733.
- Xu J, Liao L, Ning G, Yoshida-Komiya H, Deng C, O'Malley BW. 2000. The steroid receptor coactivator SRC-3 (p/CIP/RAC3/AIB1/ACTR/TRAM-1) is required for normal growth, puberty, female reproductive function, and mammary gland development. *Proc Natl Acad Sci* **97**: 6379–6384.
- Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, Saitou M. 2008. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* **40**: 1016–1022.
- Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hubner K, Scholer HR. 1996. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**: 881–894.
- Ying QL, Nichols J, Chambers I, Smith A. 2003. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**: 281–292.
- Zhang Z, Teng CT. 2000. Estrogen receptor-related receptor $\alpha 1$ interacts with coactivator and constitutively activates the estrogen response elements of the human lactoferrin gene. *J Biol Chem* **275**: 20837–20846.
- Zhang X, Zhang J, Wang T, Esteban MA, Pei D. 2008. Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* **283**: 35825–35833.

Ncoa3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming

Michelle Percharde, Fabrice Lavial, Jia-Hui Ng, et al.

Genes Dev. 2012 26: 2286-2298 originally published online September 26, 2012

Access the most recent version at doi:[10.1101/gad.195545.112](https://doi.org/10.1101/gad.195545.112)

Supplemental Material

<http://genesdev.cshlp.org/content/suppl/2012/09/25/gad.195545.112.DC2.html>

References

This article cites 58 articles, 23 of which can be accessed free at:
<http://genesdev.cshlp.org/content/26/20/2286.full.html#ref-list-1>

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genes & Development* go to:
<http://genesdev.cshlp.org/subscriptions>

In Vivo Epigenomic Profiling of Germ Cells Reveals Germ Cell Molecular Signatures

Jia-Hui Ng,^{1,4,8} Vibhor Kumar,^{2,8} Masafumi Muratani,^{1,8} Petra Kraus,³ Jia-Chi Yeo,^{1,7} Lai-Ping Yaw,¹ Kun Xue,¹ Thomas Lufkin,^{3,6} Shyam Prabhakar,^{2,*} and Huck-Hui Ng^{1,4,5,6,7,*}

¹Gene Regulation Laboratory

²Computational and Systems Biology

³Stem Cell and Developmental Biology

Genome Institute of Singapore, Singapore 138672

⁴NUS Graduate School for Integrative Sciences and Engineering, Singapore 117456

⁵Department of Biochemistry, National University of Singapore, Singapore 117597

⁶Department of Biological Sciences, National University of Singapore, Singapore 117543

⁷School of Biological Sciences, Nanyang Technological University, Singapore 637551

⁸These authors contributed equally to this work

*Correspondence: prabhakars@gis.a-star.edu.sg (S.P.), nghh@gis.a-star.edu.sg (H.-H.N.)

<http://dx.doi.org/10.1016/j.devcel.2012.12.011>

SUMMARY

The limited number of *in vivo* germ cells poses an impediment to genome-wide studies. Here, we applied a small-scale chromatin immunoprecipitation sequencing (ChIP-seq) method on purified mouse fetal germ cells to generate genome-wide maps of four histone modifications (H3K4me3, H3K27me3, H3K27ac, and H2BK20ac). Comparison of active chromatin state between somatic, embryonic stem, and germ cells revealed promoters and enhancers needed for stem cell maintenance and germ cell development. We found the nuclear receptor Nr5a2 motif to be enriched at a subset of germ cell *cis*-regulatory regions, and our results implicate Nr5a2 in germ cell biology. Interestingly, in germ cells, the H3K27me3 histone modification occurs more frequently at regions that are enriched for retrotransposons and MHC genes, indicating that these loci are specifically silenced in germ cells. Together, our study provides genome-wide histone modification maps of *in vivo* germ cells and reveals the molecular chromatin signatures of germ cells.

INTRODUCTION

Germ cells are unipotent cells that differentiate into either male or female gametes. The fusion of the mature male and female gametes generates a zygote that gives rise to all cell types of the embryo proper and extraembryonic tissues. To date, germ cells are the only cell type that can revert back to pluripotent embryonic germ cells (EGCs) simply by providing the appropriate culture system (Matsui et al., 1992; Resnick et al., 1992). Remarkably, this conversion does not require the overexpression of any transcription factor, occurs rapidly (within 7 to 10 days), and the conversion efficiency drastically declines as germ cells progress from embryonic day 8.5 (E8.5) to E12.5

(Labosky et al., 1994). Therefore, these studies suggest that germ cells undergo developmental changes when they arrive at the genital ridge such that the ability to regain pluripotency is lost.

In vivo, germ cells undergo extensive epigenetic reprogramming that is pivotal for parental imprint erasure and X chromosome reactivation, and potentially primes germ cells for postfertilization development into the early embryo (Sasaki and Matsui, 2008). Prior to E13.5, germ cells reduce stable repressive marks such as histone H3 lysine 9 dimethylation (H3K9me2) and gain histone H3 lysine 27 trimethylation (H3K27me3), which is a repressive mark associated with poised developmental genes (Bernstein et al., 2006; Seki et al., 2005). Germ cells also undergo two rounds of DNA demethylation at E8.0 and E12.5 (Hajkova et al., 2002; Seki et al., 2005). By E13.5, germ cells are exceedingly hypomethylated, with DNA methylation reaching levels that are lower than DNA methylation-deficient mouse embryonic stem cells (ESCs) (Popp et al., 2010).

Because combinatorial chromatin modifications can identify *cis*-regulatory regions (Heintzman et al., 2009), analyzing the epigenome of germ cells will provide insights into the biology of these cells. Given that chromatin immunoprecipitation coupled to sequencing (ChIP-seq) typically requires millions of cells, it is impractical to purify the limited number of germ cells on such a large scale. Here, we devised a small-scale ChIP-seq strategy to map histone modifications in fetal germ cells. Using the genome-wide maps of four histone modifications (H3K4me3, H3K27me3, H3K27ac, and H2BK20ac), coupled with transcriptional profiling and motif analyses, we identified genomic loci with germ cell promoter and enhancer signatures, and demonstrated that Nr5a2 is implicated in germ cell biology. In addition, we found classes of H3K27me3-marked sites that are enriched for H3K27me3 in germ cells compared to ESCs and other somatic cell types, which suggest extensive silencing of key developmental pathways in germ cells. These germ-cell-specific H3K27me3 marks are also enriched for retrotransposon repeats and major histocompatibility complex (MHC) genes, indicating an additional importance of H3K27me3 in suppressing retrotransposons in the DNA methylation-deficient germ cells. Lastly, our data reveal classes of differentially marked genes in ESCs and germ cells, which provide insight into major processes

Figure 1. Robustness of Modified Small-Scale ChIP-Seq Method to Map Histone Modifications using a Small Number of Purified Cells

(A) Experimental workflow of small-scale ChIP-seq. Germ cells were FACS-purified based on *Pou5f1*-GFP expression. Following ChIP, the reverse-crosslinked immunoprecipitated DNA was amplified by 15 cycles of WGA. Amplification adaptors were cleaved by two rounds of Bpml restriction digestion prior to ChIP-seq library preparation.

(B) qPCR results of small-scale H3K4me3 ChIP performed on 10,000, 25,000, and 100,000 mouse ESCs after WGA amplification. *Pou5f1* DE and *Pou5f1* Pro refer to the distal enhancer and proximal promoter of the *Pou5f1* gene, respectively. Data represent the mean $\pm$ SEM ($n = 3$).

(C) UCSC browser view showing the similarity of mouse ESC H3K4me3 ChIP-seq tag count signals without (standard) and with 15 cycles (technical duplicates: WGA1 and WGA2) of WGA.

(D) Scatterplot of mouse ESC H3K4me3 ChIP-seq signals and Venn diagram illustrating overlap of peaks (using top 10,000 peaks) between amplified (WGA1 and WGA2) and nonamplified samples (Standard). Each point represents total tag count in a randomly chosen 1-kb genomic region; 100,000 bins are shown in the scatterplot.

See also Figure S1 and Table S1.

occurring during germ cell development and distinguish germ cells from ESCs.

RESULTS

A Robust Small-Scale ChIP-Seq Method for a Small Number of Cells

Given the relatively small number of germ cells in vivo, it is necessary to modify the standard ChIP protocol (Figure 1A; see Experimental Procedures). We first tested our small-scale ChIP method on ESCs using H3K4me3 ChIP quantitative PCR (qPCR), and detected enrichment at the *Pou5f1* enhancer and promoter as well as at known bivalent genes when as few as 10,000 cells were used (Figure 1B). Notably, whole-genome amplification (WGA) was essential to achieve sufficient material for ChIP-seq, and without it the amount of DNA could not be accurately measured. As a strategy to remove the WGA adaptors

that would interfere with subsequent reference genome mapping, WGA was performed using primer pairs that contain Bpml recognition sites; Bpml cleaves DNA 16 bp after its recognition site. We prepared two technical replicates of WGA/Bpml-treated H3K4me3 ChIP-seq libraries using 50 pg ChIP DNA that was isolated from a standard protocol. WGA technical duplicates (WGA1 and WGA2) are similar to each other and highly correlated to their standard counterpart ($r = 0.898$ and 0.896), with a peak overlap of 86.4% and 86.1%, respectively (Figures 1C and 1D). We then tested our small-scale ChIP-seq methodology directly on 1,000, 10,000, and 100,000 ESCs for the H3K27me3 histone modification. The small-scale ChIP-seq data also show high similarity, correlation, and sensitivity to data from a standard ChIP-seq (Figures S1A–S1C available online). Therefore, our small-scale ChIP-seq method can be used to generate genome-wide data from a small number of cells, which will be valuable for studying developmental processes in vivo.

Figure 2. Identification of Active Promoters and Enhancers in Germ Cells

(A) UCSC browser view of active promoter regions (indicated by orange horizontal bar) marked by H3K4me3 and H3K27ac in E13.5 germ cells.

(legend continued on next page)

Distinct Classes of Active Promoters and Enhancers Distinguish ESCs and Germ Cells from Somatic Cells

To ensure that a pure population of germ cells is used, green fluorescent protein-positive (GFP+) germ cells from the gonads of *Pou5f1*-GFP transgenic mouse embryos were purified by fluorescence-activated cell sorting (FACS) (Yeom et al., 1996). Using microarray, we profiled transcriptional changes in postmigratory germ cells that were purified from E11.5, E13.5, and E15.5 embryos, and used GFP-negative gonadal nongerm cells as control. We found that E13.5 germ cells are undergoing a transitional phase whereby stem cell maintenance and cell-cycle genes are downregulated while spermatogenesis genes are upregulated (Figures S1D and S1E). Therefore, E13.5 germ cells serve as a paradigm to study events that trigger the loss of ability to revert back to a pluripotent state in vitro and activation of fetal germ cell differentiation.

Applying the small-scale ChIP-seq method on FACS-purified E13.5 male germ cells, we generated ChIP-seq libraries for H3K4me3, H3K27me3, H3K27ac, and H2BK20ac (Table S1). More than 90% of the detected H3K4me3 peaks mapped to promoters and CpG islands (Figure S1F), suggesting a high accuracy of the prediction method using our H3K4me3 ChIP-seq data set. In addition, we observed enrichment of H3K4me3 and H3K27ac at highly expressed genes in E13.5 germ cells (Figures S1G and S1H), and ChIP duplicates were well correlated (Figure S1I). For downstream analyses, only peaks ($p < 0.01$) that overlap in biological replicates were used (see Supplemental Experimental Procedures).

As expected, expressed genes (e.g., *Zic3*, *Dazl*, *Nr5a2*, and *Tcf1*) exhibit active marks of H3K27ac and H2BK20ac, while known bivalent genes (e.g., *Dlx* and *Hox*) are marked by both H3K4me3 and H3K27me3 (Figures 2A and S2A). To understand the similarities and differences of germ cells and ESCs, K-mean clustering was performed on active promoter and enhancer sites (defined by H3K27ac peaks in either ESCs or germ cells) (Creyghton et al., 2010; Hawkins et al., 2011). For this analysis, H3K27ac ChIP-seq signals from neural progenitor cells (NPC), adult liver, and ProB cells were also included (Creyghton et al., 2010), and we obtained nine major classes (Figure S2B; Table S2). Germ-cell-specific H3K27ac peaks (classes 5, 6, and 9) were enriched for genes (single nearest genes within 30 kb) with reproduction-related annotations such as insemination, fertility, and sex determination (e.g., *Tex13* and *Dazl*) (Figures 2A and S2C). Notably, the H3K27ac peaks shared between ESCs and germ cells (class 1), but absent in somatic cells, were enriched for stem cell maintenance genes (e.g., *Sox2*, *Tcf1*, and *Nanog*). However, ESC-specific peaks

(class 2) or those shared between ESCs and other somatic cell types but not germ cells (class 3) are adjacent to genes involved in various developmental organ systems (e.g., *Klf9* and *Sox17*) (Figure S2C). In addition, the transforming growth factor β (TGF β) signaling pathway and cell-cycle checkpoint genes are enriched in E13.5 germ cells as compared to E11.5 germ cells.

Identification of Germ Cell *cis*-Regulatory Elements

In order to identify active promoters that are marked by the H3K27ac histone modification in germ cells only (germ cell promoter signatures), we clustered regions with both H3K4me3 and H3K27ac in E13.5 germ cells by comparison with other cell types such as liver, ProB, NPC, and ESC (Figure 2B) (Creyghton et al., 2010). Genomic Regions Enrichment of Annotations Tool (GREAT) analysis reveals that the single nearest genes within 30 kb of these germ cell promoter signatures are highly enriched for mouse phenotypes related to reproduction and male fertility (Figure 2C).

Besides promoters, enhancers are also tightly regulated in a cell-type-specific manner (Creyghton et al., 2010; Heintzman et al., 2009; Visel et al., 2009). Here, we define enhancers as regions with H3K27ac peaks ($p < 0.01$), and neither lying on any known gene promoter nor exhibiting H3K4me3 peak ($p < 0.01$) in both replicates. Based on these criteria, a total of 3,494 active enhancers were identified for E13.5 germ cells. Corroborating an enhancer identity, these H3K27ac-marked regions are also enriched for H2BK20ac and p300 binding in E13.5 germ cells (Figures 2D and S2D) (Visel et al., 2009; Wang et al., 2008), and enhanced *luciferase* reporter gene expression (Figure S2E, see Supplemental Experimental Procedures).

Among the germ cell enhancers, 2,234 regions (germ cell enhancer signatures) are marked by H3K27ac and H2BK20ac specifically in germ cells but not in other cell types (Creyghton et al., 2010) (Figure 2E). GREAT analysis on the germ cell enhancer signatures revealed enriched mouse phenotypes related to small testis, infertility, abnormal gametogenesis, and, surprisingly, abnormal embryonic growth and embryonic lethality (Figure 2F). One possibility is that germ cells and ESCs utilize different enhancers to regulate a common set of genes associated with embryonic lethality. To examine this further, we first identified genes with H3K27ac-marked promoters in both germ cells and ESCs. Among genes adjacent to germ cell enhancer signatures, 36.1% were also adjacent to an ESC enhancer signature (Table S3; Figures S2F and S2G; “differential” enhancers). Significant Gene Ontology (GO) enrichment

(B) Plot of H3K4me3 (meK4) and H3K27ac (ack27) ChIP-seq signals for E13.5 germ cells, adult liver, ProB cells, ESC, and NPC. Top regions for H3K4me3 and H3K27ac peaks in E13.5 germ cells are shown. A window of 8 kb (−4 kb to +4 kb) around the peak center is shown. Promoters that are marked by the H3K27ac histone modification in E13.5 germ cells only (germ cell promoter signatures) are indicated below the dotted line.

(C) Enriched mouse phenotypes for nearest genes within 30 kb of E13.5 germ cell promoter signatures ($q < 9E-09$).

(D) UCSC browser view of enhancer regions (indicated by red horizontal bar) marked by H2BK20ac and H3K27ac in E13.5 germ cells.

(E) Plot of H3K27ac (ack27) and H2BK20ac (ack20) ChIP-seq signals, which are hallmarks of active enhancers, for E13.5 germ cells, liver, ProB, ESC, and NPC. A window of 8 kb (−4 kb to +4 kb) around the peak center is shown. Enhancers that are marked by the H3K27ac histone modification in E13.5 germ cells only (germ cell enhancer signatures) are indicated below the dotted line.

(F) Enriched mouse phenotypes for nearest genes within 30 kb of E13.5 germ cell enhancer signatures ($q < 0.001$).

(G) The position weight matrix of the top three enriched motifs found at germ cell enhancer signatures.

See also Figure S2 and Tables S2, S3, and S4.

(legend on next page)

terms for this class of overlapping genes include developmental process, regulation of transcription, cell differentiation, and cell division (data not shown). However, we did not get enrichment for stem cell maintenance genes, suggesting that most of them, such as *Pou5f1* (Yeom et al., 1996), *Zic3*, and *Tcf1*, are regulated by “conserved” enhancers in ESCs and germ cells (Figure S2G). Further comparison of the enriched GO terms among the different enhancer categories reveals that enhancers can be differently utilized for functional subcategories of commonly expressed genes (Figure S2H). We then used the locations of germ cell enhancer signatures for motif enrichment analysis, and the FOXP1/ FAC1, DMRT, and RREB1 motifs emerged as the top three motifs that are enriched at germ cell enhancer signatures (Figure 2G; Table S4). This is consistent with previous reports that members of the DMRT family of transcription factors are important for sex determination and germ cell proliferation (Hong et al., 2007).

Motif Enrichment Analyses Uncover Potential Regulators of Germ Cell Development

We expect that if a transcription factor is important for germ cell development, then it should bind to germ cell regulatory elements to mediate downstream transcriptional changes. Based on microarray expression, we identified downregulated (class 1: 836 early germ cell genes) and upregulated genes (class 2: 701 late germ cell genes) from E11.5 to E15.5 (Figure 3A). For both classes of genes, we selected H3K27ac and H2BK20ac peaks within 10 kb of their transcriptional start site (TSS). As expected, class 2 gene regulatory elements were enriched for motifs of transcription factors (e.g., *Egr4* and *Movob*) that were associated with male germ cell development (Li et al., 2002; Tourtellotte et al., 1999) (Figure 3B). Interestingly, the class 1 gene regulatory elements had a higher frequency of ESC-associated transcription factor motifs (e.g., *Znf143*, *Tcf3*, *Nr5a1*/*Nr5a2*, *Myc*, and *E2f1*) (Figure 3B). Given that many ESC-associated genes are also expressed in early germ cells (Figure S1E), this result suggests that ESC-associated factors may also have functional roles in germ cells. Using published ChIP-seq data sets of mouse ESCs (Chen et al., 2008; Heng et al., 2010), we found a significantly greater overlap between *Nr5a2* ChIP-seq peaks and E13.5 germ cell enhancers as compared to enhancers of NPC and ProB ($p < 0.002$) (Figure 3C). This germ cell enrichment was not observed for *n-Myc* and *E2f1*, which also have motifs enriched at class 1 gene regulatory elements, despite having similar expression patterns as *Nr5a2* in germ cells (Figure S1E). Interestingly, we also found a higher overlap of

germ cell enhancer regions with *Sox2* and *Nanog* peaks (Figure S3A). We validated that *Nr5a2* is expressed and binds to these germ cell enhancer regions (also p300-bound) in E13.5 germ cells (Figures 3D, S3B, and S3C), suggesting that *Nr5a2* may be a potential regulator of germ cell development.

Nr5a2 Is Implicated in Germ Cell Biology

Knockout of *Nr5a2* results in embryonic lethality by 7.5 days postcoitum (Paré et al., 2004), so it is challenging to ascertain its function in germ cells in vivo. To investigate if *Nr5a2* has any function in germ cells, we tested the effect of *Nr5a2* knockdown on in vitro germ cell differentiation (West et al., 2006). Using the EOS-C(3+) reporter comprising three copies of the distal *Pou5f1* enhancer (Hotta et al., 2009), GFP+ germ cell-like cells that were enriched for germ cell marker expression could be purified from day 7 embryoid bodies (EB) (Figures S3D and S3E). The GFP+ cells of both *Luci* control knockdown and *Nr5a2* knockdown day 7 EBs (Figure S3F) were similarly FACS-purified and cultured in EGC-conversion medium (EGM) in the presence of retinoic acid (RA) for 5–7 days prior to detection of converted EGCs by alkaline phosphatase (AP) staining (West et al., 2006). Consistent with published reports that *Nr5a2* is dispensable for ESC maintenance (Wagner et al., 2010), *Nr5a2* knockdown resulted in a marginal reduction in AP+ colonies in ESC medium (ESM) before EB differentiation (Figures 3E, S3G, and S3H). In contrast, a drastic reduction in AP+ colonies was observed for *Nr5a2* knockdown cells after EB differentiation and culture in EGM (RA) conditions. We harvested RNA from GFP+ cells of day 7 EBs and found that postmigratory germ cell markers, but not early germ cell genes, were expressed at reduced levels in the *Nr5a2* knockdown EBs (Figure S3I), thus implicating *Nr5a2* in germ cell biology. However, we do not exclude the possibility that *Nr5a2* could also affect EGC conversion. Further study, such as conditional knockout, is required to validate *Nr5a2*'s role in germ cell development in vivo and to determine whether *Nr5a2* functionally regulates its bound genes in germ cells in vivo.

Germ Cells Have the H3K27me3 Histone Modification at Sites Enriched for Retrotransposon Repeats

To examine the status of poised or repressed regions, we plotted the H3K4me3 and H3K27me3 ChIP-seq signals for the union of H3K27me3-marked regions in ESCs, germ cells, and other somatic cell types (NPC, mouse embryonic fibroblast [MEF], and whole brain) (Mikkelsen et al., 2007). As compared to other somatic cell types, germ cells share more bivalent regions with

Figure 3. Identification of *Nr5a2* as a Potential Regulator of Germ Cell Biology

- (A) Microarray heatmaps of class 1 (early) and class 2 (late) genes that are differentially expressed in E11.5, E13.5, and E15.5 germ cells.
- (B) Frequency of top ten enriched motifs found at H3K27ac-marked regions that are within 10 kb class 1 and class 2 genes. Motifs were searched using the TRANSFAC and JASPER databases.
- (C) A higher overlap was observed between *Nr5a2* binding (mouse ESC ChIP-seq peaks) and germ cell enhancer regions. The frequencies of transcription factor binding at cell-type-specific enhancer regions are expressed as fold relative to ProB's frequency. Asterisks indicate significant ($p < 0.002$) enrichment of binding at germ cell enhancer regions compared to NPC and ProB. *Nr5a2* binding of germ cell enhancers in ESCs represents a combination of (1) enhancers that are shared between ESCs and germ cells and (2) germ cell enhancers that could be primed or “bookmarked” by *Nr5a2* in ESCs, but are not active in ESCs.
- (D) *Nr5a2* and p300 ChIP-qPCR at germ cell enhancer regions (N1–N10) that overlap with *Nr5a2* peaks. Data represent the mean $\pm$ SEM.
- (E) *Nr5a2* knockdown negatively affects in vitro germ cell differentiation. *Luciferase* and *Nr5a2* knockdown cells purified from day 0 and day 7 EBs were cultured in ESM (ES cell medium) and EGM (RA) (embryonic germ cell media supplemented with retinoic acid), respectively. Quantitation of AP+ colonies is expressed as a ratio of the respective control frequencies. Data represent the mean $\pm$ SEM for ESM ($n = 2$) and EGM (RA) ($n = 3$).

See also Figure S3.

(legend on next page)

ESCs (Figure S4A). Upon clustering, seven classes of H3K27me3-marked regions were obtained (Figure 4A), with classes 4 and 7 marked by H3K27me3 more prominently or specifically in germ cells. This result is congruent with immunohistochemistry-based studies that show elevated levels of H3K27me3 in germ cells compared to neighboring somatic cells (Seki et al., 2005). Notably, ChIP-seq enabled the identification of specific H3K27me3-marked genomic regions in germ cells. GREAT analyses of class 7 regions (germ cell specific) revealed enrichment terms related to immune system activation like antigen processing and presentation, MHC class-1 complex, cytokine-mediated signaling pathway, and leukocyte-mediated cytotoxicity (Figure S4B). Indeed, immunity-related genes (e.g., *TAP*) at the *MHC* gene cluster are specifically marked by H3K27me3 in germ cells (Figure 4B), unlike the *Hox* clusters, which are marked by H3K27me3 in all examined cell types (Figure S4C). Although class 4 regions may not be germ cell specific, it is interesting that the single nearest genes, such as *Snai1* and *Hnf1b*, are enriched for developmental processes and cell differentiation functions (Figure S4B and S4D). This suggests that some developmental genes are repressed by H3K27me3 in E13.5 germ cells.

Given that germ cells are DNA methylation deficient, we hypothesized that H3K27me3 may be important to keep silent genes repressed. We analyzed the occurrence of transposon repeats in the top H3K27me3-marked regions (above 95th percentile) of each cell type, and found that germ cell H3K27me3-marked regions contain more SINE-B2, SINE-B4, LINE-L1, LTR-ERV, and LTR-MaLR repeats (Figures 4C and S4E). Therefore, germ cells have an epigenetic landscape that potentially represses retrotransposons and protect genomic integrity of the germline.

DISCUSSION

Chromatin profiling can be a powerful technology to uncover regulatory sequences in which sample material is limiting (Adli et al., 2010). Here, we demonstrate that our modified small-scale ChIP-seq method can be robustly performed on a small number of purified cells (less than 100,000) and provides a way to study early mammalian development in vivo. We expect that this method is ideally suited to study early developmental processes whereby the subject material is limiting.

We also provide genome-wide histone modification data for in vivo fetal germ cells, which facilitated the determination of active and inactive *cis*-regulatory elements. By comparison with other cell types, we identified germ cell promoter and

enhancer signatures that are implicated in reproduction-related phenotypes. Taking a step further, we found that the Nr5a2 motif is enriched at active regulatory elements that are located adjacent to differentially expressed genes in postmigratory germ cells. We tested the effect of *Nr5a2* knockdown using an in vitro differentiation model, and our results implicate *Nr5a2* in germ cell biology. Given that the EB-differentiated germ cell-like cells may not provide the most ideal representation of in vivo germ cells, it is important to further study the role of *Nr5a2* using in vivo animal models.

Comparison of H3K27ac-marked sites among germ cells, ESCs, and three somatic cell types defined clusters of genes that are differentially marked and predictive of functional categories. For example, regions active in both ESCs and germ cells (Figure S2B, class 1) could be the minimal set of promoters or enhancers essential for activation of genes that prevent differentiation (stem cell maintenance) (Figure 4D). This epigenetic similarity between ESCs and germ cells may explain why the latter can regain pluripotency more readily than other cell types. On the other hand, ESC-specific active regions (class 2), which could be needed for the multiple lineage differentiation potential, are not marked by H3K27ac in germ cells. Our data also reveal other germ cell features that potentially serve as barriers to pluripotency. Apart from the active status of reproductive and spermatogenesis genes, E13.5 germ cells have activated the TGF- β signaling pathway that inhibits germ cell proliferation (Moreno et al., 2010). Furthermore, we found that germ cells have higher levels of H3K27me3 at genes enriched for developmental and differentiation functions (Figure 4A, class 4), suggesting that lineage-specific differentiation options are blocked.

Although it is already known that germ cells lose DNA methylation and gain H3K27me3 during migration (Seki et al., 2005), our study uncovers the sites that are marked by H3K27me3 in germ cells only. Our analyses also reveal that H3K27me3 could be important to keep retrotransposons repressed in the DNA hypomethylated germ cells, so as to preserve genomic integrity of the germline. Interestingly, germ-cell-specific H3K27me3 regions are also enriched for genes involved in activation of immunogenicity. In postnatal mice, the blood-testis barrier and other immune mechanisms have evolved to suppress immune response to spermatozoa (Meinhardt and Hedger, 2011). Our results suggest that fetal germ cells have utilized H3K27me3 to suppress the expression of genes that could trigger immune activation and cell death. Future studies are required to examine whether other repressive histone modifications, such as H3K9me3, which is present in fetal germ cells (Henckel et al., 2012), also have similar functions as H3K27me3.

Figure 4. H3K27me3-Marked Regions that Are Present in Germ Cells

(A) Plot of H3K4me3 (mK4) and H3K27me3 (mK27) ChIP-seq signals for germ cells (E11.5 and E13.5), ESC, NPC, MEF, and whole brain (Mikkelsen et al., 2007). A window of 12 kb (−6 kb to +6 kb) around the peak center is shown. The union of H3K27me3-marked regions in the examined cell types was clustered into seven distinct classes, with class 7 representing regions with H3K27me3 enriched in germ cells.

(B) UCSC browser view of a class 7 region. Highlighted regions indicate genomic locations that have H3K27me3 in E13.5 germ cells, but not in ESC, MEF, NPC, and whole brain.

(C) The frequency of top H3K27me3 sites (above 95th percentile in each cell type) that overlap with repeats. This fraction can be written as (bins with H3K27me3 above 95 percentile and lying on a repeat family) / (bins with H3K27me3 above 95 percentile). Asterisks indicate significantly lower frequencies than E13.5 germ cells (two proportion z-test: $p < 0.001$).

(D) Model illustrating major categories of genes that distinguish ESCs from E11.5 and E13.5 germ cells.

See also Figure S4.

In summary, through our germ cell histone modification ChIP-seq data, we provide insight into germ cell promoter and enhancer signatures, potential regulators, and epigenetic features of germ cells in vivo. We envisage that small-scale ChIP-seq will provide an invaluable tool to study in vivo development on a genome-wide scale.

EXPERIMENTAL PROCEDURES

Germ Cell Purification

GFP+ germ cells were FACS purified (FACSARIA II SORP, BD Bioscience) from trypsinized E13.5 gonads (*Pou5f1*-GFP JAX^R mice stock number 004654 X WT-CD1). All animal procedures were performed according to the Singapore A*STAR Biopolis Biological Resource Center Institutional Animal Care and Use Committee (IACUC Protocol No: 110648).

ChIP

For small-scale ChIP, 0.2 ml PCR tubes (Axygen) were used to reduce sample loss. FACS-purified germ cells were fixed in 1% formaldehyde (room temperature, 10 min), quenched with 1 vol 250 mM glycine (room temperature, 5 min), and rinsed with chilled TBSE buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA) twice before storage at -80°C . After thawing on ice, fixed cells were pooled into a PCR tube for each ChIP and centrifuged (2,000 rpm, 10 min). Cell pellet is not visible, hence supernatant was carefully removed such that $\sim 10\ \mu\text{l}$ remains. Cells were lysed with 100 μl 1% SDS Lysis Buffer (on ice, 5 min) and then centrifuged (2,000 rpm, 10 min). Remove supernatant as described above and resuspend samples in 100 μl of 0.1% SDS buffer. Samples were sonicated 9 times (30 s pulses with 30 s break interval) using the Bioruptor water bath sonicator (Diagenode). Chromatin extracts were then precleared with Dynal Magnetic Beads (Invitrogen) (4°C , 1 hr) followed by centrifugation (2,000 rpm, 30 min). Supernatant (precleared chromatin) was immunoprecipitated overnight with Dynal Magnetic Beads coupled with anti-H3K4me3 (04-745, Millipore), anti-H3K27me3 (07-449, Millipore), anti-H3K27ac (ab4729, Abcam), and H2BK20ac (ab52988, Abcam) antibodies. On the next day, beads were washed (nutate in 0.1% SDS buffer for 4 min thrice then once in TE buffer) and elution was performed in a PCR machine (68°C , 10 min). After reverse crosslinking (with pronase at 42°C for 2 hr and 67°C for 6 hr) and DNA purification (phenol-chloroform extraction), small-scale ChIP DNA was subjected to 15 cycles of amplification with the GenomePlex Single Cell Whole Genome Amplification Kit (WGA4, Sigma) using universal primers linked to BpmI restriction site. Amplified ChIP DNA was column-purified (QIAquick) and restriction digested with BpmI (NEB). Samples were column-purified again, ligated to BpmI-containing adaptors (T4 DNA Ligase, NEB; 16°C for 2 hr) and subjected to a second round of BpmI digestion prior to ChIP-seq library preparation (ChIP-Seq DNA Sample Prep Kit IP-102-1001, Illumina; Supplemental Experimental Procedures). Real-time qPCR was performed (ABI PRISM 7900 Sequence Detection System) and relative occupancy values were determined by the immunoprecipitation efficiency (amount of immunoprecipitated DNA relative to input). For Figures 1C and 1D, WGA1 and WGA2 ChIP-seq libraries were generated using 50 pg normal-scale ChIP DNA (quantified using Quant-iT PicoGreen, Invitrogen) for WGA, post-ChIP processing, and ChIP-seq library preparation as technical replicates.

Microarray

RNA was extracted from FACS-purified cells using the Arcturus PicoPure RNA Isolation Kit (Applied Biosystems) and amplified using the Ovation RNA Amplification System V2 (NuGEN). Microarray data (Illumina MouseWG-6 Expression BeadChip version 2.0) were generated for E11.5, E13.5, and E15.5 germ cells, and GFP-negative nongerm cells of the gonad.

EB-Mediated In Vitro Differentiation

Mouse ESCs were trypsinized and cultured in noncoated 96 wells overnight in EB medium (ESC medium without β -mercaptoethanol and LIF; with $1.0\ \mu\text{g}\ \text{ml}^{-1}$ puromycin for selection). On day 1, the EBs were transferred into noncoated Petri dishes in EB medium. On day 7, EBs were dissociated (West et al., 2006) and FACS-purified GFP+ cells were cultured on feeder

MEFs in ESC medium containing $15\ \text{ng}\ \text{ml}^{-1}$ bFGF (Invitrogen), $30\ \text{ng}\ \text{ml}^{-1}$ SCF (R&D), $5\ \mu\text{M}$ RA (Sigma), and $1.0\ \mu\text{g}\ \text{ml}^{-1}$ puromycin. After 5–7 days, AP staining was performed using a commercial ESC characterization kit (Chemicon) according to the manufacturer's protocol.

Computational Analysis

All ChIP-seq and input control tag data were binned in the form of 200-bp bins. Guanine-cytosine (GC) bias correction was done for ChIP-seq data using respective input controls. Peak detection was done on binned GC-corrected signals after convolving it with a kernel just as in F-Seq; however here, kernels were first learned using enriched regions for each ChIP-seq data. ChIP-seq peaks overlapping repeats (RNA repeats, satellites, and segmental duplications) and peaks in input control were removed as false positives. For E13.5 germ cells, only peaks that were present in both biological replicates were included in the final set. Sites with overlapping peaks from H3K27ac ChIP-seq but not overlapping H3K4me3 or H3K27me3 peaks were taken as active enhancers in E13.5 germ cells. The UCSC browser view figures were plotted using GC-corrected signals after kernel convolution.

Clustering of binned ChIP-seq data was done after blurring them with rectangular mean kernel and taking log-transformation in order to have less effect from noise and outliers (Supplemental Experimental Procedures).

To find enrichment of motifs in different classes of peaks, genes were first classified according to their expression patterns in E11.5, E13.5, and E15.5 germ cells. H3K27ac and H2BK20ac peaks within 10 kb of TSS of genes were then used to find enrichment of motifs in different classes (Supplemental Experimental Procedures).

ACCESSION NUMBERS

The Gene Expression Omnibus accession number for the microarray and ChIP-seq (GC corrected) data reported in this paper is GSE38165.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.devcel.2012.12.011>.

ACKNOWLEDGMENTS

We are grateful to the Biomedical Research Council (BMRC), Agency for Science, Technology and Research (A*STAR) for funding. J.N., J.Y., and K.X. are supported by the A*STAR Graduate Scholarship, NTU Research Scholarship, and the National Science Scholarship, respectively. We acknowledge the GTB group for sequencing. We thank laboratory members for technical support and critical comments on the manuscript.

Received: May 17, 2012

Revised: October 17, 2012

Accepted: December 19, 2012

Published: January 24, 2013

REFERENCES

- Adli, M., Zhu, J., and Bernstein, B.E. (2010). Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat. Methods* 7, 615–618.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.

- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15–23.
- Hawkins, R.D., Hon, G.C., Yang, C., Antosiewicz-Bourget, J.E., Lee, L.K., Ngo, Q.-M., Klugman, S., Ching, K.A., Edsall, L.E., Ye, Z., et al. (2011). Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res.* 21, 1393–1409.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Henckel, A., Chebli, K., Kota, S.K., Arnaud, P., and Feil, R. (2012). Transcription and histone methylation changes correlate with imprint acquisition in male germ cells. *EMBO J.* 31, 606–615.
- Heng, J.-C.D., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.-H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., et al. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 6, 167–174.
- Hong, C.-S., Park, B.-Y., and Saint-Jeannet, J.-P. (2007). The function of Dmrt genes in vertebrate development: it is not just about sex. *Dev. Biol.* 310, 1–9.
- Hotta, A., Cheung, A.Y.L., Farra, N., Vijayaragavan, K., Séguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., et al. (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. *Nat. Methods* 6, 370–376.
- 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 insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* 120, 3197–3204.
- Li, B., Dai, Q., Li, L., Nair, M., Mackay, D.R., and Dai, X. (2002). Ovol2, a mammalian homolog of Drosophila ovo: gene structure, chromosomal mapping, and aberrant expression in blind-sterile mice. *Genomics* 80, 319–325.
- 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.
- Meinhardt, A., and Hedger, M.P. (2011). Immunological, paracrine and endocrine aspects of testicular immune privilege. *Mol. Cell. Endocrinol.* 335, 60–68.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.
- Moreno, S.G., Attali, M., Allemand, I., Messiaen, S., Fouchet, P., Coffigny, H., Romeo, P.H., and Habert, R. (2010). TGFbeta signaling in male germ cells regulates gonocyte quiescence and fertility in mice. *Dev. Biol.* 342, 74–84.
- Paré, J.F., Malenfant, D., Courtemanche, C., Jacob-Wagner, M., Roy, S., Allard, D., and Bélanger, 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.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101–1105.
- 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.
- Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* 9, 129–140.
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* 278, 440–458.
- Tourtellotte, W.G., Nagarajan, R., Auyeung, A., Mueller, C., and Milbrandt, J. (1999). Infertility associated with incomplete spermatogenic arrest and oligozoospermia in Egr4-deficient mice. *Development* 126, 5061–5071.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858.
- Wagner, R.T., Xu, X., Yi, F., Merrill, B.J., and Cooney, A.J. (2010). Canonical Wnt/ β -catenin regulation of liver receptor homolog-1 mediates pluripotency gene expression. *Stem Cells* 28, 1794–1804.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903.
- West, J.A., Park, I.-H., Daley, G.Q., and Geijsen, N. (2006). In vitro generation of germ cells from murine embryonic stem cells. *Nat. Protoc.* 1, 2026–2036.
- Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hübner, K., and Schöler, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122, 881–894.

REVIEW

The transcriptional regulation of pluripotency

Jia-Chi Yeo^{1,2}, Huck-Hui Ng^{1,2,3,4,5}

¹Gene Regulation Laboratory, Genome Institute of Singapore, 60 Biopolis Street, #02-01 Genome Building, Singapore 138672, Singapore; ²School of Biological Sciences, Nanyang Technological University, Singapore 637551, Singapore; ³Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597, Singapore; ⁴Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore; ⁵NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 117456, Singapore

The defining features of embryonic stem cells (ESCs) are their self-renewing and pluripotent capacities. Indeed, the ability to give rise into all cell types within the organism not only allows ESCs to function as an ideal *in vitro* tool to study embryonic development, but also offers great therapeutic potential within the field of regenerative medicine. However, it is also this same remarkable developmental plasticity that makes the efficient control of ESC differentiation into the desired cell type very difficult. Therefore, in order to harness ESCs for clinical applications, a detailed understanding of the molecular and cellular mechanisms controlling ESC pluripotency and lineage commitment is necessary. In this respect, through a variety of transcriptomic approaches, ESC pluripotency has been found to be regulated by a system of ESC-associated transcription factors; and the external signalling environment also acts as a key factor in modulating the ESC transcriptome. Here in this review, we summarize our current understanding of the transcriptional regulatory network in ESCs, discuss how the control of various signalling pathways could influence pluripotency, and provide a future outlook of ESC research.

Keywords: embryonic stem cells; pluripotency; ESCs; EpiSCs; transcriptional regulation; gene expression; signaling pathways; naïve; primed

Cell Research (2013) **23**:20–32. doi:10.1038/cr.2012.172; published online 11 December 2012

Introduction

Embryonic stem cells (ESCs) are derived from the inner cell mass (ICM) of pre-implantation embryos [1–3]; and the two defining features of ESCs are their ability to self-renew, as well as give rise to all cell lineages of the organism. This unique feature of ESCs to form the various tissue types, termed ‘pluripotency’, besides enabling

the *in vitro* study of early mammalian development, has also facilitated the advancement of regenerative medicine, whereby ESCs could be used to create clinically relevant cell types for the replacement of worn-out tissue. However, it is this same remarkable developmental plasticity of ESCs which poses a major challenge towards the efficient control of ESC differentiation into the desired lineage. Therefore, to better understand mammalian development, as well as to exploit the tremendous therapeutic potential of ESCs, it is necessary to identify the molecular mechanisms governing a pluripotent or differentiated ESC fate. Here in this review, we will summarize the current progress towards understanding the ESC transcriptional regulatory network, and also discuss how modulation of the various signalling pathways in ESCs could influence pluripotency.

Dissecting the ESC transcriptional regulatory network

The core ESC pluripotency factors: Oct4, Sox2 and

Correspondence: Huck-Hui NG

E-mail: nghh@gis.a-star.edu.sg

Abbreviations: 2C (2-cell); 2i (two inhibitor); BMP4 (bone morphogenic protein 4); ChIP (chromatin immunoprecipitation); ChIP-seq (chromatin immunoprecipitation and sequencing); ESCs (embryonic stem cells); EpiSCs (Epiblast-derived stem cells); Fgf4 (fibroblast growth factor 4); hESCs (human embryonic stem cells); ICM (inner cell mass); iPSCs (induced pluripotent stem cells); LIF (leukemia inhibitory factor); lincRNA (large intergenic non-coding RNA); mESCs (mouse embryonic stem cells); miRNA (micro RNA); ncRNA (non-coding RNA); PcG (Polycomb group proteins); PGCs (primordial germ cells); POU (Pit-Oct-Unc); RNAi (RNA interference); RNA-seq (RNA sequencing); SCF (stem cell factor); siRNA (short interfering RNA); XaXa (double X-chromosome active); XaXi (single X-chromosome inactivated)

Nanog

In ESCs, the pluripotent state is mainly regulated by the core transcription factor trio of Oct4, Sox2 and Nanog [4].

Oct4 (encoded by the *Pou5f1* gene), belongs to the Pit-Oct-Unc (POU) family of homeodomain proteins, and is exclusively expressed within the totipotent mouse blastomeres, pluripotent epiblast as well as primordial germ cells (PGCs) [5, 6]. Importantly, Oct4 plays a critical role in the establishment and maintenance of pluripotency, as *Pou5f1*-null embryos do not form a pluripotent ICM, but rather, differentiate into trophectodermal tissue [7]. Similarly, Oct4 is also critical for maintaining mouse ESCs (mESCs) in an undifferentiated state and has to be tightly regulated. Depletion of Oct4 mRNA by 50% is sufficient to result in the formation of trophectoderm cells, while Oct4 overexpression by 50% will promote mesodermal and endodermal differentiation [8].

Sox2, which contains the high-mobility group box DNA-binding domain, is expressed within the ICM and extraembryonic ectoderm of pre-implantation blastocysts [9]. Like Oct4 knockout mice, *Sox2*-null blastocysts fail to form a pluripotent ICM [9], and mESCs deficient in Sox2 differentiated primarily into trophectoderm [10]. This similarity of phenotypes produced by Sox2 and Oct4 loss is attributed to the synergistic action of Oct4/Sox2 in the regulation of various ESC-specific genes [11–16], including themselves [17–19]. Indeed, ectopic Oct4 expression was found to be sufficient in rescuing the differentiation phenotype of *Sox2*^{-/-} mESCs [10].

Nanog, the third member of the core ESC transcription factors, was discovered through a screen for pluripotency factors that could sustain mESC self-renewal in the absence of leukemia inhibitor factor (LIF) [20, 21]. Although *Nanog*-null pre-implantation embryos do not possess a pluripotent ICM [21, 22], *Nanog*^{-/-} mESCs could be established through the *in vitro* disruption of both endogenous *Nanog* alleles [21, 23]. Importantly, these *Nanog*-null mESCs although prone to differentiation, could still be kept pluripotent [21, 23]. Therefore, it is believed that Nanog, while necessary for the acquisition of pluripotency, is dispensable once pluripotency is achieved [23].

Given the important roles of the core ESC transcription factors in establishing a pluripotent stem cell fate, chromatin immunoprecipitation (ChIP) technologies have been used to map the genomic-binding sites of these core ESC factors in mESCs and hESCs. Together, these studies found extensive Oct4, Sox2 and Nanog co-binding at numerous active, as well as silent genomic target sites [24, 25]. Together, it is proposed that the core ESC transcription factors serve to establish a pluripotent

state by: (1) activating the expression of other pluripotency-associated factors while simultaneously repressing lineage-specific genes, and (2) by activating their own gene expression and that of each other [4]. Importantly, this model may account for how ESCs can sustain self-renewal and pluripotency, while still remain poised for differentiation.

The expanded ESC pluripotency network

To uncover additional novel ESC regulators, one method is to perform RNA interference (RNAi)-mediated gene knockdown and to observe for any loss of pluripotency. Indeed, large-scale RNAi knockdown studies have led to the discovery of important mESC factors such as *Esrrb*, *Tbx3* and *Tcl1* [26], as well as the chromatin regulators *Tip60*-*p400* [27] and *SetDB1* [28]. Similarly, unbiased genome-wide siRNA screens were able to identify *Cnot3* and *Trim28* [29], *Paf1C* [30] and the mediator and cohesin complex [31] as important mESC transcriptional cofactors. Extending this approach into hESCs, Chia *et al.* [32] used a similar genome-wide siRNA screening to identify components of the INO80 chromatin remodeling complex, the mediator and TAF transcriptional regulatory complexes, and the COP9 signalosomes as important hESC factors. Importantly, the PRDM14 transcription factor was found to be an essential factor for hESCs, but not in mESCs, hence highlighting that critical species-specific differences exist between mouse and human ESCs (hESCs) [32].

Determining the protein-binding partners of known pluripotency factors is another method of identifying novel ESC regulators. Through the coupling of affinity purification methods with mass spectrometry technology, numerous co-binding proteins of the core pluripotency transcription factors have been identified [33–40]. Taken together, these studies reveal an extensive protein-protein interaction network which includes other ESC transcription regulators, chromatin remodeling and modifying factors, DNA methyltransferases and Polycomb group proteins (PcG). This therefore suggests that the core ESC factors may regulate gene expression through the modulation of chromatin states. Importantly, this large and intricate network of protein interactions could suggest how small changes in the levels of core ESC factors, like Oct4 or Sox2, are sufficient to perturb the ESC self-renewal programme to trigger differentiation [8, 41].

Using ChIP with massively parallel DNA sequencing (ChIP-seq) [42], or *in vivo* biotinylated ChIP with DNA microarray [43], two independent groups examined the context in which these additional pluripotency factors could play in the specification of an ESC identity. Together, these studies observed the binding of multiple

transcription factors onto a similar genomic region; and importantly, these factors could also be grouped into either an Oct4- or Myc-centric module based on their genomic targets [43]. These data thus indicate that within ESCs, the Myc-cluster appears to function independently from the core pluripotency network [29, 42, 43].

The role of Myc in ESC transcriptional regulation

The Myc module which consists of c-Myc, n-Myc, Rex1, Zfx and E2f1, is known to be involved in self-renewal and cell metabolism [42-44]. Although approximately one-third of all active ESC genes are bound by both c-Myc and the core ESC pluripotency factors [45], the functions of these two modules in gene regulation appear to differ. For instance, Oct4, Sox2 and Nanog in concert with the Mediator complex are able to recruit RNA polymerase II (RNA Pol II) for gene transcription [31], while c-Myc is believed to control the transcriptional pause release of RNA Poly II via the p-TEFb cyclin-dependant kinase [45]. It is thus believed that the core ESC factors will select ESC genes for expression through the recruitment of RNA Pol II, while c-Myc functions to control gene expression through the release of transcriptional pause [4]. Importantly, it should also be highlighted that initial reports of cancer cells possessing an ESC-like transcriptional program [46, 47] were later attributed to be a consequence of Myc-module activity, as opposed to the core ESC factors [48].

Recent data by Lin *et al.* [49] and Nie *et al.* [50] suggest that rather than activating new genes, Myc acts only to amplify the existing transcriptional output of active genes. By increasing the levels of Myc, both groups found the loading of Myc proteins onto the promoters of active genes, but not for promoters of silent genes. This therefore indicates that Myc is unable to initiate *de novo* gene activation. Indeed, higher levels of p-TEFb occupancy with increased levels of elongation-associated RNA Polymerase II phosphorylation were found at Myc-bound sites [49], consistent with the previously proposed idea that c-Myc controls transcriptional pause release [45]. Taken together, this ability to amplify existing active gene transcription may in part suggest how the addition of c-Myc during reprogramming is able to increase the efficacy of induced pluripotent stem cell (iPSC) formation [51].

Transcriptional control of non-coding RNAs in ESCs

MicroRNAs (miRNAs), which post-transcriptionally regulate mRNA levels, are important for proper ESC function [52, 53]. Importantly through ChIP-seq, Marson *et al.* [54] found the binding of Oct4, Sox2, Nanog and Tcf3 transcription factors at ESC-related miRNA genes,

as well as transcriptionally silent PcG-occupied tissue-specific miRNAs, suggesting that the core ESC transcriptional network is able to regulate miRNA expression. A notable example of these inactive miRNAs includes *let-7*, which is known to target c-Myc and the pluripotency factors Sall4 and Lin28 [55]. Interestingly, the ESC-related *miRNA-302/367* cluster which is also regulated by the ESC transcriptional circuit [54], is able to reprogramme fibroblasts into iPSCs without the need for additional protein factors [56].

Large intergenic non-coding RNAs (lincRNAs) expressed in mESCs are also known to be the targets of several pluripotency-associated transcription factors, and have been demonstrated to be essential for maintaining a pluripotent state and suppress lineage specification [57]. Upon knockdown, many of these ESC lincRNAs were shown to induce gene expression changes similar to the depletion of known ESC factors. Importantly, it was found that lincRNAs could associate with multiple chromatin complexes, hence suggesting the possibility that lincRNAs may serve as scaffolds for the recruitment of different protein complexes for specific functions. In that regard, it would be interesting to examine how the ESC-specific lincRNA expression signature, through the assembly of various protein complexes, is able to sustain ESC pluripotency.

Extrinsic signalling and ESC transcriptional regulation

Signalling in mouse ESCs

Traditionally, mESCs were cultured and kept pluripotent on a layer of mitotically inactivated feeder cells in serum-supplemented media [1, 2] (Figure 1). However, the undefined nature of feeder co-culture posed a significant challenge in mapping the specific extrinsic signalling factors, cellular pathways involved and their effects upon gene transcription which ultimately control ESC self-renewal and differentiation.

After the initial discovery of mESCs in 1981, work over the next 20 years then identified the IL-6 family cytokine LIF as the key active component produced by feeder cells which sustains mESC growth [58, 59], as well as Bone morphogenic protein 4 (BMP4) which could substitute for the use of serum in mESC culture [60]. Critically, the combinatorial use of LIF and BMP4 alone in a defined feeder- and serum-free culture was sufficient to derive and maintain germ-line transmittable mESCs [60], hence delineating the signalling pathways controlling mESC self-renewal and pluripotency. At the transcriptional level, LIF and BMP4 signalling is able to induce phosphorylation and activation of their down-

Figure 1 Interconversion between mouse pluripotent states. mESCs are derived from the inner cell mass (ICM) of mouse pre-implantation blastocysts, while EpiSCs are isolated from the epiblast compartment of post-implantation embryos. The conversion between the pluripotent mESC and EpiSC states is possible through cell culture; introduction of transcription factors (TFs), or application of chemical inhibitors. Unipotent PGCs are derived from the embryonic gonad and can also be differentiated from EpiSCs using BMP4; or with BMP4, BMP8b, LIF, and stem cell factor (SCF). Subsequently, PGCs can be restored to pluripotency through the culture with LIF, FGF2 and SCF to form mouse embryonic germ cells (EGCs). mESCs are also known to transiently cycle into a 2-cell (2C)-like state with expanded developmental potential, giving rise to both embryonic and extraembryonic tissues. The process and mechanism by which this 2C transition occurs are still not known, and whether 2C-like cells could be directly isolated from 2C embryos has not been established.

stream transcription factors Stat3 and Smad1 respectively [60, 61]. Importantly, genome-wide mESC transcription factor mapping studies reveal the co-binding of Stat3 and Smad1 transcription factors at Oct4, Sox2 and Nanog regions, thus demonstrating how LIF and BMP4 signalling may sustain the core ESC transcriptional network [42].

In contrast to the self-renewing effect of LIF and BMP4, mESC autocrine stimulation by fibroblast growth factor 4 (Fgf4) [62, 63], working through the Mek/Erk signalling pathway, is known to induce mESCs to exit self-renewal and initiate differentiation [64, 65]. This pro-differentiation effect of Fgf/Mek/Erk signalling can be further inferred through complementary experiments, whereby chemical inhibition or genetic knockout of Fgf/Mek/Erk signalling cascade components caused impaired mESC differentiation [64–68]. The exact mechanism by which active Mek/Erk signalling induces mESCs to exit pluripotency is still not known. Given that terminal signalling kinases such as Erk have been previously shown in other cell types to regulate gene expression activity via transcription factor phosphorylation [69, 70], phosphorylation of chromatin remodelling complexes [71], and direct binding onto DNA as a transcriptional repressor [72], it would be interesting to determine if Erk may adopt similar mechanisms in regulating mESC pluripotency.

mESCs are also responsive to Wnt-signalling, but whether Wnt signalling promotes self-renewal or causes mESC differentiation is still being actively debated [73,

74]. Active canonical Wnt-signalling leads to the stabilisation of β -catenin, hence allowing for the association of β -catenin with the TCF/LEF family of transcription factors for gene activation [73]. Tcf3, which is the most abundant Tcf protein in mESCs [75], has been previously shown to co-localize with the core Oct4, Sox2 and Nanog-binding sites [54, 76], therefore suggesting that Wnt signalling, like LIF and BMP pathways, is integrated into the ESC transcriptional regulatory core. In mESCs, Tcf3 functions primarily as a transcriptional repressor [75, 77, 78] and *Tcf3*-null mESCs are resistant towards differentiation [75, 79]. However, as the process in which Tcf3 regulates gene expression is still not clear, and with multiple models being proposed to account for the mechanism of how β -catenin and Tcf3 interaction could initiate gene expression [73], additional studies are clearly needed to define the role of Wnt-signalling in ESCs. Additionally, given that Wnt-signalling in hESCs has not been well explored, it would also be interesting to determine if similar Wnt-regulatory pathways are conserved in hESCs.

Signalling in hESCs

Although hESCs may share the same Oct4-Sox2-Nanog core ESC transcriptional regulatory circuit [24, 25]; they differ from mESCs in their extrinsic signalling requirement to maintain self-renewal and pluripotency. Previous studies have shown that LIF is dispensable for

hESC culture [80], while the presence of BMP4 will cause hESC differentiation into trophoblast [81]. Rather, FGF2 and Activin/Nodal signalling is critical for the maintenance of an undifferentiated hESC state [82-84].

The exact mechanism by which FGF2-signalling sustains hESC pluripotency is still not clear and is complicated by the fact that FGF-signalling could activate multiple signalling cascades [85]. Issues pertaining to signalling ‘crosstalk’ by other receptor kinases onto FGF-receptor-associated pathways [86, 87], as well as varying hESC culture conditions employed by different groups, have made the dissection of hESC FGF2-signalling more difficult. However, previous studies have indicated that FGF-signalling through its downstream protein ERK is necessary to prevent extra-embryonic differentiation of hESCs [88, 89], and that FGF2 may cooperate with other growth factors (like Activin signalling, through SMAD2/3 transcription factors), to upregulate the expression of the core pluripotency gene *NANOG* [84, 90].

There have already been several large-scale attempts at profiling the global hESC phosphoproteome via mass spectrometry techniques [91-94], with two studies seeking to specifically address the dynamics of FGF2-dependent tyrosine and serine/threonine phosphorylation [95, 96]. Together, FGF2-signalling in hESCs not only resulted in phosphorylation of proteins of various signalling cascades such as that of PI3K, MAPK, Wnt, but could also lead to phosphorylation of pluripotency-associated transcription regulators like OCT4, SOX2, SALL4 and DPPA4 [95, 96]. These studies while informative in revealing a possible phospho-interactome downstream of FGF2-FGFR, unfortunately do not factor in signalling ‘crosstalk’ by other receptor kinases onto FGFR-associated pathways [86, 87]. Therefore, future attempts at studying the hESC signalling pathways, aside from adopting defined culture conditions, should also seek to utilize specific kinase inhibitors, gene knockdown or gene deletion strategies to more accurately delineate the cell-signalling events.

Unlike the ambiguity surrounding FGF2 signalling, Smad2/3, the downstream effectors of Activin/Nodal signaling are previously known to directly bind and regulate the expression of *NANOG* [97]. Recent ChIP-seq in hESCs also found the binding of SMAD2/3 at *OCT4*, *TERT*, *MYC* and *DPPA4* genes, with SMAD2/3 sharing approximately one-third overlap with *NANOG* genomic targets [98]. Importantly, the authors found that *NANOG* overexpression could sustain SMAD2/3 target gene expression even in the absence of Activin/Nodal signalling, thus suggesting that *NANOG* may function as a SMAD2/3 transcriptional co-factor during active Activin/Nodal signalling in hESCs [98]. In a separate study,

ChIP-seq revealed extensive co-occupancy of SMAD3 along with the hESC genome with OCT4, although it was not verified if SMAD3 could bind alongside other hESC pluripotency factors [99]. Taken together, these data provide important insight into the potential mechanisms by which Activin/Nodal signalling helps sustain hESC pluripotency.

Transcriptional regulation of ground state mESCs

As previously mentioned, the conventional culture of mESCs requires the presence of LIF and BMP4 [58-60]. However, even in the presence of LIF/BMP4, the autocrine pro-differentiation Fgf4/Mek/Erk signalling in mESCs is still active [60, 100]. This therefore suggests that pluripotency is sustained by LIF and BMP4 acting downstream of the Erk pathway to prevent mESC differentiation.

Expanding upon this observation, Ying *et al.* [100] surprisingly found that the requirement of external LIF/BMP4 in mESC culture could be abrogated via the usage of small molecule kinase inhibitors. This two inhibitor (2i) culture, consisting of PD0325901 and CHIR99021 to respectively target Mek and Gsk3, serves to protect pluripotent mESCs from the pro-differentiation effect of Fgf4 stimulation [64], as well as the transcriptional repressor activity of Tcf3 [101]. More importantly, by using the 2i culture, germ-line transmitting ESCs could be generated from previously recalcitrant mouse strains [100, 102], and from the rat [103, 104].

When mESCs are cultured under conventional LIF conditions, there is a heterogeneous expression of pluripotency transcription factors such as Nanog, Rex1 and Stella [23, 105, 106]. Interestingly, after FACS-purification, these sorted cells rather than maintaining a pure cell population, will revert into a heterogeneous population. These data therefore suggest that gene transcription of mESCs under LIF culture conditions exists in a fluctuating and dynamic state. Upon transfer into 2i conditions, Nanog and Rex1 expression will become homogeneously high [107]. As Nanog-low cells are prone to differentiate, and Rex-low mESCs do not contribute to chimera formation upon blastocyst microinjection [23, 105], the capture of a Nanog/Rex1-high mESC state indicates that pluripotency may have been stabilised under 2i conditions [107]. Interestingly, under 2i or Mek-inhibition, the *Nanog* gene expression in mESCs switches into a biallelic expression mode, as opposed to the monoallelic expression status in conventional LIF/serum culture [108]. With these unique characteristics, 2i-cultured mESCs are said to reside in a novel and distinct ‘ground state’ pluripotency [100].

Although the protein levels of pluripotency regulators like Nanog and Rex1 are known to become uniformly-

high when mESCs are switched from LIF/serum into 2i media [107], Marks *et al.* [109] however found through RNA-sequencing (RNA-seq) that apart from *Tcl1*, the mRNA levels of most other pluripotency genes like *Pou5f1*, *Sox2*, *Nanog*, *Esrrb*, *Klf2*, *Klf4* and *Tbx3* did not change between the two culture conditions. The reason behind this discrepancy is still unclear, and may have arisen due to the presence of LIF with the 2i media used in this study [109]. It is also possible that translational and/or post-translational regulatory mechanisms are involved.

This study by Marks *et al.* [109] also found that lineage genes were suppressed under 2i/LIF conditions, and that 2i/LIF mESCs had higher expression of metabolic genes. Taken together, these data argue against the hypothesis proposed by Efroni *et al.* [110] that an innate global transcriptional hyperactivity results in pluripotent developmental plasticity. Rather, it is believed that the presence of lineage gene expression in conventional mESC culture may be the result of serum stimulation [109]. Additionally, unlike cells cultured under LIF/serum conditions, 2i/LIF mESCs were found to contain more proximal-promoter pausing by RNA Polymerase II (RNA Pol II), especially at many lineage-specific genes. It was also proposed that this RNA Pol II pausing could be important towards establishing the ground state pluripotency.

Interconversion between alternative pluripotent states

Mouse epiblast stem cells

While the presence of LIF is necessary for mESCs to maintain pluripotency, the discovery of a novel LIF-independent pluripotent stem cell population derived from E5.5 to E7.5 post-implantation mouse embryos [111, 112], suggests that other states of pluripotency may exist (Figure 1). As these cells were isolated from the post-implantation epiblast tissue of the developing embryo, they were hence termed mouse epiblast stem cells (EpiSCs). While EpiSCs can self-renew and were demonstrated to be pluripotent through teratoma assays and through *in vitro* differentiation into germ cells [111–113], it should be noted that EpiSCs exhibit limited developmental potential and are generally considered inefficient in the formation of chimeras [111, 112, 114].

Similarly, while EpiSCs do express the core Oct4, Sox2 and Nanog pluripotency factors, they differ markedly from mESCs with regards to their gene expression profile, epigenetic status and usage of signalling pathways to maintain a stem cell identity [106, 111, 112, 115, 116]. Notably, EpiSC cultures like that of hESCs, require

the addition of Activin and Fgf2, but not LIF or BMP4 for self-renewal [117]. The expression levels of pluripotency markers such as Rex1, Stella, Klf2, Klf4 are lower in EpiSCs, with a concomitant higher expression of differentiation markers [112]. It should also be highlighted that while female mESCs are in the pre-inactivation state with two active X-chromosomes (XaXa), EpiSCs have already undergone X-inactivation (XaXi), consistent with a more developmentally advanced stage [117]. Therefore, these findings indicate that EpiSCs have already been ‘primed’ for differentiation, as opposed to mESCs which exist in a more developmentally ‘naive’ state.

Overall, EpiSCs appear to partially resemble hESCs, and may potentially be the counterpart of hESCs. Therefore, EpiSCs could serve as a more relevant mouse model for the study of early human embryonic development, as compared with mESCs. However, despite the aforementioned similarities with hESCs, inhibition of Fgf2/Erk signalling in EpiSCs does not result in loss of Nanog expression, and no evidence of Oct4-mediated transcriptional regulation at *Fgf2* promoters was observed for EpiSCs [118]. Similarly, certain key features of EpiSCs like the absence of Rex1, a mESC pluripotency marker, and the expression of the Fgf5 epiblast marker, are not shared by hESCs [118]. Further research is thus needed to establish whether these dissimilarities are due to species- or developmental-specific differences.

The transition between naive and primed pluripotency

As mESCs are derived from a developmentally earlier timepoint (pre-implantation embryos) as compared to EpiSCs (post-implantation embryos), the conversion of mESCs into EpiSCs would therefore correspond to a differentiation step along the normal developmental pathway (Figure 1). In this regard, culturing mESCs with FGF2 and Activin A readily results in the formation of EpiSCs [114]. Similarly, FGF treatment together with LIF/Stat3 blockade could also convert mESCs into EpiSCs [118].

In contrast, the reverse transition of EpiSCs into mESCs is more difficult, occurring at an extremely low frequency through culture with LIF [102, 114]. The reversion rates may be improved by 10^4 – 10^5 fold if the pluripotency factors Klf2, Klf4, Nanog, Nr5a2 or Esrrb were to be ectopically expressed [22, 102, 114, 119, 120]. Similarly, a chemical approach can also be adopted to convert EpiSCs into naive mESCs using a combination of LSD1, ALK5, MEK, FGFR and GSK3 inhibitors [121] (Figure 1).

Recently, Prdm14 and Klf2 were found to synergize with each other to rapidly induce the reprogramming of EpiSCs into mESCs within 3–4 days [122], in contrast to

the usual process of 8-10 days. By itself, the introduction of Prdm14 does not have much effect on the reversion process. However, transcriptomic analysis revealed that Prdm14 actually serves to prime EpiSCs for conversion through the simultaneous repression of lineage-associated genes and activation of early epiblast genes. The synergistic effect of Prdm14-Klf2 can also be attributed to the ability of Prdm14 to enhance Klf2 recruitment onto key mESC pluripotency gene loci such as the *Nr5a2* promoter and the *Oct4* distal enhancer [122].

Establishing a mESC-like hESC state

In addition to post-implantation embryos, EpiSCs can also be readily isolated from pre-implantation stage mouse blastocysts [123], which have been traditionally used to derive mESCs (Figure 1). This meant that modulation of the signalling environment could influence the formation of naive or primed pluripotency states. Because hESCs resemble EpiSCs more than mESCs, despite also having been derived from pre-implantation blastocysts, it is believed that during the hESC derivation process, pre-hESCs may have progressed into a more developmentally advanced EpiSC-like state [124]. In support of this ‘primed’ pluripotent state of hESCs, non-human primate ESCs from rhesus monkeys, like mouse EpiSCs, are unable to contribute to chimera formation [125].

Currently, hESCs suffer from very poor gene-targeting efficacy unlike mESCs, making the genetic manipulation of hESCs for research or therapeutic purposes extremely difficult [126]. Critically, the developmental stage differences between the two cell types imply that many of the protocols for mESC differentiation into various lineages may not work in a hESC system. In light of these problems, the creation of a mESC-like hESC state would not only make hESCs more amenable to gene targeting, but also enable the transfer of existing mouse differentiation protocols into hESCs.

There have been several attempts by various groups to generate mESC-like human pluripotent stem cells. The methods include either a direct conversion from conventional hESCs [127, 128] or the reprogramming of somatic cells into mESC-like hESCs [128-131] (Figure 2). Unfortunately, these cells either were dependent upon transgene expression for long-term culture, or have not been thoroughly characterised for features such as naive mESC characteristics and complete transgene independence. Taken together, while these studies have demonstrated the feasibility of creating hESCs which exhibit characteristics of naive pluripotency, future efforts should focus on improving the culture conditions to enable transgene-free long-term maintenance of these

A Direct conversion approach

B Reprogramming approach

Figure 2 Methods to create mESC-like hESCs. **(A)** hESCs with mESC-like characteristics can be generated directly from conventional hESCs using OCT4 with SOX2, or, KLF4 with KLF2, in the culture with LIF/PD/CH [128]. Similarly, the generation of hESCs which resemble mESC colonies could be achieved through culture with LIF/PD/SB [127]. **(B)** Human fibroblasts can also be reprogrammed into mESC-like human iPSCs through expression of OCT4, SOX2, NANOG, LIN28 with LIF/PD/CH/A83 [129]; through OCT4, SOX2, KLF4, c-MYC, NANOG with LIF [130]; using OCT4, SOX2, KLF4 with LIF/PD/CH [128]; or OCT4, SOX2, KLF4, c-MYC, NR5A2, RAR-γ with LIF/PD/CH [131]. PD: MEK inhibitor PD0325901. CH: GSK3 inhibitor CHIR99021. SB: p38 inhibitor SB203580. A83: ALK4,5,7 inhibitor A83-01.

mESC-like hESCs.

Several studies have also explored the possibility of generating mESC-like hESCs directly from pre-implantation human embryos. Female hESCs derived from the

culture of human blastocysts in hypoxic (5% O₂) conditions were found to be in an XaXa status, a characteristic of mouse naive pluripotency [132]. However, these cells were maintained in conventional FGF-containing media and were not tested further for features of naive pluripotency.

The generation of naive hESCs from embryos using 2i media was reported to be unsuccessful [133]. In mice, culturing 8-cell stage embryos under MEK inhibition will result in the failure to form the hypoblast compartment with a reciprocal expansion of the pluripotent epiblast [134]. However, culturing early human embryos with MEK inhibitor does not block hypoblast formation, and neither would the development of the NANOG-positive epiblast compartment be affected [133, 135]. Thus taken together, these studies suggest that additional signalling pathways could be involved in the segregation of the human ICM into the epiblast and hypoblast, or that species-specific differences exist between mouse and human embryos in the biological functions of FGF/MEK signalling.

The totipotential of 2C-like mESCs?

Recently, Macfarlan *et al.* [136] reported the discovery of a rare transient population of mESCs, which can give rise to both embryonic and extraembryonic tissues. By comparing the gene expression signature between mouse oocytes and 2-cell (2C)-staged embryos using RNA-seq, the authors found that the 2C transcriptome contained many genes, which were driven by retroviral elements. Among these repeats, the MuERV-L family of retroelements were found to be the most abundant. Indeed, by combining this MuERV-L regulatory sequence with a td-Tomato red fluorescence gene, the authors demonstrated through zygotic injection that strong MuERV-L reporter fluorescence was detected at the 2C stage, which would then gradually decrease and become undetectable by the blastocyst stage.

Surprisingly, stable integration of this MuERV-L td-Tomato reporter into mESCs led to the detection of a similar red fluorescence within a very small population of cells (approximately 0.2%-1.5%); and gene expression profiling of these tdTomato+ mESCs revealed that they resembled 2C embryos (Figure 1). Like Nanog, Rex1 and Stella [23, 105, 106], MuERV-L tdTomato expression was also found to exist in a dynamic state. Subsequent genetic labelling experiments suggested that all mESCs within the culture could transiently pass through this 2C-like state. Intriguingly, these 2C-like mESCs do not contain any Oct4, Sox2 and Nanog proteins, even though the transcript levels of these genes were unaffected. Perhaps the most fascinating aspect of this study, is the

demonstration that injection of these 2C-like mESCs into morula stage embryos could result in the contribution of these 2C-like donor cells towards both embryonic and extraembryonic tissues, hence suggesting a totipotent-like capability.

Together, this study raises several new interesting questions. First, what is the significance of this 2C-like stage, and why do mESCs transiently enter this phase? Second, as these 2C-like mESCs do not possess Oct4, Sox2 and Nanog proteins, how would the transcriptional network of these 2C-like mESCs be regulated? Third, as only a very rare population of mESCs are expressing MuERV-L at a given time, what would be the signalling pathways or cellular mechanisms that regulate the entry or exit of this phase? And lastly, do hESCs possess a similar transient 2C-like phase, and if so, would it also be regulated by endogenous retroviral elements?

Future outlook

Deciphering the ESC transcriptional network is essential towards understanding the cellular mechanisms that govern pluripotency. In this regard, the different studies involving various experimental approaches have now enabled researchers to appreciate the processes by which the core ESC transcription factors establish an overall ESC identity.

To bring the potential of hESCs one step closer towards its application in regenerative medicine, future research should likely focus upon dissecting the pathways that regulate lineage commitment. Recently, by probing the temporal gene expression and chromatin changes during the directed differentiation of ESCs into cardiac lineages, several novel regulators of cardiac tissue formation have been identified [137, 138]. This is achieved either by (1) determining the stage-specific activation of gene enhancers, and applying a DNA binding motif search to predict the transcription factors that are involved during cardiac differentiation [137]; or (2) by predicting key regulatory genes based on the induction of RNA expression, loss of repressive H3K27me3 marks and reciprocal increase of active H3K4me3 modification [138]. Therefore, it would be interesting to test if other novel tissue-specific regulators could be identified for different somatic lineages through analysis of temporal chromatin changes. In this regard, the ability to genetically modify and introduce hESCs reporter genes into that specify certain tissue lineages would be of tremendous value for the purpose of studying directed differentiation. For instance, by coupling these lineage-specific hESC reporters with high-throughput genome-wide siRNA screens or miRNA mimic libraries, numerous factors

that regulate differentiation into specific lineages could be identified.

Improving the efficacy of directed ESC differentiation into the desired cell-type, as well as overcoming the immature phenotype of hESC-derived differentiated cells [139] represent key issues to be tackled in the future. Resolution of these problems likely will require the optimization of important cell culture parameters such as the extracellular matrix, growth factor and cytokine-signalling environment, three-dimensional (3D) cell organization and cell culture duration. Similarly, the capture of ESC-differentiated cells at the progenitor stage may enable easier expansion and more efficient differentiation into the desired cell type of choice. This can be achieved either through the co-culture of ESC-differentiated tissue with the appropriate mesenchymal cells [140], or through the isolation of self-renewing progenitors within a heterogeneous cell population [141].

While directed differentiation is useful for the derivation of a single, or a few cell types for tissue replacement, the complete generation of complex organs comprising many cell types that work in a coordinated fashion presents a greater challenge. Most remarkably, it was recently demonstrated that pluripotent stem cells retain a self-organising ability to differentiate into 3D organoids, which resemble optic cups [142, 143] and adenohypophysis tissues [144]. Similarly, intestinal organoids have been reported to develop from hESC-derived posterior endoderm monolayers upon exposure to the appropriate signalling factors [145], and functional thyroid gland tissue has been generated from mESCs [146]. Therefore, if one were to harness this powerful self-developing property of pluripotent stem cells for the generation of novel complex organ types for research or medical purposes, it would be important to first determine the mechanisms underlying these processes. In this regard, studying the temporal transcriptional changes occurring at the whole-organoid, or at the single-cell level, may be a good starting point to dissect the pathways involved.

Acknowledgments

We would like to thank Winston Chan and Kevin Andrew Uy Gonzales from the Genome Institute of Singapore for their comments on this manuscript.

References

- Evans MJ, Kaufman MH. Establishment in culture of pluripotent cells from mouse embryos. *Nature* 1981; **292**:154-156.
- Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* 1981; **78**:7634-7638.
- Thomson JA, Itskovitz-Eldor J, Shapiro SS, *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* 1998; **282**:1145-1147.
- Young RA. Control of the embryonic stem cell state. *Cell* 2011; **144**:940-954.
- Rosner MH, Vigano MA, Ozato K, *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* 1990; **345**:686-692.
- Scholer HR, Dressler GR, Balling R, Rohdewohld H, Gruss P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J* 1990; **9**:2185-2195.
- Nichols J, Zevnik B, Anastassiadis K, *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998; **95**:379-391.
- Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 2000; **24**:372-376.
- Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003; **17**:126-140.
- Masui S, Nakatake Y, Toyooka Y, *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* 2007; **9**:625-635.
- Yuan H, Corbi N, Basilico C, Dailey L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 1995; **9**:2635-2645.
- Nishimoto M, Fukushima A, Okuda A, Muramatsu M. 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* 1999; **19**:5453-5465.
- Tokuzawa Y, Kaiho E, Maruyama M, *et al.* Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Mol Cell Biol* 2003; **23**:2699-2708.
- Nakatake Y, Fukui N, Iwamatsu Y, *et al.* Klf4 cooperates with Oct3/4 and Sox2 to activate the *Lefty1* core promoter in embryonic stem cells. *Mol Cell Biol* 2006; **26**:7772-7782.
- Kuroda T, Tada M, Kubota H, *et al.* Octamer and Sox elements are required for transcriptional *cis* regulation of *Nanog* gene expression. *Mol Cell Biol* 2005; **25**:2475-2485.
- Rodda DJ, Chew JL, Lim LH, *et al.* Transcriptional regulation of *nanog* by OCT4 and SOX2. *J Biol Chem* 2005; **280**:24731-24737.
- Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F. Oct-3/4 and Sox2 regulate *Oct-3/4* gene in embryonic stem cells. *J Biol Chem* 2005; **280**:5307-5317.
- Tomiooka M, Nishimoto M, Miyagi S, *et al.* Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res* 2002; **30**:3202-3213.
- Chew JL, Loh YH, Zhang W, *et al.* Reciprocal transcriptional regulation of *Pou5f1* and *Sox2* via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* 2005; **25**:6031-6046.
- Chambers I, Colby D, Robertson M, *et al.* Functional expression cloning of *Nanog*, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003; **113**:643-655.
- Mitsui K, Tokuzawa Y, Itoh H, *et al.* The homeoprotein

- Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003; **113**:631-642.
- 22 Silva J, Nichols J, Theunissen TW, *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* 2009; **138**:722-737.
- 23 Chambers I, Silva J, Colby D, *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* 2007; **450**:1230-1234.
- 24 Loh YH, Wu Q, Chew JL, *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 2006; **38**:431-440.
- 25 Boyer LA, Lee TI, Cole MF, *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005; **122**:947-956.
- 26 Ivanova N, Dobrin R, Lu R, *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* 2006; **442**:533-538.
- 27 Fazio TG, Huff JT, Panning B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 2008; **134**:162-174.
- 28 Bilodeau S, Kagey MH, Frampton GM, Rahl PB, Young RA. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* 2009; **23**:2484-2489.
- 29 Hu G, Kim J, Xu Q, Leng Y, Orkin SH, Elledge SJ. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* 2009; **23**:837-848.
- 30 Ding L, Paszkowski-Rogacz M, Nitzsche A, *et al.* A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* 2009; **4**:403-415.
- 31 Kagey MH, Newman JJ, Bilodeau S, *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 2010; **467**:430-435.
- 32 Chia NY, Chan YS, Feng B, *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* 2010; **468**:316-320.
- 33 Wang J, Rao S, Chu J, *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* 2006; **444**:364-368.
- 34 Liang J, Wan M, Zhang Y, *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* 2008; **10**:731-739.
- 35 Wu Q, Chen X, Zhang J, *et al.* Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* 2006; **281**:24090-24094.
- 36 van den Berg DL, Snoek T, Mullin NP, *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 2010; **6**:369-381.
- 37 Pardo M, Lang B, Yu L, *et al.* An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 2010; **6**:382-395.
- 38 Ding J, Xu H, Faiola F, Ma'ayan A, Wang J. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Res* 2012; **22**:155-167.
- 39 Gao Z, Cox JL, Gilmore JM, *et al.* Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. *J Biol Chem* 2012; **287**:11384-11397.
- 40 Mallanna SK, Ormsbee BD, Iacovino M, *et al.* Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells* 2010; **28**:1715-1727.
- 41 Kopp JL, Ormsbee BD, Desler M, Rizzino A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* 2008; **26**:903-911.
- 42 Chen X, Xu H, Yuan P, *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008; **133**:1106-1117.
- 43 Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 2008; **132**:1049-1061.
- 44 Kidder BL, Yang J, Palmer S. Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. *PLoS One* 2008; **3**:e3932.
- 45 Rahl PB, Lin CY, Seila AC, *et al.* c-Myc regulates transcriptional pause release. *Cell* 2010; **141**:432-445.
- 46 Ben-Porath I, Thomson MW, Carey VJ, *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* 2008; **40**:499-507.
- 47 Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* 2008; **2**:333-344.
- 48 Kim J, Woo AJ, Chu J, *et al.* A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* 2010; **143**:313-324.
- 49 Lin CY, Loven J, Rahl PB, *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* 2012; **151**:56-67.
- 50 Nie Z, Hu G, Wei G, *et al.* c-Myc Is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* 2012; **151**:68-79.
- 51 Nakagawa M, Koyanagi M, Tanabe K, *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 2008; **26**:101-106.
- 52 Kanellopoulou C, Muljo SA, Kung AL, *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* 2005; **19**:489-501.
- 53 Wang Y, Medvid R, Melton C, Jaenisch R, Blüthgen R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 2007; **39**:380-385.
- 54 Marson A, Levine SS, Cole MF, *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 2008; **134**:521-533.
- 55 Melton C, Judson RL, Blüthgen R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* 2010; **463**:621-626.
- 56 Anokye-Danso F, Trivedi CM, Juhr D, *et al.* Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* 2011; **8**:376-388.
- 57 Guttman M, Donaghey J, Carey BW, *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 2011; **477**:295-300.
- 58 Williams RL, Hilton DJ, Pease S, *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 1988; **336**:684-687.
- 59 Smith AG, Heath JK, Donaldson DD, *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 1988; **336**:688-690.

- 60 Ying QL, Nichols J, Chambers I, Smith A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 2003; **115**:281-292.
- 61 Niwa H, Burdon T, Chambers I, Smith A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* 1998; **12**:2048-2060.
- 62 Rathjen J, Lake JA, Bettess MD, Washington JM, Chapman G, Rathjen PD. Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci* 1999; **112 (Part 5)**:601-612.
- 63 Ma YG, Rosfjord E, Huebert C, *et al.* Transcriptional regulation of the murine k-FGF gene in embryonic cell lines. *Dev Biol* 1992; **154**:45-54.
- 64 Kunath T, Saba-El-Leil MK, Almousailleakh M, Wray J, Meloche S, Smith A. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* 2007; **134**:2895-2902.
- 65 Stavridis MP, Lunn JS, Collins BJ, Storey KG. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development* 2007; **134**:2889-2894.
- 66 Cheng AM, Saxton TM, Sakai R, *et al.* Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* 1998; **95**:793-803.
- 67 Burdon T, Stracey C, Chambers I, Nichols J, Smith A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol* 1999; **210**:30-43.
- 68 Chen S, Do JT, Zhang Q, *et al.* Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci USA* 2006; **103**:17266-17271.
- 69 Johnson GL, Lapadat R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 2002; **298**:1911-1912.
- 70 Ramos JW. The regulation of extracellular signal-regulated kinase (ERK) in mammalian cells. *Int J Biochem Cell Biol* 2008; **40**:2707-2719.
- 71 Simone C, Forcales SV, Hill DA, Imbalzano AN, Latella L, Puri PL. p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet* 2004; **36**:738-743.
- 72 Hu S, Xie Z, Onishi A, *et al.* Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* 2009; **139**:610-622.
- 73 Sokol SY. Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development* 2011; **138**:4341-4350.
- 74 Wray J, Hartmann C. WNTing embryonic stem cells. *Trends Cell Biol* 2012; **22**:159-168.
- 75 Pereira L, Yi F, Merrill BJ. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* 2006; **26**:7479-7491.
- 76 Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* 2008; **22**:746-755.
- 77 Yi F, Pereira L, Merrill BJ. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* 2008; **26**:1951-1960.
- 78 Tam WL, Lim CY, Han J, *et al.* T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* 2008; **26**:2019-2031.
- 79 Guo G, Huang Y, Humphreys P, Wang X, Smith A. A Piggy-Bac-based recessive screening method to identify pluripotency regulators. *PLoS One* 2011; **6**:e18189.
- 80 Daheron L, Opitz SL, Zaehres H, *et al.* LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* 2004; **22**:770-778.
- 81 Xu RH, Chen X, Li DS, *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* 2002; **20**:1261-1264.
- 82 Xu RH, Peck RM, Li DS, Feng X, Ludwig T, Thomson JA. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods* 2005; **2**:185-190.
- 83 James D, Levine AJ, Besser D, Hemmati-Brivanlou A. TGF-beta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* 2005; **132**:1273-1282.
- 84 Vallier L, Alexander M, Pedersen RA. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* 2005; **118**:4495-4509.
- 85 Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. *Nat Rev Cancer* 2010; **10**:116-129.
- 86 Katoh M. Cross-talk of WNT and FGF signaling pathways at GSK3beta to regulate beta-catenin and SNAIL signaling cascades. *Cancer Biol Ther* 2006; **5**:1059-1064.
- 87 Guo X, Wang XF. Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res* 2009; **19**:71-88.
- 88 D'Amour KA, Gage FH. Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells. *Proc Natl Acad Sci USA* 2003; **100 Suppl 1**:11866-11872.
- 89 Li J, Wang G, Wang C, *et al.* MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. *Differentiation* 2007; **75**:299-307.
- 90 Vallier L, Mendjan S, Brown S, *et al.* Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* 2009; **136**:1339-1349.
- 91 Swaney DL, Wenger CD, Thomson JA, Coon JJ. Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci USA* 2009; **106**:995-1000.
- 92 Van Hoof D, Munoz J, Braam SR, *et al.* Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* 2009; **5**:214-226.
- 93 Brill LM, Xiong W, Lee KB, *et al.* Phosphoproteomic analysis of human embryonic stem cells. *Cell Stem Cell* 2009; **5**:204-213.
- 94 Rigbolt KT, Prokhorova TA, Akimov V, *et al.* System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal* 2011; **4**:rs3.
- 95 Ding VM, Boersema PJ, Foong LY, *et al.* Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells. *PLoS One* 2011; **6**:e17538.
- 96 Zoumaro-Djayoon AD, Ding V, Foong LY, Choo A, Heck AJ,

- Munoz J. Investigating the role of FGF-2 in stem cell maintenance by global phosphoproteomics profiling. *Proteomics* 2011; **11**:3962-3971.
- 97 Xu RH, Sampsel-Barron TL, Gu F, *et al.* NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell* 2008; **3**:196-206.
- 98 Brown S, Teo A, Pauklin S, *et al.* Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* 2011; **29**:1176-1185.
- 99 Mullen AC, Orlando DA, Newman JJ, *et al.* Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* 2011; **147**:565-576.
- 100 Ying QL, Wray J, Nichols J, *et al.* The ground state of embryonic stem cell self-renewal. *Nature* 2008; **453**:519-523.
- 101 Wray J, Kalkan T, Gomez-Lopez S, *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol* 2011; **13**:838-845.
- 102 Hanna J, Markoulaki S, Mitalipova M, *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* 2009; **4**:513-524.
- 103 Buehr M, Meek S, Blair K, *et al.* Capture of authentic embryonic stem cells from rat blastocysts. *Cell* 2008; **135**:1287-1298.
- 104 Li P, Tong C, Mehrian-Shai R, *et al.* Germline competent embryonic stem cells derived from rat blastocysts. *Cell* 2008; **135**:1299-1310.
- 105 Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* 2008; **135**:909-918.
- 106 Hayashi K, Lopes SM, Tang F, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 2008; **3**:391-401.
- 107 Wray J, Kalkan T, Smith AG. The ground state of pluripotency. *Biochem Soc Trans* 2010; **38**:1027-1032.
- 108 Miyazari Y, Torres-Padilla ME. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* 2012; **483**:470-473.
- 109 Marks H, Kalkan T, Menafrá R, *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 2012; **149**:590-604.
- 110 Efroni S, Duttagupta R, Cheng J, *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2008; **2**:437-447.
- 111 Brons IG, Smithers LE, Trotter MW, *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 2007; **448**:191-195.
- 112 Tesar PJ, Chenoweth JG, Brook FA, *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 2007; **448**:196-199.
- 113 Hayashi K, Surani MA. Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming *in vitro*. *Development* 2009; **136**:3549-3556.
- 114 Guo G, Yang J, Nichols J, *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* 2009; **136**:1063-1069.
- 115 Bao S, Tang F, Li X, *et al.* Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* 2009; **461**:1292-1295.
- 116 Sun B, Ito M, Mendjan S, *et al.* Status of genomic imprinting in epigenetically distinct pluripotent stem cells. *Stem Cells* 2012; **30**:161-168.
- 117 Hanna JH, Saha K, Jaenisch R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* 2010; **143**:508-525.
- 118 Greber B, Wu G, Bernemann C, *et al.* Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* 2010; **6**:215-226.
- 119 Hall J, Guo G, Wray J, *et al.* Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* 2009; **5**:597-609.
- 120 Festuccia N, Osorno R, Halbritter F, *et al.* *Esrrb* is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* 2012; **11**:477-490.
- 121 Zhou H, Li W, Zhu S, *et al.* Conversion of mouse epiblast stem cells to an earlier pluripotency state by small molecules. *J Biol Chem* 2010; **285**:29676-29680.
- 122 Gillich A, Bao S, Grabole N, *et al.* Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* 2012; **10**:425-439.
- 123 Najm FJ, Chenoweth JG, Anderson PD, *et al.* Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell* 2011; **8**:318-325.
- 124 De Los Angeles A, Loh YH, Tesar PJ, Daley GQ. Accessing naïve human pluripotency. *Curr Opin Genet Dev* 2012; **22**:272-282.
- 125 Tachibana M, Sparman M, Ramsey C, *et al.* Generation of chimeric rhesus monkeys. *Cell* 2012; **148**:285-295.
- 126 Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells. *Nat Biotechnol* 2003; **21**:319-321.
- 127 Xu Y, Zhu X, Hahm HS, *et al.* Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc Natl Acad Sci USA* 2010; **107**:8129-8134.
- 128 Hanna J, Cheng AW, Saha K, *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci USA* 2010; **107**:9222-9227.
- 129 Li W, Wei W, Zhu S, *et al.* Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell Stem Cell* 2009; **4**:16-19.
- 130 Buecker C, Chen HH, Polo JM, *et al.* A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell* 2010; **6**:535-546.
- 131 Wang W, Yang J, Liu H, *et al.* Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc Natl Acad Sci USA* 2011; **108**:18283-18288.
- 132 Lengner CJ, Gimelbrant AA, Erwin JA, *et al.* Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* 2010; **141**:872-883.
- 133 Roode M, Blair K, Snell P, *et al.* Human hypoblast formation is not dependent on FGF signalling. *Dev Biol* 2012; **361**:358-363.

- 134 Nichols J, Silva J, Roode M, Smith A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* 2009; **136**:3215-3222.
- 135 Kuijk EW, van Tol LT, Van de Velde H, *et al.* The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* 2012; **139**:871-882.
- 136 Macfarlan TS, Gifford WD, Driscoll S, *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 2012; **487**:57-63.
- 137 Wamstad JA, Alexander JM, Truty RM, *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 2012; **151**:206-220.
- 138 Paige SL, Thomas S, Stoick-Cooper CL, *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* 2012; **151**:221-232.
- 139 Cohen DE, Melton D. Turning straw into gold: directing cell fate for regenerative medicine. *Nat Rev Genet* 2011; **12**:243-252.
- 140 Sneddon JB, Borowiak M, Melton DA. Self-renewal of embryonic-stem-cell-derived progenitors by organ-matched mesenchyme. *Nature* 2012; **491**:765-768.
- 141 Cheng X, Ying L, Lu L, *et al.* Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell* 2012; **10**:371-384.
- 142 Eiraku M, Takata N, Ishibashi H, *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* 2011; **472**:51-56.
- 143 Nakano T, Ando S, Takata N, *et al.* Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* 2012; **10**:771-785.
- 144 Suga H, Kadoshima T, Minaguchi M, *et al.* Self-formation of functional adenohypophysis in three-dimensional culture. *Nature* 2011; **480**:57-62.
- 145 Spence JR, Mayhew CN, Rankin SA, *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue *in vitro*. *Nature* 2011; **470**:105-109.
- 146 Antonica F, Kasprzyk DF, Opitz R, *et al.* Generation of functional thyroid from embryonic stem cells. *Nature* 2012; **491**:66-71.

REVIEW

Transcriptomic analysis of pluripotent stem cells: insights into health and disease

Jia-Chi Yeo^{1,2} and Huck-Hui Ng^{1,2,3,4,5,*}

Abstract

Embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) hold tremendous clinical potential because of their ability to self-renew, and to differentiate into all cell types of the body. This unique capacity of ESCs and iPSCs to form all cell lineages is termed pluripotency. While ESCs and iPSCs are pluripotent and remarkably similar in appearance, whether iPSCs truly resemble ESCs at the molecular level is still being debated. Further research is therefore needed to resolve this issue before iPSCs may be safely applied in humans for cell therapy or regenerative medicine. Nevertheless, the use of iPSCs as an *in vitro* human genetic disease model has been useful in studying the molecular pathology of complex genetic diseases, as well as facilitating genetic or drug screens. Here, we review recent progress in transcriptomic approaches in the study of ESCs and iPSCs, and discuss how deregulation of these pathways may be involved in the development of disease. Finally, we address the importance of these advances for developing new therapeutics, and the future challenges facing the clinical application of ESCs and iPSCs.

Keywords Embryonic stem cells, gene expression, induced pluripotent stem cells, pluripotency, regenerative medicine, therapy, transcriptional regulation, transcriptomics.

Stem cell transcriptomics and transcriptional networks

Embryonic stem cells (ESCs) have the unique ability to self-renew and differentiate into cells of all three germ layers of the body. This capacity to form all adult cell

types, termed 'pluripotency', allows researchers to study early mammalian development in an artificial setting and offers opportunities for regenerative medicine, whereby ESCs could generate clinically relevant cell types for tissue repair. However, this same malleability of ESCs also renders it a challenge to obtain *in vitro* differentiation of ESCs to specific cell types at high efficacy. Therefore, harnessing the full potential of ESCs requires an in-depth understanding of the factors and mechanisms regulating ESC pluripotency and cell lineage decisions.

Early studies on ESCs led to the discovery of the core pluripotency factors Oct4, Sox2 and Nanog [1], and, increasingly, the use of genome-level screening assays has revealed new insights by uncovering additional transcription factors, transcriptional cofactors and chromatin remodeling complexes involved in the maintenance of pluripotency [1]. The study of ESC transcriptional regulation is also useful in the understanding of human diseases. ESCs, for instance, are known to share certain cellular and molecular signatures similar to those of cancer cells [2], and deregulation of ESC-associated transcriptional regulators has been implicated in many human developmental diseases.

Despite the promising potential, the use of human ESCs (hESCs) in clinical applications has been slow because of ethical, immunological and tumorigenicity concerns [3]. These ethical and immunogenicity issues were seemingly overcome by the creation of induced pluripotent stem cells (iPSCs), whereby exogenous expression of Oct4, Sox2, Klf4 and c-Myc in differentiated cells could revert them to pluripotency [4]. However, the question of whether these iPSCs truly resemble ESCs is still actively debated and remains unresolved [5]. Nevertheless, the application of iPSCs as an *in vitro* human genetic disease model has been successful in revealing novel molecular disease pathologies, as well as facilitating genetic or drug screenings [6].

In this review, we describe recent advances in understanding the ESC and iPSC transcriptional network, and also discuss how deregulation of ESC pathways is implicated in human diseases. Finally, we address how the knowledge gained through transcriptional studies of ESCs and iPSCs has impacted translational medicine.

*Correspondence: nghh@gis.a-star.edu.sg

¹Gene Regulation Laboratory, Genome Institute of Singapore, 60 Biopolis Street, Genome, Singapore 138672

Full list of author information is available at the end of the article

Transcriptomic approaches for studying stem cells

The transcriptome is the universe of expressed transcripts within a cell at a particular state [7]; and understanding the ESC transcriptome is key towards appreciating the mechanism behind the genetic regulation of pluripotency and differentiation. The methods used to study gene expression patterns can be classified into two groups: (1) those using hybridization-based approaches, and (2) those using sequencing-based approaches (Table 1).

For hybridization-based methods, the commonly used 'DNA microarray' technique relies on hybridization between expressed transcripts and microarray printed oligonucleotide (oligo) probes from annotated gene regions [7]. In addition to allowing the identification of highly expressed genes, microarrays also enable the study of gene expression changes under various conditions. However, microarrays have their limitations, whereby prior knowledge of genomic sequences is required, and cross-hybridization of oligo probes may lead to false identification [7]. Subsequently, later versions of microarrays were modified to include exon-spanning probes for alternative-spliced isoforms, as well as 'tiling arrays,' which comprise oligo probes spanning large genomic regions to allow for the accurate mapping of gene transcripts [7,8]. Indeed, conventional microarrays and tiling arrays have been instrumental in advancing our understanding of ESC transcriptional regulation (Table 1) through the mapping of ESC-associated transcription-factor binding sites (chromatin immunoprecipitation (ChIP)-chip) [9,10], identification of microRNA (miRNA) regulation in ESCs [11], as well as the identification of long non-coding RNA (lncRNA) [12] and long intergenic non-coding RNA (lincRNA) [13,14].

Sequence-based transcriptomic analysis on the other hand involves direct sequencing of the cDNA. Initially, Sanger sequencing techniques were used to sequence gene transcripts, but these methods were considered expensive and low throughput [7]. However, with the development of next-generation sequencing (NGS), such as the 454, Illumina and SOLiD platforms, it is now possible to perform affordable and rapid sequencing of massive genomic information [8]. Importantly, NGS when coupled with transcriptome sequencing (RNA-seq) offers high-resolution mapping and high-throughput transcriptome data, revealing new insights into transcriptional events such as alternative splicing, cancer fusion-genes and non-coding RNAs (ncRNAs). This versatility of NGS for ESC research is evident through its various applications (Table 1), such as chromatin immunoprecipitation coupled to sequencing (ChIP-seq) [15], methylated DNA immunoprecipitation coupled to sequencing (DIP-seq) [16], identification of long-range chromatin interactions [17], miRNA profiling [18], and RNA-binding protein immunoprecipitation coupled to sequencing (RIP-seq) [19].

Table 1. Transcriptomic approaches for studying stem cells

Objective	Method	Reference
DNA sequencing	NGS	[65]
mRNA expression analysis	Microarray	[58]
	RNA-seq	[93]
miRNA expression analysis	Microarray	[11]
	RNA-seq	[18]
lncRNA expression analysis	Microarray	[12,13]
Identification of alternative splicing isoforms	Microarray	[22,94]
	RNA-seq	[95]
Mapping of protein-DNA binding	ChIP-chip	[9,10,24]
	ChIP-PET	[23]
	ChIP-seq	[15]
DNA methylation profiling	BS-seq	[68]
	MethylC-seq	[68]
	DIP-seq	[16]
Mapping of long-range chromatin interactions	ChIA-PET	[17]
	3C	[29]
Identification of RNA-protein interactions	RIP-seq	[19]
	RIP and direct RNA quantification	[14]

BS-seq, bisulfite sequencing; ChIA-PET, chromatin interaction analysis with paired-end tag sequencing; ChIP-chip, chromatin immunoprecipitation on chip; ChIP-PET, chromatin immunoprecipitation with paired-end tag sequencing; ChIP-seq, chromatin immunoprecipitation and sequencing; DIP-seq, DNA immunoprecipitation and sequencing; MethylC-seq, methylcytosine sequencing; NGS, next-generation sequencing; RIP, RNA-binding protein immunoprecipitation; RIP-seq, RNA-binding protein immunoprecipitation and sequencing; RNA-seq, RNA sequencing; 3C, chromosome conformation capture.

Transcriptomics has been instrumental in the study of alternative splicing events. It has been suggested that around 95% of all multi-exon human genes undergo alternative splicing to generate different protein variants for an assortment of cellular processes [20], and that alternative splicing contributes to higher eukaryotic complexity [21]. In mouse ESCs (mESCs) undergoing embryoid body formation, exon-spanning microarrays have identified possible alternative splicing events in genes associated with pluripotency, lineage specification and cell-cycle regulation [22]. More interestingly, it was found that alternative splicing of the *Serca2b* gene during ESC differentiation resulted in a shorter *Serca2a* isoform with missing miR-200 targeting sites in its 3'-UTR. Given that miR-200 is highly expressed in cardiac lineages, and that *Serca2a* protein is essential for cardiac function, the results suggest that during mESC differentiation some genes may utilize alternative splicing to bypass lineage-specific miRNA silencing [22]. With the largely uncharacterized nature of alternative splicing in ESCs, and the availability of high-throughput sequencing tools, it would be of interest to further dissect these pathways.

Transcriptional networks controlling ESCs

The core transcriptional regulatory network

In ESCs, the undifferentiated state is maintained by the core transcription factors Oct4, Sox2 and Nanog [1]. Early mapping studies revealed that Oct4, Sox2 and Nanog co-bind gene promoters of many mESC and hESC genes [23,24]. Importantly, the core transcription factors were found to maintain pluripotency by: (1) activating other pluripotency factors, while simultaneously repressing lineage-specific genes via Polycomb group proteins; and (2) activating their own gene expression, as well as that of each other. Therefore, with this autoregulatory and feed-forward system, Oct4, Sox2 and Nanog constitute the ESC core transcriptional network (Figure 1) [23,24]. Subsequent studies on additional ESC-related

transcription factors using ChIP-based transcriptomics led to the discovery of transcription factors associating into an 'Oct4' or 'Myc' module [10,15].

The expanded pluripotency network

Apart from Oct4, Sox2 and Nanog, the Oct4 module also includes the downstream transcription factors of the LIF, BMP4 and Wnt signaling pathways: Stat3, Smad1 and Tcf3 [15,25]. Indeed, Stat3, Smad1 and Tcf3 co-occupy certain regulatory regions with Oct4, Sox2 and Nanog, thus establishing the pathway in which external signaling can affect ESC transcriptional regulation [15,25]. Mass spectrometry has also facilitated the study of protein-protein interaction networks of core transcription factors [26,27], revealing that Oct4 can interact with a diverse

population of proteins, including transcriptional regulators, chromatin-binding proteins and modifiers, protein-modifying factors, and chromatin assembly proteins. Importantly, knockdown of Oct4 protein levels is known to cause the loss of co-binding activity of other transcription factors [15,27], suggesting that Oct4 serves as a platform for the binding of its interacting protein partners onto their target genes.

The Myc module consists of transcription factors such as c-Myc, n-Myc, Zfx, E2f1 and Rex1, and is associated with self-renewal and cellular metabolism [10,15]. Approximately one-third of all active genes in ESCs are bound by both c-Myc and the core transcription factors [28]. However, unlike Oct4, Sox2 and Nanog, which can recruit RNA polymerase II via coactivators such as the Mediator complex [29], c-Myc rather appears to control the transcriptional pause release of RNA polymerase II, via recruitment of a cyclin-dependent kinase, p-TEFb [28]. It is therefore proposed that Oct4-Sox2-Nanog selects ESC genes for expression by recruiting RNA polymerase II, while c-Myc serves to regulate gene expression efficiency by releasing transcriptional pause [1]. This may thus account for the reason why overexpression of c-Myc is able to improve the efficiency of iPSC generation, and how c-Myc could be oncogenic. In fact, the Myc module rather than the Oct4 module in ESCs was recently found to be active in various cancers, and may serve as a useful tool in predicting cancer prognosis [9].

Besides targeting transcription factors to regulate gene expression, Oct4 is also known to affect the ESC chromatin landscape. Jarid2 [30-34] and Pcl2/Mtf2 [30,31,34-35] have been identified as components of the Polycomb Repressive Complex 2 (PRC2) in ESCs, and regulated by the core ESC transcription factors [10,15]. From these studies, Jarid2 is suggested to recruit PRC2 to its genomic targets, and can also control PRC2 histone methyltransferase activity [30-34]. The second protein Pcl2 shares a subset of PRC2 targets in ESCs [34-35] and appears to promote histone H3 lysine 27 trimethylation [35]. Knockdown of Pcl2 promotes self-renewal and impairs differentiation, suggesting a repressive function of Pcl2 by suppressing the pluripotency-associated factors Tbx3, Klf4 and Foxd3 [35]. Oct4 has also been demonstrated to physically interact with Wdr5, a core member of the mammalian Trithorax complex, and cooperate in the transcriptional activation of self-renewal genes [36]. As Wdr5 is needed for histone H3 lysine 4 trimethylation (H3K4me3), Oct4 depletion notably caused a decrease in both Wdr5 binding and H3K4me3 levels at Oct4-Wdr5 co-bound promoters. This indicates that Oct4 may be responsible for directing Wdr5 to ESC genes and maintaining H3K4me3 open chromatin [36]. As chromatin structure and transcriptional activity can be altered via addition or removal of histone modifications [37], the

ability of Oct4, Sox2 and Nanog to regulate histone modifications expands our understanding of how the core transcriptional factors regulate chromatin structure to ultimately promote a pluripotent state.

Pluripotent transcription factor regulation of non-coding RNA

ncRNAs are a diverse group of transcripts, and are classified into two groups: (a) lncRNAs for sequences more than 200 nucleotides in length; and (b) short ncRNAs for transcripts of less than 200 bases [38].

miRNAs that are about 22 nucleotides in length are considered to be short ncRNAs. In ESCs, miRNA expression is also regulated by the core transcription factors (Figure 1), whereby the promoters of miRNA genes, which are preferentially expressed in ESCs, are bound by Oct4, Sox2, Nanog and Tcf3 factors. Similarly, miRNA genes involved in lineage specification were occupied by core transcription factors in conjunction with Polycomb group proteins, to exert transcriptional silencing [39]. Examples of these silenced miRNA genes include *let-7*, which targets pluripotency factors Lin28 and Sall4 [11], as well as miR-145, which is expressed during hESC differentiation to suppress the pluripotency factors OCT4, SOX2 and KLF4 in hESCs [40].

The lncRNA *Xist*, which performs a critical role in X-chromosome inactivation, is silenced by the core ESC factors along intron 1 of the mESC *Xist* gene (Figure 1) [41]. Similarly, ESC transcription factors also regulate the expression of the *Xist* antisense gene *Tsix* [42,43]. However, it was found that deletion of *Xist* intron 1 containing the Oct4-binding sites in ESCs did not result in *Xist* derepression [44]. Epiblast-derived stem cells and hESCs that express Oct4 are known to possess an inactive X-chromosome [45], and interestingly, pre-X inactivation hESCs have been derived from human blastocysts cultured under hypoxic conditions [46]. Therefore, it is likely that the ESC transcriptional network indirectly regulates X-chromosome activation status via an intermediary effector.

Recently, lincRNAs have been demonstrated to both maintain pluripotency and suppress lineage specification, hence integrating into the molecular circuitry governing ESCs [14]. Pluripotency factors such as Oct4, Sox2, Nanog and c-Myc have also been found to co-localize at lincRNA promoters, indicating that lincRNA expression is under the direct regulation of the ESC transcriptional network. Interestingly, mESC lincRNAs have been found to bind multiple ubiquitous chromatin complexes and RNA-binding proteins, leading to the proposal that lincRNAs function as 'flexible scaffolds' to recruit different protein complexes into larger units. By extension of this concept, it is possible that the unique lincRNA signature of each cell type may serve to bind protein

complexes to create a cell-type-specific gene expression profile.

Cellular reprogramming and iPSCs

The importance of the transcriptional regulatory network in establishing ESC self-renewal and pluripotency was elegantly demonstrated by Takahashi and Yamanaka [4], whereby introduction of four transcription factors Oct4, Sox2, Klf4 and c-Myc (OSKM) could revert differentiated cells back to pluripotency as iPSCs. iPSCs were later demonstrated to satisfy the highest stringency test of pluripotency via tetraploid complementation to form viable 'all-iPSC' mice [47].

However, reprogramming is not restricted to the four OSKM factors only. Closely related family members of the classical reprogramming factors such as Klf2 and Klf5 can replace Klf4, Sox1 can substitute for Sox2, and c-Myc can be replaced by using N-myc and L-myc [48]. However, Oct4 cannot be replaced by its close homologs Oct1 and Oct6 [48], but can be substituted using an unrelated orphan nuclear receptor, Nr5a2, to form mouse iPSCs [49]. Similarly, another orphan nuclear receptor, Esrrb, was demonstrated to replace Klf4 during iPSC generation [50]. Human iPSCs (hiPSCs), aside from the classical OSKM factors [51], can also be generated using a different cocktail of factors comprising OCT4, SOX2, NANOG and LIN28 [52]. Recently, the maternally expressed transcription factor Glis1 replaced c-Myc to generate both mouse iPSCs and hiPSCs [53]. Glis1 is highly expressed in unfertilized eggs and zygotes but not in ESCs; thus, it remains to be determined if other maternally expressed genes could similarly reinstate pluripotency.

While certain transcription factors may be replaced with chemicals during the reprogramming process, they all still require at least one transcription factor [54]. Recently, however, the creation of hiPSCs and mouse iPSCs via miRNA without additional protein-encoding factors was reported [55,56]. By expressing the miR-302-miR-367 clusters, iPSCs can be generated with two orders of magnitude higher efficiency compared with conventional OSKM reprogramming [55]. Similarly, iPSCs could be formed by transfecting miR-302, miR-200 and miR-369 into mouse adipose stromal cells, albeit at lower efficiency [56]. The ability of miRNAs to reprogram somatic cells is intriguing, and it would be of great interest to determine the gene targets of these reprogramming miRNAs.

Expression profiling of ESCs and iPSCs

The question of whether pluripotent iPSCs truly resemble ESCs is an actively debated and evolving field, with evidence arguing both for and against iPSC-ESC similarity. As such, further research using better controlled

studies is needed to resolve this issue. Here, we summarize and present the key findings that address this topic.

Initially, it was believed that hiPSCs were similar to hESCs [52,57], but subsequent studies argued otherwise as differential gene expression [58], as well as DNA methylation patterns [59], could be distinguished between hiPSCs and hESCs (Table 2). However, these differences were proposed to be a consequence of comparing cells of different genetic origins [60], laboratory-to-laboratory variation [61], and the iPSC passage number [62]. Later, hiPSCs were described to contain genomic abnormalities, including gene copy number variation [63,64], point mutations [65] and chromosomal duplications [66] (Table 2). However, whether these genomic instabilities are inherent in hiPSCs only, or a consequence of culture-induced mutations, as previously described in hESCs, is still not certain [67]. Extended passages of iPSCs appeared to reduce such aberrant genomic abnormalities, possibly via growth outcompetition by healthy iPSCs [64], but this was contradicted by a separate study that found that parental epigenetic signatures are retained in iPSCs even after extended passaging [68]. Indeed, this 'epigenetic memory' phenomenon was also reported in two earlier studies, whereby donor cell epigenetic memory led to an iPSC differentiation bias towards donor-cell-related lineages [62,69]. The mechanism behind this residual donor cell memory found in iPSCs was attributed to incomplete promoter DNA methylation [70]. Surprisingly, knockdown of incompletely reprogrammed somatic genes was found to reduce hiPSC generation, suggesting that somatic memory genes may play an active role in the reprogramming process [70]. Differences in ncRNA expression were also found between iPSCs and ESCs (Table 2). For instance, the aberrantly silenced imprinted *Dlk1-Dio3* gene locus in iPSCs results in the differential expression of its encoded ncRNA *Gtl2* and *Rian*, and 26 miRNAs, and consequent failure to generate 'all-iPSC' mice [60]. Upregulation of lincRNAs specifically in hiPSCs was also reported [13]. Expression of *lincRNA-RoR* with OSKM could also enhance iPSC formation by twofold, suggesting a critical function of lincRNA in the reprogramming process [13].

As these reported variations between hESCs and hiPSCs could be attributed to small sample sizes, a recent large-scale study by Bock *et al.* [71] profiled the global transcription and DNA methylation patterns of 20 different hESC lines and 12 hiPSC lines. Importantly, the study revealed that hiPSCs and hESCs were largely similar, and that the observed hiPSC differences were similar to normally occurring variation among hESCs. Additionally, Bock *et al.* established a scoring algorithm to predict lineage and differentiation propensity of hiPSCs. As traditional methods of screening hiPSC quality rely on time-consuming and low-throughput teratoma assays,

Table 2. Transcriptomic comparisons between induced pluripotent stem cells and embryonic stem cells

Characteristic	Mouse iPSCs	Human iPSCs
mRNA expression	Distinct from mESCs at lower passages, donor cell gene expression still present [62,69]; closely resemble mESCs at late passages [62]	Distinct from hESCs at lower passages [58], with residual donor gene expression [70,96,97]; closely resemble hESCs at late passages [58,98]
miRNA expression	miRNA encoded within the imprinted <i>Dlk1-Dio3</i> locus is aberrantly silenced [60]	Small number of differences reported [58,99], but variation between hESCs and hiPSCs comparable to somatic and cancer cells [100]
lncRNA expression	Not determined	Differences in lncRNA expression reported. <i>lincRNA-RoR</i> enhances reprogramming by twofold [13]
DNA methylation status	Distinct from mESCs at lower passages, donor cell DNA methylation pattern still present [62,69]; closely identical to mESCs at late passages [62]	Differences in DNA methylation reported [59,68,70], but not in all hiPSCs [71]
Genome status	Not determined	Possess gene copy number deletions and duplications [63-64], somatic coding mutations [65], and chromosomal duplications [66]

hESC, human embryonic stem cell; hiPSC, human induced pluripotent stem cell; iPSC, induced pluripotent stem cell; lincRNA, long-intergenic non-coding RNA; lncRNA, long non-coding RNA; mESC, mouse embryonic stem cell; miRNA, microRNA.

the hiPSC genetic scorecard offers researchers a quick assessment of the epigenetic and transcriptional status of pluripotent cells. This may be especially useful for the rapid monitoring of cell-line quality during large-scale production of iPSCs [71].

Deregulation of transcriptional networks in disease

Blastocyst-derived ESCs possess an innate ability for indefinite self-renewal, and can be considered a primary untransformed cell line. Unlike primary cell cultures with limited *in vitro* lifespans, or immortalized/tumor-derived cell lines that do not mimic normal cell behavior, ESCs thus offer a good model for studying cellular pathways. ESC transcriptomics have indeed advanced our understanding into the molecular mechanisms affecting certain human diseases.

For instance, it was previously reported that cancer cells possess an ESC-like transcriptional program, suggesting that ESC-associated genes may contribute to tumor formation [72]. However, this expression signature was shown to be a result of c-Myc, rather than from the core pluripotency factors (Table 3) [9]. As c-Myc somatic copy-number duplications are the most frequent in cancer [73], the finding that c-Myc releases RNA polymerase II from transcriptional pause [28] offers new understanding into the transcriptional regulatory role of c-Myc in ESCs and cancer cells. Another pluripotency-associated factor, Lin28, which suppresses the maturation of pro-differentiation *let-7* miRNA, is also highly expressed in poorly differentiated and low prognosis tumors [74]. Importantly, *let-7* silences several oncogenes, such as *c-Myc*, *K-Ras*, *Hmga2* and the gene encoding cyclin-D1, suggesting that Lin28 deregulation may promote oncogenesis [74].

Aside from cancer, mutations in ESC-associated transcriptional regulators can cause developmental

abnormalities. The Mediator-cohesin complex, which occupies 60% of active mESC genes, is responsible for regulating gene expression by physically linking gene enhancers to promoters through chromatin loops [29]. Notably, the binding pattern of Mediator-cohesin onto gene promoters differs among cell types, indicative of cell-type-specific gene regulation [29]. In hESCs, Mediator was also revealed to be important in the maintenance of pluripotent stem cell identity during a genome-wide siRNA screen, suggesting an evolutionarily conserved role [75]. Given this important gene regulatory function of the Mediator-cohesin complex in mESCs and hESCs, mutations in these proteins are associated with disorders such as schizophrenia, and Opitz-Kaveggia and Lujan syndromes [29]. Interestingly, the Cornelia de Lange syndrome, which causes mental retardation due to gene dysregulation rather than chromosomal abnormalities, is associated with mutations in cohesin-loading factor Nipbl [29]. Therefore, it is proposed that such developmental syndromes may arise as a result of the failure to form appropriate enhancer-promoter interactions.

Mutations in core ESC transcription factor SOX2 and the ATP-chromatin remodeler CHD7 result in developmental defects such as SOX2 anophthalmia (congenital absence of eyeballs) and CHARGE syndrome, respectively [76]. Although a direct association between CHARGE syndrome and ESCs is not known, mESC studies revealed that Chd7 co-localizes with core ESC factors and p300 protein at gene enhancers to modulate expression of ESC-specific genes [77]. It is thus possible that CHARGE syndrome may arise due to CHD7 enhancer-mediated gene dysregulation. In neural stem cells, Chd7 is able to bind with Sox2 at the *Jag1*, *Gli3* and *Mycn* genes, which are mutated in the developmental disorders Alagille, Pallister-Hall and Feingold syndromes [78]. Similarly, Chd7 has been described to interact with the PBAF

Table 3. Dysregulation of transcriptional networks in stem cells and disease

Gene/protein	Role in ESCs	Role in disease
<i>c-MYC</i>	Involved in the expression of self-renewal genes [101]; recruits p-TEFb to initiate transcriptional pause release of RNA polymerase II [29]	Most common gene duplication in cancer [73]; c-Myc appears to be responsible for the gene expression signature of cancer cells [9]
<i>LIN28</i>	Maintains ESC pluripotency by binding and inhibiting the maturation of pro-differentiation <i>let-7</i> miRNA; <i>LIN28</i> is also a hiPSC reprogramming factor [74]	Highly expressed in poorly differentiated and low prognosis tumors; as <i>let-7</i> silences the expression of oncogenes <i>c-Myc</i> , <i>K-Ras</i> , <i>Hmga2</i> and the gene encoding cyclin-D1, <i>LIN28</i> suppression of <i>let-7</i> miRNA may thus promote oncogenesis [74]
<i>SOX2</i>	A core ESC transcription factor together with Oct4 and Nanog. Regulates the expression of pluripotency genes, and suppresses lineage-specific genes [23,24]; Sox2 is also an iPSC reprogramming factor [4]	Mutation in <i>SOX2</i> causes anophthalmia (congenital loss of eyeballs) in humans. Proposed to cooperate with <i>CHD7</i> to regulate genes involved in Alagille, Pallister-Hall and Feingold syndromes [76]
<i>CHD7</i>	Binds with core ESC factors and p300 at gene enhancers to modulate ESC-specific gene expression [77]	Mutations in <i>CHD7</i> result in CHARGE syndrome; proposed to cooperate with <i>SOX2</i> to regulate genes involved in Alagille, Pallister-Hall and Feingold syndromes [76]
Mediator	Physically links the Oct4/Sox2/Nanog-bound gene enhancers to active gene promoters via chromatin looping [29]; necessary for normal gene activity	Mutations in Mediator are associated with Opitz-Kaveggia, Lujan, and transposition of the great arteries syndromes; also implicated in schizophrenia, colon cancer progression [1] and uterine leiomyomas [102]
Cohesin	Proposed to bind and stabilize the Oct4/Sox2/Nanog enhancer-promoter chromatin loops [1]; necessary for normal gene activity	Cohesin mutations implicated in Cornelia de Lange syndrome, whereby patients exhibit developmental defects and mental retardation due to dysregulation of gene expression [29]
Nipbl	Binds with mediator complex to allow loading of cohesin and formation of stable chromatin loop [29]	Nipbl mutations implicated in Cornelia de Lange syndrome, whereby patients exhibit developmental defects and mental retardation due to dysregulation of gene expression [29]

ESC, embryonic stem cell; hiPSC, human induced pluripotent stem cell; iPSC, induced pluripotent stem cell; miRNA, microRNA.

complex to control neural crest formation [79]. Therefore, these data hint that *Chd7* may partner different proteins to cooperatively regulate developmental genes. Although the mechanism behind gene regulation by *Chd7* and its interacting partners is not well understood, the use of ESCs may serve as a useful system to further probe *Chd7* function during development and disease.

Clinical and therapeutic implications

The development of hiPSC technology offers the unique opportunity to derive disease-specific hiPSCs for the *in vitro* study of human disease pathogenesis (Figure 2). A major advantage of using disease-specific hiPSCs is that they allow the capture of the patient's genetic background and, together with the patient's medical history, will enable the researcher to uncover the disease genotypic-phenotypic relationship [6]. A number of patient-derived hiPSC disease models have been established, including those for Hutchinson Gilford Progeria, Timothy syndrome, schizophrenia and Alzheimer's disease [5,80-83], and these have been useful in understanding the cellular mechanisms behind these illnesses. For example, transcriptional profiling of schizophrenia neurons derived from iPSCs have identified 596 differentially expressed genes, 75% of which were not previously implicated in schizophrenia [82]. This highlights the potential of disease-specific iPSCs in unlocking hidden pathways. Additionally, the use of disease cell lines can facilitate

drug design and screening under disease conditions (Figure 2) [6]. One such example is the drug roscovitine, which was found to restore the electrical and Ca^{2+} signaling in Timothy syndrome cardiomyocytes [81].

The self-renewing ability of hiPSCs means that a potentially unlimited source of patient-specific cells can be generated for regenerative purposes (Figure 2). Importantly, hiPSCs, when coupled with gene targeting approaches to rectify genetic mutations, can be differentiated into the desired cell type and reintroduced to the patient (Figure 2) [5]. However, unlike mESCs, hESCs and hiPSCs cannot be passaged as single cells and have very poor homologous recombination ability [84]. Circumventing this problem may require the conversion of hiPSCs into a mESC-like state, which is more amenable to gene targeting [85]. Alternatively, recent reports of successful gene targeting in human pluripotent stem cells using zinc-finger nucleases (ZFNs) [86], and transcription activator-like effector nucleases (TALENs) [87], presents another option for genetically altering hiPSCs for cell therapy. Albeit that there are concerns of off-target effects, the advantage of using nuclease-targeting approaches is that they do not necessitate the conversion of hESCs and hiPSCs into mESC-like states prior to genomic manipulation.

While it has been assumed that iPSCs generated from an autologous host should be immune-tolerated, Zhao *et al.* [88] recently demonstrated that iPSCs were immunogenic

and could elicit a T-cell immune response when transplanted into syngeneic mice. However, it should be distinguished that in the Zhao *et al.* study undifferentiated iPSCs were injected into mice, rather than differentiated iPSC-derived cells, which are the clinically relevant cell type for medical purposes. Furthermore, the immune system is capable of 'cancer immunosurveillance' to identify and destroy tumorigenic cells [89]. Hence, it may be possible that the observed iPSC immunogenicity could have arisen through cancer immunosurveillance against undifferentiated tumor-like iPSCs, and that

iPSC-derived differentiated cells may not be immunogenic. It would thus be necessary to experimentally verify if iPSC-derived differentiated cells are immunogenic in syngeneic hosts.

Conclusions and future challenges

Understanding and exploiting the mechanisms that govern pluripotency are necessary if hESCs and hiPSCs are to be successfully translated to benefit clinical and medical applications. One approach for understanding hESCs and hiPSCs would be to study their transcriptomes,

and, through various approaches, we have learnt how the core pluripotency factors create an ESC gene expression signature by regulating other transcription factors and controlling chromatin structure and ncRNA expression.

Current methodologies to generate iPSCs are inefficient, suggesting that significant and unknown epigenetic barriers to successful reprogramming remain [90]. However, defining these barriers is difficult, as existing transcriptomic studies rely on average readings taken across a heterogeneous cell population. This therefore masks essential rate-limiting transcriptional and epigenetic remodeling steps in iPSC formation. Future studies in elucidating the iPSC generation process may thus adopt a single-cell approach [91], which will offer the resolution needed to define key reprogramming steps. Future efforts should also be focused upon improving hiPSC safety for human applications, through the use of stringent genomic and functional screening strategies on hiPSCs and their differentiated tissues [3]. Only with well-defined and non-tumorigenic iPSC-derived tissue would we then be able to assess the transplant potential of iPSCs in personalized medicine.

In addition to generating disease-specific iPSCs from patients, the use of gene-modifying nucleases to create hESCs harboring specific genetic mutations may be a forward approach towards studying human disease pathogenesis [86]. With the recent creation of approximately 9,000 conditional targeted alleles in mESCs [92], it would be of tremendous scientific and clinical value to likewise establish a hESC knockout library to study the role of individual genes in disease and development. Furthermore, while SNP and haplotype mapping may be useful in associating diseases with specific genetic loci, the use of ZFNs or TALENs to recreate these specific gene variations in hESCs may offer an experimental means of verifying the relationship of SNPs or haplotypes with diseases.

Abbreviations

CHARGE, Coloboma of the eye, Heart defects, Atresia of the choanae, Retardation of growth and/or development, Genital and/or urinary abnormalities, and Ear abnormalities and deafness; ChIP, chromatin immunoprecipitation; ChIP-chip, chromatin immunoprecipitation on chip; ChIP-seq, chromatin immunoprecipitation and sequencing; DIP-seq, DNA immunoprecipitation and sequencing; ESC, embryonic stem cell; hESC, human embryonic stem cell; hiPSC, human induced pluripotent stem cell; H3K4me3, histone H3 lysine 4 trimethylation; iPSC, induced pluripotent stem cell; lincRNA, long intergenic non-coding RNA; lncRNA, long non-coding RNA; mESC, mouse embryonic stem cell; miRNA, microRNA; NGS, next-generation sequencing; ncRNA, non-coding RNA; oligo, oligonucleotide; OSKM, Oct4, Sox2, Klf4 and c-Myc; PRC2, Polycomb repressive complex 2; RIP-seq, RNA-binding protein immunoprecipitation and sequencing; RNA-seq, RNA sequencing; siRNA, short interfering RNA; SNP, single nucleotide polymorphism; TALEN, transcription activator-like effector nuclease; UTR, untranslated region; ZFN, zinc-finger nuclease.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors thank all the members of the Ng laboratory for their comments on this manuscript.

Author details

¹Gene Regulation Laboratory, Genome Institute of Singapore, 60 Biopolis Street, Genome, Singapore 138672. ²School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551. ³Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597. ⁴Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117597. ⁵NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, 28 Medical Drive, Singapore 117456.

Published: 27 October 2011

References

- Young RA: **Control of the embryonic stem cell state.** *Cell* 2011, **144**:940-954.
- Ben-David U, Benvenisty N: **The tumorigenicity of human embryonic and induced pluripotent stem cells.** *Nat Rev Cancer* 2011, **11**:268-277.
- Barrilleaux B, Knoepfler PS: **Inducing iPSCs to escape the dish.** *Cell Stem Cell* 2011, **9**:103-111.
- Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.** *Cell* 2006, **126**:663-676.
- Wu SM, Hochedlinger K: **Harnessing the potential of induced pluripotent stem cells for regenerative medicine.** *Nat Cell Biol* 2011, **13**:497-505.
- Zhu H, Lensch MW, Cahan P, Daley GQ: **Investigating monogenic and complex diseases with pluripotent stem cells.** *Nat Rev Genet* 2011, **12**:266-275.
- Wang Z, Gerstein M, Snyder M: **RNA-seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
- Marguerat S, Bahler J: **RNA-seq: from technology to biology.** *Cell Mol Life Sci* 2010, **67**:569-579.
- Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, Cantor AB, Orkin SH: **A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs.** *Cell* 2010, **143**:313-324.
- Kim J, Chu J, Shen X, Wang J, Orkin SH: **An extended transcriptional network for pluripotency of embryonic stem cells.** *Cell* 2008, **132**:1049-1061.
- Melton C, Judson RL, Blalock R: **Opposing microRNA families regulate self-renewal in mouse embryonic stem cells.** *Nature* 2010, **463**:621-626.
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS: **Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation.** *Genome Res* 2008, **18**:1433-1445.
- Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL: **Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells.** *Nat Genet* 2010, **42**:1113-1117.
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES: **lincRNAs act in the circuitry controlling pluripotency and differentiation.** *Nature* 2011, **477**:295-300.
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH: **Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.** *Cell* 2008, **133**:1106-1117.
- Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappaport J, Helin K: **TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.** *Nature* 2011, **473**:343-348.
- Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan CS, Kurnarso G, Shahab A, Bourque G, Cacheux-Rataboul V, Sung WK, Ruan Y, Wei CL: **CTCF-mediated functional chromatin interactome in pluripotent cells.** *Nat Genet* 2011, **43**:630-638.
- Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.** *Genome Res* 2008, **18**:610-621.
- Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE,

- Borowsky M, Lee JT: **Genome-wide identification of polycomb-associated RNAs by RIP-seq.** *Mol Cell* 2010, **40**:939-953.
20. Chen M, Manley JL: **Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.** *Nat Rev Mol Cell Biol* 2009, **10**:741-754.
21. Johnson JM, Castle J, Garrett-Engel P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD: **Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.** *Science* 2003, **302**:2141-2144.
22. Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zamboni AC, Vranizan K, Spindler MJ, Pico AR, Cline MS, Clark TA, Williams A, Blume JE, Samal E, Mercola M, Merrill BJ, Conklin BR: **Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation.** *Proc Natl Acad Sci U S A* 2010, **107**:10514-10519.
23. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nat Genet* 2006, **38**:431-440.
24. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA: **Core transcriptional regulatory circuitry in human embryonic stem cells.** *Cell* 2005, **122**:947-956.
25. Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA: **Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells.** *Genes Dev* 2008, **22**:746-755.
26. Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J: **An expanded Oct4 interaction network: implications for stem cell biology, development, and disease.** *Cell Stem Cell* 2010, **6**:382-395.
27. van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA: **An Oct4-centered protein interaction network in embryonic stem cells.** *Cell Stem Cell* 2010, **6**:369-381.
28. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA: **c-Myc regulates transcriptional pause release.** *Cell* 2010, **141**:432-445.
29. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467**:430-435.
30. Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jorgensen HF, Pereira CF, Leleu M, Piccolo FM, Spivakov M, Brookes E, Pombo A, Fisher C, Skarnes WC, Snoek T, Bezstarosti K, Demmers J, Klose RJ, Casanova M, Tavares L, Brockdorff N, Merckenschlager M, Fisher AG: **Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators.** *Nat Cell Biol* 2010, **12**:618-624.
31. Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D: **Jarid2 and PRC2, partners in regulating gene expression.** *Genes Dev* 2010, **24**:368-380.
32. Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappasilber J, Helin K: **JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells.** *Nature* 2010, **464**:306-310.
33. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J: **Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells.** *Cell* 2009, **139**:1290-1302.
34. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH: **Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells.** *Cell* 2009, **139**:1303-1314.
35. Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K, Torchia J, Krogan NJ, Reiter JF, Stanford WL: **Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation.** *Cell Stem Cell* 2010, **6**:153-166.
36. Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, Wang J, Rendl M, Bernstein E, Schaniel C, Lemischka IR: **Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network.** *Cell* 2011, **145**:183-197.
37. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M: **Open chromatin in pluripotency and reprogramming.** *Nat Rev Mol Cell Biol* 2011, **12**:36-47.
38. Pauli A, Rinn JL, Schier AF: **Non-coding RNAs as regulators of embryogenesis.** *Nat Rev Genet* 2011, **12**:136-149.
39. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, Sharp PA, Bartel DP, Jaenisch R, Young RA: **Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.** *Cell* 2008, **134**:521-533.
40. Xu N, Papagiannakopoulos T, Pan G, Thomson JA, Kosik KS: **MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.** *Cell* 2009, **137**:647-658.
41. Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P: **Molecular coupling of Xist regulation and pluripotency.** *Science* 2008, **321**:1693-1695.
42. Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT: **The pluripotency factor Oct4 interacts with Ctf and also controls X-chromosome pairing and counting.** *Nature* 2009, **460**:128-132.
43. Navarro P, Oldfield A, Legoupi J, Festuccia N, Dubois A, Attia M, Schoorlemmer J, Rougeulle C, Chambers I, Avner P: **Molecular coupling of Tsix regulation and pluripotency.** *Nature* 2010, **468**:457-460.
44. Barakat TS, Gunhanlar N, Pardo CG, Achame EM, Ghazvini M, Boers R, Kenter A, Rentmeester E, Grootegoed JA, Gribnau J: **RNF12 activates Xist and is essential for X chromosome inactivation.** *PLoS Genet* 2011, **7**:e1002001.
45. Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A: **Klf4 reverts developmentally programmed restriction of ground state pluripotency.** *Development* 2009, **136**:1063-1069.
46. Lengner CJ, Gimelbrant AA, Erwin JA, Cheng AW, Guenther MG, Welstead GG, Alagappan R, Frampton GM, Xu P, Muffat J, Santagata S, Powers D, Barrett CB, Young RA, Lee JT, Jaenisch R, Mitalipova M: **Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations.** *Cell* 2010, **141**:872-883.
47. Zhao XY, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Guo CL, Ma QW, Wang L, Zeng F, Zhou Q: **iPS cells produce viable mice through tetraploid complementation.** *Nature* 2009, **461**:86-90.
48. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochizuki Y, Takizawa N, Yamanaka S: **Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.** *Nat Biotechnol* 2008, **26**:101-106.
49. Heng JC, Feng B, Han J, Jiang J, Kraus P, Ng JH, Orlov YL, Huss M, Yang L, Lufkin T, Lim B, Ng HH: **The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells.** *Cell Stem Cell* 2010, **6**:167-174.
50. Feng B, Jiang J, Kraus P, Ng JH, Heng JC, Chan YS, Yaw LP, Zhang W, Loh YH, Han J, Vega VB, Cacheux-Rataboul V, Lim B, Lufkin T, Ng HH: **Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb.** *Nat Cell Biol* 2009, **11**:197-203.
51. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors.** *Cell* 2007, **131**:861-872.
52. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA: **Induced pluripotent stem cell lines derived from human somatic cells.** *Science* 2007, **318**:1917-1920.
53. Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, Kawamura Y, Mochizuki H, Goshima N, Yamanaka S: **Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1.** *Nature* 2011, **474**:225-229.
54. Gonzalez F, Boue S, Izpisua Belmonte JC: **Methods for making induced pluripotent stem cells: reprogramming a la carte.** *Nat Rev Genet* 2011, **12**:231-242.
55. Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, Morrissey EE: **Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.** *Cell Stem Cell* 2011, **8**:376-388.
56. Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, Nishikawa S, Tanemura M, Mimori K, Tanaka F, Saito T, Nishimura J, Takemasa I, Mizushima T, Ikeda M, Yamamoto H, Sekimoto M, Doki Y, Mori M: **Reprogramming of mouse and human cells to pluripotency using mature microRNAs.** *Cell Stem Cell* 2011, **8**:633-638.
57. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ: **Reprogramming of human somatic cells to pluripotency with defined factors.** *Nature* 2008, **451**:141-146.
58. Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavan N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegriani M, Plath K, Lowry WE: **Induced pluripotent stem cells and embryonic stem**

- cells are distinguished by gene expression signatures. *Cell Stem Cell* 2009, **5**:111-123.
59. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41**:1350-1353.
 60. Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K: **Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells.** *Nature* 2010, **465**:175-181.
 61. Newman AM, Cooper JB: **Lab-specific gene expression signatures in pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:258-262.
 62. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K: **Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.** *Nat Biotechnol* 2010, **28**:848-855.
 63. Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, Lynch C, Harness JV, Lee S, Barrero MJ, Ku S, Martynova M, Semechkin R, Galat V, Gottesfeld J, Izpisua Belmonte JC, Murry C, Keirstead HS, Park HS, Schmidt U, Laslett AL, Muller FJ, Nievergelt CM, Shamir R, Loring JF: **Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture.** *Cell Stem Cell* 2011, **8**:106-118.
 64. Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Närvä E, Ng S, Sourour M, Hämmäläinen R, Olsson C, Lundin K, Mikkola M, Trokovic R, Peitz M, Brüstle O, Bazett-Jones DP, Alitalo K, Lahesmaa R, Nagy A, Otonkoski T: **Copy number variation and selection during reprogramming to pluripotency.** *Nature* 2011, **471**:58-62.
 65. Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, Canto I, Giorgetti A, Israel MA, Kiskinis E, Lee JH, Loh YH, Manos PD, Montserrat N, Panopoulos AD, Ruiz S, Wilbert ML, Yu J, Kirkness EF, Izpisua Belmonte JC, Rossi DJ, Thomson JA, Eggan K, Daley GQ, Goldstein LS, Zhang K: **Somatic coding mutations in human induced pluripotent stem cells.** *Nature* 2011, **471**:63-67.
 66. Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, Plath K, Lowry WE, Benvenisty N: **Identification and classification of chromosomal aberrations in human induced pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:521-531.
 67. Maitra A, Arking DE, Shivapurkar N, Ikeda M, Stastny V, Kassaei K, Sui G, Cutler DJ, Liu Y, Brimble SN, Noaksson K, Hyllner J, Schulz TC, Zeng X, Freed WJ, Crook J, Abraham S, Colman A, Sartipy P, Matsui S, Carpenter M, Gazdar AF, Rao M, Chakravarti A: **Genomic alterations in cultured human embryonic stem cells.** *Nat Genet* 2005, **37**:1099-1103.
 68. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR: **Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.** *Nature* 2011, **471**:68-73.
 69. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI, Yabuuchi A, Takeuchi A, Cuniff KC, Hongguang H, McKinney-Freeman S, Naveiras O, Yoon TJ, Irizarry RA, Jung N, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, Daley GQ: **Epigenetic memory in induced pluripotent stem cells.** *Nature* 2010, **467**:285-290.
 70. Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M, Hochedlinger K, Costello JF, Song JS, Ramalho-Santos M: **Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPSCs.** *Nat Cell Biol* 2011, **13**:541-549.
 71. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K, Meissner A: **Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines.** *Cell* 2011, **144**:439-452.
 72. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Reggev A, Weinberg RA: **An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors.** *Nat Genet* 2008, **40**:499-507.
 73. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, McHenry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, et al.: **The landscape of somatic copy-number alteration across human cancers.** *Nature* 2010, **463**:899-905.
 74. Viswanathan SR, Daley GQ: **Lin28: a microRNA regulator with a macro role.** *Cell* 2010, **140**:445-449.
 75. Chia NY, Chan YS, Feng B, Lu X, Orlov YL, Moreau D, Kumar P, Yang L, Jiang J, Lau MS, Huss M, Soh BS, Kraus P, Li P, Lufkin T, Lim B, Clarke ND, Bard F, Ng HH: **A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity.** *Nature* 2010, **468**:316-320.
 76. Puc J, Rosenfeld MG: **SOX2 and CHD7 cooperatively regulate human disease genes.** *Nat Genet* 2011, **43**:505-506.
 77. Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, Adams DJ, Flicek P, Crawford GE, Laframboise T, Tesar P, Wei CL, Scacheri PC: **CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression.** *PLoS Genet* 2010, **6**:e1001023.
 78. Engelen E, Akinci U, Bryne JC, Hou J, Gontan C, Moen M, Szumska D, Kockx C, van Ijcken W, Dekkers DH, Demmers J, Rijkers EJ, Bhattacharya S, Philipsen S, Pevny LH, Grosveld FG, Rottier RJ, Lenhard B, Poot RA: **Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes.** *Nat Genet* 2011, **43**:607-611.
 79. Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, Helms J, Chang CP, Zhao Y, Swigut T, Wysocka J: **CHD7 cooperates with PBAF to control multipotent neural crest formation.** *Nature* 2010, **463**:958-962.
 80. Liu GH, Barkho BZ, Ruiz S, Diep D, Qu J, Yang SL, Panopoulos AD, Suzuki K, Kurian L, Walsh C, Thompson J, Boue S, Fung HL, Sancho-Martinez I, Zhang K, Yates J 3rd, Izpisua Belmonte JC: **Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome.** *Nature* 2011, **472**:221-225.
 81. Yazawa M, Hsueh B, Jia X, Pasca AM, Bernstein JA, Hallmayer J, Dolmetsch RE: **Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome.** *Nature* 2011, **471**:230-234.
 82. Brennan KJ, Simone A, Jou J, Gelboin-Burkhardt C, Tran N, Sangar S, Li Y, Mu Y, Chen G, Yu D, McCarthy S, Sebat J, Gage FH: **Modelling schizophrenia using human induced pluripotent stem cells.** *Nature* 2011, **473**:221-225.
 83. Qiang L, Fujita R, Yamashita T, Angulo S, Rhinn H, Rhee D, Doege C, Chau L, Aubry L, Vanti WB, Moreno H, Abeliovich A: **Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons.** *Cell* 2011, **146**:359-371.
 84. Zwaka TP, Thomson JA: **Homologous recombination in human embryonic stem cells.** *Nat Biotechnol* 2003, **21**:319-321.
 85. Buecker C, Chen HH, Polo JM, Daheron L, Bu L, Barakat TS, Okwieka P, Porter A, Gribnau J, Hochedlinger K, Geijsen N: **A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells.** *Cell Stem Cell* 2010, **6**:535-546.
 86. Soldner F, Laganière J, Cheng AW, Hockemeyer D, Gao Q, Alagappan R, Khurana V, Golbe LI, Myers RH, Lindquist S, Zhang L, Guschin D, Fong LK, Vu BJ, Meng X, Urnov FD, Rebar EJ, Gregory PD, Zhang HS, Jaenisch R: **Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations.** *Cell* 2011, **146**:318-331.
 87. Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassidy JP, Cost GJ, Zhang L, Santiago Y, Miller JC, Zeitler B, Cherone JM, Meng X, Hinkley SJ, Rebar EJ, Gregory PD, Urnov FD, Jaenisch R: **Genetic engineering of human pluripotent cells using TALE nucleases.** *Nat Biotechnol* 2011, **29**:731-734.
 88. Zhao T, Zhang ZN, Rong Z, Xu Y: **Immunogenicity of induced pluripotent stem cells.** *Nature* 2011, **474**:212-215.
 89. Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ: **Natural innate and adaptive immunity to cancer.** *Annu Rev Immunol* 2011, **29**:235-271.
 90. Plath K, Lowry WE: **Progress in understanding reprogramming to the induced pluripotent state.** *Nat Rev Genet* 2011, **12**:253-265.
 91. Smith ZD, Nachman I, Reggev A, Meissner A: **Dynamic single-cell imaging of direct reprogramming reveals an early specifying event.** *Nat Biotechnol* 2010, **28**:521-526.
 92. Skarnes WC, Rosen B, West AP, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A: **A conditional knockout resource for the genome-wide study of mouse gene function.** *Nature* 2011, **474**:337-342.
 93. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S: **Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq.** *Genome Res* 2011, **21**:1160-1167.
 94. Yeo GW, Xu X, Liang TY, Muotri AR, Carson CT, Coufal NG, Gage FH: **Alternative splicing events identified in human embryonic stem cells and neural progenitors.** *PLoS Comput Biol* 2007, **3**:1951-1967.
 95. Wu JQ, Habegger L, Noisa P, Szekely A, Qiu C, Hutchison S, Raha D, Egholm M, Lin H, Weissman S, Cui W, Gerstein M, Snyder M: **Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by**

- short, long, and paired-end sequencing. *Proc Natl Acad Sci U S A* 2010, **107**:5254-5259.
96. Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR: **Transcriptional signature and memory retention of human-induced pluripotent stem cells.** *PLoS One* 2009, **4**:e7076.
 97. Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC: **Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells.** *PLoS One* 2010, **5**:e8975.
 98. Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R, Young RA: **Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells.** *Cell Stem Cell* 2010, **7**:249-257.
 99. Wilson KD, Venkatasubrahmanyam S, Jia F, Sun N, Butte AJ, Wu JC: **MicroRNA profiling of human-induced pluripotent stem cells.** *Stem Cells Dev* 2009, **18**:749-758.
 100. Neveu P, Kye MJ, Qi S, Buchholz DE, Clegg DO, Sahin M, Park IH, Kim KS, Daley GQ, Kornblum HI, Shraiman BI, Kosik KS: **MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states.** *Cell Stem Cell* 2010, **7**:671-681.
 101. Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K: **Role of the murine reprogramming factors in the induction of pluripotency.** *Cell* 2009, **136**:364-377.
 102. Mäkinen N, Mehine M, Tolvanen J, Kaasinen E, Li Y, Lehtonen HJ, Gentile M, Yan J, Enge M, Taipale M, Vahteristo P, Aaltonen LA: **MED12, the Mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas.** *Science* 2011.

doi:10.1186/gm284

Cite this article as: Yeo J-C, Ng H-H: Transcriptomic analysis of pluripotent stem cells: insights into health and disease. *Genome Medicine* 2011, **3**:68.