

THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Functional analysis of the Nanog tryptophan repeat

Matúš Vojtek

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Matúš Vojtek

Lay summary

Embryonic stem cells (ESCs) have a potential to generate any tissue of the developing embryo including germ cells which end products are eggs in female and sperms in male. ESCs are special in their ability to maintain their identity after cell division for many generations which makes them virtually immortal. This capacity is termed as self-renewal.

When ESCs specialise to a specific cell type, for example a muscle cell, self-renewal and the potential to form any tissue of a developing organism are lost. The function of a particular cell is determined by activity of specific genes encoded in DNA. Different cell types contain different sets of molecules regulating gene activity. These molecules work together in order to switch specific genes on or off. A gene is turned on when a specialised enzyme is recruited to this gene and initiate transcription. This enzyme is called RNA polymerase II (RNAPII).

Nanog is one of the main regulatory molecules in ESCs. Nanog secures the self-renewal of ESCs and plays an important role in their specialisation into germ cells. Nanog contains a tryptophan repeat region (WR) which is required for Nanog function. Nanog regulates specific genes together with other molecules like Oct4 or Sox2. Oct4, Sox2 and Nanog bind DNA close to these genes and regulate their activity through RNAPII. However, how they do it and what is the role of the Nanog WR in this process remains unclear. Here, mechanism of the Nanog function and the role of the WR in this process was examined.

Nanog interacts with a handful of proteins trough WR. One of them is uncharacterised protein Resf1. Requirements of Resf1 for Nanog function and ESC self-renewal were examined by removing Resf1 gene from the mouse ESCs. Resf1 is shown not to be essential for Nanog function or ESC self-renewal. However, Resf1 has been shown to be important for germ cell specification.

Moreover, the capacity of Nanog to influence binding of Oct4, Sox2 and RNAPII was analysed. NERT cells which keep Nanog away from the DNA were used. After a drug treatment, Nanog moves quickly to the nucleus and is able to regulate its gene targets. Oct4, Sox2 and RNAPII binding before and after Nanog release was compared. The analysis revealed two distinct behaviours associated with gene activation or gene repression. The sites where Nanog increases Oct4, Sox2 and RNAPII binding are often close to the genes activated by Nanog whereas the sites where Nanog decreases RNAPII DNA binding associate with repressed genes.

Furthermore, the effects of WR disruption on Nanog capacity to bind DNA and on Oct4, Sox2 and RNAPII binding were examined. Nanog, Oct4, Sox2 and RNAPII DNA binding in NERT cells after activation of a Nanog mutant with a disrupted WR was examined. When the Nanog WR is disrupted, Nanog cannot bind DNA very well and is unable to support Oct4, Sox2 and RNAPII binding. This prevents Nanog to activate its gene targets efficiently. The WR disruption also prevents Nanog to repress genes efficiently which is linked with increased Oct4 binding.

In summary, this thesis proposes that Nanog activates expression by recruiting the core regulatory complex to the regulated genes and represses gene expression by decreasing RNAPII occupancy at associated sites. The WR is essential for proper direction of the core regulatory complex. Disruption of the WR leads to gene misreguation and loss of the self-renewal. In addition, the WR-interacting protein Resf1 is not essential for Nanog function and ESC self-renewal but it is important for germ cell specification. These results expand our understanding of the molecular mechanism underlying Nanog function and the role of Resf1 in ESC self-renewal and germ cell specification.

Abstract

The homeodomain transcription factor NANOG promotes self-renewal of embryonic stem cells (ESCs) and specification of primordial germ cells (PGCs). NANOG binds thousands of sites within the ESC genome together with OCT4, SOX2 and RNA polymerase II (RNAPII) regulating transcription of its target genes. Replacing tryptophans within the NANOG tryptophan repeat (WR) region for alanines (W10A) causes NANOG to lose its capacity to homodimerise and to interact with other proteins including RNAPII, SOX2 and RESF1. $NANOG_{W10A}$ mutant is also unable to sustain LIF-independent self-renewal. However, the requirements of the WR to interact with other proteins for NANOG function remains unclear.

To characterise the cellular localisation of RESF1, endogenously tagged *Resf1* mouse embryonic stem cell lines were produced. RESF1 is a widely expressed nuclear protein which interacts with NANOG in nuclei of ESCs. To quantify the requirements of *Resf1* for ESC self-renewal and NANOG function, *Resf1*$^{-/-}$ ES cell lines were derived. In serum/LIF culture, *Resf1* deletion promotes cell differentiation and decreases expression of *Pou5f1, Nanog* and *Esrrb*. Episomal expression of NANOG sustains LIF-independent self-renewal independently of *Resf1*. *Resf1*$^{-/-}$ ESCs are also capable to differentiate into epiblast stem cells. To investigate the role of *Resf1* in PGC specification, *Resf1*$^{-/-}$ ESCs were differentiated into PGC-like cells (PGCLCs). Deletion of *Resf1* decreased PGCLC specification 4-fold and prevented efficient activation of *Blimp1, Prdm14* and *Ap2*γ.

The effect of NANOG on OCT4, SOX2 and RNAPII chromatin binding was evaluated by inducing NANOG in *Nanog*$^{-/-}$ mouse ESCs. OCT4, SOX2 and RNAPII occupancy before and shortly after NANOG induction was measured by analysing chromatin immunoprecipitation followed by sequencing data sets. The k-mean clustering of Nanog binding sites revealed seven clusters with different response patterns of OCT4, SOX2 and RNAPII to NANOG. The sites where NANOG increases OCT4, SOX2 and RNAPII

chromatin binding are mostly associate with the genes activated by NANOG, whereas sites where RNAPII binding decreases after NANOG induction associate with genes repressed by NANOG. To assess the requirements of the NANOG WR and the NANOG DNA binding for directing OCT4, SOX2 and RNAPII chromatin binding, NANOG variants with debilitating mutations in the WR or the homeodomain were induced in *Nanog*$^{-/-}$ ESCs and chromatin binding of NANOG, OCT4, SOX2 and RNAPII was assessed. The $NANOG_{W10A}$ has reduced chromatin binding at most of the NANOG binding sites. A reduced DNA binding capacity of recombinant $NANOG_{W10A}$ was also suggested by electrophoretic mobility shift assay. Moreover, $NANOG_{W10A}$ is unable to increase OCT4, SOX2 and RNAPII binding at the sites associated with gene activation. At the sites associated with gene repression, $NANOG_{W10A}$ induction leads to increased OCT4 binding. NANOG variant incapable to bind DNA ($NANOG_{N51A}$) is unable to direct OCT4 and RNAPII binding. However, SOX2 binding temporarily decreases after $NANOG_{N51A}$ induction.

In summary, *Resf1* is not essential for ESC self-renewal and NANOG function but it is important for efficient specification of PGCLCs. NANOG directs the core transcription complex regulating self-renewal via the WR. This thesis contributes to understanding of NANOG function in ESC self-renewal and describes the function of previously uncharacterised RESF1 in ESC self-renewal and germ cell specification.

Acknowledgements

I am grateful to my supervisor Prof. Ian Chambers for giving me the opportunity to carry out interesting research projects in his lab. I thank Ian for invaluable feedback and support during my studies. Ian gave me the opportunity to freely explore my scientific interests and to develop as an independent scientist. His support to attend scientific conferences, courses and meetings benefited me greatly.

I would like to express my gratitude to Nick Mullin for being always available to answer my questions, guide me and help me. I thank Nick for teaching me many experimental techniques and for constructive discussions which helped me to conduct my research projects.

I am grateful to Dougie Colby for teaching me tissue culture techniques and for invaluable technical support. I thank Dougie for creating a friendly environment in the tissue culture facility.

I am grateful to Man Zhang for supporting me as a scientist, being my mentor and a friend. I am indebted to Man for introducing me to the PGC field and providing me with opportunities to contribute to exciting research projects.

I also thank previous lab members: Raphaël Pantier, Tulin Tatar, Jingchao Zhang and Andrea Corsinotti for creating friendly and supportive environment in the lab. It was my pleasure and privilege to work with you all.

I thank Dr. Abdenour Soufi and Prof. Steve Pollard for their feedback and suggestions as members of my PhD committee.

I am grateful to Dr. Simon Tomlinson who gave me access to computational resources. I also thank James Ashmore for guidance and feedback on my data analysis.

I acknowledge the assistance provided by the core facilities within our institute. Matthew Vermeren supported me with confocal microscopy and image analysis. Clair Cryer and Fiona Rossi supported me with FACS sorting and flow cytometry.

I would like to thank and acknowledge people who contributed to the work present in this thesis. Nick Mullin and Jingchao Zhang performed NANOG, SOX2 and RNAPII ChIP-seq experiments. Nick Mullin provided me with recombinant proteins for the EMSA experiments. Alison McGarvery analysed the microarray data and provided me with the expression tables and pre-processed aforementioned ChIP-seq data sets.

Finally, I am entirely grateful to my family for being always there for me and for believing in me. I would have not achieved what I have without their love and support.

Abbreviations

2i	two inhibitors, PD0325901 and CHIR99021
ALKs	activin-like receptor kinases
AP	alkaline phosphatase
BMP	bone morphogenic protein
bp	base pair
ChIP-seq	chromatin immunoprecipitation followed by sequencing
CK1	casein kinase
CRISPR	clustered regulatory interspaced short palindromic repeats
DAPI	4,6-diamidino-2-phenylindole
ddH2O	double distilled water
DEGs	differentially expressed genes
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucletide triphosphates
dsDNA	double stranded deoxyribonucleotic acid
DTT	dithiotreitol
E	embryonic day
EDTA	ethylenediaminetetraacetate
EMSA	electrophoretic mobility shift assay
EPI	epiblast
EpiLCs	epiblast-like cells
EpiSCs	epiblast stem cells
ERVs	endogenous retroviruses
ESC	embryonic stem cells
EV	empty vector
ExE	extra embryonic endoderm
FACS	fluorescent activated cell sorting
FC	fold change
FCS	fetal calf serum
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor

GET	gonad expressed transcript
GFP	green fluorescent protein
GRN	gene regulatory network
gRNA	guide RNA
GSK3	glycogen synthase kinase 3
HD	homeodomain
HMG	high mobility group
ICM	inner cell mass
IRES	internal ribosomal entry site
kb	kilo base
Kd	dissociation constant
kDa	kilo daltons
KO	knock out
LIF	leukemia inhibitory factor
Mbp	mega base pair
mRNA	messenger RNA
MSCV	murine stem cell virus
MZT	maternal to zygotic transition
N51A	substitution of asparagin 51 for alanine within a homeodomain
NC	negative control
NERT	fusion protein of NANOG and ER^{T2}
PAM	protospacer adjacent motif
ParE	parietal endoderm
PBS	phosphate buffer saline
PBS-T	phosphate buffer saline supplemented with Tween-20
PCR	polymerase chain reaction
PGCLCs	primordial germ cell-like cells
PGCs	primordial germ cells
POU	short of Pit, Oct, Unc
PrE	primitive endoderm
PURO	puromycin resistance cassette
RESF1	retroelement silencing factor 1

RNA	ribonucleic acid
RNAPII	RNA polymerase II
rpm	rotation per minute
RPM	reads per million
RT-qPCR	reverse transcription followed by quantitative polymerase chain reaction
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TAD	topologically associating domain
Tam	4-hydroxytamoxifen
TE	trophectoderm
TF	transcription factor
TFAS	transcription association score
TSS	transcriptional start site
U	units of activity
VE	visceral endoderm
W10A	10 tryptophans within the WR substituted for alanines
WR	NANOG tryptophan repeat
WT	wild type

Table of Contents

Lay summary I
Abstract III
Acknowledgements V
Abbreviations VII
List of Figures XV
List of Tables XVII
1 Introduction 1
1.1 Emergence of pluripotency . . . 1
1.2 Specification of primordial germ cells . . . 5
1.3 Pluripotent states of the epiblast . . . 7
1.3.1 Naïve pluripotency . . . 7
1.3.1.1 "Serum/LIF" culture condition . . . 7
1.3.1.2 "2i/LIF" culture condition . . . 8
1.3.2 Primed pluripotency . . . 9
1.3.2.1 Formative pluripotency . . . 9
1.4 Extrinsic stimuli regulating pluripotency . . . 10
1.4.1 LIF signalling . . . 10
1.4.2 TGF-β signalling . . . 11
1.4.3 FGF signalling . . . 12
1.4.4 WNT/GSK3 signalling . . . 13
1.5 Gene regulation in embryonic stem cells . . . 14
1.5.1 Oct4 and Sox2 . . . 17
1.5.2 Nanog . . . 20
1.5.2.1 Molecular features of NANOG . . . 20
1.5.2.2 NANOG function in mouse development . . . 22
1.5.2.3 Model of NANOG function . . . 22
1.5.3 Resf1 . . . 24
1.6 Aims of the thesis . . . 26
2 Materials and methods 27

2.1 Materials and reagents 27
2.2 Routine culture of mouse ESCs 33
2.2.1 Cell thawing 33
2.2.2 Cell freezing 33
2.3 Colony forming assay 34
2.4 Alkaline phosphatase staining 35
2.5 Differentiation of ESCs 35
2.5.1 EpiSC differentiation 35
2.5.2 EpiLC differentiation 36
2.5.3 PGCLC differentiation 37
2.6 Lipofection of embryonic stem cells 37
2.7 Fluorescent activated cell sorting (FACS) 38
2.8 Analysis of PGCLCs by flow cytometry 38
2.9 Derivation of clonal ESC lines 39
2.9.1 Generation of clonal *Resf1*$^{-/-}$ ESC lines 39
2.9.2 Generation of clonal ESC lines with endogenously tagged *Resf1* 39
2.10 Molecular cloning 40
2.10.1 Polymerase chain reaction (PCR) 40
2.10.2 Gibson assembly 43
2.10.3 DNA ligation 43
2.10.4 Transformation of bacterial cells 43
2.10.5 Cloning of *Resf1* coding sequence 44
2.10.6 Cloning of CRISPR/eSpCas9 plasmids 45
2.10.7 Sanger sequencing 46
2.10.8 Cloning of a targeting vector for endogenous tagging of *Resf1* 47
2.11 Protein analysis 47
2.11.1 Chromatin immunoprecipitation 47
2.11.2 Preparation of ChIP-seq libraries 49
2.11.3 Co-immunoprecipitation 50

2.11.4 Protein electrophoresis and western blotting 51
2.11.5 Immunostaining 52
2.11.6 Bradford assay 53
2.11.7 Electrophoretic mobility shift assay (EMSA) 53
2.12 Quantitative PCR of reverse transcribed RNA (RT-qPCR) 54
2.13 Computational analysis 56
2.13.1 ChIP-seq analysis 56
2.13.2 Identification of reproducible NANOG binding sites 57
2.13.3 Analysis of DNAseI data set 58
2.13.4 K-means clustering of NANOG binding sites 58
2.13.5 Calculation of association between NANOG peaks and differentially expressed genes 59
3 Requirements of *Resf1* for stem cell identity **61**
3.1 RESF1 localisation in ESCs 61
3.2 Requirements of *Resf1* for ESC self-renewal 63
3.2.1 Deletion of *Resf1* in ESCs 63
3.2.2 Effects of *Resf1* deletion on expression of pluripotency genes 67
3.2.3 Quantification of self-renewal capacity of *Resf1*$^{-/-}$ ESCs . 71
3.2.4 Effects of forced *Resf1* expression on ESC self-renewal . 71
3.3 Requirements of *Resf1* for NANOG function 73
3.4 Requirement of *Resf1* for transition from naïve to primed pluripotency 77
3.5 Discussion 82
3.5.1 *Resf1* has a minor effect on ESC self-renewal 82
3.5.2 RESF1 is not required for NANOG to sustain LIF-independent self-renewal 84
3.5.3 *Resf1* is not required for primed pluripotency 85
3.5.4 Conclusion 85
4 RESF1 function in specification of primordial germ cells **87**
4.1 Effects of *Resf1* deletion on PGCLC specification 87

4.2 Discussion 92
5 NANOG directs the core transcriptional network to control gene expression 95
5.1 Clustering of NANOG binding sites reveals distinct patterns of OCT4, SOX2 and RNAPII response to NANOG induction 97
5.2 Specific responses to NANOG induction associate with gene activation and gene repression 101
5.3 Discussion 105
6 NANOG requires WR to efficiently direct the core regulatory factors and regulate gene expression 111
6.1 $NANOG_{W10A}$ has impaired chromatin binding 114
6.2 $NANOG_{W10A}$ fails to increase OCT4, SOX2 and RNAPII binding at cis-regulatory elements 116
6.3 $NANOG_{W10A}$ fails to regulate its gene targets efficiently 122
6.4 Discussion 122
7 Concluding remarks and future directions 127
7.1 *Resf1* function in mouse development 127
7.2 The molecular model of NANOG function 129
References 135

List of Figures

1.1 Scheme of early mouse embryogenesis 2
1.2 BMP4 induced specification of primordial germ cells 6
1.3 OCT4 and SOX2 protein structures and OCT:SOX DNA binding motif 19
1.4 NANOG protein structure and DNA binding motif 21

3.1 Episomally expressed RESF1 localises predominantly to nucleus of ESCs 62
3.2 Tagging of endogenous RESF1 64
3.3 RESF1 localises to nuclei of ESCs 65
3.4 NANOG and RESF1 expression do not correlate 66
3.5 Creation of *Resf1*$^{-/-}$ ES cell lines 68
3.6 Sequencing of *Resf1*$^{-/-}$ cell lines 69
3.7 *Resf1*$^{-/-}$ ESCs cultured in serum/LIF express lower levels of pluripotency markers 70
3.8 *Resf1*$^{-/-}$ ESCs have decreased capacity to self-renew 72
3.9 Forced RESF1 expression cannot sustain LIF-independent self-renewal of ESCs 74
3.10 NANOG interacts with RESF1 in ESCs 75
3.11 Creation of *Resf1*$^{-/-}$ ESCs stably expressing NANOG transgene . 78
3.12 RESF1 is not required for NANOG function in ESC self-renewal . 79
3.13 *Resf1* is not required for "naïve" pluripotency 80
3.14 *Resf1*$^{-/-}$ EpiSCs have decreased *Otx2* expression 81

4.1 A scheme of PGCLC differentiation protocol 88
4.2 *Resf1*$^{-/-}$ EpiLCs express wild type levels of *Fgf5*, *Pou3f1* and *Otx2* 89
4.3 *Resf1* is required for efficient PGCLC specification 90
4.4 *Resf1* is required for activation of *Blimp1* during PGCLC specification 91

5.1 Summary of the analysed data sets 96
5.2 OCT4, SOX2 and RNAPII response differently to NANOG induction at specific NANOG binding sites 98
5.3 NANOG increases OCT4, SOX2 and RNAPII binding at distal regulatory elements 99
5.4 Distinct clusters of NANOG binding sites associate with gene activation or gene repression 102
5.5 NANOG increases binding of OCT4, SOX2 and RNAPII at the *Esrrb* locus 103
5.6 Changes of OCT4, SOX2 and RNAPII binding at the *Otx2* locus after NANOG induction 104
5.7 NANOG gene targets have distinct distribution patterns of NANOG peaks 106

6.1 Clonal variation between NERT ES cell lines 112
6.2 OCT4, SOX2 and NANOG levels in $NERT_{WT}$, $NERT_{W10A}$ and $NERT_{N51A}$ ESCs 113
6.3 The WR is required for efficient NANOG chromatin binding . . . 115
6.4 $NANOG_{W10A}$ has decreased affinity to DNA 117
6.5 The WR is required for a specific OCT4 recruitment to NANOG binding sites 118
6.6 The NANOG WR is required for efficient SOX2 recruitment to the NANOG binding sites 119
6.7 Mutation of the WR prevents efficient recruitment of RNAPII by NANOG to specific NANOG binding sites 120
6.8 $NANOG_{W10A}$ fails to regulate its gene targets efficiently 123

7.1 A proposed model of NANOG function in gene regulation of ESC self-renewal 131

List of Tables

2.1 Cell culture media and solutions 27
2.2 Cell culture flasks, dishes and plates 30
2.3 Mouse embryonic stem cell lines 30
2.4 Used buffers 31
2.5 Amount of the solutions used for AP staining 35
2.6 Lipofection of ESCs 37
2.7 Polymerase chain reaction protocol 40
2.8 Primers used for genotyping and molecular cloning 41
2.9 Oligonucleotides used for gRNA cloning 46
2.10 Recommended amounts of DNA for Sanger sequencing 46
2.11 PCR program for Sanger sequencing 47
2.12 Antibodies used for Western blotting 52
2.13 Antibodies used for immunostaining 53
2.14 Primers used for RT-qPCR 55
2.15 Summary of used sequencing-based data sets. 56

1 Introduction

Species survival depends on the propagation of the genome to the next generations. Sexually reproducing organisms propagate through highly specialised haploid cells called gametes. At fertilisation, the egg and sperm are unified into a zygote which initiates a cascade of events leading to formation of a new individual carrying the germ line essential for further propagation (Cinalli et al. 2008). How can a single cell give rise to a complex organism? This question has fascinated people since ancient times (Clift et al. 2013). Ernest Haeckel first termed a cell with the capacity to form a new organism as a stem cell in 1868 (Haeckel 1868). Today, stem cells are defined by their capacity to differentiate into more specialised cells and by the capacity to self-renew (Ramalho-Santos et al. 2007). The potential of stem cells to form any cell type (totipotency) is reduced during the development until it is lost by terminal differentiation. Understanding stem cell biology allows us to capture and manipulate various cell fates which provides not only a great tool for studying developmental and molecular processes but it also opens avenues to novel medical treatments.

1.1 Emergence of pluripotency

In the first three days after the fertilisation, the zygote is transformed into a totipotent embryo which undergoes two rounds of cell specification, preparing the embryo for implantation into the uterine wall (Fig. 1.1) (Frum et al. 2015).

Upon fertilisation, the oocyte completes the second meiosis and transition into the first mitotic division (Clift et al. 2013). The two formed daughter cells, blastomeres, undergo a series of cell divisions without a cell growth. As the genome is initially untranscribed, this process is govern by messenger RNAs (mRNAs) and proteins inherited from the egg while the embryonic genome reactivates (Schulz et al. 2018). In mice, the maternal-to-zygotic transition occurs in two waves. The minor wave finishes before the first

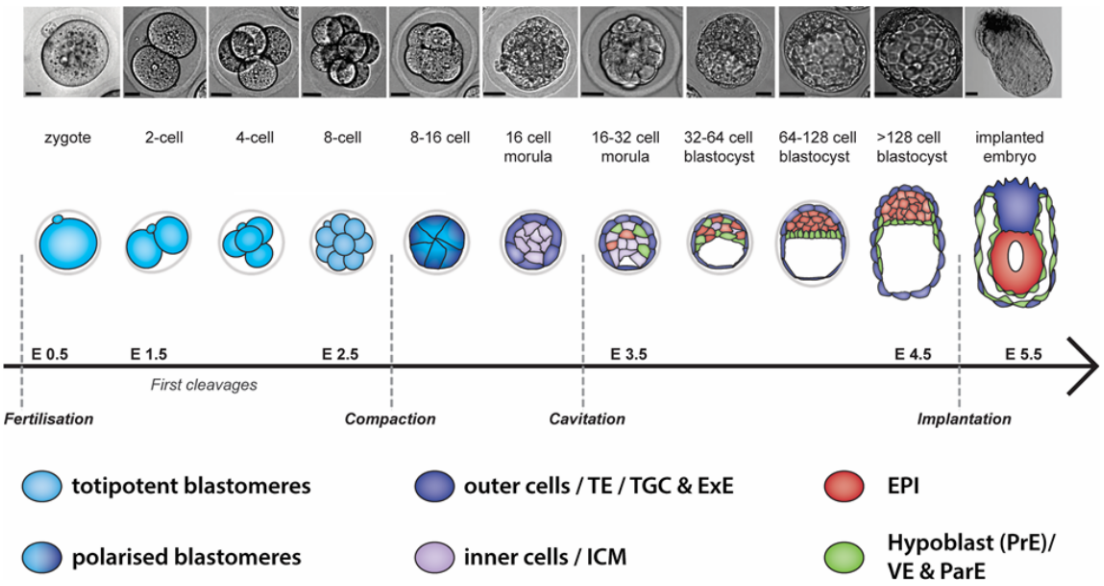


Figure 1.1: Scheme of early mouse embryogenesis. Schematic representation and microphotographies of the major events during the first 5.5 embryonic days (E) of the mouse embryogenesis. TE - trophectoderm, TGC - trophoblast giant cells, PrE - primitive endoderm, EPI - epiblast, ICM - inner cell mass, VE - visceral endoderm, ParE - parietal endoderm. Adapted from Piliszek et al.

division. The major wave finishes at the late two-cell stage. After this point the gene expression is driven solely by the forming embryo (Rossant et al. 2017; Schulz et al. 2018). The maternal-to-zygotic transition is accompanied by gene activation (Hamatani et al. 2004) including activation of multiple classes of transposable elements (Evsikov et al. 2004; Peaston et al. 2004), genome-wide genome demethylation (Smallwood et al. 2011; Smith et al. 2012) and chromatin remodelling (Wu et al. 2016; Lu et al. 2016; Eckersley-Maslin et al. 2018) presumably resetting the genome to totipotency (Rossant et al. 2017).

The blastomeres undergo a series of symmetric and asymmetric cell divisions resulting in a cell compaction and polarisation of the outer layer at the 16-cell stage forming a morula (Fig. 1.1) (Johnson et al. 1981). The differences in the cell polarity and cell-cell junctions between the inner and the outer cells result in a position-dependent activation of the HIPPO signalling pathway (Frum et al. 2015) which leads to restriction of *Sox2* expression to the inner cells (Wicklow et al. 2014) and *Cdx2* expression to the outer cells (Nishioka et al. 2009). This process ultimately segregates the inner cell mass (ICM) and the trophectoderm (Fig. 1.1).

At the 32-cell morula, water influx to the embryo creates a fluid-filled cavity called blastocoel marking the formation of a blastocyst and initiation of the second cell specification (Fig. 1.1). Similar to the first lineage separation, asymmetric cell signalling initiates separation of the epiblast and the primitive endoderm which is enforced by a specific transcriptional program (Fig. 1.1). In this case, FGF/ERK pathway drives epiblast and primitive endoderm segregation (Chazaud et al. 2006; Kang et al. 2013; Yamanaka et al. 2010; Nichols et al. 2009b).

Before the blastocyst formation, ICM expresses both epiblast marker *Nanog* and primitive endoderm marker *Gata6* ubiquitously (Kang et al. 2013; Yamanaka et al. 2010). Most likely the combination of limiting amount of FGF4 and stochastic changes in the gene expression lead to "salt and pepper" expression patterns of NANOG and GATA6 (Simon et al. 2018). The resolution

of primitive endoderm and epiblast is a gradual process revolving around antagonist effects of *Nanog* and *Gata6*. Downregulation of either *Gata6* or *Nanog* below a certain threshold shifts the gene regulatory network (GRN) towards a specific lineage (Chickarmane et al. 2008; Schröter et al. 2015). The FGF/ERK signalling is crucial in this process, communicating intracellular state to the neighbouring cells. NANOG activates *Fgf4* expression which stimulates FGF receptor 2 (FGFR2) on neighbouring cells leading to NANOG downregulation and activation of GATA6 (Frankenberg et al. 2011). Lack of FGF stimulation or *Gata6* deletion prevents primitive endoderm formation and cells collapse into epiblast-like state (Schrode et al. 2014; Bessonnard et al. 2014; Yamanaka et al. 2010; Chazaud et al. 2006; Nichols et al. 2009b; Arman et al. 1998; Cheng et al. 1998; Rappolee et al. 1994; Wilder et al. 1997). On the other hand, the formation and maintenance of epiblast relies on *Oct4, Sox2* and *Nanog* (Nichols et al. 1998; Avilion et al. 2003; Mitsui et al. 2003; Silva et al. 2009). Resolved epiblast and primitive endoderm then undergo cell sorting separating epithelial layer of primitive endoderm (also called hypoblast) and epiblast preparing the blastocyst for the implantation (Fig. 1.1). The proper segregation of the hypoblast and the epiblast is enforced by elimination of the wrongly positioned cells (Plusa et al. 2008).

During implantation, the blastocyst transforms into an elongated cup-like structure called the "egg cylinder" by outgrowth of trophectoderm and epiblast cavitation creating the proamniotic cavity (Fig. 1.1). The trophectoderm becomes extra-embryonic endoderm which together with the visceral endoderm (derived from the primitive endoderm) supports epiblast specification (Bedzhov et al. 2014). The epiblast is adjacent to the extraembryonic endoderm (Fig. 1.1) which is required for instructing primitive streak formation, anterior-posterior patterning and germ cell induction (Rossant et al. 2017).

1.2 Specification of primordial germ cells

Primordial germ cells arise as a group of approximately 40 cells in the extraembryonic mesoderm around E7.25 marked by the expression of STELLA (*Pgc7*) and alkaline phosphatase (AP) (Ginsburg et al. 1990; Lawson et al. 1994; Saitou et al. 2002; Sato et al. 2002). Primordial germ cells subsequently migrate through the developing hindgut and colonise the embryonic gonads where they ultimately differentiate into either spermatozoa or oocytes. This process involves repression of somatic mesodermal program (Saitou et al. 2002), re-establishment of a gene regulatory network (GRN) similar to the GRN of pre-implantation epiblast (Yabuta et al. 2006; Leitch et al. 2013) and genome-wide epigenetic reprogramming (Seki et al. 2007).

Competence of the post-implantation epiblast to specify into the primordial germ cells fate is restricted to a narrow time window between E5.5 and E6.25. During this period, cells become responsive to BMP4 (Ohinata et al. 2009), downregulate OTX2 (Zhang et al. 2018a) and initiate expression of the key primordial germ cells factors *Blimp1* (*Prdm1*), *Prdm14* and AP2γ (Fig. 1.2) (Ohinata et al. 2005; Vincent et al. 2005; Yamaji et al. 2008; Weber et al. 2010). BMP4 induces *Prdm14* and *Blimp1* in a dose dependent manner (Ohinata et al. 2009). Additional to BMP4, visceral endoderm expressed BMP2 (Ying et al. 2001) can also induce *Blimp1* and *Prdm14* with a reduced efficiency (Ohinata et al. 2009). Rapid elimination of OTX2 is crucial for germ line entry as forced OTX2 expression prevents induction of primordial germ cell programme *in vitro* (Zhang et al. 2018a). On the other hand, deletion of *Otx2* increases number of primordial germ cells *in vivo* and bypasses cytokine requirements for primordial germ cells specification *in vitro* (Zhang et al. 2018a). Rapid downregulation of *Otx2* by BMP4 also indicates that BMP signalling directly represses *Otx2* before activation of *Blimp1* and *Prdm14* (Zhang et al. 2018a).

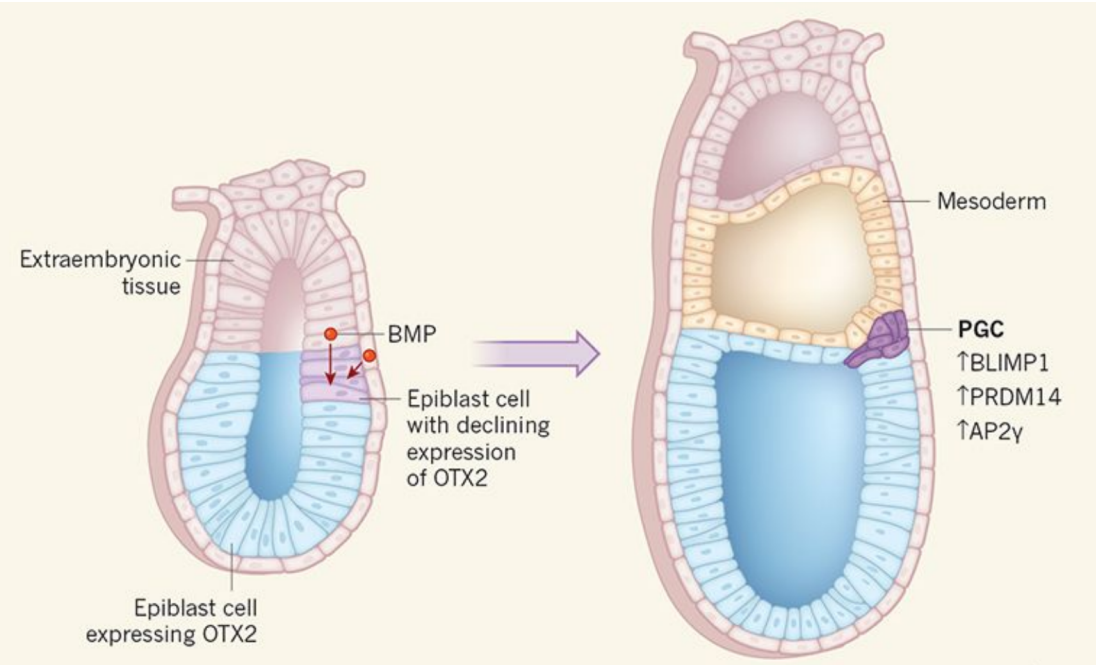


Figure 1.2: BMP4 induced specification of primordial germ cells. Scheme of a mouse embryo at the embryonic day E5.5-6.25 and E7.25 adapted from Laird 2018. Extraembryonic tissue secrets BMP signals to initiate specification of the primordial germ cells (PGCs) from the posterior proximal epiblast, marked by reduced expression of OTX2. Small cluster of PGCs marked by BLIMP1, PRDM14 and AP2γ emerges at E7.25.

1.3 Pluripotent states of the epiblast

In the mouse, the pluripotent epiblast emerges from the inner cell mass around the 32-cell stage (E3.25) and undergo massive physiological and morphological changes until it gives rise to endoderm, mesoderm, ectoderm and the germ line. During this time, the epiblast changes its morphology, transcriptome, epigenome, metabolism and signalling dependencies. Our understanding of this process mainly depends on our ability to capture the intermediate stages *in vitro*. To this day, two main pluripotency states has been described: "naïve" and "primed" pluripotency. In addition, "formative" pluripotent state was proposed as an intermediate state between the naïve and primed states. These states are very dynamic and can be manipulated by adjusting extrinsic or intrinsic factors *in vitro*. Here, the main features of "naïve", "formative" and "primed" pluripotency and the *in vitro* culture systems representing these stages are summarised.

1.3.1 Naïve pluripotency

1.3.1.1 "Serum/LIF" culture condition

Epiblast cells of the pre-implantation mouse blastocyst can be cultured *in vitro* as embryonic stem cells (ESCs) (Evans et al. 1981; Martin 1981). Classically, ESCs are grown on a layer of inactivated fibroblast cells (feeders) in fetal calf serum (serum). The feeder cells are unnecessary if the ESCs are supported by a layer of gelatin and stimulated by the leukemia inhibitory factor (LIF) (Smith et al. 1988; Williams et al. 1988). ESCs self-renew virtually indefinitely *in vitro* while maintaining pluripotency. Transplanted ESCs are able to colonise the blastocyst and contribute to all the embryonic lineages including the germ line (Bradley et al. 1984). Their "naïve" state is highlighted by the capacity to generate entire animal by a complementation of a inviable tetraploid embryo (Nagy et al. 1993).

ESCs form round colonies with a typical dome shaped morphology expressing the key pluripotency genes *Oct4* and *Sox2*. Autocrine secretion

of pro-differentiation factors including fibroblast growth factors (FGF) creates a conflicting signalling environment (Hackett et al. 2014). This results in heterogeneous expression of pluripotency factors including *Nanog, Esrrb, Prdm14* and *Rex1* (Chambers et al. 2003; Martello et al. 2012; Yamaji et al. 2013; Toyooka et al. 2008). Therefore, ESCs cultured in serum/LIF condition represent a spectrum of pluripotent states (Hayashi et al. 2008).

1.3.1.2 "2i/LIF" culture condition

More recently, improved culture conditions were described utilising two small molecule inhibitors PD0325901 and CHIR99021 (2i) (Boroviak et al. 2014). These inhibitors attenuates differentiation via FGF and WNT signalling pathways (more details in sections 1.4.3 and 1.4.4) (Wray et al. 2011). The combination of 2i and LIF (2i/LIF) allows substitution of serum for a chemically defined medium. 2i/LIF culture also allowed derivation of ESCs from mouse strains from which derivation of ESCs failed in serum/LIF medium (Nichols et al. 2009a; Kiyonari et al. 2010). Rat ESCs were also successfully derived using 2i culture as well (Buehr et al. 2008; Li et al. 2008). 2i/LIF also enforces the pluripotency regulatory network resulting in a more homogeneous population of cells similar to the E3.75-E4.5 epiblast (Boroviak et al. 2014; Boroviak et al. 2015). The capacity of *Hex*$^{+}$ ESC population to differentiate into trophoblast suggests that these ESCs cultured in 2i/LIF resemble a population of early epiblast including a small population of totipotent cells (Morgani et al. 2013). Because the ESCs in 2i/LIF resemble early stages of epiblast, this state is also referred to as a "ground state" pluripotency. Compared to serum/LIF conditions, ESCs cultured in 2i/LIF display higher and more homogeneous expression of pluripotency factors *Nanog, Esrrb, Klf4, Klf2* and *Tbx3* (Wray et al. 2010; Marks et al. 2012), lower levels of DNA-methylation and trimethylation of the lysine 27 of histone 3 (H3K27me3) which represents the pre-implantation epiblast more closely (Leitch et al. 2013; Tee et al. 2014).

1.3.2 Primed pluripotency

Stem cells derived from the postimplantation epiblast resemble lineage primed pluripotent state transcriptionally similar to the anterior primitive streak (Kojima et al. 2014). These cells are termed as epiblast stem cells (EpiSCs) and depend on bFGF and Activin A signalling (Brons et al. 2007; Tesar et al. 2007). EpiSCs can be derived from the postimplantation epiblast up until the onset of somatogenesis when pluripotency is lost (Osorno et al. 2012). EpiSCs, similarly to ESCs, are capable of multilineage differentiation and contribute to the all three germ layers of the post-implantation embryo (Huang et al. 2012). However, EpiSCs are unable to contribute to the preimplantation blastocyst and contribute poorly to the germ line (Tesar et al. 2007; Hayashi et al. 2011). Even though EpiSCs express the core pluripotency factors *Oct4* and *Sox2*, expression of other factors including *Nanog, Esrrb, Klf4, Klf5, Prdm14* and *Rex1* is reduced or absent (Hackett et al. 2014). Compared to the "naïve" state, promoters of pluripotency genes in EpiSCs are methylated and female EpiSCs have additionally undergone X chromosome inactivation (Guo et al. 2009; Habibi et al. 2013; Meyenn et al. 2016).

1.3.2.1 Formative pluripotency

Naïve and primed pluripotency states have different molecular features and potencies to contribute to the germ layers and germ cells (Hayashi et al. 2011). Recently, Smith proposed that naïve ESCs are not directly capable of lineage specification but need to transition through a distinct "formative" state which grants them this capacity. In agreement with this hypothesis, epiblast-like cells (EpiLCs), closely resembling E5.5-E6.5 epiblast, have a distinct transcriptional profile from ESCs and EpiSCs (Hayashi et al. 2011).

Replacing 2i/LIF medium for a medium containing bFGF and Activin A initiates dissolution of naïve pluripotency which is marked by *Nanog* downregulation and upregulation of EpiLC markers including *Otx2, Fgf5* and *Pou3f1* (Hayashi et al. 2011). These changes are accompanied by repurposing OCT4 to bind new regulatory elements and by changes in the chromatin

landscape (Buecker et al. 2014; Respuela et al. 2016). EpiLCs transiently gain PGC competence at the day two, similarly to the time restricted PGC competence *in vivo* (Hayashi et al. 2011). At these stage, EpiLCs can be differentiated into germ cell competent PGC-like cells (PGCLCs) by addition of a cytokine cocktail including BMP4, LIF, SCF and EGF (Hayashi et al. 2011; Hayashi et al. 2013). After four days of the differentiation, PGCLCs can be isolated based on expression of SSEA1 and CD61 surface markers and express the key PGC markers *Blimp1, Prdm14* and *Ap2*γ (Hayashi et al. 2012).

1.4 Extrinsic stimuli regulating pluripotency

1.4.1 LIF signalling

Leukemia inhibitory factor (LIF) is a key component of the ESC culture medium and its absence leads to differentiation (Ying et al. 2003). LIF is recognised by LIF receptor which responds by heterodimerisation with gp130 (Yoshida et al. 1994). The LIF receptor/gp130 activates JAK kinase which in turn activates STAT3 by phosphorylation. The STAT3 is the main effector of the LIF pathway. Absence of STAT3 prevents LIF mediated self-renewal whereas overexpression of STAT3 sustains LIF-independent self-renewal (Niwa et al. 1998; Matsuda et al. 1999). *In vivo*, STAT3 is required to sustain pluripotency of the ICM (Do et al. 2013). Phospho-STAT3 is then translocated into the nucleus where it activates transcription of several pluripotency promoting factors including *Klf4, Gbx2* and *Tfcp2l1* (Hall et al. 2009; Niwa et al. 2009; Tai et al. 2013). Of these, only the *Tfcp2l1* overexpression can bypass LIF-signalling and sustain ESC self-renewal without LIF (Martello et al. 2013; Ye et al. 2013). Even though, *Tfcp2l1* is likely the main mediator of LIF/STAT3 pathway, LIF action also involves non-canonical activation of MEK/ERK and PIK3/AKT pathways (Niwa et al. 2009).

1.4.2 TGF-β signalling

Transforming growth factor-β (TGF-β) signalling is mostly involved in embryonic development, tissue homeostasis and immune response. TGF-β family ligands include TGF-β 1-3, bone morphogenic proteins (BMPs), growth and differentiation factors (GDFs), Activins and Nodal (Oshimori et al. 2012). Bone morphogenic proteins (BMPs) are the key component of the serum supporting ESC self renewal. In fact, BMP4 supplementation can replace serum in ESC culture (Ying et al. 2003). The canonical BMP4-SMAD pathway involves heterodimeristaion of TGF-β receptors 1 and 2 which leads to phosphorylation of downstream targets. Canonical effectors of TGF-β are R-smad 1, 2, 3, 5 and 8 which are phosphorylated by Activin-receptor like kinases (ALKs). Phospho-SMAD2/3 form a heterotrimeric complex with SMAD4 and relocates to the nucleus where together with master transcription factors regulate gene expression (Oshimori et al. 2012; Shi et al. 2003). SMAD proteins bind to DNA weakly and require interactions with other transcription factors to stabilise their chromatin binding (Ross et al. 2008). SMAD1 has been found to share many sites with OCT4, SOX2 and NANOG in ESC genome (Chen et al. 2008). Moreover, SMAD3 directly interacts with OCT4 and regulate many shared gene targets (Mullen et al. 2011). Phospho-SMAD2/3 has been shown to support naïve pluripotecy through activation of inhibitors of differentiation (IDs), *Dusp9* and *E-cadherin* (Ying et al. 2003; Malaguti et al. 2013). Inhibitors of differentiation interact with basic helix-loop-helix transcription factors preventing them to bind DNA and drive neural differentiation (Benezra et al. 1990). Direct activation of dual-specificity phosphatase 9 *Dusp9* by BMP, inhibits the activity of ERK1/2 in ESCs (Li et al. 2012). BMP activation have been show to activate *Klf2/4* through non-canonical activation of ERK5 (Morikawa et al. 2016).

In contrast to ESCs, BMP4 induces differentiation in primed pluripotent cells (Bernardo et al. 2011). Instead, EpiSC self-renewal depends on Nodal/Activin signalling (Brons et al. 2007; Tesar et al. 2007). EpiSCs activate

Nanog as a key target of Activin/Nodal pathway (Sun et al. 2014; Vallier et al. 2009). In the absence of Nodal or Activin, EpiSCs spontaneously differentiate towards neural lineage both *in vitro* and *in vivo* (Vallier et al. 2004; Camus et al. 2006).

1.4.3 FGF signalling

FGF signalling is involved in various developmental processes including organogenesis, differentiation, cell survival and cell patterning. It acts through secreted fibroblast growth factors (FGFs) which interact with tyrosine kinase FGF receptors (FGFRs) (Ornitz et al. 2015). Among other effectors of FGF signalling, MEK/ERK pathway has been shown to drive differentiation of ESCs. Upon FGFR activation, PKCζ is activated and phosphorylates MEK which in turn phosphorylates ERK1/2 proteins (Miura et al. 2018). In ESCs, the phosphorylation site of PKCζ is blocked by O-linked N-acetylglucosamine preventing MEK/ERK activation and cell differentiation (Miura et al. 2018). Inhibition of MEK promotes self-renewal by blocking differentiation (Ying et al. 2008). However, MEK inhibition alone cannot support LIF-independent self-renewal. Deletion of *Erk2* compromises neural and mesoderm lineage specification similarly to the *Fgf4* knockout (Kunath et al. 2007; Stavridis et al. 2007). This demonstrates that ERK1/2 is the main effector downstream of FGF4 in this context. In contrast, it has been suggested that intermediate levels of ERK1/2 activity support ESC self-renewal (Ma et al. 2016) as ERK1/2 depletion decreases proliferation rate and expression of *Pou5f1* and *Sox2* (Chen et al. 2015). Regulation of ERK activity by FGF and other signalling pathways including LIF (Burdon et al. 1999; Matsuda et al. 1999; Niwa et al. 1998; Ernst et al. 1996) might therefore maintain balance between self-renewal and pluripotency.

ERK2 regulates differentiation on multiple levels. Phosphorylation of KLF2 by ERK2 leads to KLF2 degradation and destabilisation of the pluripotency network (Yeo et al. 2014). Moreover, ERK1/2 has been shown to "poise" developmental genes through direct phosphorylation of RNAPII serine

5 and recruitment of repressive polycomb complex (Tee et al. 2014). This maybe allows a quick activation of developmental genes in naïve stem cells. In contrast, EpiSCs require ERK1/2 activation to sustain self-renewal (Brons et al. 2007; Tesar et al. 2007; Greber et al. 2010).

1.4.4 WNT/GSK3 signalling

Canonical WNT signalling acts through activation of a destruction complex which targets the key WNT effector β-CATENIN. Without the WNT signal, β-CATENIN is phosphorylated and targeted for destruction by the destruction complex. The destruction complex contains two main kinases: glycogen synthase kinase 3 (GSK3), casein kinase 1 (CK1) and a scafold protein AXIN. After β-CATENIN recruitment by the destruction complex, β-CATENIN is sequentially phosphorylated by CK1 and GSK3 which leads to polyubiqitination and proteosomal degradation of β-CATENIN (Routledge et al. 2019; Gao et al. 2010; Rao et al. 2010).

In the presence of WNT ligand molecules, WNT ligand forms a trimeric complex with Frizzled (Fz) and LDL-receptor 5/6 (LRP5/6) receptors (Nusse et al. 2017; DeBruine et al. 2017). The WNT/Fz/LRP complex recruits DISHEVELED (Dvl) and GSK3 phosphorylates LRP. This leads to sequestration of AXIN and a decrease of the destruction complex activity (Schaefer et al. 2019). In mouse ESCs, WNT stimulation or GSK3 inhibition restricts progression from naïve pluripotency through stabilisation of β-CATENIN (Ten Berge et al. 2011; Yi et al. 2011; Ogawa et al. 2006; Sato et al. 2004). β-*Catenin*$^{-/-}$ ESCs do not responds to GSK3 inhibitor CHIRON, demonstrating that the main effector of the WNT in this context is β-CATENIN (Lyashenko et al. 2011; Wray et al. 2011).

Stabilisation of β-CATENIN leads to its translocation to the nucleus and interaction with T cell factor (TCF) and lymphoid enhancer factor (LEF) transcription factors (Schaefer et al. 2019). Interaction of β-CATENIN with TCF3 prevents repression of pluripotency genes by TCF3 and fosters the ESC self-renewal (Faunes et al. 2013; Wray et al. 2011; Yi et al. 2008). Deletion

of *Tcf3* also supports ESC self-renewal (Guo et al. 2011; Yi et al. 2011). The main downstream target of WNT/β-CATENIN pathway repressed by TCF3 is *Esrrb*. ESRRB overexpression substitutes for GSK3 inhibition wherease *Esrrb* deletion prevents effect of CHIRON (Martello et al. 2012). However, ability of the Wnt3a to stimulate expression of pluripotency genes in *Tcf3*$^{-/-}$ ESCs indicates that there are additional mechanisms of WNT/β-CATENIN function in ESC self-renewal (Yi et al. 2011).

1.5 Gene regulation in embryonic stem cells

Specification of cell function is governed by changes in gene expression. Lineage specific genes are activated, whereas genes not required for the specific function are silenced. The DNA code around the gene dictates when, where and at what level the gene should be expressed. This code is divided into three functional units: the core promoter, the region proximal to the core promoter and more distal enhancers (Fuda et al. 2009). The sequence of the core promoter specifies chromatin binding of general transcription factors (TFs) which form the preinitiation complex (Juven-Gershon et al. 2008). The sequence of the promoter-proximal region and enhancers determines binding of specific TFs which can have either positive (activators) or negative (repressors) effects on transcription.

Activators and repressors can regulate transcription by a direct interaction with RNA polymerase II (RNAPII), general TFs and/or by recruitment of coregulators. Coregulators are usually multiprotein complexes unable to specifically bind DNA but with an ability to induce (coactivators) or repress (corepressors) transcription. The majority of coactivators and corepressors function through one of the multiple mechanisms including covalent modification of the core histone tails, nucleosome organisation or DNA modification. In addition, some coregulators, especially kinases, directly modify RNAPII or general TFs. However, in most cases, the relationships between coregulator activity, TF binding and gene expression are only

correlative and the causal mechanisms are unknown. Moreover, many TFs simultaneously repress and activate genes in the same cell (Latchman 2001). For most of the TFs behaving in this way, including NANOG, the process determining activatory or repressive action on specific genes is unknown.

Enhancers are particularly important for cell-type-specific spatiotemporal regulation of gene expression. Classically, enhancers were defined as DNA regions enhancing gene expression independently of their orientation and position from a gene (Banerji et al. 1981). Recently, most cell-type specific enhancers have been predicted from genome-wide profiles of TF binding (Johnson et al. 2007; Robertson et al. 2007), DNAseI hypersensitivity (Boyle et al. 2008) or enhancer associated histone marks including H3K27ac or H3K4me1. However, functional validation of the identified putative enhancers remains challenging (Arnold et al. 2013). In addition, a distinct group of enhancers with a high occupancy of a mediator complex subunit MED1 was identified in mouse ESCs (Whyte et al. 2013). These enhancers are often located in large enhancer domains termed "super-enhancers" and display high occupancy of the key pluripotency TFs OCT4, SOX2, NANOG, ESRRB and KLF4 (Whyte et al. 2013). Enhancers are often located ~20-50 kb from the regulated gene but there are examples of enhancers acting on a distance up to 1 Mbp (Lettice et al. 2014). However, enhancers and the promoters of the genes they regulate are often in a close proximity within three dimensional space (Dowen et al. 2014; Hnisz et al. 2016; Ji et al. 2016). Observations of action of multiple enhancers on a single gene and simultaneous activation of two genes by a single enhancer (Lim et al. 2018) support the notion that enhancers and promoters with mutual communication are organised to spatially restricted compartments. Moreover, formation of clusters of active RNAPII, called "transcriptional factories", were observed in fixed cells (Jackson et al. 1998). Later, dynamic formation of RNAPII clusters was observed by a super-resolution imaging in living cells (Cisse et al. 2013). These clusters of active RNAPII associate with mRNA production (Cho et al. 2016). This suggests that distinct transcriptional units, containing promoters

and enhancers, form dynamic transcriptional hubs (Furlong et al. 2018). In addition to RNAPII, other transcriptional factors including BRD4, MED1 and OCT4 were observed to form puncta in cells (Sabari et al. 2018; Cho et al. 2018; Boija et al. 2018). BRD4, MED1, RNAPII and OCT4 have been found to form clusters at super-enhancers in ESCs (Sabari et al. 2018; Boija et al. 2018). These clusters display a liquid-like behaviour including fast dynamics of fluorescent recovery after fotobleaching and fusion of multiple cluster (Sabari et al. 2018; Boija et al. 2018; Cho et al. 2018). It has been proposed that TFs accumulate at enhancers to regulate gene expression via multi protein-protein interactions of low complexity domains and liquid-liquid demixing (Hnisz et al. 2017; Boija et al. 2018). This is supported by the capacity of multiple TFs to form phase-separated droplets alone or in the presence of a mediator complex *in vitro* (Boija et al. 2018; Sabari et al. 2018). Moreover, disruption of the MED1 low complexity domain prevents MED1 phase-separation *in vitro* and *in vivo* (Sabari et al. 2018). Incorporation of RNAPII to low-complexity domain hubs was also shown as well as sensitivity of these hubs to 1,6-hexandiol which supports this hypothesis (Chong et al. 2018; Sabari et al. 2018; Boija et al. 2018).

ESCs contain a number of TFs safeguarding pluripotency and self-renewal. Many of these TFs bind to the same regions and regulate gene expression cooperatively (Young 2011). Regulation of multiple gene targets creates a highly interconnected gene regulatory network (GRN) (Young 2011). Moreover, binding of these TFs correlates with STAT3, β-CATENIN and SMAD1 binding which directly connects the TF regulatory network with the signalling pathways (Boyer et al. 2005; Chen et al. 2008; Loh et al. 2006; Cole et al. 2008). OCT4, SOX2 and NANOG have been argued to play the central role in the pluripotency GRN (Chambers et al. 2009; Young 2011).

1.5.1 Oct4 and Sox2

OCT4 (also known as OCT3), encoded by the *Pou5f1* gene, is a member of POU (Pit, Oct, Unc) transcription factor family. OCT4 together with other OCT proteins was first identified and named based on its capacity to bind consensus octameric DNA motif "ATGCAAAT" which was previously found in various promoter and enhancer sequences (reviewed by Zhao 2013). The POU DNA binding domain consists of a homeodomain (POU_{HD}) and a POU specific domain (POU_S) connected by a linker peptide (Fig. 1.3). Structural studies of OCT1, OCT4 and OCT6 showed that POU_{HD} and POU_S independently bind half sites of the consensus motif positioning each DNA binding domain on both sides of the helix (Klemm et al. 1994; Reményi et al. 2001; Esch et al. 2013; Jauch et al. 2011). Quantitative DNase I footprinting experiments showed that OCT1 POU_S binds the 5' half of the octamer motif with a low affinity (1.7 mM) whereas POU_{HD} binds the 3' half of the motif more tightly (0.15 mM) (Klemm et al. 1996). The peptide linker joining both DNA binding domains of OCT4 forms a structured α-helix unlike OCT1 linker which is unstructured (Esch et al. 2013). The additional α-helix in OCT4 POU domain was suggested to be important for heterodimerisation (Esch et al. 2013). Besides the structured DNA binding domains, OCT4 contains two disordered trans-activation domains flanking POU binding domain (Brehm et al. 1997) (Fig. 1.3).

SOX2 is a member of the Sry box (Sox) family which contains a High-Mobility-Group (HMG) box DNA-binding domain (Pevny et al. 1997; Kamachi et al. 2013). SOX2 also contains a C-terminal transactivation domain with a serine-rich region (Fig. 1.3) (Wong et al. 2016). The SOX2 HMG recognises the "(A/T)(A/T)CAAAG" consensus motif (Bowles et al. 2000). SOX2 has been found to bind in a close proximity to OCT4 binding sites at promoters of multiple pluripotency genes including *Fgf4* (Yuan et al. 1995), *Utf1* (Nishimoto et al. 1999), *Sox2* (Chew et al. 2005; Tomioka et al. 2002), *Oct4* (Chew et al. 2005; Okumura-Nakanishi et al. 2005) and *Nanog* (Kuroda et al.

2005; Rodda et al. 2005). Genome wide motif analysis of OCT4 and SOX2 chromatin binding sites identified a similar Oct/Sox composite motif (Chen et al. 2008) (Fig. 1.3). Synergistic binding of SOX2 and OCT4 to the *Fgf4* enhancer has been shown to activate enhancer function (Ambrosetti et al. 1997). These results argue that OCT4 and SOX2 function is highly cooperative (Ambrosetti et al. 1997; Ambrosetti et al. 2000). Structural analysis showed that interaction of POU_S with the major groove and the HMG with the adjacent minor groove position OCT/SOX protein complex on the same side of the helix (Ambrosetti et al. 1997; Chambers et al. 2009).

The requirements of *Sox2* and *Pou5f1* in mouse development were examined in knockout mice. *Pou5f1*$^{-/-}$ embryos fail to establish pluripotent epiblast and development fails at the peri-implantation stage (Nichols et al. 1998). Deletion of embryonic *Sox2* does not prevent acquisition of pluripotency due to the persistence of maternally inherited SOX2 protein and mRNA (Avilion et al. 2003). When both embryonic and maternal *Sox2* mRNA is targeted, embryos display defects in morula cavitation and trophectoderm specification (Keramari et al. 2010). Deletion of either *Sox2* or *Pou5f1* in ESCs induces trophoblast differentiation highlighting their requirements for pluripotency maintenance (Niwa et al. 2000; Masui et al. 2007). In contrast to ESCs, *Sox2* is not required for maintenance of primed EpiSCs due to compensation by the closely related SOXB1 family member *Sox3* (Corsinotti et al. 2017). OCT4 and SOX2 are uniformly expressed in both naïve and primed pluripotent stem cells. Tight regulation of OCT4 expression is required for maintenance of pluripotency. Decrease or increase of OCT4 levels leads to differentiation (Niwa et al. 2000). Counterintuitively, 50% reduction of OCT4 levels increase OCT4 chromatin binding and stabilises the pluripotent gene regulatory network (Karwacki-Neisius et al. 2013). This suggests that OCT4 not only supports ESC self-renewal but also supports responsiveness of ESCs to differentiation stimuli when expressed at wild-type levels (Karwacki-Neisius et al. 2013).

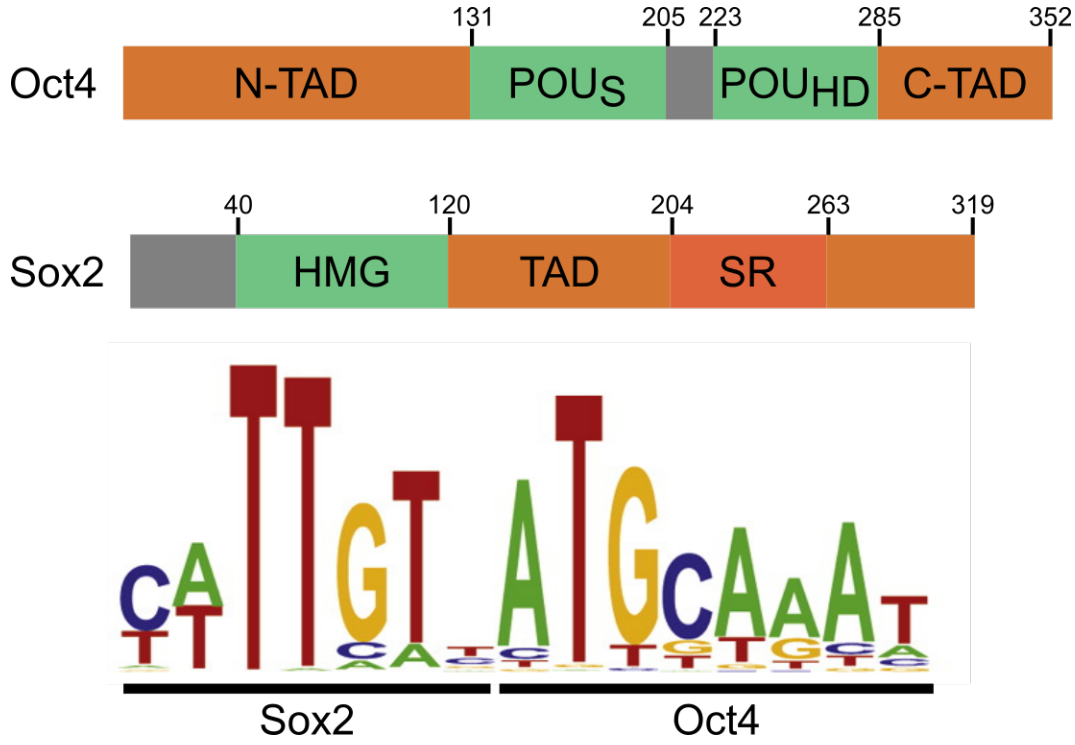


Figure 1.3: OCT4 and SOX2 protein structures and OCT:SOX DNA binding motif. Schematic representation of OCT4 and SOX2 proteins. Each box represent a domain: HMG - high mobility group DNA-binding domain, (N/C) TAD - (N-terminal/C-terminal) transactivation domain, POU_{HD} - POU homeodomain, POU_S - POU specific domain. The logo representation of the composite OCT:SOX DNA motif adapted from Chen et al. 2008 is shown.

1.5.2 Nanog

1.5.2.1 Molecular features of NANOG

NANOG is a homeodomain protein with the highest similarity to the Nkx protein family in mouse (Chambers et al. 2003). NANOG homeodomain recognises the consensus "TAAT" motif as many other homeodomain proteins (Fig. 1.4) (Jauch et al. 2008; Mitsui et al. 2003). NANOG-DNA interaction can be disrupted by a single point mutation of the asparagine 51 (N51) within the homeodomain (Jauch et al. 2008; Noyes et al. 2008). The second described domain of NANOG termed as the tryptophan repeat (WR) is composed by ten pentapeptide repeats with highly conserved tryptophan residues (Fig. 1.4) (Mullin et al. 2011). The WR is required for NANOG homodimerisation and NANOG function in ESC self-renewal (Mullin et al. 2008). Mutation of the individual tryptophan residues within the WR, gradually decreases the capacity of NANOG to drive LIF-independent self-renewal and to form homodimers (Mullin et al. 2017). Therefore, NANOG homodimerisation could be required for DNA binding specificity or avidity. An alternative hypothesis suggests that the WR can act as a protein hub enabling recruitment of specific proteins to the NANOG binding sites (Mullin et al. 2011). In fact, NANOG interacts with more than 100 proteins in ESCs including transcription factors, coactivator and corepressor complexes (Gagliardi et al. 2013). However, only SOX2, RNAPII, NCOR1, RESF1, ELMSAN1 and ZIC3 require the WR for the interaction with NANOG (Mullin et al., unpublished). Nevertheless, direct interaction of NANOG with these proteins might be absolutely critical for NANOG function in ESC self-renewal. Formation of NANOG-protein complexes could be regulated by posttranslational modifications. NANOG has been shown to be ubiquitinated and phosphorylated at multiple sites affecting its stability (Jin et al. 2016; Brumbaugh et al. 2014; Moretto-Zita et al. 2010; Yates et al. 2005). However, further studies are required to determine the requirements of individual protein-protein interactions for NANOG function.

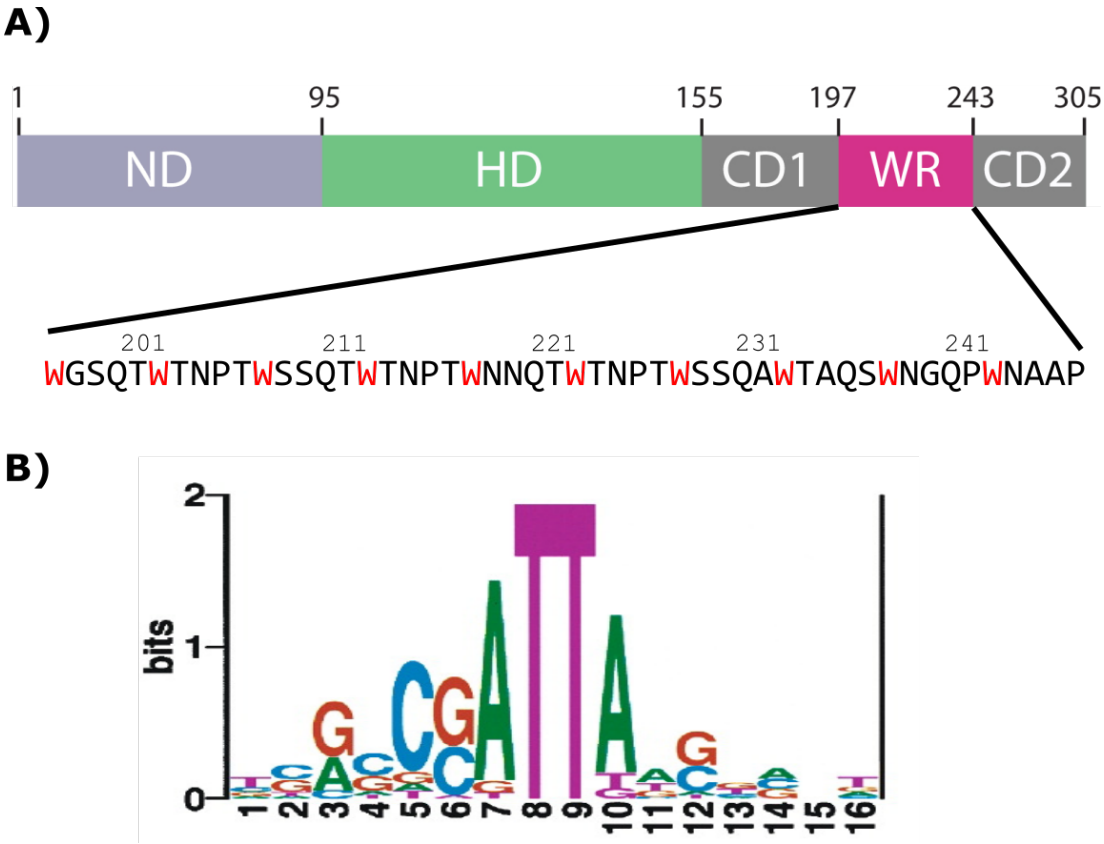


Figure 1.4: NANOG protein structure and DNA binding motif. A) Schematic representation of NANOG protein structure. Each box represent a domain: ND - N-terminal domain, HD - homeodomain, WR - tryptophan repeat, CD1/2 - C-terminal domain. The sequence of the WR with highly conserved tryptophans (W) is shown. **B)** The logo representation of the consensus DNA motif bound by NANOG adapted from Mitsui et al. 2003.

1.5.2.2 NANOG function in mouse development

Nanog is essential for acquisition of pluripotency and its deletion prevents epiblast specification (Mitsui et al. 2003; Silva et al. 2009). However, *Nanog* is not essential for the pluripotency maintenance, demonstrated by the capacity of *Nanog*$^{-/-}$ ESCs to self-renew in the presence of LIF, the capacity to contribute to all three germ layers (Chambers et al. 2007; Carter et al. 2014; Schwarz et al. 2014), and the capacity to establish and maintain EpiSCs (Osorno et al. 2012). Moreover *Nanog*$^{-/-}$ mouse embryos have a decreased efficiency of primordial germ cell (PGC) specification (Chambers et al. 2007; Zhang et al. 2018b). In contrast, increased expression of NANOG in the PGC competent EpiLCs, initiates PGC-like cell (PGCLC) specification even without otherwise necessary BMP4 (Murakami et al. 2016). This is particularly interesting as increased NANOG expression in EpiSCs drives reprogramming to the ground pluripotent state (Silva et al. 2009). This maybe due to redirected NANOG binding to PGC-fate inducers *Prdm1* and *Prdm14* in the PGC competent state (Murakami et al. 2016).

1.5.2.3 Model of NANOG function

NANOG binds tens of thousands sites within the ESC genome (Karwacki-Neisius et al. 2013; Chen et al. 2008; Heurtier et al. 2019; Chronis et al. 2017; Festuccia et al. 2018; Galonska et al. 2015; Kim et al. 2018; King et al. 2017; Whyte et al. 2013). Many of these sites are close to sites bound by other pluripotency transcription factors, including OCT4, SOX2, ESRRB as an evidence of combinatorial effect of TFs in gene regulation (Chen et al. 2008; Loh et al. 2006; Young 2011).

Despite a vast number of genes targeted by NANOG, only a small proportion is directly regulated by NANOG (Festuccia et al. 2012; Festuccia et al. 2013; Festuccia et al. 2018; Heurtier et al. 2019). Among the genes with the highest activation after NANOG induction in *Nanog*$^{-/-}$ ESCs are *Essrb* and *Klf4* (Festuccia et al. 2012). *Esrrb* pre-mRNA levels increase after 20 min of NANOG induction, highlighting the direct regulation of *Esrrb* by NANOG

(Festuccia et al. 2012). In contrast, a recent study claimed that the expression of *Klf4* and *Esrrb* do not change upon NANOG induction (Heurtier et al. 2019). This result was observed in both *Nanog*$^{+/+}$ and *Nanog*$^{-/-}$ inducible NANOG ESCs. As response of ESRRB expression to NANOG follows a saturation curve (Festuccia et al. 2018), elevating NANOG levels in *Nanog*$^{+/+}$ ESCs may have had small effect on *Esrrb* expression. This however do not explain why Heurtier et al. 2019 did not identified *Klf4* and *Essrb* among other NANOG responsive genes after NANOG induction in *Nanog*$^{-/-}$ ESCs.

Esrrb and *Klf4* overexpression can sustain LIF-independent self-renewal and can revert EpiSCs to naïve pluripotency similarly to *Nanog* (Festuccia et al. 2012; Guo et al. 2009; Silva et al. 2009). Moreover, *Esrrb* can substitute for *Nanog* in multiple functions. *Nanog*$^{-/-}$ ESCs can sustain LIF-independent self-renewal when ESRRB is overexpressed, whereas NANOG cannot sustain LIF-independent self-renewal without *Esrrb* (Festuccia et al. 2012). Similarly, *Esrrb* can substitute for *Nanog* function in PGC specification (Zhang et al. 2018b). This indicates that *Esrrb* is the main factor downstream *Nanog*.

Among the most prominent genes negatively regulated by NANOG is *Otx2*. *Otx2* and *Nanog* has been shown to have an antagonist function in ESCs and EpiLCs (Acampora et al. 2017; Zhang et al. 2018a). While *Nanog* safeguards naïve pluripotency, *Otx2* primes ESCs for differentiation (Acampora et al. 2017). Moreover, *Nanog* induces PGC specification in EpiLCs (Murakami et al. 2016) whereas *Otx2* blocks entry into the germ line (Zhang et al. 2018a). These results show that NANOG mediated *Otx2* repression is crucial in multiple biological processes. NANOG has been proposed to regulate *Otx2* through H3K27me3 deposition by a polycomb complex (Heurtier et al. 2019). However, whether NANOG directly recruits polycomb complex to *Otx2* and whether the dynamics of H3K27me3 acquisition matches the rate of *Otx2* repression remains to be determined.

1.5.3 Resf1

RESF1 (also known as KIAA1551) is a ~170 kDa protein conserved from chicken to human with no known functional domains (Hudson et al. 2005; Fukuda et al. 2018). *Resf1* was firstly described as a gonad expressed transcript (GET) due to its expression in the urogenital systems of newly developing chick and mouse embryos (Hudson et al. 2005). Hudson et al. proposed that GET might play a role in development of many parts of the urogenital system including germ cells and gonad cortex. This hypothesis has been supported by the results of the International Mouse Phenotyping Consortium (IMPC) showing female and male infertility of *Resf1*$^{-/-}$ mice (Dickinson et al. 2016).

Recently, RESF1 has been shown to be involved in repression of retrotransposon elements in mouse ESCs (Fukuda et al. 2018). The authors named this protein as "Retroelement Silencing Factor 1" (RESF1) and this name was adopted in this thesis as well. RESF1 has a preference to silence the L1 group of Long Interspersed Elements (LINEs) and type K of Endogenous Retroviruses (ERVK) (Fukuda et al. 2018). The authors observed depletion of DNA-methylation and H3K9me3 at the murine stem cell virus (MSCV) promoter and its simultaneous activation after *Resf1* deletion in ESCs. Depletion of H3K9me3 could be explained by a reduction of Histone-lysine N-methyltransferase SETDB1 binding at these sites upon *Resf1* deletion (Fukuda et al. 2018). Transcriptional profile of *Resf1*$^{-/-}$ cells revealed upregulation of 22 repeat families from which 19 were Endogeneous Retroviruses (ERVs) (Fukuda et al. 2018). Surprisingly, over one thousand genes were differentially expressed in *Resf1* knock out mouse ESCs. The authors also reported 74 chromatin binding sites of RESF1 determined by a chromatin immunoprecipitation followed by a next generation sequencing (ChIP-seq). ERVs were enriched among these sites (Fukuda et al. 2018). RESF1 binding sites were often associated with misregulated imprinted genes and differentially methylated regions (Fukuda et al. 2018).

RESF1 has been also identified to interact with NANOG (Gagliardi et al. 2013) and OCT4 (Berg et al. 2010) in mouse ESCs. Interestingly, RESF1-NANOG interaction depends on the functionally essential WR region of NANOG (Mullin et al. unpublished). However, the role of RESF1 in self-renewal and pluripotency remains unknown.

1.6 Aims of the thesis

The aim of this thesis is to examine the role of the tryptophan repeat (WR) in NANOG function. The WR has been identified to be essential for the defining function of NANOG to sustain LIF-independent self-renewal (Mullin et al. 2008). Even though it is known that debilitating mutations of the WR tryptophans prevents NANOG homodimerisation and interaction with RNAPII, SOX2, ZIC3, NCOR1, RESF1 and ELMSAN1 (Mullin et al. unpublished), the mechanism of the WR action remains unclear. RESF1 is an uncharacterised protein interacting with NANOG and its function in ESCs is unknown. Therefore, RESF1 is a promising target to reveal novel mechanisms in maintenance of ESC self-renewal.

In this thesis, the function of RESF1 protein in ESC self-renewal and early germ line specification was explored. Firstly, RESF1 localisation in ESCs was determined by endogenous tagging of *Resf1* in mouse ESCs. Requirements of *Resf1* for ESC self-renewal, NANOG function to confer LIF-independent self-renewal and the requirement of RESF1 for early germ line specification were examined in *Resf1* knockout ESC lines.

As, WR might be also required for avidity and specificity of NANOG DNA binding (Mullin et al. 2011) and binding of the core ESC TFs OCT4, SOX2 and RNAPII, immediate changes in gene expression and in OCT4, SOX2 and RNAPII chromatin binding after induction of NANOG variants with debilitating mutations in the homeodomain or the WR were examined. K-means clustering was used to separate NANOG binding sites into distinct clusters based on the OCT4, SOX2 and RNAPII changes. Clustering revealed association between specific patterns of TF binding and gene activity after NANOG induction.

2 Materials and methods

2.1 Materials and reagents

Table 2.1 Cell culture media and solutions

Glutamate/pyruvate solution

5.5 ml	100 mM Sodium pyruvate (Invitrogen, cat. 11360-039)
5.5 ml	200 mM L-glutamine (Invitrogen, cat. 25030-024)

GMEMβ (serum medium)

500 ml	GMEM (Sigma, cat. G5154)
51 ml	FCS (Gibco, cat 10270-106)
5.5 ml	100x NEAA (Invitrogen, cat. 11140-036)
11 ml	Glutamate/pyruvate
1.14 ml	50 mM β-mercaptoethanol (Gibco, cat. 31350-010)

Homemade N2

0.5 ml	100 mg ml^{-1} Apo-transferrin (Sigma, cat. T-1147)
16.5 µl	0.6 mg ml^{-1} Progesterone (Sigma, cat. P8783)
50 µl	160 mg ml^{-1} Putrescine (Sigma, cat. P5780)
5 µl	3 nM Sodium selenite (Sigma, cat. S5261)
0.33 ml	7.5% w/v BSA fraction V (Invitrogen, cat. 15260-037)
3.6 ml	DMEM/F-12 without HEPES (Invitrogen, cat. 21041-025)

N2B27 medium

100 ml	DMEM/F-12 (1:1 v/v, no L-glutamine) (GIBCO cat. 12634010)
100 ml	Neurobasal (GIBCO cat. 21103049)
2 ml	200 mM L-glutamine (Invitrogen cat. 25030024)
2 ml	100x NEAA (Invitrogen, cat. 11140-036)
1 ml	100x N2 (GIBCO cat. 17502048, or homemade)
2 ml	50x B27 (GIBCO cat. 17504044)
400 µl	50 mM β-mercaptoethanol (Invitrogen, cat. 21985-023)

PGCLC medium

41 ml	GMEM (Sigma, cat. G5154)
7.5 ml	KSR (Invitrogen, cat. 10828-028)
0.5 ml	100x NEAA (Invitrogen, cat. 11140-036)
1 ml	Glutamate/pyruvate (homemade)
100 µl	50 mM β-mercaptoethanol (Gibco, cat. 31350-010))
500 µl	Penicillin/Streptomycin
50 µl	50 µg ml^{-1} Bmp4 (R&D Systems, cat. 314-BP-010)
50 µl	50 µg ml^{-1} Bmp8a (R&D Systems, cat. 1073-BP-010)
50 µl	10 µg ml^{-1} SCF (R&D Systems, cat. 455-MC-010)
10 µl	500 µg ml^{-1} EGF (R&D Systems, cat. 2028-EG-010)
50 µl	1×10^6 U ml^{-1} ESGRO (Millipore, ESG1106)

Trypsin solution

186 mg	EDTA (Sigma, cat. E5134)
500 ml	1x PBS (Sigma, cat. D8537)
5 ml	Chicken serum (Sigma, cat. C5405)
5 ml	2.5% w/v Trypsin (Invitrogen, cat. 15090-046)

Wash buffer

50 ml	1x PBS (Sigma, cat. D8537)
670 µl	7.5% w/v BSA fraction V (Invitrogen, cat. 15260-037)

AP fixative solution

65 ml	Acetone
22 ml	Citrate Solution (Sigma, cat. R86)
8 ml	Formaldehyde

AP staining solution

94 ml	ddH_2O
2 ml	FRV-Alkaline Solution (Sigma, cat. R86)
2 ml	Sodium Nitrite Solution (Sigma, cat. R86)
2 ml	Naphthol AS-BI Alkaline Solution (Sigma cat. R86)

Table 2.2 Cell culture flasks, dishes and plates

Type	Area [cm^2]	Volume of medium [ml]
96-well plate	0.25	0.1-0.2
48-well plate	0.95	0.2
24-well plate	1.89	1
12-well plate	3.8	1
6-well plate	9.62	2-3
10 cm dish	59.2	10
15 cm dish	149.6	20-50
T25 flask	25	10
T75 flask	73.7	30
T150 flask	149.1	50

Table 2.3 Mouse embryonic stem cell lines

Name	Description	Reference
E14Tg2a	Wild type ESCs	(Handyside et al. 1989)
E14/T	E14Tg2a ESCs which constitutively express polyoma large T	(Aubert et al. 2002)
Resf1$^{-/-}$	E14Tg2a ESCs lacking both *Resf1* alleles	(Results 3.2.1)
Rv5	E14Tg2a ESCs with *Resf1* coding sequence extended with 3x v5-tag, eGFP and flanked by a loxP site	(Results 3.1)
$NERT_{WT}$	*Nanog*$^{-/-}$ ESCs with randomly integrated Nanog-ER^{T2} expression cassette	(Festuccia et al. 2012)
$NERT_{W10A}$	*Nanog*$^{-/-}$ ESCs with randomly integrated W10A Nanog-ER^{T2} expression cassette	(Zhang 2016)
$NERT_{N51A}$	*Nanog*$^{-/-}$ ESCs with randomly integrated N51A Nanog-ER^{T2} expression cassette	(Zhang 2016)

Table 2.4 Used buffers

ChIP Swelling buffer	25 mM HEPES pH 7.9
	1.5 mM $MgCl_2$
	10 mM KCl
	0.1% NP-40
Sonication buffer	50 mM HEPES pH 7.9
	140 mM NaCl
	1 mM EDTA
	1% Triton X-100
	0.1% Na-deoxycholate
	0.1% SDS
ChIP Wash buffer A	50 mM HEPES pH 7.9
	500 mM NaCl
	1 mM EDTA
	1% Triton X-100
	0.1% Na-deoxycholate
	0.1% SDS
ChIP Wash buffer B	20 mM Tris-HCl pH 8.0
	1 mM EDTA
	250 mM LiCl
	0.5% NP-40
	0.5% Na-deoxycholate
TE buffer	10 mM Tris-HCl pH 8.0
	1 mM EDTA
ChIP Elution buffer	50 mM Tris-HCl pH 8.0
	1 mM EDTA

Swelling buffer	5 mM Pipes pH 8
	85 mM KCl
NE buffer	20 mM HEPES pH 7.6
	350 mM KCl
	0.2 mM EDTA pH 8
	1.5 mM $MgCl_2$
	20% Glycerol
Gibson assembly	100 mM Tris-HCl pH 8.0
mastermix (2x)	10 mM $MgCl_2$
	0.2 mM dNTPs
	50 mU μl^{-1} Phusion polymerase
	16 mU μl^{-1} T5 exonuclease
EMSA buffer (5x)	50 mM Hepes pH 7.9
	250 mM KCl
	50 mM NaCl
	50% Glycerol
	2 mM EDTA
	0.5 mg ml^{-1} BSA
Transfer buffer	25 mM Tris-HCl
	0.21 M Glycine
	10% Methanol
Blocking buffer	0.1% Triton X-100 in PBS
	1% BSA
	3% Donkey serum

2.2 Routine culture of mouse ESCs

ESCs (Tab. 2.3) were routinely cultured at 37 °C and 7% CO_2 in serum medium (Tab. 2.1) supplemented with LIF (serum/LIF) at final concentration 1000 U ml^{-1} unless specified otherwise. Every other day, cells were washed twice with PBS and detached by the Trypsin solution (Tab. 2.1) for 3 min at 37 °C. Cells were resuspended in serum medium and centrifuged at 1,100 rpm for 3 min. Medium was aspirated and the cell pellet was resuspended in fresh culture medium. Cells were counted with a haemocytometer. Cells were passaged 1:5 to 1:15 v/v on 0.1% gelatin coated plate or flask containing appropriate volume of the culture medium (Tab. 2.2).

2.2.1 Cell thawing

Frozen cell suspension was rapidly thawed at 37 °C and diluted in warm culture medium or the Wash buffer (Tab. 2.1). Cells were collected by centrifugation and the supernatant was aspirated. The cell suspension was resuspended in a warm culture medium and cultured as described in methods 2.2.

2.2.2 Cell freezing

Cultured cells were washed twice with PBS and detached by Trypsin solution for 3 min at 37 °C. Cell pellets were resuspended in 10% v/v DMSO in KSR and transferred into CryoTubes (NUNC, cat. 377224). The cells were placed immediately on ice and stored at –80 °C. Cells were transferred to a liquid nitrogen tank for a long term storage the next day.

2.3 Colony forming assay

To quantify self-renewal capacity of *Resf1*$^{-/-}$ ESCs, 600 of wild-type E14Tg2a and *Resf1*$^{-/-}$ ESCs were plated in separate gelatin coated wells in a 6-well plate. The cells were cultured in the presence or absence of LIF for 6 days. For each replicate, three wells were seeded for each condition. Formed colonies were stained for alkaline phosphatase (AP) as described in method section 2.4 and quantified based on their morphology.

To quantify the self-renewal capacity of *Resf1*$^{-/-}$ ESCs with episomal NANOG expression, pPyCAG-$(Flag)_3$-Nanog-IRES-Puro plasmid (Chambers et al. 2003) was linearised by ScaI-HF (NEB, cat. R3122) and purified by ethanol precipitation. The *Resf1*$^{-/-}$ and E14Tg2a ES cell lines (Tab. 2.3) were transfected with the linearised plasmid as described in the method section 2.6. The transfected cells were passaged every other day in serum/LIF medium supplemented with puromycin. After more than 10 days, the expression of the Flag-tagged NANOG was quantified by western blot as described in the method section 2.11.4. To quantify self-renewal capacity, the prepared cells (600) were plated on a gelatin coated 6-well plate in triplicates and cultured in 3 ml serum medium with or without LIF. The medium was replaced every other day. After 8 days, formed colonies were stained for AP (methods 2.4) and were quantified based on their morphology.

To determine if forced RESF1 expression supports ESC self-renewal, pPyCAG-$(Flag)_3$-Resf1-PGK-PURO (see method section 2.10.5), pPyCAG-$(Flag)_3$-Nanog-PGK-PURO (positive control) and pPyCAG-PGK-PURO (negative control) were transfected into E14/T ESCs as described in method section 2.6. The next day, each transfected cell line was plated into six gelatin coated 10 cm plates ($5x10^4$ cells/plate). The cells were cultured for 8 days in serum medium supplemented with puromycin and G418 in presence or absence of LIF. The medium was replaced every other day. Formed colonies were stained for AP (methods 2.4) and quantified based on their morphology.

2.4 Alkaline phosphatase staining

The cultured cells were fixed with the Fixative solution (Tabs. 2.1, 2.5) for at least 1 min at the room temperature. The fixed cells were washed with H_2O and stained with the AP Staining solution (Tabs. 2.1, 2.5) for 20 min in dark at the room temperature. The solution was washed with H_2O and the plates were dried before analysis.

Table 2.5 Amount of the solutions used for AP staining

Reagent	6-well plate	10 cm plate
Fixative solution	0.5 ml	2 ml
Staining solution	1 ml	3 ml

2.5 Differentiation of ESCs

2.5.1 EpiSC differentiation

EpiSCs were derived from ESCs *in vitro* as described by Guo et al. 2009. ESCs ($3x10^4$) cultured in serum/LIF were plated on a gelatin coated 6-well plate. The cells were cultured in serum/LIF medium for one day. Cells were washed twice with PBS. N2B27 medium (homemade N2) (Tab. 2.1) supplemented with 20 ng ml^{-1} bFGF (R&D Systems, cat. 233-FB-025/CF) and 20 ng ml^{-1} Activin A (PeproTech, cat. 120-14E) was added to the cells which were cultured at 37 °C, 7% CO_2. After 24 h, cells were washed twice with PBS and dissociated by 200 µl of TrypLE Express (Gibco, cat. 12604013) for 2 min at 37 °C. Cells were resuspended in 1 ml of the N2B27/bFGF/Activin medium and pelleted by centrifugation (1,100 rpm, 3 minutes). The cells were resuspended in 600 µl of the N2B27/ActivinA/bFGF medium and 100 µl of the cell suspension was plated on a 6-well plate pre-treated with 7.5 µg ml^{-1} Fibronectin (Sigma, cat. F1141) for 1 h. EpiSCs were passaged 1:5 to 1:15 ratio every other day following the same procedure. EpiSCs were analysed after 12 days of the culture.

2.5.2 EpiLC differentiation

EpiLC differentiation protocol was adapted from Hayashi et al. 2013. Modified N2B27 medium was used in this protocol. The N2B27 medium was prepared as described before (Tab. 2.1) with the exception that phenol-red free DMEM/F-12 (Invitrogen, cat. 21041-025), phenol-red free Neurobasal (Invitrogen, cat. 12348-017) and N2 (GIBCO cat. 17502048) were used instead.

ESCs cultured in serum/LIF medium were washed twice with PBS and dissociated using Trypsin solution for 3 min at 37 °C. The cells were washed twice with PBS or the Washing buffer and resuspended in 3 ml of the N2B27 medium supplemented with 0.4 µM PD0325901, 3 µM CHIR99021 and 100 U ml^{-1} ESGRO (Millipore, cat. ESG1106) referred to as "2i/LIF medium". The cell suspension containing 3x10^5 cells was plated on a 6-well plate pre-treated with 0.01% Poly-L-ornithine (Sigma-Aldrich, cat. no. P3655) and 10 ng ml^{-1} Laminin (BD Biosciences, cat. no. 354232) as described by Hayashi et al. 2013. Cells were cultured at 37 °C, 5% CO_2. Medium was replaced every day and cells were passaged every two days as described above. ESCs were washed with 1 ml of the Wash buffer and detached by 500 µl TrypLE Express (Invitrogen, cat. 12604-021) at 37 °C for 3 min. Cell suspension was diluted in 800 µl Wash buffer and cells were counted.

For EpiLC differentiation, ESCs (1x10^5) adapted in 2i/LIF medium for at least three passages were washed with the Wash buffer and collected as described above. Cells were resuspended in 1 ml of N2B27 medium supplemented with 20 ng ml^{-1} Activin A, 12 ng ml^{-1} bFGF and 1% KSR (EpiLC medium). Cell suspension was transferred to a 12-well plate coated with 41.7 µg ml^{-1} human plasma fibronectin (Millipore, cat. no. FC010). Cells were swirled well and cultured at 37 °C at 5% CO_2. The next day the medium was replaced with 1.5 ml of fresh EpiLC medium. After 48h, cells were washed with the Wash buffer, dissociated with 300 µl of TrypLE Express (Invitrogen, cat.

12604-021) at 37 °C for 2 min and resuspended in 800 µl of the Wash buffer. EpiLCs were analysed by RT-qPCR (method section 2.12) or were used for PGCLC differentiation (method section 2.5.3).

2.5.3 PGCLC differentiation

EpiLCs were differentiated into PGCLCs based on previously published protocol (Hayashi et al. 2013). EpiLCs ($1x10^5$) (see 2.5.2) were resuspended in 1 ml of PGCLC medium (Tab. 2.1) and aggregated in 20 µl drops hanging over PBS at 37 °C, 7% CO_2 for two days. Formed embryonic bodies were collected into 3 ml to 5 ml of PGCLC medium and cultured in a suspension at 37 °C, 7% CO_2, 72 rpm for two days. At the day four of the differentiation, expression of SSEA-I and CD61 was analysed by a flow cytometry (see 2.8). The remaining cells were lysed and the mRNA levels were analysed by RT-qPCR (see 2.12).

2.6 Lipofection of embryonic stem cells

ESCs were resuspended in serum/LIF medium and plated on a 6-well plate or a 10 cm dish coated with 0.1% gelatin (Tab. 2.6). DNA was diluted in Opti-MEM reduced serum medium (GIBCO, cat. 31985062) and mixed with P3000 reagent (Invitrogen cat. L3000001) in a separate tube (Tab. 2.6). Lipofectamine 3000 (Invitrogen cat. L3000001) was diluted in the same volume of the Opti-MEM medium in a separate tube (Tab. 2.6). DNA and Lipofectamine solutions were mixed and incubated for 5 min at the room temperature. The solution was added to the cell suspension in a drop-wise fashion. Cells were incubated at 37 °C, 7% CO_2 over night.

Table 2.6 Lipofection of ESCs

Component	6-well plate	10 cm dish
Number of cells	$1x10^6$	$6x10^6$
Opti-MEM medium	125 µl	250 µl
Lipofectamine	1.5 µl	3 µl
DNA	3 µg	6 µg
P3000 Reagent	3 µg	6 µg

2.7 Fluorescent activated cell sorting (FACS)

Cells cultured in a 6-well plate were washed twice with PBS and dissociated by accutase (StemPro, cat. A1110501) for 5 min at 37 °C. Cells were resuspended in 5 ml of 2% FCS in PBS and centrifuged (1,100 rpm, 3 min). Cells were resuspended in 1 ml of 2% FCS in PBS to obtain a suspension of $1x10^6$ cells/ml. The cell suspension was passed through a cell strainer. DAPI (100 ng ml^{-1}) was added to each suspension and single cells were sorted using FACS Aria II (BD Biosciences) into gelatin coated 96-well plates containing 200 µl of warm serum/LIF medium supplemented with penicillin and streptomycin. The cells were grown at the standard culture conditions and expanded as required.

2.8 Analysis of PGCLCs by flow cytometry

Embryonic bodies were collected in a 1.5 ml tubes, washed with 1 ml of PBS and dissociated in 400 µl of 0.1% Trypsin solution at 37 °C for 10 min. Cell suspension was diluted with 800 µl of serum medium and passed through a cell strainer. A small proportions of each analysed suspension were combined in a separate tube for control samples. Cells were centrifuged (1,100 rpm, 3 min) and resuspended in 100 µl of serum medium supplemented with 0.5 µl SSEA-I (BioLegend, cat: 125607) and 0.15 µl PE-CD61 antibodies (BioLegend, cat. 104307). The negative control was resuspend in 300 µl of the serum medium and divided into three 100 µl fractions. SSEA-I and CD61 antibodies were added to one fraction resulting in two single-stain control samples and one negative control. Cells were incubated for 15 min at the room temperature in dark and washed twice with 1 ml PBS. Cell pellets were resuspended in 200 µl of 2% KSR in PBS and analysed on 5 laser LSR Fortesa analyser (BD Biosciences). Single cells were gated based on forward and side scatters. Live cells were gated based on DAPI signal and SSEA-I, CD61 double positive population was gated based on the negative and single stain controls.

2.9 Derivation of clonal ESC lines

2.9.1 Generation of clonal *Resf1*$^{-/-}$ ESC lines

Prepared eSpCas9 plasmids encoding gRNAs targeting either 3' or 5' flanking regions of the *Resf1* locus (see method section 2.10.6) were mixed together and used to transfect E14Tg2a ESCs as described in the method section 2.6. On the second day after the transfection, single cells positive for GFP and mCherry signal were sorted into 96-well plates as described in the method section 2.7. DNA from individual clonal cell lines was isolated using DNeasy Blood & Tissue Kits (QIAGEN, cat. 69504) and genotyped using Resf1_WT and Resf1_KO genotyping primer pairs (see method section 2.10.1). The Resf1_KO PCR products were cleaned using QIAquick PCR Purification Kit (QIAGEN, cat. 28104). Five microlitres of the cleaned PCR products were cloned into pCR-Blunt II-TOPO plasmid using Zero Blunt TOPO PCR Cloning Kit (Invitrogen, cat. K2800-20SC). The cloned plasmids were used to transform *E. coli* (method section 2.10.4) and single colonies were expanded in 25 ml LB supplemented with 50 µl ml^{-1} kanamycin at 37 °C and 220 rpm over night. Plasmids were isolated using QIAprep Spin Miniprep Kit (QIAGEN, cat. 27106) and sequenced by Sanger sequencing (method section 2.10.7).

2.9.2 Generation of clonal ESC lines with endogenously tagged *Resf1*

To create endogenously tagged *Resf1* cell lines, E14Tg2a ESCs were transfected with a targeting vector (see method section 2.10.8) and eSpCas9 plasmid targeting *Resf1* stop codon (see method section 2.10.6) as described in the method section 2.6. The next day, culture medium was replaced with 3 ml of serum/LIF medium and cultured for one day. The cells were washed twice with PBS and dissociated using 200 µl of Trypsin solution for 5 min at 37 °C. Single cells with high GFP signal were sorted and expanded as described in

the method section 2.7. DNA from the individual clonal cell lines was isolated using DNeasy Blood & Tissue Kits (QIAGEN, cat. 69504) and genotyped using Rv5 genotyping primer pair (see 2.10.1).

2.10 Molecular cloning

2.10.1 Polymerase chain reaction (PCR)

PCR was performed using high fidelity Q5 polymerase (NEB) according to the recommended protocol (NEB, M0491). The reaction mixture was prepared as indicated (Tab. 2.7) and DNA was amplified using T100 Thermal Cycler (BIO-RAD) (Tab. 2.7).

Table 2.7 Polymerase chain reaction protocol

Component	Volume [µl]
5X Q5 Reaction Buffer	5 µl
dNTPs (10 mM each)	0.5 µl
Forward primer	1.25 µl
Reverse primer	1.25 µl
Template DNA	1 ng to 1 µg
Q5 DNA Polymerase	0.25 µl
5x Q4 High GC Ehancer	5 µl
Nuclease-Free Water	to 25 µl

Step	Temperature [°C]	Time
Initial Denaturation	98	1 min
25-35 cycles	98	5-10 s
	50-72	10-30 s
	72	20-30 s/kb
Final Extension	72	2 min
Hold	12	∞

Table 2.8 Primers used for genotyping and molecular cloning

Name	Sequence (5' to 3')	Comment
Resf1_WT-F	ggggagctgtgatgtgaagt	Genotyping primer amplifying part of *Resf1* gene
Resf1_WT-R	ccaggaaataattctggctca	Genotyping primer amplifying part of *Resf1* gene
Resf1_KO-F	ccccaatgccgtgaagttg	Genotyping primers flanking *Resf1* gene
Resf1_KO-R	tcagcagcttggcaaactgtt	Genotyping primers flanking *Resf1* gene
Rv5_geno-F	ccatcgaggtgggtgtcaaa	Genotyping primers for Rv5 ESC
Rv5_geno-R	cctccttctccagttgcagg	Genotyping primers for Rv5 ESC
5'homology-F	agaactagtggatccgctggacttg aactttgtgtatagttg	Used to create 5' homology arm for Rv5 targeting vector
5'homology-R	tttaactggactactcccgagg	Used to create 5' homology arm for Rv5 targeting vector
3'homology-F	cttcgtatagcatacattatacgaagt tattcccaggaagtcatggtttttg	Used to create 3' homology arm for Rv5 targeting vector
3'homology-R	cttgatatcgaattcaaatgctctttatt gttgcatagtttg	Used to create 3' homology arm for Rv5 targeting vector

3v5IPloxP-F	agtagtccagttaaaggatccggcg gcggcggcggcgggaagcccatt ccaaaccctctcctcggcctggact ctacaggcaagcctattcccaaccc cttgctgggtctcgattcaacaggca agcccattcctaacccactcctggg gcttgattcaacatgaggccggcct cgataagcttgatatcgaattcc	Used to create fragment containing 3x v5-tag, IRES-Puro and loxP site for creating the Rv5 targetting vector
3v5IPloxP-R	gcctgataacttcgtatagcatac	Used to create fragment containing 3x v5-tag, IRES-Puro and loxP site for creating the Rv5 targetting vector
R_CDS1-F	atgaattggaatacaaaacaagag	Used for amplification of 5' half of *Resf1* conding sequence
R_CDS1-R	aggtccactcacaatggtgc	Used for amplification of 5' half of *Resf1* conding sequence
R_CDS2-F	gcaccattgtgagtggacct	Used for amplification of 3' half of *Resf1* conding sequence
R_CDS2-R	tatttaactggactactcccgag	Used for amplification of 3' half of *Resf1* conding sequence
FlagR-F	cattttggcaaagaattcctcgagcc gccaccatggactacaaggacga cgacaaggactacaaggacgacg acaaggactacaaggacgacgac aagggatccggcggcggcggcgg cgccatgatacaaaacaagag	Used to add 3x Flag-tag and a glycine linker to the 5' end of the *Resf1* conding sequence

FlagR-R	cgatatcaagcttatcgagcgcggc cgcctatttaactggactactcccga g	Used to extend 3' end of *Resf1* conding sequence with a complementary sequence for cloning into a pPyCAG plasmids and introduce NotI cutting site

2.10.2 Gibson assembly

Linearised DNA vector (50 ng) was mixed with molar excess of DNA fragments and topped up to 10 µl with ddH_2O. Gibson reaction mix (Tab. 2.4) was added in 1:1 (v/v) ratio and the sample was incubated for 1 h at the room temperature. The negative control was prepared by replacing DNA vector for ddH_2O.

2.10.3 DNA ligation

DNA fragments with complementary sticky ends were ligated using T4 ligase (NEB) according to the recomended protocol (NEB, M0202). Briefly, 50 ng of DNA vector was mixed with molar excess of insert, T4 ligation buffer (10x, NEB), T4 ligase and topped up to the recommended volume by nuclease-free water. Reaction was incubated for 1 h at the room temperature.

2.10.4 Transformation of bacterial cells

Competent *E. coli* were thawed on ice and mixed with plasmid DNA. The cells were incubated for 30 min on ice and heat shocked at 42 °C. The cells were recovered in 500 µl of SOC medium at 37 °C for 1 h. The cell suspension (10 µl to 250 µl) was plated on agar plates supplemented with penicillin (100 µl ml^{-1}) or kanamycin (50 µg ml^{-1}) and grown at 37 °C over night. Single clones were isolated and expanded in LB medium supplemented with particular antibiotic.

2.10.5 Cloning of *Resf1* coding sequence

E14Tg2a ESCs cultured in serum/LIF condition were lysed and the total RNA was extracted using RNeasy Plus Mini kit (QIAGEN, cat. 74136) according to the instructions. RNA was eluted in 100 µl of nuclease-free H_2O and quantified using Nanodrop ND-1000 (Thermo Scientific). One microgram of RNA was mixed with 1 µl of 50 µM oligo(dT) primers (Invitrogen, cat. 18418020), 1 µl of 50 mM dNTPs (Invitrogen, cat. 10297018) and topped to 12.5 µl with nuclease-free H_2O. The reaction was incubated at 65 °C for 5 min and cooled down in ice for at least 2 min. Subsequently, 2 µl 100 mM dithiotreitol, 20 U of RNAseOUT (Invitrogen, cat. 10777019) and 100 U of SuperScript III reverse transcriptase (Invitrogen, cat. 18080044) were added. The reaction mix was adjusted to the final volume of 20 µl with Nuclease-free H_2O and incubated 10 min at 25 °C, 50 min at 50 °C and 15 min at 70 °C.

The *Resf1* coding sequence was amplified from the prepared cDNA in two halves. The 5' halve was amplified using R_CDS1-F and R_CDS1-R primers as described in the method section 2.10.1. The 3' halve was amplified using R_CDS2-F and R_CDS2-R primers as described in the method section 2.10.1. The PCR products were purified using QIAquick PCR Purification Kit (QIAGEN, cat. 28104) and cloned into pCR-Blunt II-TOPO plasmid using Zero Blunt TOPO PCR Cloning Kit (Invitrogen, cat. K2800-20SC). The cloned plasmids were used to transform *E. coli* (method section 2.10.4) and single colonies were expanded in 25 ml of Lysogeny broth medium supplemented with 50 µl ml^{-1} kanamycin at 37 °C and 220 rpm over night. Plasmids were isolated using QIAprep Spin Miniprep Kit (QIAGEN, cat. 27106) and sequenced by Sanger sequencing (method section 2.10.7). The cloned 5' halve of the coding sequence was amplified by PCR using FlagR-F and R_CDS1-R primers extending the 5' end with a complementary overhang, 3x Flag-tag and glycine linker (method section 2.10.1). The cloned 3' halve of the coding sequence was amplified by PCR using R_CDS2-F and FlagR-R primers extending the 3' end with a complementary overhang (method section

2.10.1). The template DNA was digested using DpnI (NEB, cat. R0176S) and amplified DNA was purified using QIAquick PCR Purification Kit (QIAGEN, cat. 28104). The two halves of the prepared *Resf1* coding sequence were cloned into XhoI, NotI digested pPyCAG-IRES-Puro (Chambers et al. 2003) and pPyCAG-PGK-PURO plasmids by Gibson assembly (method section 2.10.2). The episomal expression of the Flag-tagged RESF1 from the prepared plasmids in ESCs was assessed by immunofluorescence (method section 2.11.5) and western blot (method section 2.11.4).

2.10.6 Cloning of CRISPR/eSpCas9 plasmids

CRISPR/eSPCas9 plasmids were prepared according to the published protocol (Ran et al. 2013). eSpCas9(1.1)-T2A-Puro, eSpCas9(1.1)-T2A-eGFP and eSpCas9(1.1)-T2A-mCherry (Adgene #71814) plasmids were cut with BbsI (NEB, cat. no. R0539) and gel purified using QIAquick Gel Extraction Kit (QIAGEN, cat. 28704). The designed complementary pair of single-stranded DNA (Tab. 2.9) were mixed to the final concentration 10 µM and incubated at 95 °C for 5 min. The temperature was gradually decreased to 25 °C (5 °C min^{-1}). Prepared DNA was ligated into eSpCas9 plasmids and used for transformation of *E. coli* as described in method sections 2.10.3 and 2.10.4. The "Resf1_STOP" oligonucleotide was ligated into the eSpCas9(1.1)-T2A-Puro plasmid. The "3'_Resf1" oligonucleotides were ligated into eSpCas9(1.1)-T2A-eGFP plasmid and the "5'_Resf1" oligonucleotides were ligated into eSpCas9(1.1)-T2A-mCherry plasmid. Single clones were selected and expanded in 25 ml of Lysogeny broth medium supplemented with 100 µg ml^{-1} over night at 37 °C, 220 rpm. DNA plasmids were purified by QIAprep Spin Miniprep Kit (QIAGEN, cat. 27106) and validated by Sanger sequencing. Correct plasmids were purified by ethanol precipitation, resuspended in ddH_2O and stored at –20 °C.

Table 2.9 Oligonucleotides used for gRNA cloning

Name	Top oligonucleotide	Bottom oligonucleotide
*Resf1*_STOP	caccacttcctgggactatttaac	aaacgttaaatagtcccaggaagt
3'_*Resf1*_1	cacccaaactggcgcagtgcaaaa	aaacttttgcactgcgccagtttg
3'_*Resf1*_2	caccgaagtgactagcagatcatg	aaaccatgatctgctagtcacttc
3'_*Resf1*_3	caccatactgtgtagttagtactc	aaacgagtactaactacacagtat
3'_*Resf1*_4	cacccaggtgtatatcatagcacc	aaacggtgctatgatatacacctg
5'_*Resf1*_1	cacctaccctattaataaaccttc	aaacgaaggtttattaatagggta
5'_*Resf1*_2	cacctgttcagtaagttgcttaac	aaacgttaagcaacttactgaaca
5'_*Resf1*_3	caccagggaagcagtaggcttact	aaacagtaagcctactgcttccct
5'_*Resf1*_4	caccctcagggtaggctttttata	aaactataaaaagcctaccctgag

2.10.7 Sanger sequencing

To sequence DNA by Sanger sequencing, reaction mix was prepared by mixing DNA (Table 2.10), 5x Sequencing Buffer (Applied Biosystems, cat. 4337455), 1 µl sequencing primers (5 µM each) and 1 µl BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems, cat. 4337455). The reaction at the final volume of 10 µl was incubated according to the program below (Tab. 2.11). The prepared reactions were sequenced by a third party company.

Table 2.10 Recommended amounts of DNA for Sanger sequencing

DNA Type	Recommendation [ng]
dsDNA from standard plasmid	300-400
100-200bp PCR product	1.5-3
200-500bp PCR product	4.5-10
500-1000bp PCR product	7.5-20
1000-2000bp PCR product	15-40
>2000bp PCR product	60-100

Table 2.11 PCR program for Sanger sequencing

Temperature	Time
96 °C	1min
96 °C	15sec
50 °C	10sec
60 °C	2min
Repeat steps 2-4 x24	
4 °C	∞

2.10.8 Cloning of a targeting vector for endogenous tagging of *Resf1*

Genomic DNA was extracted from wild-type E14Tg2a ESCs using DNeasy Blood & Tissue Kit (QIAGEN, cat. 69504). The 5' and the 3' homology arms were amplified from the genomic DNA by PCR using 3' and 5' homology primer pairs (see method section 2.10.1). The $(v5)_3$-IRES-GFP-loxP fragment was amplified using pBS-Tet2-$(Flag)_3$-IRES-GFP (Pantier 2018) and the "3v5IPloxP" primer pair (see method section 2.10.1). Overlapping fragments were cloned into linearised (EcoRI + BamHI) pBlueScript plasmid using Gibson assembly as described in the method section 2.10.2. The cloned plasmid was transformed into *E. coli* (method section 2.10.4) and single colonies were expanded in LB medium supplemented with 100 µg ml^{-1} ampicilin. Plasmid was extracted using QIAprep Spin Miniprep Kit (QIAGEN, cat. 27106) and veryfied by restriction digest and Sanger sequencing (method section 2.10.7).

2.11 Protein analysis

2.11.1 Chromatin immunoprecipitation

Chromatin immunoprecipitation protocol was adapted from Stock et al. 2019. Each of the WT, W10A and N51A NERT ESCs (see tab. 2.3) were plated on three gelatin coated 15 cm plates at density of 25x10^6 cells per

plate. Cells were cultured in 30 ml of serum/LIF medium supplemented with G418 at 37 °C, 7% CO_2 for ~16-17 h. The medium was replaced with 18 ml of the fresh serum/LIF/G418 medium. WT, W10A and N51A NERT cells (Tab. 2.3) were induced by addition of 4-hydroxytamoxifen at the final concentration of 1 µM for 0, 1 or 3 h. Cells were crosslinked for 10 min at 37 °C by formaldehyde (1% v/v). Formaldehyde was quenched by glycine at the final concentration 125 mM for 5 min at the room temperature. Fixed cells were washed three times with PBS and treated with 10 ml of the ChIP Swelling buffer supplemented with protease inhibitor cocktail (Roche, cat. 4693159001) for 10 min at 4 °C. Cells were collected by scraping and centriguation (3,000g, 4 °C, 5 min). The nuclei were resuspended in 2 ml of the Sonication buffer supplemented with the Complete mini protease inhibitor cocktail (Roche). The nuclei were transferred into sonication tubes (Diagenode, C01020031) containing the recommended amount of the sonication beads (Diagenode, C01020031). Chromatin was sonicated in 30 s on/off intervals for 100 min by Bioruptor Pico (Diagenode, B01060010) at 4 °C. Sheared chromatin was centrifuged (16,000g, 15 min, 4 °C) and the supernatant was transferred into 2 ml DNA LoBind tubes (Eppendorf, Cat. no. 0030108078). The amount of chromatin in each sample was estimated by measuring DNA concentration of a diluted sample (1:50 v/v) in 0.1 M NaOH by Nanodrop ND-1000 (Thermo Scientific). The chromatin was flash frozen in liquid nitrogen and stored at –80 °C.

To immunoprecipitate OCT4, 700 µg of sheared chromatin was topped to 1 ml with the Sonication buffer and 10 µg of rabbit anti-mouse Oct4 antibody (Abcam ab19857, Lot: GR315684-7) was added. Samples were rotated overnight at 4 °C. Magnetic DynaBeads Protein G were washed three times with PBS and resuspended in the original volume of the Sonication buffer. Bead suspension (50 µl) was added to each of the reaction. Samples were further rotated for 3 h at 4 °C. The beads were washed once with the Soncitaion buffer, once with the Wash buffer A, once with the Wash buffer B and twice with the TE buffer using a magnet. Each of the washing step was perfomed

in 1 ml volume for 5 min on a wheel at 4 °C. The beads were resuspended in 250 µl of the Elution buffer supplemented with SDS (final concentration 1 %) and incubated on a wheel for 15 min at the room temperature. The beads were collected by centrifutaion (200 g, 3 min) and separated on a magnet. The supernatant was transferred into a new tube and the elution was repeated resulting in 500 µl of the eluted solution.To assess the input chromatin, 50 µl of the chromatin was mixed with 450 µl of the TE buffer.

To remove RNA and to reverse chromatin crosslinks, NaCl and RNase A were added to the final concentration of 160 mM and 20 µg ml^{-1}, respectively. The samples were incubated at 65 °C overnight. EDTA concentration was increased to 5 mM and 200 µg ml^{-1} of proteinase K (Roche) was added. Proteins were degraded at 45 °C for 2 h. Ultra pure Phenol:Chlorophorm:Isoamylalkohol (25:24:1 v/v, Invitrogen) was added to the sample (1:1 v/v) and vigorously vortexed. The liquid phases were separated by centrifugation (16,000g, 10 min) and the aqueous layer was extracted. Sodium acetate 1:10 v/v (3 M, pH 5.2), 2 µl glycogen (20 mg ml^{-1}, Roche) and 2 ml of ice cold ethanol (100 %) were added. DNA was precipitated at –80 °C for 2 h. Samples were centrifuged (16,000g, 30 min, 4 °C), supernatant was discarded and the residual ethanol was left to evaporate for 10 min at the room temperature. DNA was resuspended in 22 µl (immunoprecipitated samples) or 50 µl (input samples) of nuclease-free water. The DNA concentration of diluted samples (1:200 v/v) was determined by Qubit dsDNA HS Assay Kit (Invitrogen cat. Q32854). The shearing efficiency was analysed from the purified input DNA by electrophoresis in 1.6% agarose gel at constant 120 V. Efficiency of the immunoprecipitation was assessed by qPCR.

2.11.2 Preparation of ChIP-seq libraries

The ChIP-seq libraries were prepared by NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (NEB, #E6240L) according to the instructions. Briefly, 10 ng of ChIP DNA was end repaired, dA tailed and ligated with multiplex compatible NEBNext Multiplex primers (NEB, cat. E7335S, E7500S

and E7710S) with a clean up using AMPure XP beads (Beckman Coulter) after each step. DNA within 100-500 bp was separated using AMPure XP beads and DNA was amplified as described in the protocol. Amplified library was purified using AMPure XP beads and analysed using Tapestation 2200 (Agilent). Libraries were pooled together in equimolar ratios and sequenced using HiSeq 4000 sequencer (Illumina).

2.11.3 Co-immunoprecipitation

The E14/T ESCs ($6x10^6$) were transfected with pPyCAG-$(HA)_3$-Nanog-IRES-Puro and pPyCAG-$(Flag)_3$-Resf1-IRES-Puro (see method section 2.10.5) as described in the method section 2.6. In parallel, E14/T ESCs were transfected with only pPyCAG-$(HA)_3$-Nanog-IRES-Puro plasmid as a negative control. The day after the transfection, medium was replaced with fresh serum/LIF medium supplemented with G418 and puromycin and cultured for one day. Cells were washed twice with PBS and detached using Trypsin solution for 3 min at 37 °C. Trypsin was neutralised by 10 ml of serum medium. Cells were centrifuged (1,100 rpm, 3 min) and the pellet was washed twice with 3 ml of cold PBS. The cells were bursted in 250 µl of Swelling buffer with freshly added 0.5% NP-40 and 1x Protease Inhibitor Cocktail (Roche, cat. 4693159001) for 20 min in ice. Nuclei were collected by centrifugation (1,500 rpm, 5 min, 4 °C), resuspended in 1 ml of the NE buffer with freshly added 0.2% NP-40, 0.5 mM DTT and 1x Protease Inhibitor Cocktail (Roche, cat. 4693159001) and transferred into non-stic microtubes (Alpha Laboratories, Cat. LW2410AS). Nuclei were lysed for 1 h at 4 °C in a presence of 150 U of Benzonase nuclease (Novagen, Cat. 71206) on a wheel. Nuclear lysates were cleared by centrifugation (13,300 rpm, 30 min, 4 °C) and transferred into new non-stick tubes. Five percent (50 µl) from each sample was separated for input controls. Thirty microlitres of anti-FLAG M2 affinity beads slurry (Sigma A2220) was washed three times with PBS and resuspended in the original volume of the NE buffer. Thirty microlitres of the beads were added to the remaining nuclear extracts and incubated for 2 h at

the room temperature on a wheel. The beads were washed three times with PBS using magnet. The proteins were eluted by boiling the beads for 5 min in 30 µl of 1x LDS Sample Buffer (Invitroge, cat. B0008) supplemented with 250 mM DTT. The elution was repeated and the two fractions were combined. The input and immunoprecipitated samples were analysed by SDS-PAGE and Western blot as described in the method section 2.11.4.

2.11.4 Protein electrophoresis and western blotting

The protein extracts were mixed with a 4X Bolt™ LDS Sample Buffer (Invitrogen, cat. B0008) and dithiotreitol (10 mM) and boiled for 5 min. Samples together with SeeBlue™ Plus2 pre-stained protein standard (Invitrogen, cat. LC5925) were loaded on a Bolt™ 12% Bis-Tris Plus Gel (Life Biosciences, cat. NW00122BOX) and run at constant 180 V in a Bolt™ MES SDS running buffer (Invitrogen, cat. B0002).

The proteins were transferred onto a nitrocellulose membrane in the Transfer buffer at constant 180 mA over night at 4 °C or at 380 mA for 70 min in ice. The membrane was blocked in 10% milk resuspended in PBS and supplemented with 0.1% Tween-20 (PBS-T) for 1 h or over night. The primary antibodies (Tab. 2.12) were diluted in 5 ml of 5% milk in PBS-T and added on the membrane. The membrane was stained for at least 1 hour at the room temperature while swirling. The membrane was washed four times for 5 min with PBS-T. The secondary antibodies (Tab. 2.12) were diluted in 5 ml of 5% milk in PBS-T and incubated with the membrane for at least 1 h at the room temperature. The membrane was washed four times for 5 min in PBS-T. The membrane was incubated with Pierce™ ECL Western Blotting Substrate (Invitrogen, cat. 32106) for 2 min at the room temperature if an antibody with conjugated horse radish peroxidase was used.

Table 2.12 Antibodies used for Western blotting

Antibody	Dilution (v/v)	Species	Source
Flag	1:5,000	Mouse (monoclonal)	Sigma (F3165)
Nanog	1:1,000	Rabbit (monoclonal)	Bethyl (A300-397A)
Esrrb	1:1,000	Mouse (monoclonal)	R&D (PP-H6707-00)
β-actin-HRP	1:10,000	Mouse (monoclonal)	Abcam (ab2027)
Lamin B1	1:5,000	Rabbit (polyclonal)	Abcam (ab16048)
Rabbit IgG 680 RD	1:1,000	Donkey	LI-COR (926-68073)
Mouse IgG 800CW	1:1,000	Donkey	LI-COR (926-32212)

2.11.5 Immunostaining

Cells were washed with PBS and fixed by 4% paraformaldehyde at the room temperature for 10 min. Cells were washed and permeabilised in 0.3% v/v Triton X-100 in PBS for 15 min at the room temperature. The solution was discarded and cells were blocked in buffer containing 0.1% Triton X-100 and 3% v/v donkey serum for 1 h at the room temperature. Fixed and permeabilised cells were incubated with primary antibodies (Tab. 2.13) diluted in the blocking buffer over night at 4 °C. The cells were washed four times with 0.1% Triton X-100 and incubated with the fluorescent-labelled secondary antibodies diluted (1:1,000 v/v) in the blocking buffer for 1 h at the room temperature. DNA was stained with with 4',6-diamidino-2-phenylindole (DAPI) for 5-10min at the room temperature (1:2,000 dilution in PBS). DAPI was washed once with PBS for 5 min and samples were stored in a mounting solution at 4 °C in dark. The samples were analysed using SP8 Lightening confocal microscope (Leica). Intensity of RESF1 and NANOG proteins in single nuclei were quantified by Matthieu Vermeren.

Table 2.13 Antibodies used for immunostaining

Antibody	Dilution(v/v)	Species	Source
Nanog	1:250	Rabbit (polyclonal)	Bethyl (A300397A)
Nanog	1:500	Rat (monoclonal)	eBioscience (14-5761-80)
v5	1:250	Mouse (monoclonal)	eBioscience (14-6796-80)
Flag	1:500	Mouse (monoclonal)	Sigma cat. (F1804)

2.11.6 Bradford assay

The protein concentration of *E. coli* purified wild-type NANOG and $NANOG_{W10A}$ proteins was determined by a bradford assay according the "microlitre plate protocol" (Bio-rad). The standard calibration solutions were prepared by diluting bovine serum albumin solution (1 mg ml^{-1}, Bio-rad, #5000206) to 50, 100, 200, 300, 400 and 500 ng ml^{-1}. The measured proteins were diluted so their concentration was within the linear range of the assay. Two hundred microlitres of 4-fold diluted Protein Assay Reagent (Bio-Rad, #5000006) was mixed with 10 µl of a sample in a 96-well microplate. Solutions were incubated for 5 min in dark and absorbance was measured at 600 nm. Concentration of the protein samples was determined from interpolation of the linear regression of the standard curve.

2.11.7 Electrophoretic mobility shift assay (EMSA)

Pure recombinant wild-type NANOG and $NANOG_{W10A}$ proteins were kindly provided by Nick Mullin. The concentration was determined by Bradford assay (method section 2.11.6). The EMSA was performed with the 26mer "TTA AACCTGTTAATGGGAGCGCATTG" containing a Nanog binding motif (Jauch et al. 2008). Reactions were prepared by mixing 1 µl of 10 nM DNA probe, 2 µl of EMSA buffer, 1 µl of 2.5 mM DTT, 1 µl of Tween-20, 1 µl of 250 ng µl^{-1} Poly(dG:dC) competitor (InvivoGen, cat. tlrl-pgcn), 4 µl H_2O and 1 µl of a protein sample at various concentrations. A negative control was prepared by replacing protein with water. Samples were incubated for 30 min at the room

temperature and analysed by electrophoresis in 2% agarose gel. Labelled DNA probe was visualised at 700 nm. The ratio of bound and unbound DNA was quantified using Odyssey FC Imaging System (LI-COR).

2.12 Quantitative PCR of reverse transcribed RNA (RT-qPCR)

Cultured cells were washed with PBS and lysed on the plate using 350 µl RLT buffer (QIAGEN) supplemented with 2-mercaptoethanol (1.4 M; Sigma, cat M6250). The cell lysates were stored at –80 °C. The RNA was extracted using RNeasy Plus Mini kit (QIAGEN, cat. 74136) according to the instructions. The RNA was resuspended in 20 µl to 50 µl of nuclease-free H_2O and stored at –80 °C.

Purified RNA was reverse transcribed using SuperScript III (Invitrogen, cat. 18080085) according to the recommended protocol. Briefly, 0.2 µg to 1 µg of RNA was mixed with 1 µl of 50 µM random hexamers (Invitrogen, cat. N8080127) and 1 µl of 10 mM dNTPs (Invitrogen, cat. 10297018) in the final volume of 11 µl. The mix was incubated at 65 °C for 5 min and 2 min in ice. Subsequently, 2 µl of 100 mM DTT, 20 U of RNAseOUT (Invitrogen, cat. 10777019) and 100 U of SuperScript III reverse transcriptase (Invitrogen, cat. 18080044) were added. The reaction mix was adjusted to the final volume of 20 µl with nuclease-free H_2O and incubated 10 min at 25 °C, 1 h at 50 °C and 15 min at 70 °C. The cDNA was diluted 1:10 (v/v) before the quantification.

The qPCR reaction mix was prepared by mixing 5 µl of the cDNA, 4.5 µl of Takyon™ SYBR 2X qPCR Mastermix (Eurogentec) and 0.5 µl of a primer pair mix (10 mM each) (Tab. 2.14) in a 384-well plate in duplicates. Specificity of the used primers (Tab. 2.14) was determined from a meting curve and their efficiency (>90%) by a linear regression.

Table 2.14 Primers used for RT-qPCR

Target	Forward primer	Reverse primer
Tbp	ggggagctgtgatgtgaagt	ccaggaaataattctggctca
Pou5f1	ccccaatgccgtgaagttg	tcagcagcttggcaaactgtt
Pou3f1	ccatcgaggtgggtgtcaaa	cctccttctccagttgcagg
Nanog	aggatgaagtgcaagcggtg	tgctgagcccttctgaatcag
Esrrb	cgattcatgaaatgcctcaa	cctcctcgaactcggtca
Rex1	cagctcctgcacacagaaga	actgatccgcaaacacctg
Fgf5	aaaacctggtgcaccctaga	catcacattcccgaattaagc
Otx2	gacccggtacccagacatc	gctcttcgattcttaaaccatacc
Blimp1	cccctcatcggtgaagtcta	tggtggaactcctctctgga
Prdm14	tcaattcactcccgaagtacca	ccggggatggcagaagtaaa
Ap2γ	atccctcacctctcctctcc	ccagatgcgagtaatggtcgg

2.13 Computational analysis

Table 2.15 Summary of used sequencing-based data sets.

Data set	Cell lines	Tamoxifen treatment [h]	Reference
Nanog ChIP-seq	$NERT_{WT}$ $NERT_{W10A}$ $NERT_{N51A}$	0, 0.5, 1, 3, 6	Jingchao Zhang (unpublished)
RNAPII ChIP-seq	$NERT_{WT}$ $NERT_{W10A}$ $NERT_{N51A}$	0, 0.5, 1, 3, 6	Jingchao Zhang, Nick Mullin (unpublished)
Sox2 ChIP-seq (2 replicates)	$NERT_{WT}$ $NERT_{W10A}$ $NERT_{N51A}$	0, 1, 3	Jingchao Zhang, Nick Mullin (unpublished)
Oct4 ChIP-seq	$NERT_{WT}$ $NERT_{W10A}$ $NERT_{N51A}$	0, 1, 3	This thesis
Microarray	$NERT_{WT}$ $NERT_{W10A}$	0, 1, 2, 3, 4, 5, 6	Jingchao Zhang (unpublished)
DNAseI hypersensitivity	E14Tg2a	N/A	Joshi et al. 2015 (SRX1158314)

2.13.1 ChIP-seq analysis

For Oct4 data sets (see table 2.15), raw reads were trimmed and quality filtered (-q 30) using Trim Galore in the pair-end mode (Krueger 2019). Filtered reads were aligned to mm10 genome assembly using the Bowtie2 tool (Langmead et al. 2012) and duplicates were removed using the Picard toolkit (Institute 2018). Input samples were merged for each of the examined cell line to increase the genome coverage. Reads were further filtered, sorted and indexed using SAMTools (-q 30) (Li et al. 2009). The peaks were called using the inputs from the particular cell line by MACS2 (Zhang et al. 2008) with the parameters –SPMR -q 0.01 -bdg. Data for RNAPII, SOX2 and NANOG (see table 2.15) were analysed in the similar way by Alison McGarvey. The

reads overlapping mm10 black list regions and *Otx2* an *Nanog* exons were removed using BEDTools (Quinlan et al. 2010). To determine the average binding after NANOG induction, filtered BAM files of all the time point samples after the tamoxifen treatment were merged for each cell line separately using the SAMTools.

Genome tracks and heatmaps were generated using deeptools (Ramírez et al. 2016). Firstly, input and library size normalised genome tracks were generated using bamCopare function with parameters -bs 10, -smoothLength 30, -e, –effectiveGenomeSize 2652783500, -scaleFactorsmethod None, normalize Using CPM. For the pair end samples, only one read per pair was counted. Heatmaps were generated from the bigWig files using computeMatrix and plotHeatmap functions.

To visualise relative distances between ordered peaks and the nearest TSSs, a list of ordered nanog peaks corresponding to the heatmaps was annotated using ChIPpeakAnno package (Zhu et al. 2010). The absolute distance of each peak to the nearest TSS was determined and the peaks were grouped based on whether they are within 3 kb from the nearest TSS or not. The data was visualised using ggplot2 (Wickham 2016).

2.13.2 Identification of reproducible NANOG binding sites

In order to quantify changes at the NANOG binding sites, a unified list of WT, W10A and N51A NANOG binding sites was generated. Firstly, reproducible binding sites for each mutant was generated by removing sites which were found in less than three time points. Secondly, the size of each binding site was determined by extending the peak summit by 250 bp to each direction. Finally, the binding sites of the NANOG mutants were merged into a single list. The overlapping binding sites were resolved by keeping the sites with the highest confidence score. This resulted in 37,499 reproducible NANOG binding sites.

2.13.3 Analysis of DNAseI data set

The raw DNAseI data from E14Tg2a mouse ESCs cultured in serum/LIF conditions (Joshi et al. 2015) was retrieved from the SRA database (SRR2173788). Reads were trimmed and quality filtered (-q 30) using Trim Galore (Krueger 2019). Filtered reads were aligned to the mm10 genome assembly using the Bowtie2 tool (Langmead et al. 2012) and duplicates were removed using the Picard toolkit (Institute 2018). Heatmaps were generated using deeptools (Ramírez et al. 2016) using bamCoverage function with parameters -bs 10, -smoothLength 30, -e, –effectiveGenomeSize 2652783500, -scaleFactorsmethod None, normalize Using CPM.

2.13.4 K-means clustering of NANOG binding sites

To prevent biases caused by computing relative changes from very small values, NANOG bindings sites where neither OCT4, SOX2 or RNAPII bind were removed from the list of reproducible NANOG binding sites (methods 2.13.1) using the BEDTools (Quinlan et al. 2010). All the peaks within undefined chromosomes and the mitochondrial chromosome were removed as well. Reads over the peaks for each sample were counted using Rsubread package (Liao et al. 2019). To normalise the samples and account for the different sequencing depths, the number of reads per million unique and aligned reads (RPM) was calculated.

To determine combination of different patterns of OCT4, SOX2 and RNAPII responses regardless of the magnitude of binding at a particular site, the data were scaled using the min-max scaling. Formally if the signal of a transcription factor *f*, at the time point *t* and the site *p* is *X* then the scaled value X_{pftSC} is:

$$X_{pftSC} = \frac{X_{pft} - min_{pf}(X)}{max_{pf}(X) - min_{pf}(X)}$$

This scales the values for each factor at each site from 0 to 1. The scaled values were clustered using k-means function in R base with parameters iter.max = 1000, centers = 7. The number of clusters was firstly estimated by

the elbow method which determines square root differences of each centroid as a function of the number of clusters. After visual evaluation, seven was found as the optimal number of clusters.

2.13.5 Calculation of association between NANOG peaks and differentially expressed genes

First, annotations of the known genes within mm10 genome were retrieved using the BioMart R package (Kinsella et al. 2011). Based on expression matrices and a list of differentially expressed genes after NANOG induction generated by Alison McGarvey, genes were grouped into four categories: 1. Genes activated by NANOG induction, 2. Genes repressed by NANOG induction, 3. Genes unaffected by NANOG induction but expressed in ESCs, 4. Genes unexpressed in ESCs. For each reproducible NANOG peak (method section 2.13.2), the distance to every gene within 500 kb was calculated. Using a normalised RPM value of wild-type NANOG and a distances to genes for each peak, the transcriptional factor association score (TFAS) as described by Ouyang et al. was calculated. This score is directly proportional to a binding intensity at the binding site and exponentially decreases with the distance. Formally, TFAS for combination of a peak *p* and a gene *g* is equal to

$$TFAS_{gp} = i_p e^{-d_k/d_0}$$

where i_p is the intensity of the peak *p*, d_k is the distance between the peak *p* and the gene *g* in kb and d_0 is a constant. Varying the constant from 5 kb to 500 kb had negligible effect on reached conclusions. Constant equal to 50 kb was selected for figure 5.4. In order to compare association of NANOG clusters to genes, NANOG peaks were divided by the cluster (methods 2.13.4) and TFAS was calculated for each gene. Formally, TFAS for gene *g* is

$$TFAS_g = \sum_p TFAS_{gp}$$

where $TFAS_{gp}$ is association score of gene g with the peak p. To determine likelihood of a particular cluster being associated more with DEGs than the expressed genes, Wilcoxon rank test was used to compare TFAS of differentially expressed genes (induced and repressed genes separately) with the expressed genes for each cluster separately.

3 Requirements of *Resf1* for stem cell identity

RESF1 was identified as a NANOG binding protein (Gagliardi et al. 2013). The NANOG tryptophan repeat (WR) is required for NANOG function in ESC self-renewal (Mullin et al. 2008). Interestingly, disruption of the NANOG WR also prevents NANOG-RESF1 interaction (Mullin et al., unpublished). This lead me to investigate whether RESF1 is required for NANOG function. First, cellular localisation of RESF1 was examined.

3.1 RESF1 localisation in ESCs

To identify the cellular localisation of RESF1, the *Resf1* coding sequence was cloned into pPyCAG-IP (Chambers et al. 2003) with a triple Flag-tag at the N-terminus (Fig. 3.1A). Flag-tagged RESF1 was episomally expressed in ESCs and visualised by immunostaining. This showed that RESF1 is mostly co-localised with DAPI in the nucleus (Fig. 3.1B). However, lower levels of RESF1 were also found in the cytoplasm (Fig. 3.1B).

Episomal expression could lead to non-physiological levels of RESF1 and bias observed localisation of RESF1. Moreover, exogenous expression does not allow quantification of any correlation between NANOG and RESF1 expression. Therefore, endogenously tagged *Resf1* cell lines were generated using CRISPR/Cas9 (method section 2.9.2). A PAM site in the proximity of the *Resf1* stop codon was targeted (Fig. 3.2B). The targeting vector was co-transfected with the eSpCas9 plasmid to extend *Resf1* coding sequence by a triple v5 tag, stop codon, internal ribosome entry site (IRES), green fluorescent protein (GFP) and a loxP site (Fig. 3.2A, method section 2.9.2). Cells with successfully edited *Resf1* locus are expected to express v5-tagged RESF1 and GFP. Single cells with high GFP signal were obtained using Fluorescence Activated Cell Sorting (FACS) (Fig. 3.2B) and expanded in culture prior to genotyping. The genotyping primer pair flanks the inserted

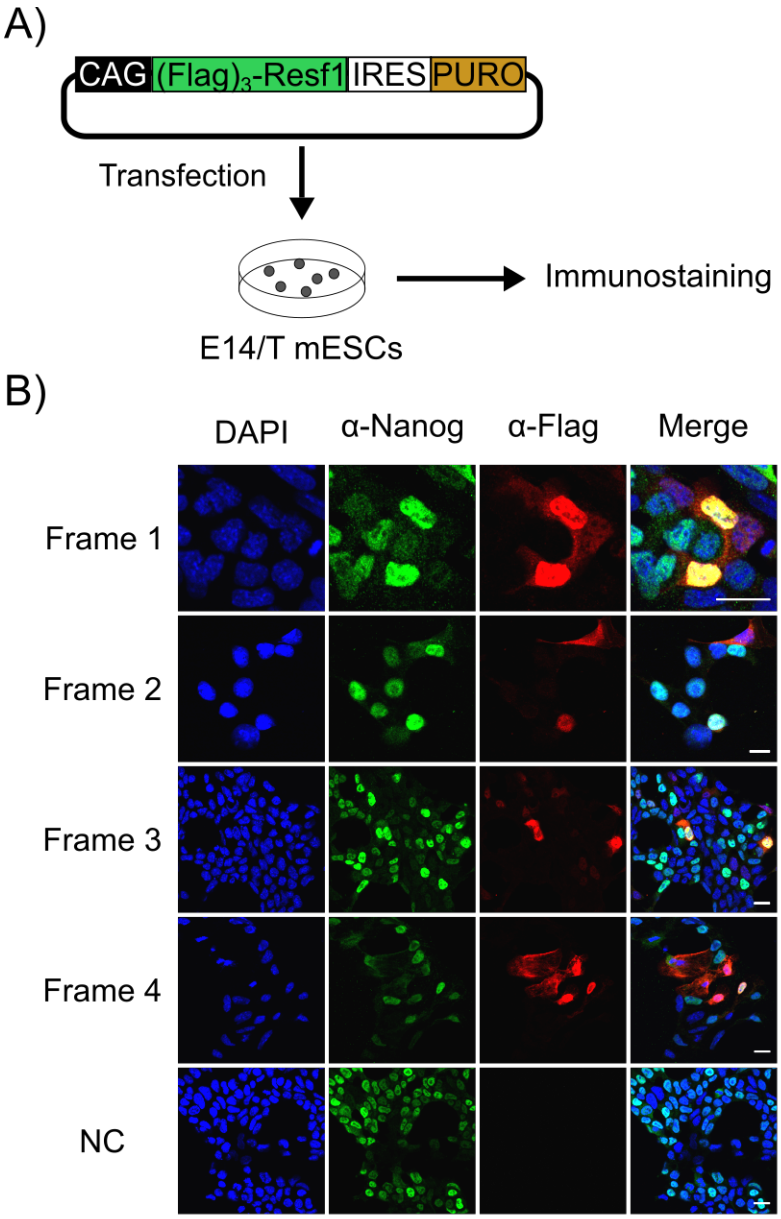


Figure 3.1: Episomally expressed RESF1 localises predominantly to nucleus of ESCs. A) pPyCAG-$(Flag)_3$-RESF1-IP was episomally expressed in E14/T ESCs and analysed by α-Flag immunostaining. **B)** ESCs transiently expressing pPyCAG-$(Flag)_3$-RESF1-IP were stained for Flag (red), NANOG (green) and DAPI (blue). Four representative frames of the examined sample at different magnifications are shown. Negative control (NC) did not contain the α-Flag antibody. White bars represent 20 µm.

region so that a successful insertion of the designed sequence increases the length of the amplified DNA from 735 bp to 2,380 bp (Fig. 3.2A). Three cell lines (c25, c26 and c30) were obtained with a single band of the correct size (Fig. 3.2C). Sanger sequencing of the amplified genomic DNA confirmed correct insertion of the designed targeting vector. Endogenously tagged *Resf1* cell lines are referred further in the text as Rv5 c25, c26 and c30.

The localisation of endogenous RESF1 in the Rv5 ESCs was analysed by v5 immunostaining and confocal microscopy (Fig. 3.3). All three analysed cell lines consistently express RESF1 in the nucleus. RESF1 expression was observed in NANOG negative cells as well. Intensity of RESF1 and NANOG signal in single nuclei was quantified. Kendall's rank correlation between RESF1 and NANOG expression in Rv5 cell lines was calculated (Fig. 3.4). The correlation between RESF1 and NANOG protein levels is low with Kendall coefficient values below 0.4. Moreover, measured correlation between RESF1 and NANOG varies between tested Rv5 cell lines which is likely caused by a limited number of measured cells. This suggests that RESF1 and NANOG protein levels do not have a high degree of monotonic correlation although the correlation is significantly different from the mean signal in all of the measured cell lines (Fig. 3.4).

3.2 Requirements of *Resf1* for ESC self-renewal

3.2.1 Deletion of *Resf1* in ESCs

The fact that RESF1 is expressed in ESCs and interacts with the core pluripotency transcription factors NANOG (Gagliardi et al. 2013) and OCT4 (Berg et al. 2010), suggests that it has a role in the ESC self-renewal. To assess the function of RESF1, *Resf1* gene in ESCs was deleted using CRISPR/Cas9 system (method section 3.2.1). The regions flanking the entire *Resf1* locus were targeted by two sets of guide RNAs (gRNAs) (Fig. 3.5A). Single clones from the targeted ESC population were expanded and the deletion of *Resf1* was validated by PCR genotyping and Sanger sequencing.

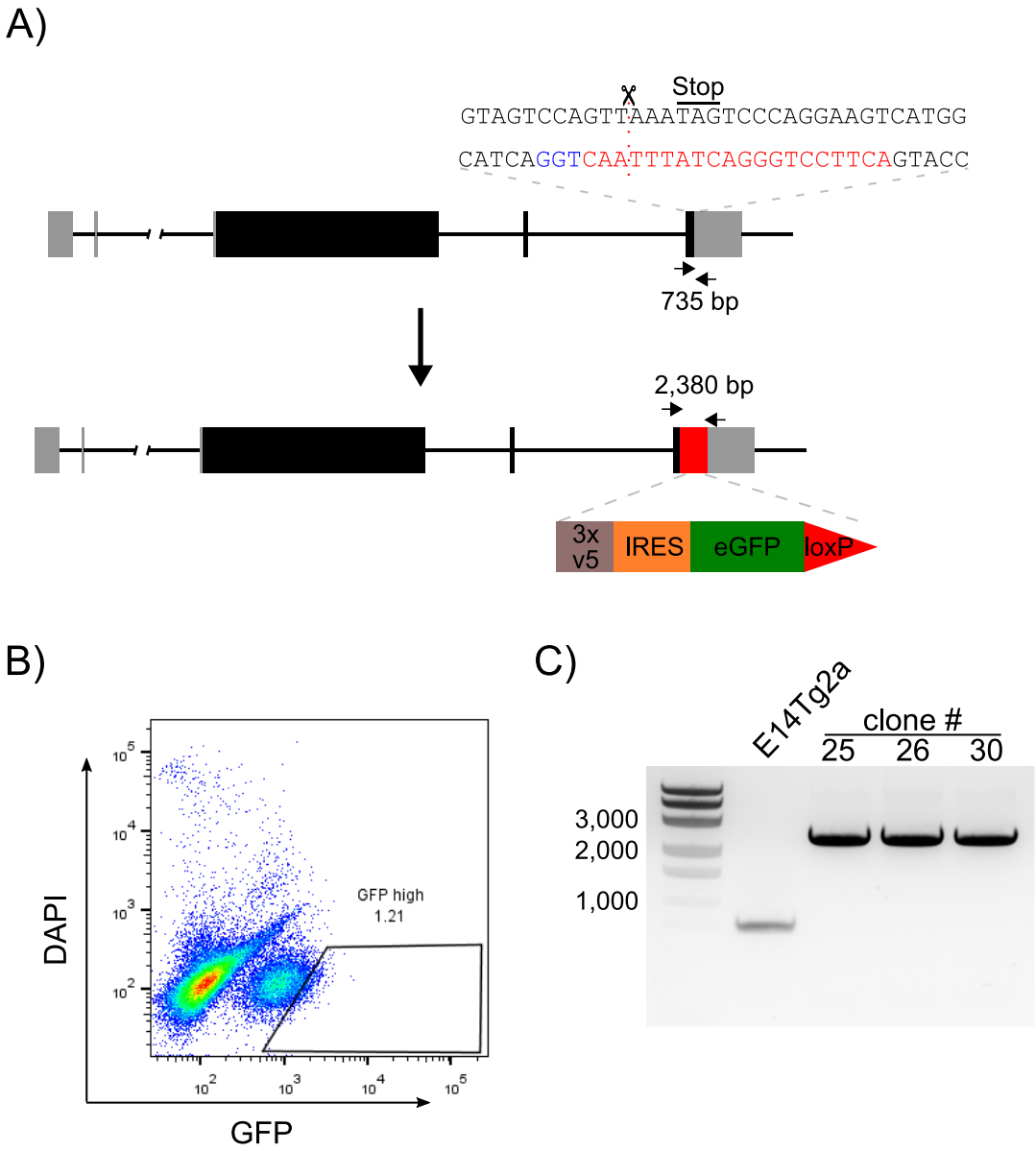


Figure 3.2: Tagging of endogenous RESF1. A) CRISPR/Cas9 system was used to insert $(v5)_3$-tag, stop codon, IRES, eGFP and a loxP sequence before the *Resf1* stop codon. The designed guide RNA (red) targets the PAM site (blue) in the proximity of the *Resf1* stop codon. **B)** One day post transfection, GFP-high cells were sorted and genotyped **(C)** using primers indicated with the arrows in **A)**.

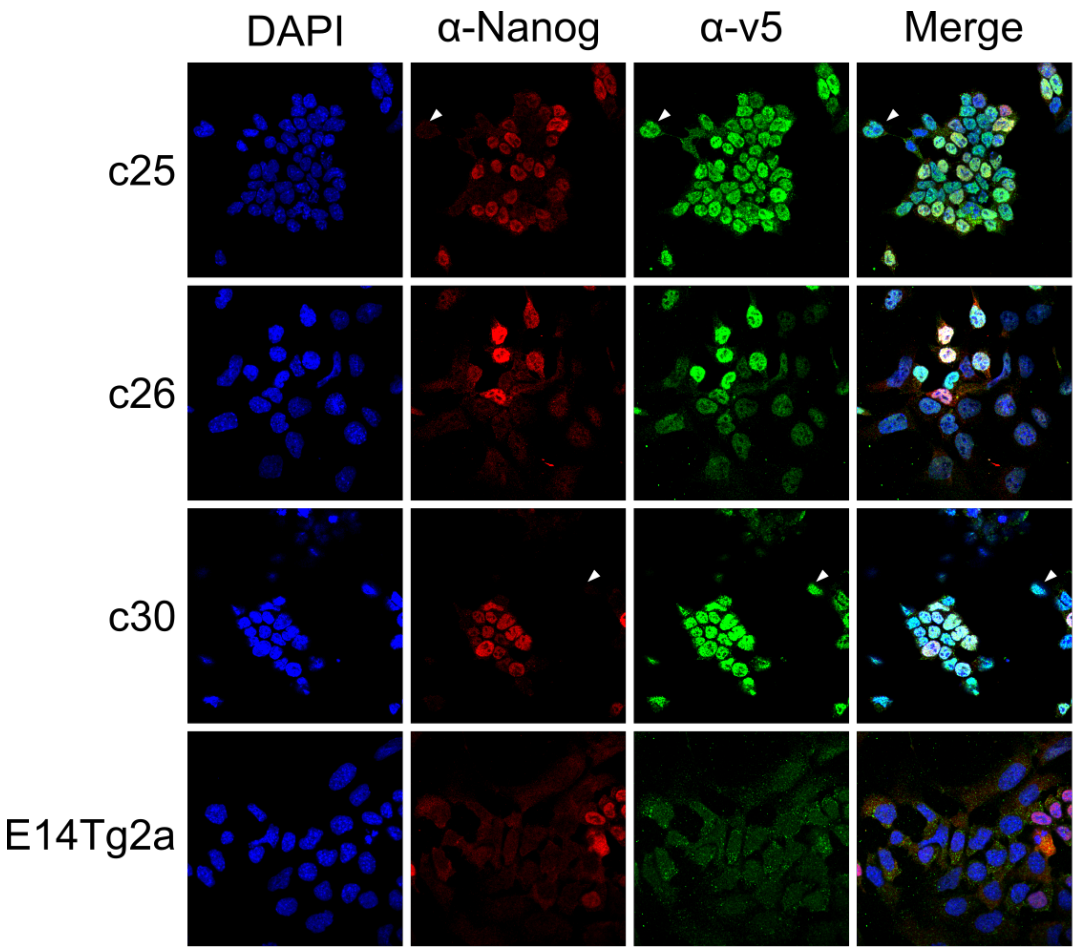


Figure 3.3: RESF1 localises to nuclei of ESCs. RESF1-$(v5)_3$-IRES-eGFP (Rv5) clones 25, 26 and 30 (c25, c26 and c30) were stained with DAPI (blue) and α-Nanog (red), α-v5 (green) antibodies. White arrow heads indicate cells expressing RESF1 but no NANOG. E14Tg2a ESCs were used as a negative control. Representative fields of the analysed samples are shown.

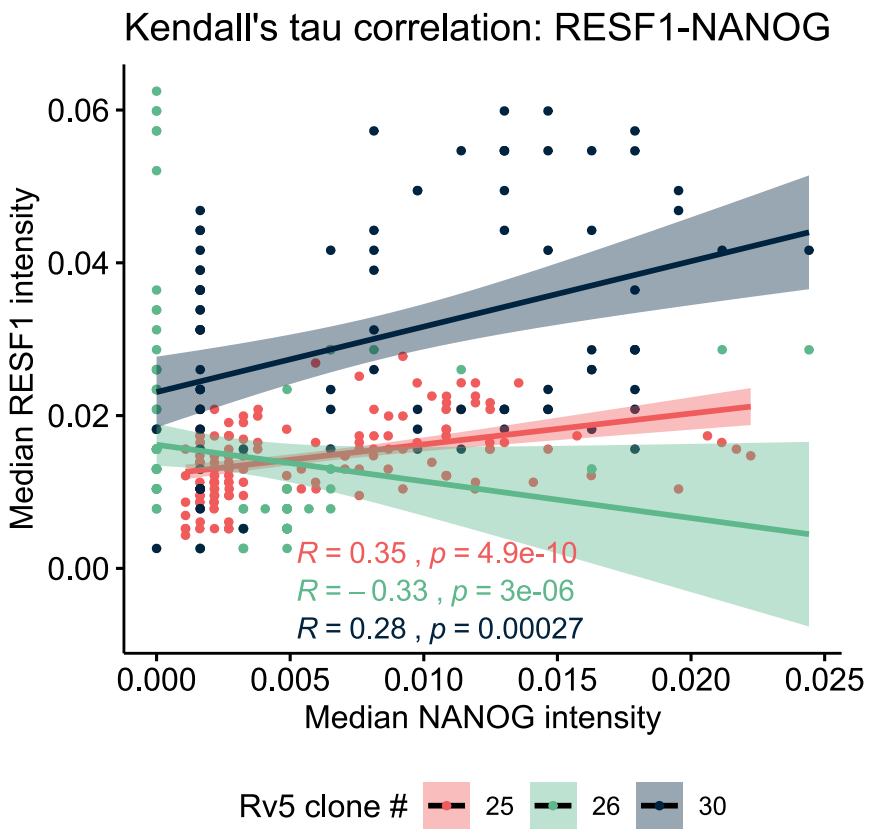


Figure 3.4: NANOG and RESF1 expression do not correlate. RESF1 and NANOG immunofluorescent intensities in individual cell nuclei of Rv5 clones 25, 26 and 30 were quantified. Kendall's tau correlation between NANOG and RESF1 expression in the individual cell lines was computed. P-values and R coefficients are reported for each cell line.

For PCR genotyping, two primer pairs were used (Fig. 3.5A). The primer pair A, amplifies a region in the intron four of *Resf1*. The primer pair B is complementary to regions flanking the *Resf1* gene and spans over 29 kb. Primer pair B do not amplify a product from wild type cells due to a large distance between primers. Upon deletion of *Resf1*, these regions come into proximity and allow amplification of ~650 bp product using the primer pair B (Fig. 3.5A).

Two clones (4 and 24) lacking *Resf1* were identified (Fig. 3.5B). DNA around the deletion junction was amplified using the primer pair B, sub-cloned and sequenced (Fig. 3.6). Two separately cloned PCR products of the *Resf1*$^{-/-}$ clone 4 produced different sequences. From three separately cloned PCR products of *Resf1*$^{-/-}$ clone 24, two had the same sequence (sample 1) and one produced a different sequence (sample 2). All sequenced PCR products map to different regions flanking *Resf1*, indicating deletions of various sizes as shown in Fig. 3.6. This pattern is consistent with deletion of both *Resf1* alleles in *Resf1*$^{-/-}$ clones 4 and 24.

3.2.2 Effects of *Resf1* deletion on expression of pluripotency genes

Resf1$^{-/-}$ ESCs cultured in serum/LIF seem to form more differentiated colonies (Fig. 3.7A). Therefore, expression of the key pluripotency markers was assessed by quantitative PCR and immunoblotting (Fig. 3.7). *Resf1*$^{-/-}$ ESCs express significantly less *Pou5f1*, *Sox2*, *Nanog* and *Esrrb* mRNA (Fig. 3.7B). The protein levels of NANOG, OCT4, SOX2 and ESRRB are also decreased in this culture condition (Fig. 3.7C, D). However, due to large sample variation, decreased proteins levels in the *Resf1*$^{-/-}$ ESCs could not be determined with a statistical confidence.

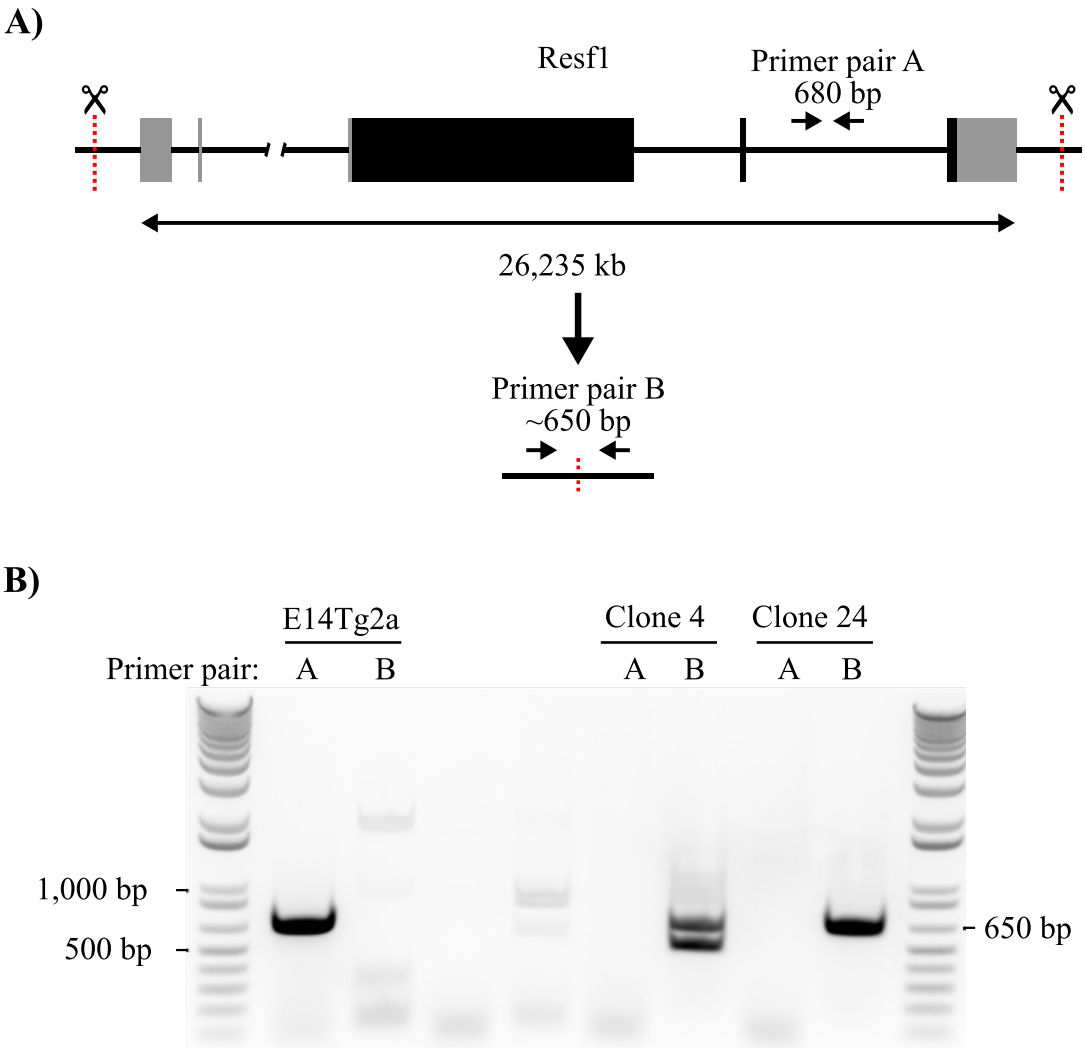


Figure 3.5: Creation of *Resf1*$^{-/-}$ ESC lines. A) Scheme of *Resf1* gene. Gray boxes - untranslated regions, black boxes - coding sequence, lines - introns. *Resf1* was targeted with two sets of gRNAs (indicated in red) to delete entire *Resf1* gene. Primer pairs A and B used for genotyping are indicated by arrows. **B)** Genotyping of the targeted ESCs clones alongside wild type E14Tg2a cells was performed by PCR and analysed by electrophoresis.

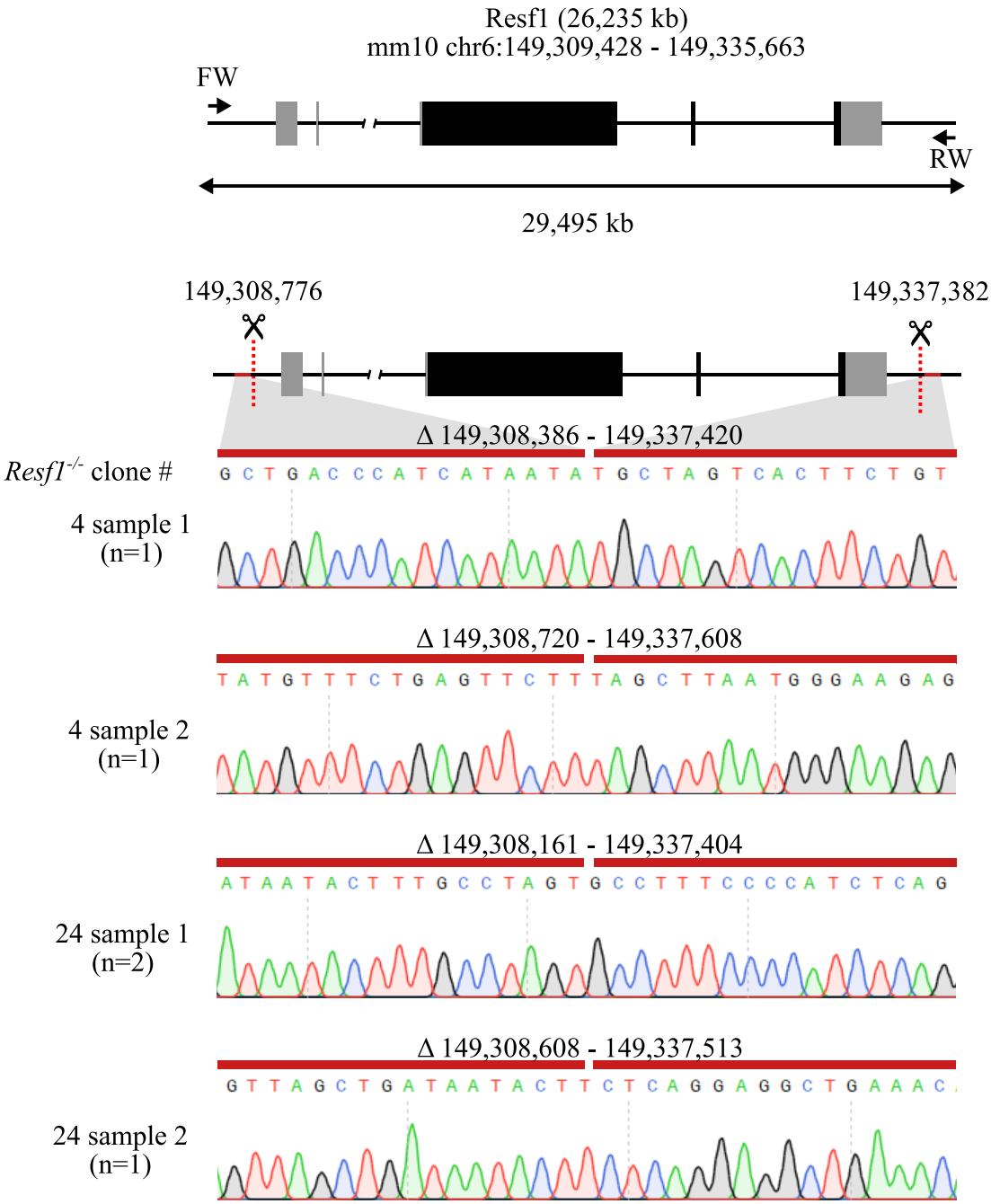


Figure 3.6: Sequencing of *Resf1*$^{-/-}$ cell lines. Deletion of *Resf1* locus in *Resf1*$^{-/-}$ clones 4 and 24 was assessed by Sanger sequencing. Forward (FW) and reverse (RV) primers (arrows) were used to amplify regions flanking *Resf1* gene. Individual PCR products were cloned and sequenced. DNA trace of each allele of each clone at the deletion junction is shown. The genome coordinates correspond to the mm10 genome assembly. Gray boxes - untranslated regions, black boxes - coding sequence, lines - introns.

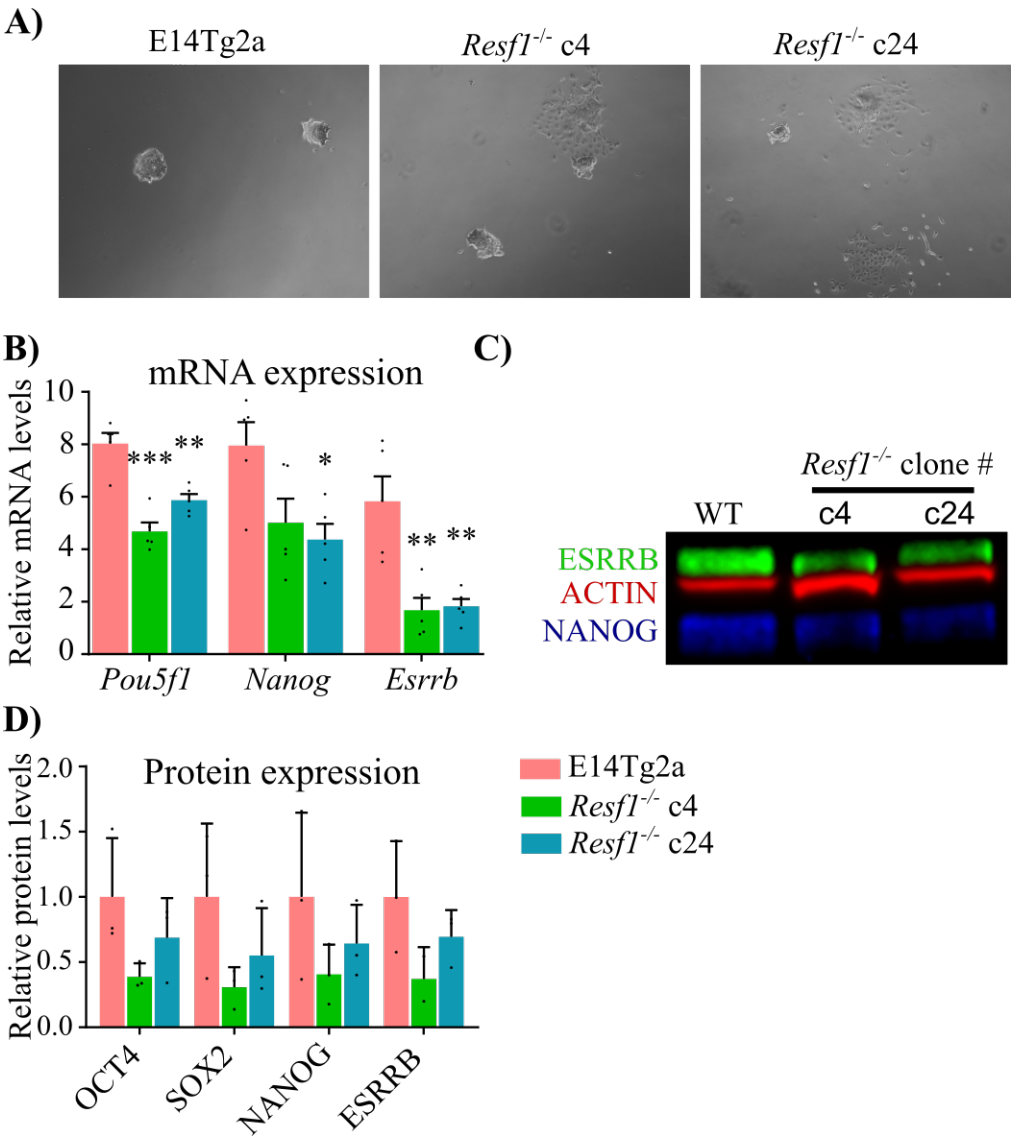


Figure 3.7: *Resf1*$^{-/-}$ ESCs cultured in serum/LIF express lower levels of pluripotency markers. **A)** Microphothographies of cell colonies formed by E14Tg2a, *Resf1*$^{-/-}$ c4 and c24 ESC lines in Serum/LIF. **B)** Relative mRNA expression levels of indicated genes over TBP measured by RT-qPCR. Bars represent mean±s.d. **C)** Representative immunoblot of ESRRB, NANOG and ACTIN levels in examined cell lines. **D** Nuclear levels of OCT4, SOX2, NANOG and ESRRB proteins relative to β-ACTIN in *Resf1*$^{-/-}$ and E14Tg2a ESCs were measured by immunoblot. Bars represent mean±s.d. values normalised to E14Tg2a samples. Statistical significance between *Resf1*$^{-/-}$ and E14Tg2a cells was measured with unpaired, two-sided T-test and p-values were corrected for multiple testing using Holm-Šídák method. n=6 (mRNA), n=3 (protein),; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

3.2.3 Quantification of self-renewal capacity of *Resf1*$^{-/-}$ ESCs

ESCs can self-renew in the presence of fetal calf serum and leukemia inhibitory factor (LIF). Self-renewing ESCs are marked by expression of alkaline phosphatase. In the absence of LIF, ESCs begin to differentiate. This process is marked by a gradual down-regulation of pluripotency markers like alkaline phosphatase and a decrease of cell proliferation. In serum/LIF, *Resf1*$^{-/-}$ ESCs have decreased expression of pluripotency markers and show signs of an increased propensity to differentiate (see section 3.2). This suggests that *Resf1* may influence the efficiency of ESC self-renewal.

To quantify *Resf1* requirements for self-renewal capacity, *Resf1*$^{-/-}$ ESCs were cultured at clonal density in fetal calf serum (FCS) with or without LIF for six days. Formed colonies were stained for alkaline phosphatase (alkaline phosphatase), categorised by their morphology and quantified (Fig. 3.8). In the presence of LIF, both E14Tg2a and *Resf1*$^{-/-}$ ESCs formed colonies with similar morphology (Fig. 3.8A). The number of undifferentiated colonies was significantly different for *Resf1*$^{-/-}$ clone 24, whereas clone 4 formed a similar number of undifferentiated colonies as the wild type control. No significant differences were observed for colony counts with mixed and differentiated morphologies. *Resf1*$^{-/-}$ ESCs formed less mixed colonies than the wild-type control (Fig. 3.8B). In the absence of LIF, E14Tg2a ESCs retained a proportion of colonies with a mixed morphology which was not the case for *Resf1*$^{-/-}$ cells (Fig. 3.8A). These results show that *Resf1* deletion has a negative effect on ESC self-renewal appearing after LIF withdrawal.

3.2.4 Effects of forced *Resf1* expression on ESC self-renewal

As the absence of RESF1 decreases ESC self-renewal capacity, it is possible that increased RESF1 expression may enhance ESC self-renewal. To assess this, E14/T ESCs were transfected with pPyCAG-$(Flag)_3$-RESF1-IP

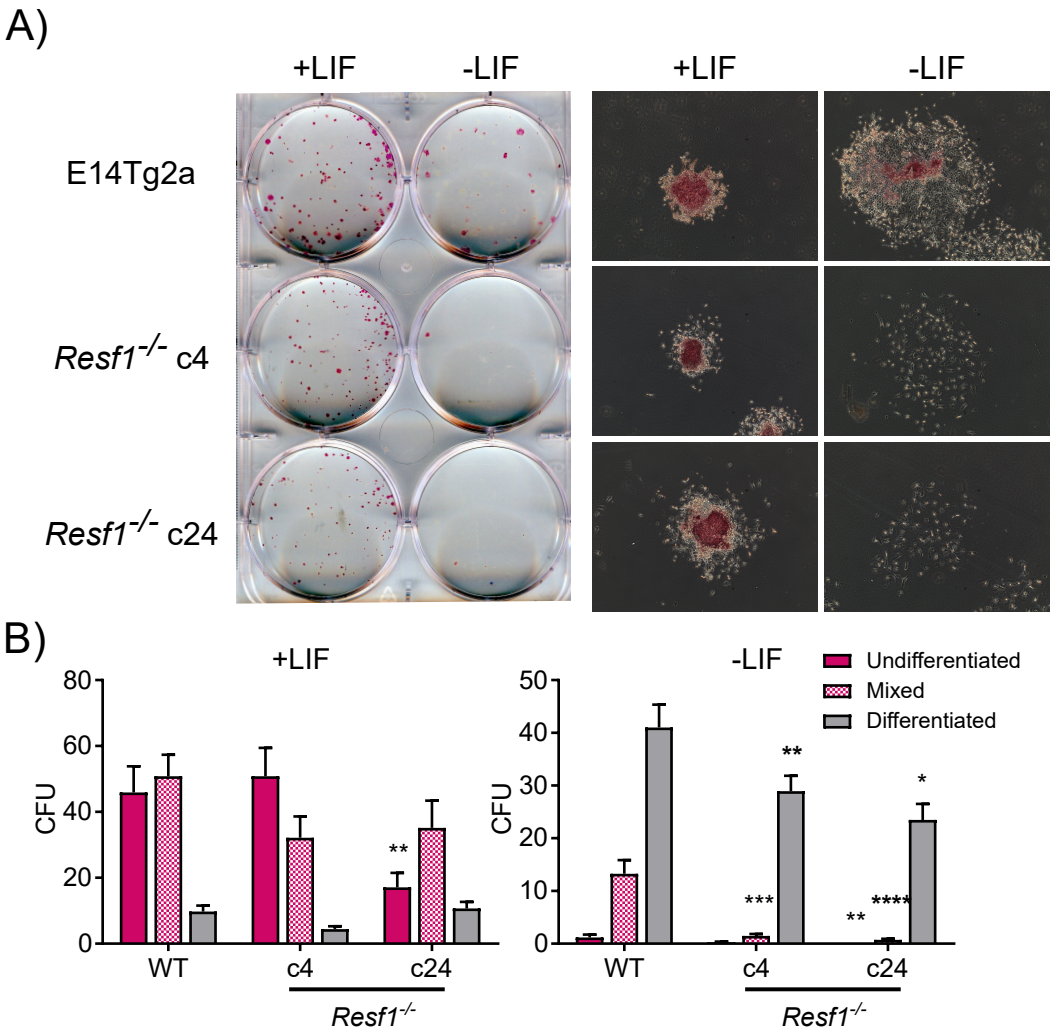


Figure 3.8: *Resf1*$^{-/-}$ ESCs have decreased capacity to self-renew. *Resf1*$^{-/-}$ ESCs and E14Tg2a ESCs were cultured with or without LIF at a clonal density. Colony morphology **(A)** and number of colonies **(B)** were assessed by alkaline phosphatase staining after 6 days. Bars represent mean±sem. Statistical significance between *Resf1*$^{-/-}$ and wild-type cells was determined by Dunn's test. n=15, * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

plasmid (Fig. 3.9A). The E14/T ESCs stably express polyoma large T antigen which enables propagation of the transfected plasmid (Chambers et al. 2003). Self-renewal of the transfected cells was quantified by alkaline phosphatase staining of formed colonies after eight day culture in presence or absence of LIF (Fig. 3.9A). As expected, ESCs transfected with an empty vector (EV) formed alkaline phosphatase positive colonies only in the presence of LIF. Cells overexpressing NANOG formed colonies both in presence and absence of LIF. Episomal expression of RESF1 did not sustain self-renewal in the absence of LIF. However, cell transfected with the RESF1 transgene produced significantly more differentiated colonies similarly to the cells expressing NANOG transgene (Fig. 3.9). In the presence of LIF, episomal RESF1 expression resulted in approximately 2-fold higher number of undifferentiated colonies compared to the negative control (Fig. 3.9B).

3.3 Requirements of *Resf1* for NANOG function

NANOG was identified based on the capacity to sustain LIF-independent ESC self-renewal (Chambers et al. 2003). The NANOG WR is essential for NANOG function (Mullin et al. 2011) and interaction with RESF1 (Mullin et al. unpublished). I showed that deletion of *Resf1* decreases ESC self-renewal efficiency (Fig. 3.7, 3.8). Therefore, RESF1 may be required for NANOG function to sustain LIF-independent self-renewal.

First, the RESF1-NANOG interaction was assessed by co-immunoprecipitation of ectopically expressed $(Flag)_3$-RESF1 and $(HA)_3$-NANOG proteins in E14/T ESCs (method section 2.11.3). $(Flag)_3$-RESF1 was enriched after Flag immunoprecipitation. NANOG co-immunoprecipitated together with RESF1. The negative control sample contained similar levels of HA-NANOG in the input but no detectable NANOG protein in the immunoprecipitate. This confirms that the RESF1-NANOG interaction is specific (Fig. 3.10).

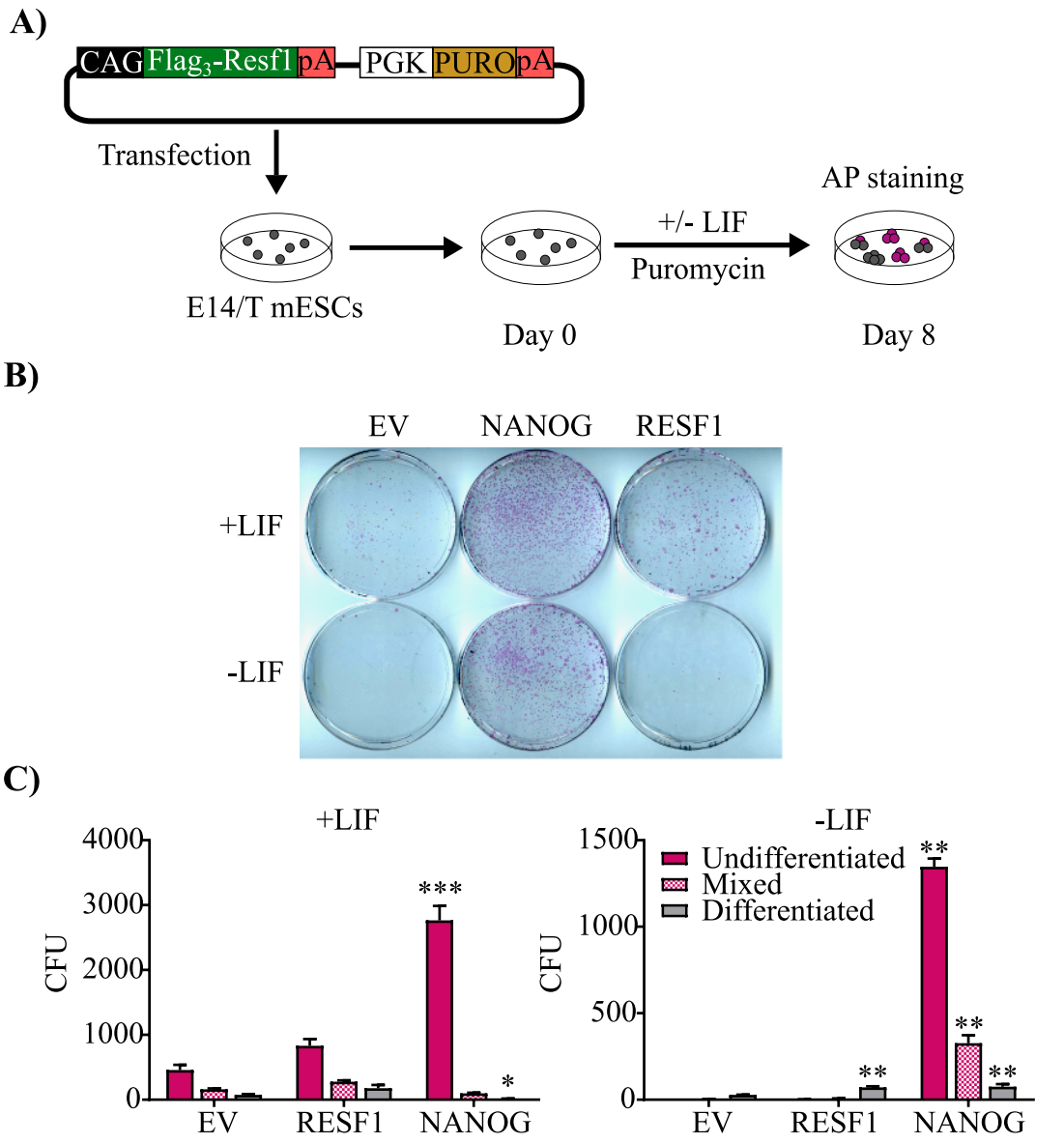


Figure 3.9: Forced RESF1 expression cannot sustain LIF-independent self-renewal of ESCs. **A)** E14/T ESCs were transfected with a plasmid containing $(Flag)_3$-RESF1 or $(Flag)_3$-NANOG under a CAG promoter and puromycin resistance gene under a PGK promoter. Empty vector (EV) was used as a negative control. Transfected cells were plated at a clonal density and cultured in the presence or absence of LIF for eight days with the puromycin selection. **B)** Formed colonies were assessed by alkaline phosphatase (AP) staining and quantified **(C)**. Representative plates are shown. Values represent mean±sem. Statistical difference between *Resf1*$^{-/-}$ and wild-type cells was assessed by a Dunn's method. n=6; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

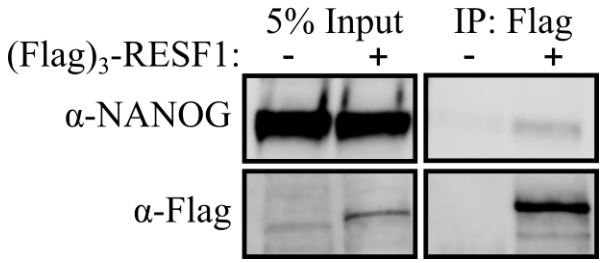


Figure 3.10: NANOG interacts with RESF1 in ESCs. $(Flag)_3$-RESF1 and $(HA)_3$-NANOG were co-expressed in ESCs. Flag was immunoprecipated (IP) from nuclear extracts. The input and the immunoprecipitate samples were analysed by immunoblot. As a negative control (-), untransfected cells lacking expression of 3$(Flag)_3$-RESF1 were used.

Requirements of RESF1 for NANOG function were tested by forced expression of NANOG in *Resf1*$^{-/-}$ ESCs. Wild type E14Tg2a and *Resf1*$^{-/-}$ ESCs were transfected with pPyCAG-NANOG-IP (Chambers et al. 2003) and selected by puromycin for one week prior clonal assay (Fig. 3.11A). The expression of transgenic NANOG was evaluated by immunoblot (Fig. 3.11B). The clonal assay was performed in absence or presence of LIF for eight days (Fig. 3.11A). Self-renewal capacity was analysed by alkaline phosphatase staining and colony counting (Fig. 3.11).

In the presence of LIF, wild type E14Tg2a and *Resf1*$^{-/-}$ ESCs formed mostly alkaline phosphatase positive colonies with mixed morphology (Fig. 3.11). The numbers or formed colonies with different morphologies was significantly different only between *Resf1*$^{-/-}$ c24 and E14Tg2a cells. The *Resf1*$^{-/-}$ c24 cells formed approximately 4-fold lower number of undifferentiated colonies compared to the wild type cells. In the absence of LIF, E14Tg2a and *Resf1*$^{-/-}$ c4 ESCs expressing DsRed differentiated and lost alkaline phosphatase expression within eight days (Fig. 3.11A, B). Unexpectedly, *Resf1*$^{-/-}$ c24 ESCs expressing DsRed formed significantly higher number of alkaline phosphatase positive cells with mixed morphology (Fig. 3.11A, B, C). The number of differentiated colonies at these conditions was similar for all three cell lines.

NANOG overexpression combined with LIF treatment increased formation of undifferentiated colonies compared to the DsRed control (Fig. 3.11A, B). E14Tg2a and *Resf1*$^{-/-}$ c4 cells formed similar number of undifferentiated colonies and almost no colonies with mixed or differentiated morphology (Fig. 3.11C). *Resf1*$^{-/-}$ c24 cells expressing transgenic NANOG formed significantly more colonies with mixed morphology compared to the wild type cells. On the other hand, the number of undifferentiated colonies formed by *Resf1*$^{-/-}$ clone 24 was approximately 2-fold lower.

NANOG overexpression sustained expression of alkaline phosphatase in E14Tg2a and *Resf1*$^{-/-}$ ESCs cultured without LIF (Fig. 3.11A, B). *Resf1*$^{-/-}$ ESCs formed mostly colonies with mixed morphology in contrast to the E14T2ga cells

which formed similar numbers of differentiated, mixed and undifferentiated colonies (Fig. 3.11C). The only statistical difference between wild type and *Resf1*$^{-/-}$ ESCs was observed in a number of undifferentiated colonies formed by *Resf1*$^{-/-}$ clone 24.

In summary, NANOG sustains LIF-independent self-renewal in *Resf1*$^{-/-}$ ESCs. Only one of the examined *Resf1*$^{-/-}$ clones showed significant differences in self-renewal compared to wild type cells. However, significant differences were observed for this clone in the control samples as well. This indicates that self-renewal of *Resf1*$^{-/-}$ c24 ESCs is likely confounded by unanticipated changes in gene expression.

3.4 Requirement of *Resf1* for transition from naïve to primed pluripotency

In the previous sections, self-renewal capacity of *Resf1*$^{-/-}$ ESCs cultured in serum/LIF conditions was assessed. ESCs cultured in serum are heterogeneous and represent a range of pluripotency states (Boroviak et al. 2014). Addition of two small inhibitors (2i) enforces naïve pluripotency and ESCs cultured in 2i/LIF are more homogeneous (Marks et al. 2012). On the other hand, stimulation of ESCs by FGF and Activin initiate transition to a primed pluripotent state, represented by epiblast stem cells (EpiSCs) (Brons et al. 2007; Tesar et al. 2007).

To assess effects of *Resf1* deletion on naïve and primed pluripotency, *Resf1*$^{-/-}$ ESCs cultured in 2i/LIF and *Resf1*$^{-/-}$ EpiSCs were evaluated. The parental wild-type E14Tg2a ESCs and *Resf1*$^{-/-}$ ESCs were adapted to "naïve" pluripotent state by culture in 2i/LIF medium for at least three passages. The *Resf1*$^{-/-}$ ESCs in 2i/LIF form typical dome shaped colonies comparable to wild-type cells (Fig. 3.13A). Expression of pluripotency markers *Pou5f1*, *Nanog*, *Esrrb* and *Rex1* is not significantly different in *Resf1*$^{-/-}$ cells compared to the wild-type control (Fig. 3.13).

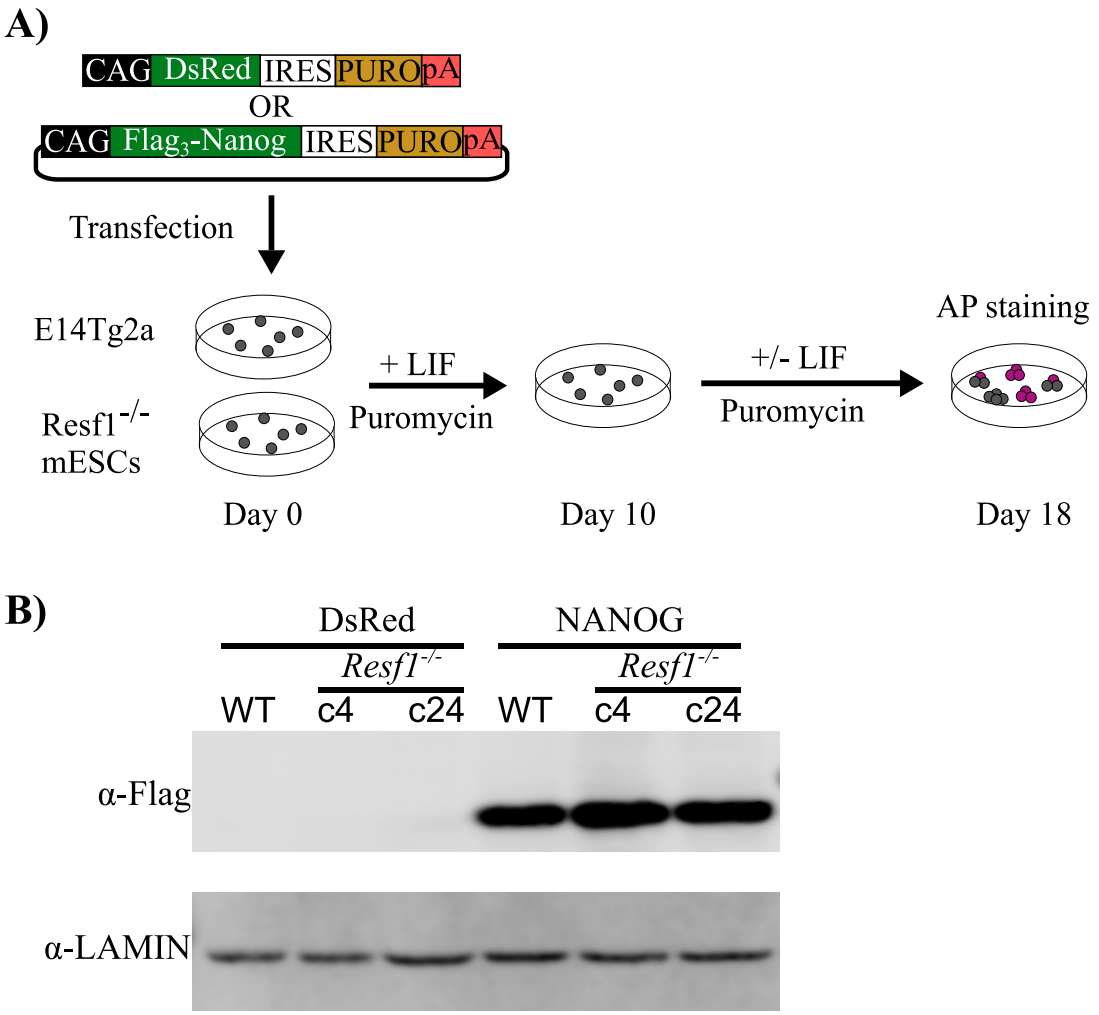


Figure 3.11: Creation of *Resf1*$^{-/-}$ ESCs stably expressing NANOG transgene. A) Wild type (WT) E14Tg2a and *Resf1*$^{-/-}$ ESCs were transfection with pPyCAG-(Flag)$_3$-Nanog-IP and pPyCAG-(Flag)$_3$-DsRed-IP and selected with puromycin for ten days. The episomal levels of Flag-NANOG were assessed by immunoblot **(B)**. LAMIN was used as a loading control. The self-renewal capacity was assessed by a clonal assay for eight days in the presence or absence of LIF. Morphology of the formed colonies was assessed by alkaline phosphatase (AP) staining.

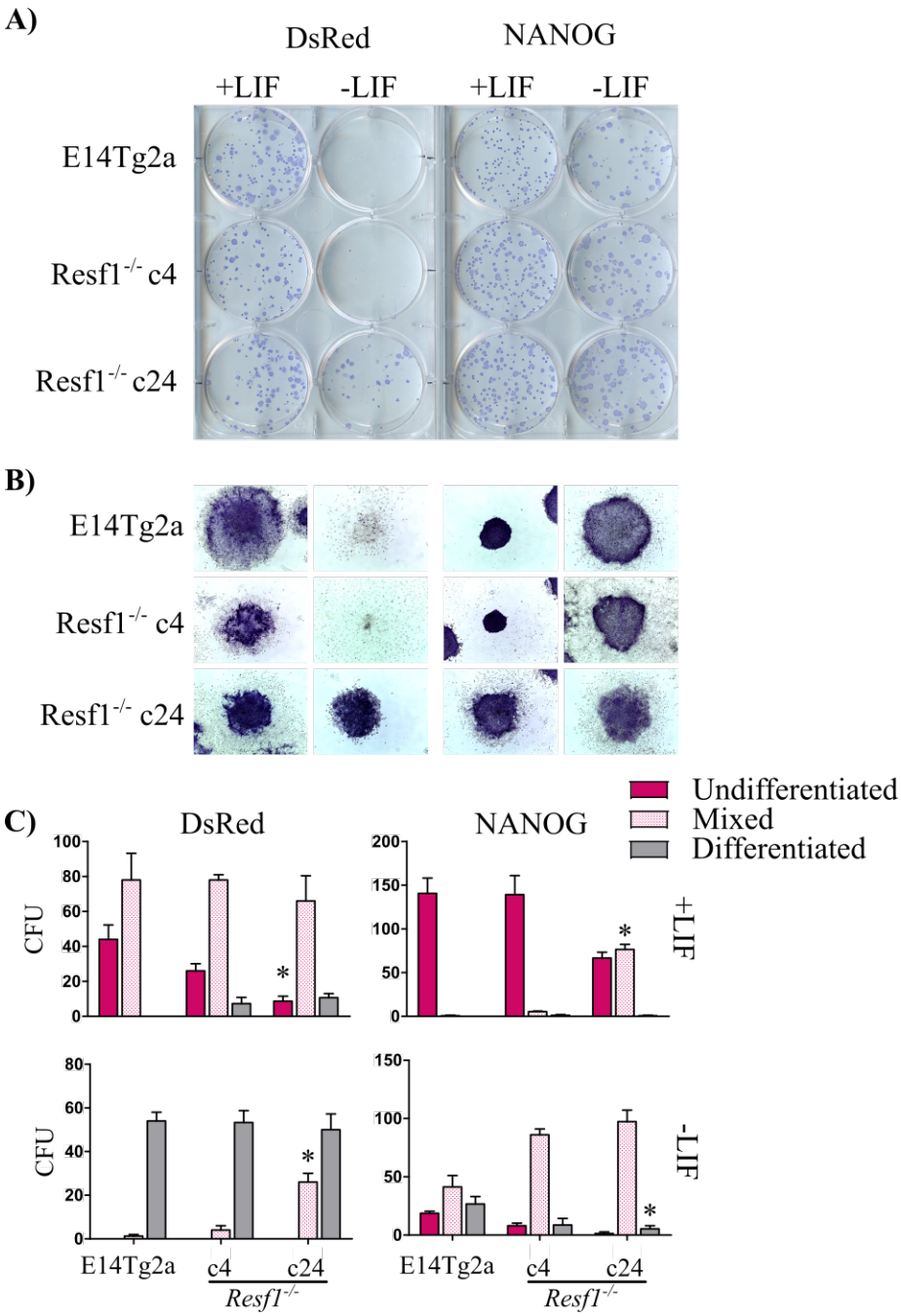


Figure 3.12: RESF1 is not required for NANOG function in ESC self-renewal. **A)** E14Tg2a and *Resf1*$^{-/-}$ ESCs with stable integrated Flag-NANOG or DsRed (negative control) transgenes were cultured at a clonal density in the presence or absence of LIF for eight days. Formed colonies were stained for alkaline phosphatase and counted according to morphology. **B)** Representative colonies are shown. **C)**. Values represent mean±sem, n=9. Statistical difference between *Resf1*$^{-/-}$ and wild-type cells was assessed by a Dunn's method. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

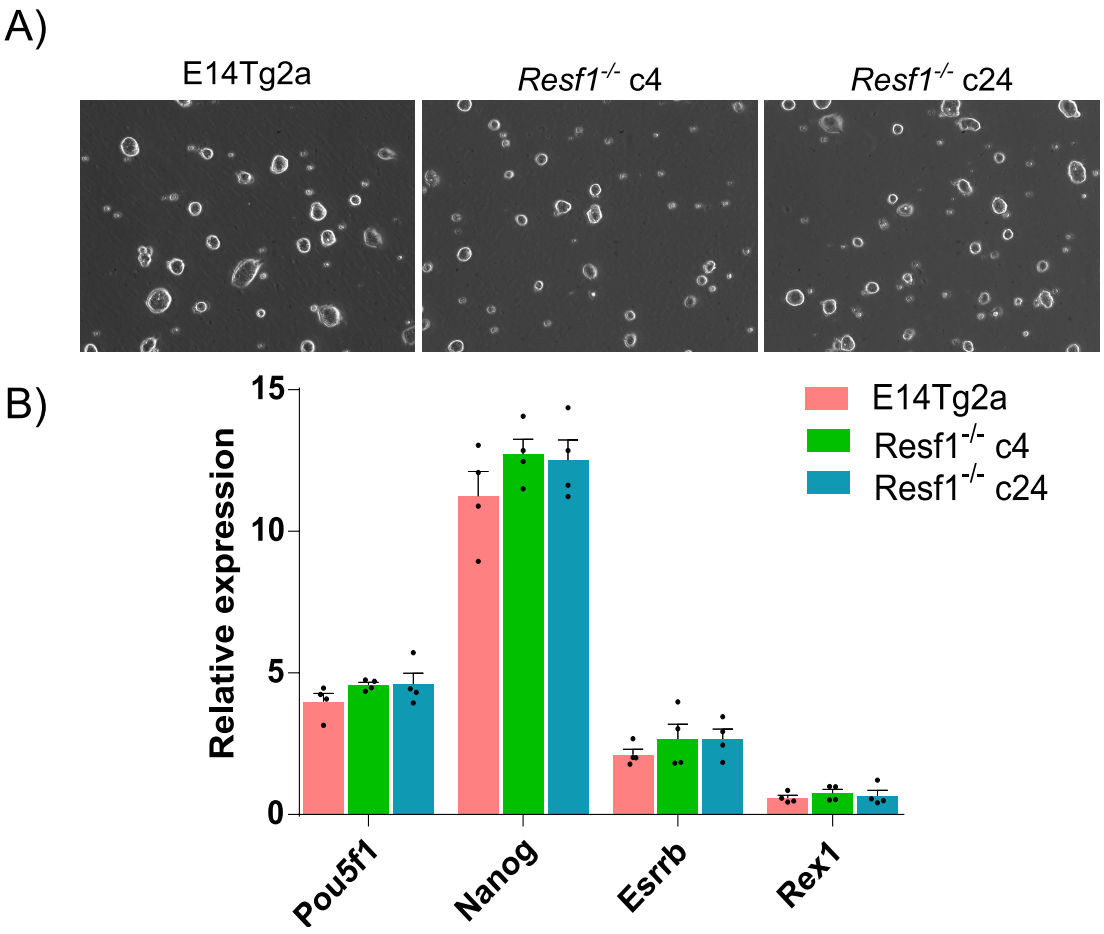


Figure 3.13: *Resf1* is not required for "naïve" pluripotency. **A)** Microphothography of cell colonies formed by E14Tg2a, *Resf1*$^{-/-}$ c4 and c24 ESC lines adapted to 2i/LIF conditions. **B)** Relative mRNA expression of indicated genes over TBP. Bars represent mean±s.d. Statistical significance between *Resf1*$^{-/-}$ and wild-type cells was measured with unpaired, two-sided T-test and p-values were corrected for multiple testing using Holm-Šídák method. n=4, * $p<0.05$; ** $p<0.01$.

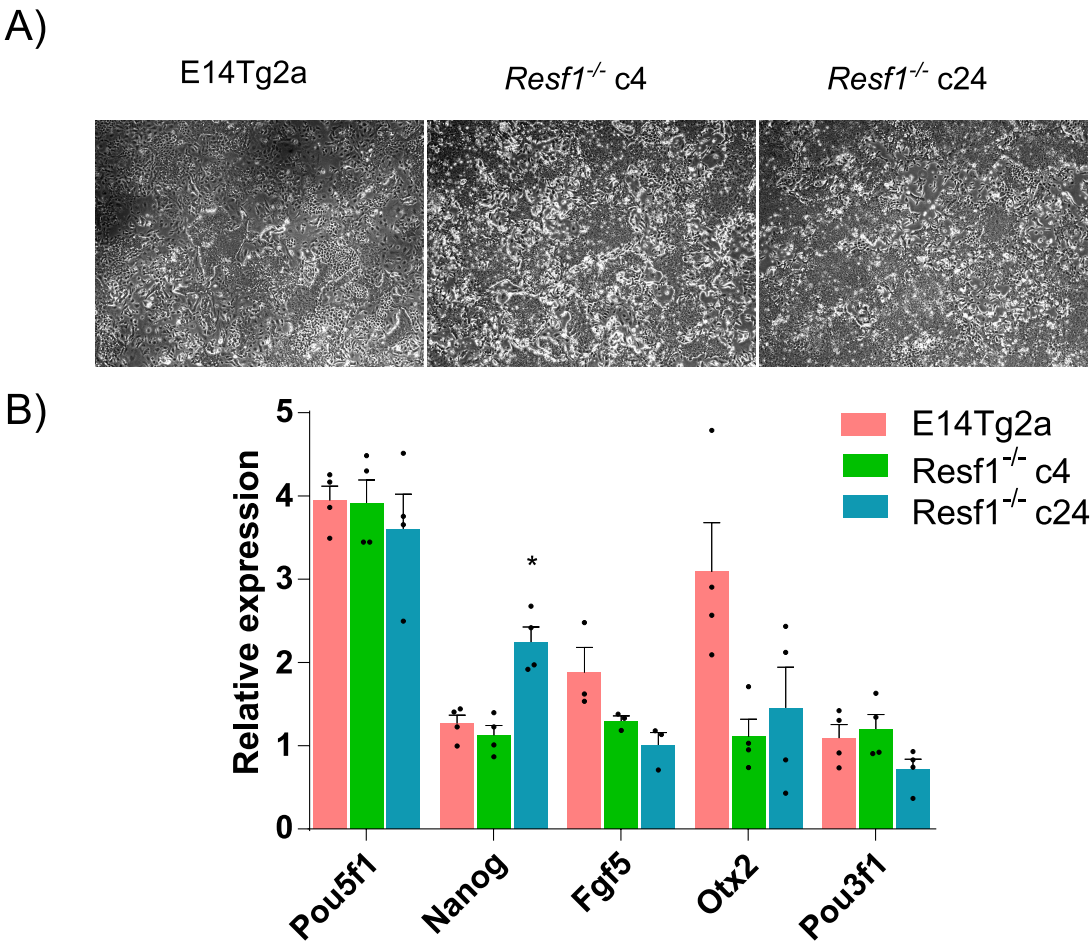


Figure 3.14: *Resf1*$^{-/-}$ EpiSCs have decreased *Otx2* expression. **(A)** Microphothography of cell colonies formed by E14Tg2a, *Resf1*$^{-/-}$ c4 and c24 EpiSC lines. **B)** Relative expression of indicated genes over TBP was measured by RT-qPCR. Bars represent mean±s.d. Statistical significance between *Resf1*$^{-/-}$ and wild-type cells was measured with unpaired, two-sided T-test and p-values were corrected for multiple testing using Holm-Šídák method. n=4; * $p<0.05$; ** $p<0.01$.

To test the *Resf1* requirement for "primed" pluripotent state, *Resf1*$^{-/-}$ ESCs were differentiated to EpiSCs. *Resf1*$^{-/-}$ EpiSCs have a typical flattened morphology (Fig. 3.14A) and express similar levels of *Pou5f1* compared to the wild-type cells. The expression of *Pou5f1*, *Fgf5* and *Pou3f1* in *Resf1*$^{-/-}$ EpiSCs is not statistically different from the wild-type EpiSCs (Fig. 3.14B). *Otx2* expression appears to be lower in *Resf1*$^{-/-}$ cells. On the other hand, *Nanog* expression in the *Resf1*$^{-/-}$ clone 24 is significantly increased compared to the E14Tg2a cells (Fig. 3.14B).

3.5 Discussion

3.5.1 *Resf1* has a minor effect on ESC self-renewal

RESF1 is a poorly characterised protein of unknown function. Only very recently, Fukuda et al. examined RESF1 function in retrotransposon silencing in ESCs (Fukuda et al. 2018). In this chapter, RESF1 function in ESC self-renewal was examined.

Due to a lack of antibodies against RESF1 protein, RESF1 localisation in ESCs was assessed by episomal expression of Flag-tagged RESF1 (Fig. 3.1) and in endogenously tagged RESF1 ES cell lines (Fig. 3.3). The expression of endogenously expressed RESF1 was localised solely to the nucleus (Fig. 3.3) whereas episomally expressed RESF1 was observed also in the cytoplasm (Fig. 3.1). It is possible that episomal expression caused ectopic localisation of RESF1 in the cytoplasm. Therefore, nuclear localisation of endogenously expressed RESF1 is more reliable. Observed nuclear localisation of RESF1 is also consistent with the previously published data showing nuclear localisation of transiently expressed Flag-tagged RESF1 in ESCs (Fukuda et al. 2018). Moreover, RESF1 expression in NANOG negative cells (Fig. 3.3) and a weak correlation between RESF1 and NANOG protein levels, suggest that NANOG and RESF1 are independently regulated.

To determine the role of RESF1 in ESC self-renewal, *Resf1* was deleted in ESCs using CRISPR/Cas9 (Fig. 3.5). *Resf1*$^{-/-}$ ESCs cultured in serum/LIF

express decreased levels of pluripotency markers and appear to have a higher propensity to differentiate compared to wild-type ESCs (Fig. 3.7). Therefore, self-renewal capacity of *Resf1*$^{-/-}$ ESCs was examined using a clonal assay. In the presence of LIF, deletion of *Resf1* caused a reduced number of undifferentiated colonies in only one of the examined clonal cell lines (Fig. 3.8). Nevertheless, the capacity of *Resf1*$^{-/-}$ ESCs to self-renew in serum/LIF culture demonstrates that *Resf1* deletion is not absolutely required for ESC self-renewal.

After LIF withdrawal, *Resf1* knockout ESCs formed significantly fewer alkaline phosphatase positive colonies compared to wild-type ESCs (Fig. 3.5). Wild-type ESCs cultures retain a small number of undifferentiated cells likely due to autocrine expression of LIF. Therefore, it is possible that RESF1 could be involved in the LIF signalling pathway. To determine if RESF1 regulates expression of LIF, LIF receptor or other components of this pathway, additional experiments are required. Moreover, the number of differentiated colonies formed in absence of LIF, was significantly lower for *Resf1*$^{-/-}$ ESCs (Fig. 3.5). Because ESC differentiation leads to proliferation arrest and cell death (Duval et al. 2004), it is possible that *Resf1*$^{-/-}$ ESCs are poised to differentiate or differentiate faster. This is supported by decreased mRNA and protein levels of NANOG, OCT4 and ESRRB in *Resf1*$^{-/-}$ ESCs (Fig. 3.7). However, in the presence of MEK/ERK and GSK3 inhibitors (2i), *Resf1*$^{-/-}$ ESCs express similar levels of *Oct4*, *Sox2*, *Nanog*, *Esrrb* and *Rex1* as wild-type ESCs (Fig. 3.13). It is therefore also possible that *Resf1*$^{-/-}$ ESCs have increased expression of FGFR2 driving NANOG downregulation and differentiation similarly to the primitive endoderm specification in mouse embryo (Frankenberg et al. 2011). To answer these questions, effects of *Resf1* deletion on differentiation kinetics and activity of signalling pathways needs to be determined.

If *Resf1* deletion has a negative effect on ESC self-renewal, it maybe expected that increased RESF1 expression could enhance ESC self-renewal. Indeed episomal expression of RESF1 has a positive effect on ESC

self-renewal but the effect is minor (Fig. 3.9). Even though episomal RESF1 expression cannot sustain LIF-independent self-renewal, a significantly higher number of differentiated colonies was retained in the culture compared to the control (Fig. 3.9). This might be due to a delayed differentiation opposite to the phenotype observed after *Resf1* deletion (Fig. 3.8). This suggests that *Resf1* has a minor positive effect on ESC self-renewal. However, small effect sizes of *Resf1* deletion and RESF1 ectopic expression on ESC self-renewal, decrease statistical confidence of these findings. Furthermore, relationship between RESF1 concentration and expression of the key pluripotency factors needs to be determined in order to determine the role of RESF1 in maintenance of ESC state. An inducible RESF1 system could be introduced to *Resf1*$^{-/-}$ ESCs to measure direct RESF1 effects on gene expression.

3.5.2 RESF1 is not required for NANOG to sustain LIF-independent self-renewal

Interestingly, RESF1 was found to interact with the key ESC self-renewal determinants NANOG (Gagliardi et al. 2013) and OCT4 (Berg et al. 2010), suggesting that RESF1 may function in ESC self-renewal. As the NANOG WR is required for NANOG function to sustain LIF-independent self-renewal (Chambers et al. 2003) and interaction with RESF1 (Mullin et al. unpublished), RESF1 may be required for NANOG function in ESC self-renewal. NANOG-RESF1 interaction in ESC nuclei was confirmed by co-immunoprecipation (Fig. 3.10). To investigate the requirements of RESF1 on NANOG function, the capacity of NANOG to sustain LIF-independent self-renewal in *Resf1*$^{-/-}$ ESCs was measured (Fig. 3.11). NANOG overexpression sustained LIF-independent self-renewal in both examined *Resf1*$^{-/-}$ cell lines, suggesting redundancy of RESF1-NANOG interaction in NANOG function. The *Resf1*$^{-/-}$ c24 expressing DsRed or NANOG showed a reduced self-renewal capacity in the presence of LIF compared to the clone 4. This is consistent with reduced expression of pluripotency markers (Fig. 3.7) and self-renewal capacity (Fig. 3.8) of the clone 24 in serum/LIF

conditions. These results, together with the unexpected capacity of *Resf1*$^{-/-}$ c24 ESCs to self-renew in absence of LIF after random integration of pPyCAG-DsRed-PGK-Puro (Fig. 3.9) suggests, that properties of this clonal cell line are confounded by unknown changes. As it has been shown that *Resf1*$^{-/-}$ ESCs have increased expression of retrotransposons (Fukuda et al. 2018), increased activity of retrotransposons could have caused unpredictable alterations in the genome and gene expression. Alternatively, unanticipated off-target activity of Cas9 during the Resf1$^{-/-}$ c24 cell line derivation, could have caused similar changes.

3.5.3 *Resf1* is not required for primed pluripotency

The potential of *Resf1*$^{-/-}$ ESCs to progress to primed pluripotency state was assessed by differentiation of *Resf1*$^{-/-}$ ESCs to epiblast stem cells (EpiSCs). Cells lacking *Resf1* cultured in presence of Activin A and FGF2 for more than six passages, formed colonies with a similar morphology to the wild-type EpiSCs (Fig. 3.14A). The expression of the key EpiSC markers *Pou5f1, Fgf5* and *Pou3f1* in *Resf1*$^{-/-}$ EpiSCs is similar to wild-type EpiSCs (Fig. 3.14B). The only significant difference between wild-type and *Resf1*$^{-/-}$ EpiSCs was observed for *Nanog* expression in the *Resf1*$^{-/-}$ clone 24. This could be caused by a clonal differences described above.

3.5.4 Conclusion

In summary, deletion of *Resf1* in ESCs demonstrated that *Resf1* is dispensable for ESC maintenance but *Resf1* deletion decreases expression of pluripotency markers in the serum/LIF culture conditions. This phenotype is rescued by 2i/LIF culture suggesting that *Resf1* deletion increases the propensity of ESCs to differentiate via FGF or Wnt signalling pathways. *Resf1*$^{-/-}$ ESCs are capable to differentiate into EpiSCs indicating that *Resf1* is not required for primed pluripotency. These results are in an agreement with the recent finding that *Resf1*$^{-/-}$ mice can reach adulthood (Dickinson et al. 2016). However, the appearance of *Resf1*$^{-/-}$ homozygous mice after *Resf1*$^{+/-}$ mice

cross is lower than the Mendelian ratio (Dickinson et al. 2016). However, the developmental stage in which *Resf1*$^{-/-}$ embryo development fails has not been determined. Defects in pre-implantation epiblast similar to defects observed in *Resf1*$^{-/-}$ ESCs cultured in serum/LIF conditions, could explain this phenotype. In addition, functional studies of RESF1 in this thesis are limited by the clonal variations of the *Resf1*$^{-/-}$ ESCs. This could be resolved by rescuing RESF1 phenotype by integrating RESF1 inducible system into *Resf1*$^{-/-}$ ESCs.

4 RESF1 function in specification of primordial germ cells

4.1 Effects of *Resf1* deletion on PGCLC specification

Although RESF1 is not essential for NANOG function in ESC self-renewal, it may be required in a different process. In addition to a role in epiblast specification, NANOG is also required for efficient generation of primordial germ cells (PGCs) in the mouse embryo (Chambers et al. 2007; Zhang et al. 2018b). RESF1 was also proposed to play a role in germ cell development (Hudson et al. 2005). In fact, both male and female *Resf1 null* mice are infertile (Dickinson et al. 2016). This led to the hypothesis that RESF1 functions together with NANOG in the early specification of PGCs. To test the requirements of *Resf1* for PGC specification, the protocol for PGC like cells (PGCLCs) differentiation developed by Hayashi et al. was used (Fig. 4.1). Following this protocol, ESCs cultured in 2i/LIF are differentiated into epiblast-like cells (EpiLCs) by Activin A and FGF2 stimulation (Fig. 4.1). EpiLCs are transiently competent to PGC specification (Hayashi et al. 2011) and initiate PGCLC differentiation after cytokine treatment. After four day differentiation, PGCLCs are marked by the expression of surface markers SSEA-I and CD61 as well as expression of the key germ cell factors *Blimp1*, *Prdm14*, *Ap2γ* and *Nanog* (Hayashi et al. 2011).

Analysis of the *Resf1*$^{-/-}$ EpiLCs at the day two of the differentiation, showed similar expression of the key EpiLC markers *Fgf5*, *Pou3f1* and *Otx2* compared to the wild-type EpiLCs (Fig. 4.2B). The efficiency of PGCLC formation was compared between wild-type and *Resf1*$^{-/-}$ ES cell lines, based on CD61 and SSEA1 expression after four days of PGCLC differentiation (Fig. 4.3). Wild-type E14Tg2a ESCs gave rise to ~2-6% of CD61^{+} SSEA1^{+} cells. *Resf1*$^{-/-}$ cells consistently produced a lower proportion of this population (Fig.

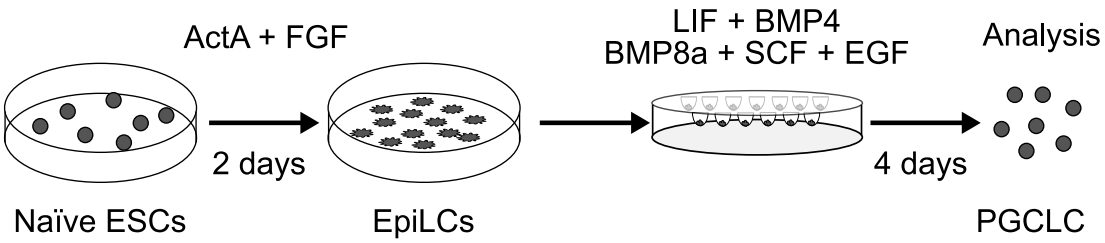


Figure 4.1: A scheme of PGCLC differentiation protocol. Naïve ESCs are differentiated into primordial germ cell (PGC) competent Epiblast like cells (EpiLCs) for two days in the presence of Activin A and FGF2. EpiLCs are aggregated in the presence of indicated cytokines which stimulates the specification of PGC-like cells (PGCLCs). PGCLCs are analysed after 4 days.

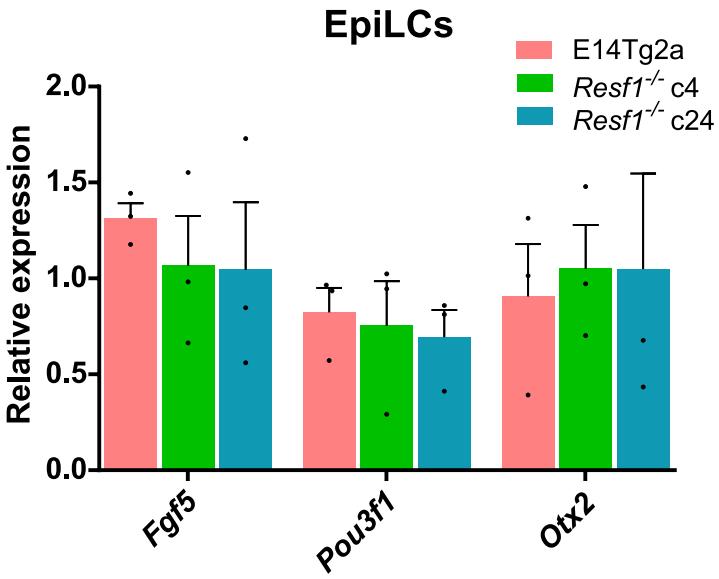


Figure 4.2: *Resf1*$^{-/-}$ EpiLCs express normal levels of *Fgf5*, *Pou3f1* and *Otx2*. Expression of *Fgf5*, *Pou3f1* and *Otx2* in E14Tg2a and *Resf1*$^{-/-}$ EpiLCs at day two was measured by RT-qPCR. Bars represent mean±s.d. of relative mRNA expression.

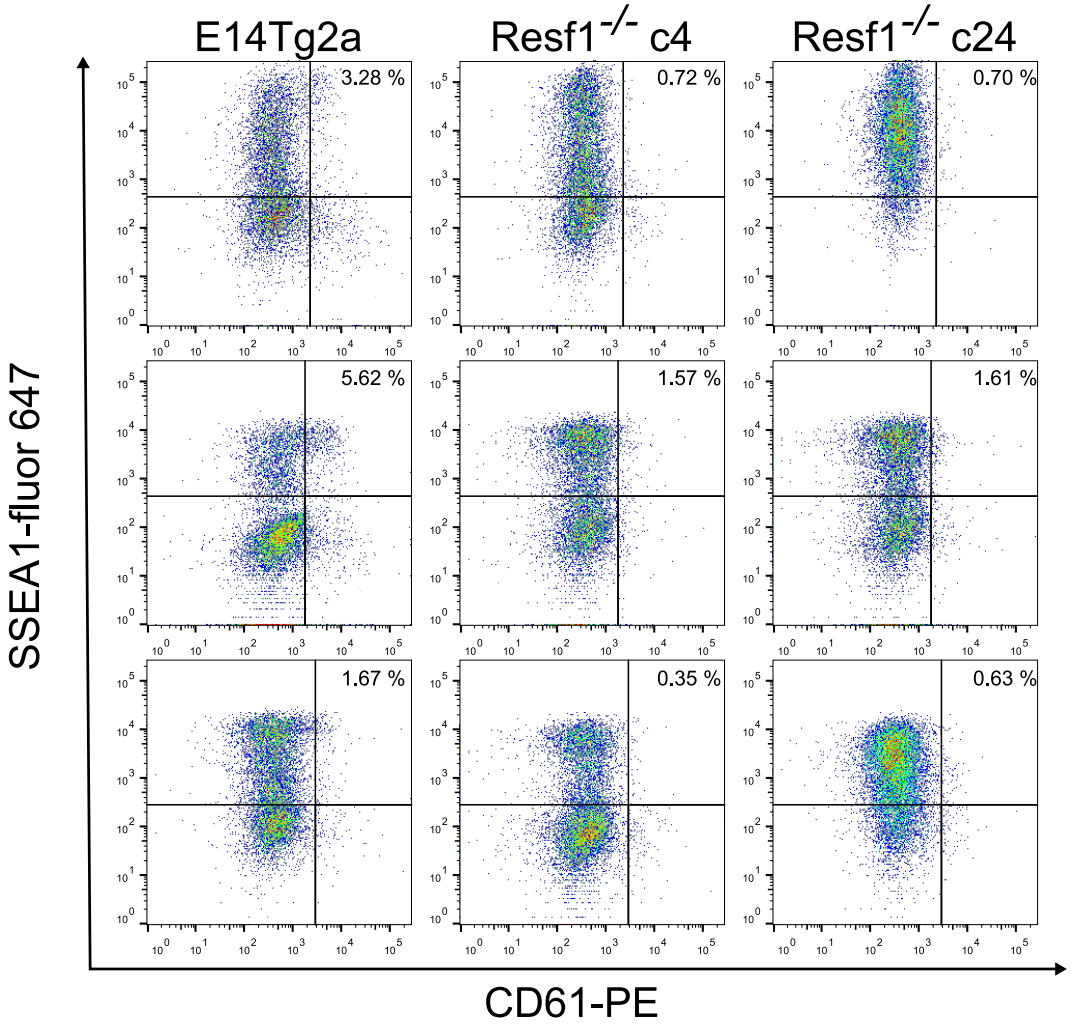


Figure 4.3: *Resf1* is required for efficient PGCLC specification. A) wild-type E14Tg2a and *Resf1*$^{-/-}$ ESC lines were differentiated to primordial germ cell like cells (PGCLCs). Expression of PGCLC surface markers SSEA1 and CD61 was analysed after four days of the differentiation by a flow cytometry. The CD61^{+}SSEA1^{+} cells were gated based on negative and single stained control samples. Three independent biological replicates are shown.

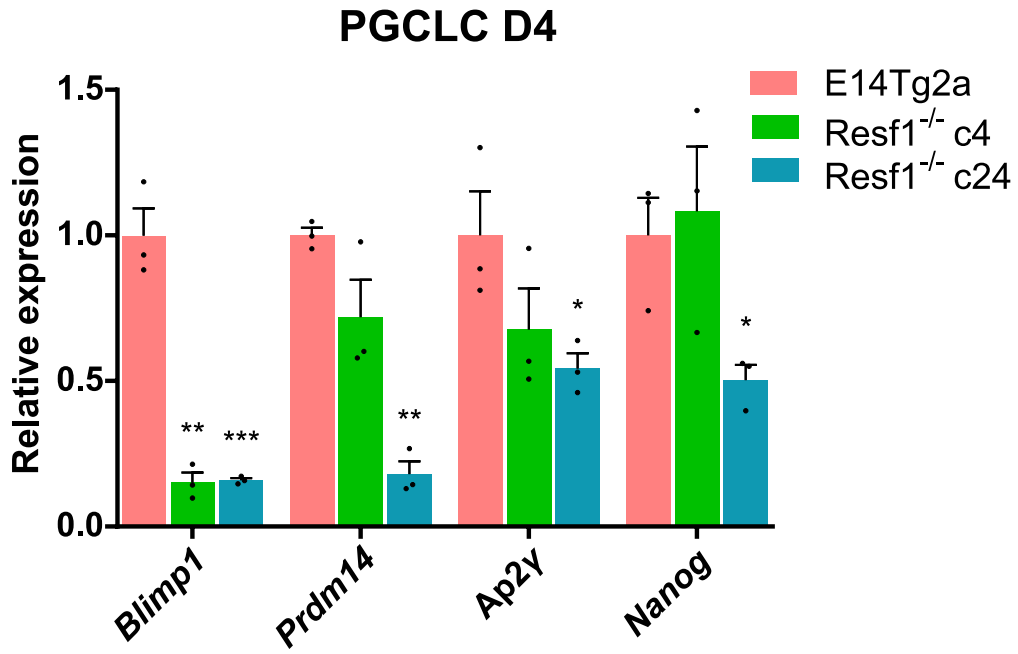


Figure 4.4: *Resf1* is required for activation of *Blimp1* during PGCLC specification. Wild type E14Tg2a and two *Resf1*$^{-/-}$ ES cell lines were differentiated into PGCLCs for four days and mRNA levels of the indicated genes were measured by RT-qPCR. Values represent the relative mRNA expression over TBP normalised to the mean expression in the wild type cells. Values represent mean±sem. The statistical significance was determined by a two-sided T-test and p-values were corrected for multiple testing using Holm-Šídák method. n=3; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

4.3B). Further analysis of the PGC markers in PGCLCs showed that *Resf1*$^{-/-}$ cells failed to upregulated *Blimp1* and have reduced expression of *Prdm14* and *Ap2γ*. *Nanog* mRNA levels were decreased only in the *Resf1*$^{-/-}$ clone 24 (Fig. 4.3B).

4.2 Discussion

NANOG partner RESF1 is expressed in developing mice and chicken gonads and likely plays a role in the development of the urogenital system (Hudson et al. 2005). Interestingly, it was recently reported by the International Mouse Phenotyping Consortium that both male and female *Resf1*$^{-/-}$ mice are infertile (Dickinson et al. 2016). However, the mechanism and the stage of germ cell development in which RESF1 plays a crucial role is unknown. One possible explanation could be that RESF1 is required for the early specification of primordial germ cells (PGCs).

NANOG was shown to be required for an efficient specification of PGCs in the mouse embryo (Chambers et al. 2007; Zhang et al. 2018b). As RESF1 interacts with NANOG in ESCs, it is possible that RESF1 participates in the early PGC specification together with NANOG. The requirement of *Resf1* for PGC specification was tested by differentiating *Resf1*$^{-/-}$ EpiLCs to PGCLCs as previously described (Hayashi et al. 2013). Proportion of the PGCLCs marked by the expression of the surface markers SSEA-I and CD61 was decreased 4 to 12-fold in the *Resf1*$^{-/-}$ cells (Fig. 4.3). *Resf1*$^{-/-}$ also failed to activate the key PGC factor *Blimp1* (Fig. 4.4). *Prdm14* and *Ap2γ* were also decreased in the *Resf1*$^{-/-}$ cells. This indicates that *Resf1* is required for efficient entry into the germ line, which could explain sterility of *Resf1*$^{-/-}$ mice.

There are multiple possible explanations of *Resf1*$^{-/-}$ phenotype in PGCLCs. Even though there is no significant difference between *Fgf5, Otx2* and *Pou3f1* expression in PGC competent wild-type and *Resf1*$^{-/-}$ EpiLCs (Fig. 4.2), it is possible that *Resf1* deletion compromises the PGC competent state by effects on other transcripts. Analysis of the global gene expression in

Resf1$^{-/-}$ EpiLCs would be required to exclude this possibility.

RESF1 could be also required for a timely downregulation of the direct NANOG target *Otx2* which is known to restrict the germ line entry (Festuccia et al. 2012; Zhang et al. 2018b). The first two days of the PGCLC differentiation are critical for OTX2 depletion and PGCLC specification (Zhang et al. 2018a). Therefore, a failure to quickly downregulate *Otx2* in *Resf1*$^{-/-}$ could explain this result. To answer this, measurement of the rate of *Otx2* mRNA and OTX2 protein depletion in the *Resf1*$^{-/-}$ cells during PGCLC differentiation could be performed.

PGCs undergo a massive DNA demethylation from E9.0 to E13.5 period (Hajkova et al. 2002). However, methylation is not fully erased from the intracisternal A-particle (IAP) transposons and some coding genes (Guibert et al. 2012; Lane et al. 2003; Lees-Murdock et al. 2003; Seisenberger et al. 2012). Transposons are silenced by *de novo* DNA methylation which finishes around E18.5 (Lees-Murdock et al. 2003). Insufficient silencing of IAP and LINEs causes DNA instability and results in infertility (Molaro et al. 2014). In fact, RESF1 has been shown to silence retrotransposon elements in ESCs likely through SETDB1 (Fukuda et al. 2018). Therefore, abnormal activity of retroelements in *Resf1*$^{-/-}$ PGCs could contribute to infertility of *Resf1* knockout mice.

Ultimately, it is required to rescue *Resf1* knockout phenotype by reintroducing RESF1, to confirm the specific effect of RESF1 in PGC specification. An inducible system could answer which stage of the PGCLC specification is dependent on RESF1 expression. Such a system could be used as a functional assay to determine the functional domains of RESF1 and potential molecular function of RESF1 which is currently unknown.

5 NANOG directs the core transcriptional network to control gene expression

NANOG binds tens of thousands sites within the ESC genome associated with thousand of genes. Only a small fraction of these genes is directly regulated by NANOG (Festuccia et al. 2012; Heurtier et al. 2019). This means that NANOG DNA binding alone is not sufficient to explain how NANOG activates and represses genes. In order to resolve this problem, we need to consider that transcriptional factors do not act alone but often co-localise with each other and recruit various coactivator and corepressor complex. OCT4 and SOX2 bind closely to NANOG in the ESC genome (Chen et al. 2008; Heurtier et al. 2019). However, it is currently unknown how NANOG influences OCT4, SOX2 and RNAPII chromatin binding.

To answer this, chromatin binding of OCT4 was measured before and after NANOG induction by chromatin immunoprecipitation followed by sequencing (ChIP-seq). OCT4 chromatin binding was assessed at zero, one and three hours of NANOG induction in *Nanog*$^{-/-}$ ESC cell line expressing NANOG-ER^{T2} fusion protein (NERT ESCs) (Festuccia et al. 2012). Modified estrogen receptor (ERT2) restricts NANOG-ER^{T2} expression to the cytoplasm and allows rapid re localisation of NANOG-ER^{T2} to the nucleus after addition of 4-hydroxytamoxifen (Festuccia et al. 2012). OCT4 ChIP-seq data was analysed together with previously generated data sets from the same cell line consisting of NANOG, SOX2 and RNAPII ChIP-seq data sets at different time points of NANOG induction (Fig. 5.1 and table 2.15).

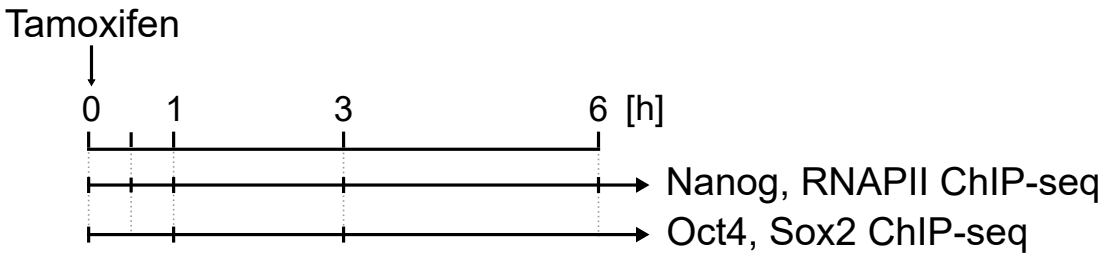


Figure 5.1: Summary of the analysed data sets. OCT4 ChIP-seq experiments were performed at 0, 1 and 3 h after NANOG induction. NANOG, SOX2 and RNAPII ChIP-seq datasets generated in the same NERT ES cell line at different time points after NANOG induction were used in the analysis.

5.1 Clustering of NANOG binding sites reveals distinct patterns of OCT4, SOX2 and RNAPII response to NANOG induction

To examine different responses of OCT4, SOX2 and RNAPII to NANOG induction, NANOG binding sites were categorised into seven clusters based on relative changes of OCT4, SOX2 and RNAPII ChIP-seq signal using k-means clustering (method section 2.13.4, Fig. 5.2). In the clusters one to three, OCT4, SOX2 and RNAPII binding increases by approximately 1.5 to 2-fold after 1 h of NANOG induction. OCT4 binding remains similar at 3 h. SOX2 binding decreases at the 3 h time point, most noticeably in the cluster three. RNAPII binding steadily increases in the cluster one, whereas in the clusters two and three RNAPII binding decreases to the original levels after 6 h. In clusters four to seven, the observed differences after NANOG induction are smaller than 2-fold. OCT4 binding is approximately 1.4-fold higher after the induction in the clusters four, five and six. In the cluster seven, OCT4 signal decreases by less than 1.5-fold. SOX2 transiently increases binding at the cluster four by ~1.4-fold and is essentially unaffected in the clusters five and seven. In the cluster six, SOX2 binding steadily decreases reaching ~1.4-fold lower values than before the induction. RNAPII responds to NANOG at the clusters four to seven similarly. In the clusters four and seven RNAPII signal rapidly decreases to approximately 1.4-fold lower values than at the 1 h. RNAPII signal then gradually increases but remains lower than before the induction. In clusters five and six, RNAPII binding is essentially unchanged across the measured period of time but there is a noticeable dip in the signal at 1h.

The identified clusters separate NANOG binding sites based on different combinations of OCT4, SOX2 and RNAPII responses to NANOG induction. However, relative changes do not reflect the true effect size and might be misleading for values of low magnitude. Therefore, read densities of NANOG, OCT4, SOX2 and RNAPII at these clusters were examined (Fig. 5.3). To reduce the complexity of the data and to increase the confidence

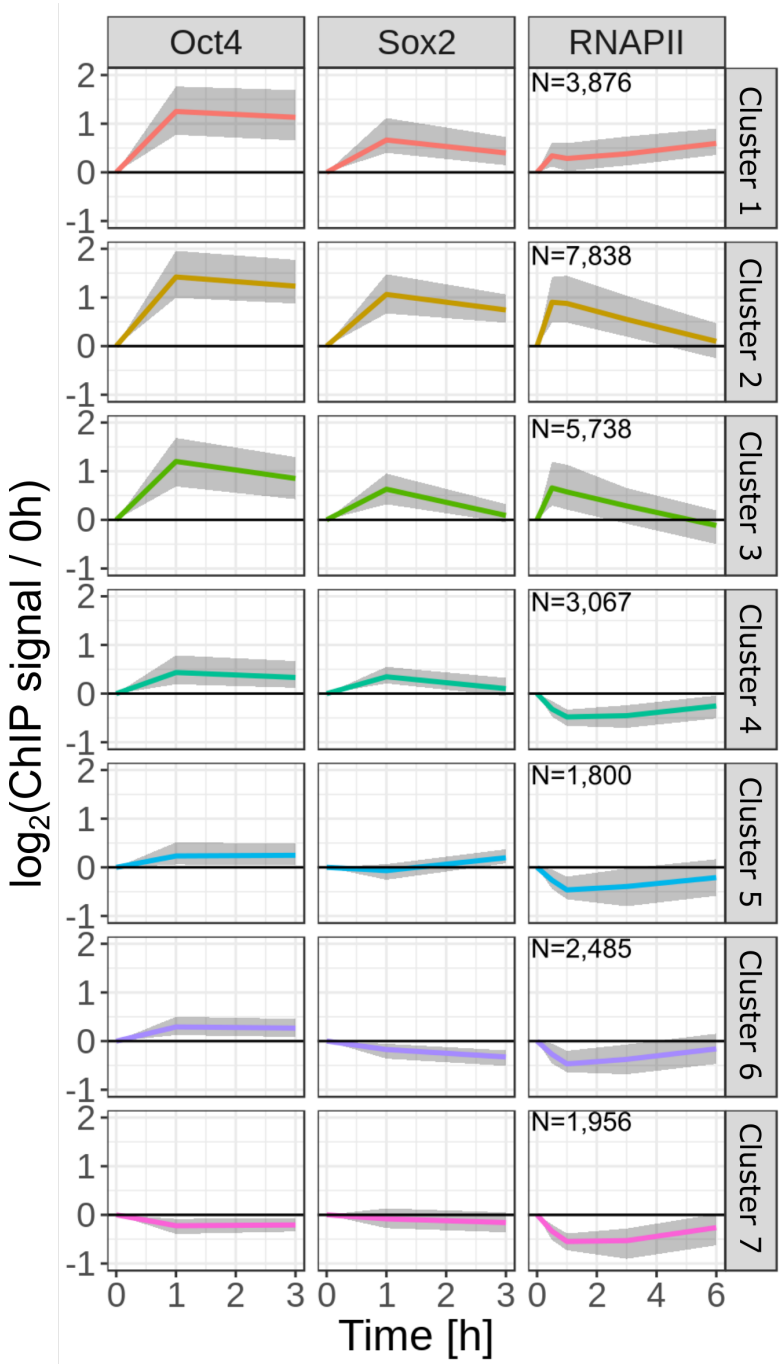


Figure 5.2: OCT4, SOX2 and RNAPII response differently to NANOG induction at specific NANOG binding sites. Nanog binding sites were clustered into 7 clusters based on OCT4, SOX2 and RNAPII chromatin binding response to wild-type NANOG induction using k-means clustering. Lines represent median of logarithm fold change differences over the 0 h time point for each factor in each cluster. Ribbon represents the first and the third quartile of the data.

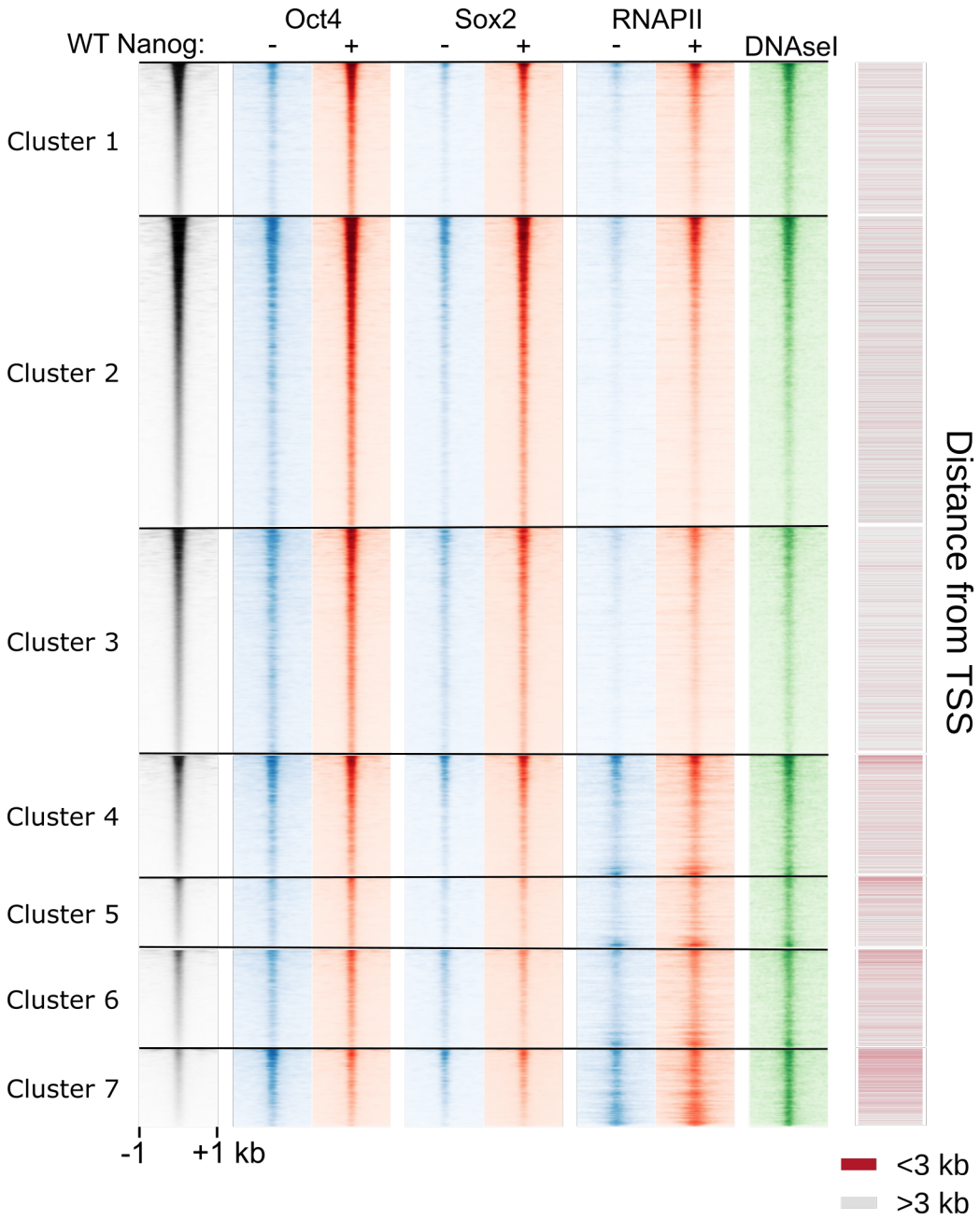


Figure 5.3: NANOG increases OCT4, SOX2 and RNAPII binding at distal regulatory elements. Heatmaps of NANOG (black), OCT4, SOX2, RNAPII (blue and red) ChIP-seq and DNAseI (green) normalised signal over a 2 kb window surrounding Nanog peaks. Nanog peaks were grouped to the identified clusters. OCT4, SOX2 and RNAPII signal before (blue) or after (red) wild-type NANOG induction is shown. Each row represent one NANOG binding site, ordered by NANOG signal. Relative distance (more or less than 3 kb from the nearest TSS) for each site is shown.

in the observed changes, the time points after the induction for each factor were combined (method section 2.13.1). To examine chromatin accessibility at different NANOG binding site clusters, previously published DNAseI hypersensitivity dataset (method section 2.13.3) was analysed. Moreover, a relative distance of each binding site to the nearest transcriptional start site (TSS) was visualised to determine possible correlation between TF binding intensity and promoter (<3 kb) or enhancer (>3 kb) binding (Fig. 5.3, methods 2.13.1).

At the clusters one, two and three, NANOG increases OCT4, SOX2 and RNAPII binding consistent with the previous analysis (Fig. 5.2). The binding intensity of the all examined factors is relatively strong at these sites (Fig. 5.3). There is an apparent correlation between NANOG, OCT4, SOX2 and RNAPII chromatin binding and chromatin accessibility. Moreover, the majority of these sites lie >3 kb away from the nearest transcriptional start site (TSS), with no apparent correlation between the binding intensity and the relative distance.

At clusters four, five, six and seven, NANOG is mostly bound more weakly compared to the clusters one, two and three (Fig. 5.3). OCT4 and SOX2 binding correlates with the NANOG intensity at these sites as well. RNAPII binding before NANOG induction at these clusters, is higher compared to other clusters. Moreover, the responses of OCT4, SOX2 and RNAPII binding to NANOG induction have lower magnitude compared to the clusters one, two and three consistent with the previous analysis. Interestingly, within the clusters four to seven, RNAPII binding is strong at sites where NANOG, OCT4 and SOX2 bind with the lowest magnitude. Moreover, there is a higher proportion of putative promoters with higher NANOG binding at these clusters (Fig. 5.3).

In summary, recruitment of OCT4 in the clusters one, two and three is similar but SOX2 and RNAPII patterns differ mainly after 1 h time point. Clusters four to seven are marked by a decreasing RNAPII signal, especially in the clusters four and seven, and display a lower magnitude of changes after NANOG induction.

5.2 Specific responses to NANOG induction associate with gene activation and gene repression

In the previous section, seven clusters of NANOG binding sites with different combined patterns of OCT4, SOX2 and RNAPII binding were identified. At most of the NANOG binding sites, NANOG has a strong positive effect on OCT4, SOX2 and RNAPII binding. At the other sites, mainly at the clusters four to seven, NANOG binds chromatin weakly and has a minor negative effect on OCT4, SOX2 and RNAPII binding (Fig. 5.2, 5.3). To describe how NANOG regulates its gene targets, relationship between observed changes in TF binding and gene expression caused by NANOG induction needs to be determined.

To achieve this, association score between different NANOG peaks in different clusters and NANOG gene targets was calculated (method section 2.13.5). Association score estimates the likelihood of a gene being regulated by a particular binding site. Using a list of differentially expressed genes (DEGs) after NANOG induction in NERT ESCs determined by Jingchao Zhang and Alison McGarvey, genes within the mouse genome were devided into four categories: 1. genes induced by NANOG, 2. genes repressed by NANOG, 3. genes not regulated by NANOG but expressed in ESCs and 4. genes unexpressed in ESCs. Subsequently, the likelihood of each cluster being associated with DEGs more than with the expressed genes was calculated (Fig. 5.4).

Clusters one associates significantly with the induced genes with the p-values $1x10^{-6.05}$ (Fig. 5.4). A good example of this association is *Esrrb*. *Esrrb* is directly activated by NANOG (Festuccia et al. 2012) and contains multiple NANOG binding sites from the cluster one in its proximity (Fig. 5.5). Cluster two is also significantly associated with the activated genes ($p = 1x10^{-3.78}$) as well as repressed genes ($p = 1x10^{-2.62}$). Clusters four and seven are significantly associated with repressed genes with the p-values $1x10^{-2.37}$

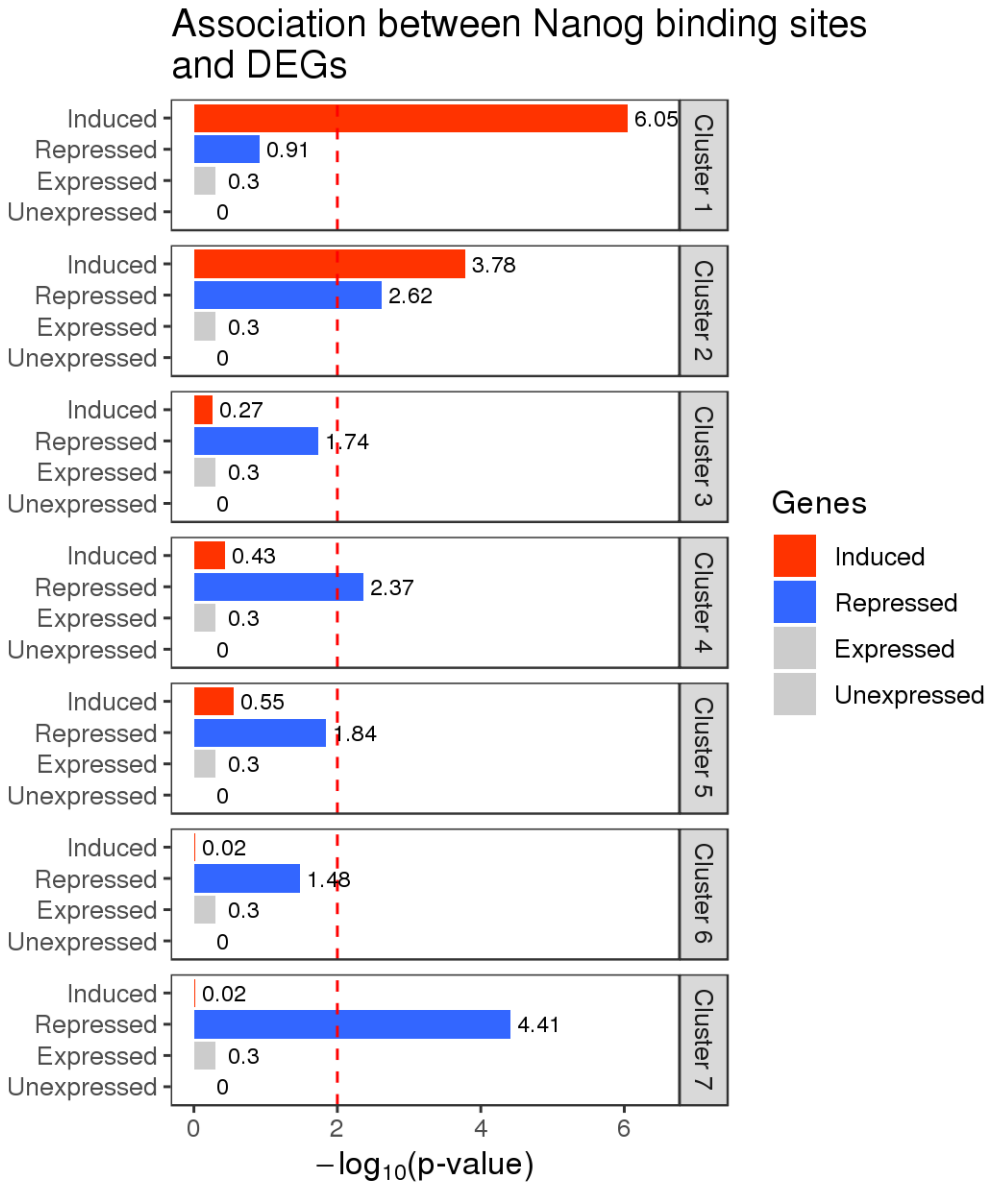


Figure 5.4: Distinct clusters of NANOG binding sites associate with gene activation or gene repression. Association between NANOG binding site clusters and genes was calculated based on microarray data from NERT ESCs (Tab. 2.3). The mouse genes were grouped into: genes induced by NANOG (red), genes repressed by NANOG (blue) or genes not affected by NANOG induction (grey). Genes which do not respond to NANOG were divided into two groups: genes expressed in ESCs and genes unexpressed in ESCs. Likelihood of the association between a cluster and a specific group of genes was calculated by Mann-Whitney U-test using expressed genes as the control.

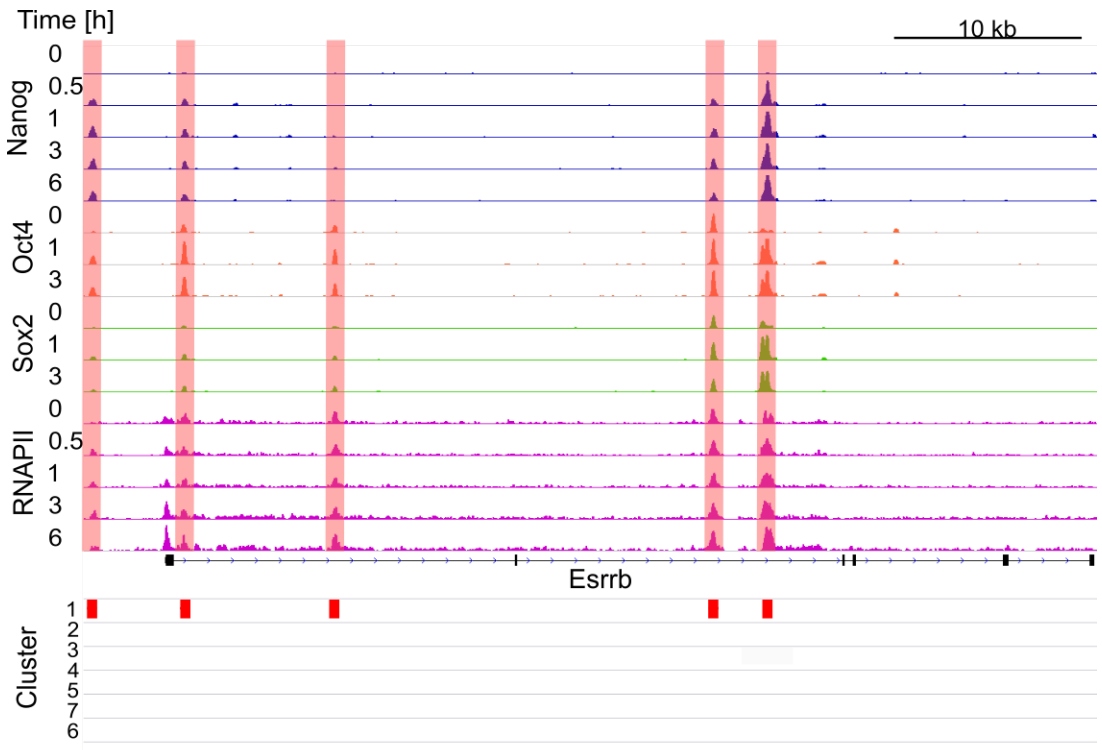


Figure 5.5: NANOG increases binding of OCT4, SOX2 and RNAPII at the *Esrrb* locus. Genome tracks of NANOG, OCT4, SOX2 and RNAPII binding at different time points after NANOG induction at the *Esrrb* locus. The NANOG binding sites at different clusters are indicated. The red boxes highlight the peaks from the cluster one.

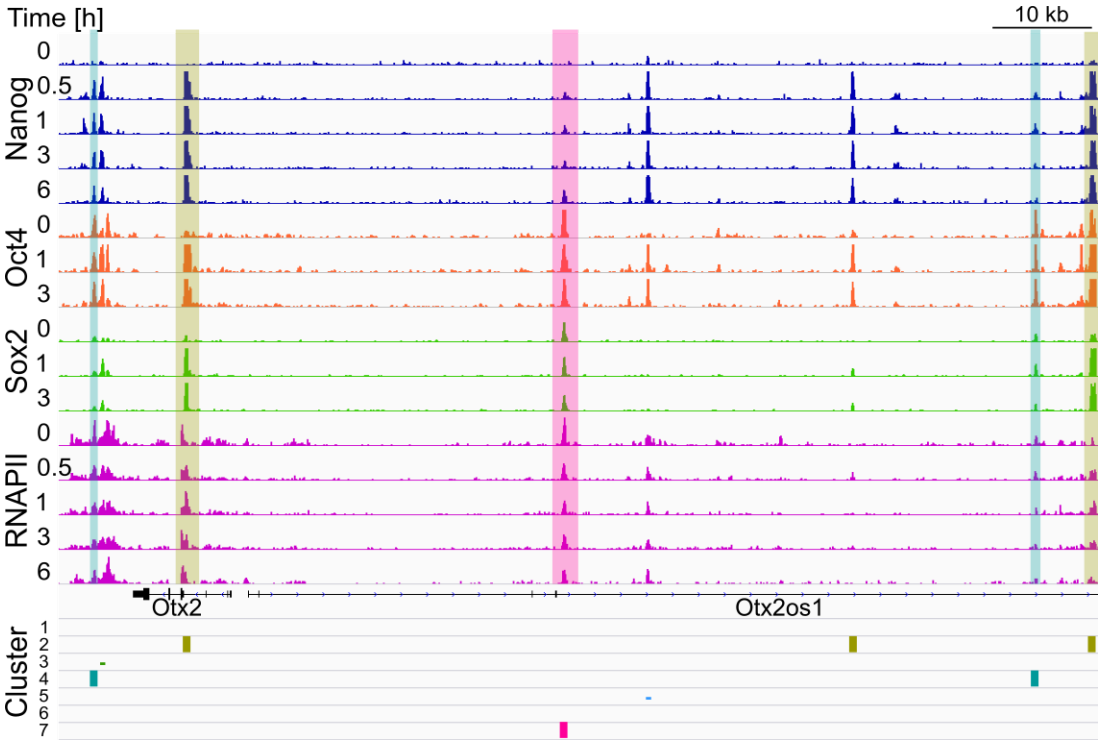


Figure 5.6: Changes of OCT4, SOX2 and RNAPII binding at the *Otx2* locus after NANOG induction. Tracks of NANOG, OCT4, SOX2 and RNAPII binding at different time points after NANOG induction at the *Otx2* locus. The NANOG binding sites at different clusters are indicated. Peaks from a cluster two (green box) and cluster seven (pink box) are highlighted.

and $1x10^{-4.41}$ respectively. For example, *Otx2* is downregulated by NANOG (Festuccia et al. 2012) and contains NANOG peaks from the clusters two, four and seven in its proximity (Fig. 5.6). Association of the clusters three, five and six with DEGs is not significant at α 0.01.

To better understand the relationship between the NANOG binding clusters and the DEGs, the frequency of these sites around the DEGs was quantified (Fig. 5.7). The frequency of NANOG peaks from each cluster being located around genes not regulated by NANOG was used as a control (Fig. 5.7). Frequency of NANOG peaks in all the examined clusters around the control genes is relatively uniform with a decreasing trend. Frequency of peaks within the cluster one around genes induced by NANOG is highest within $\sim$10 kb from the TSS. On the other hand, these clusters are positioned more distally from genes repressed by NANOG and have a lower frequency compared to induced genes. The trend is very similar for the peaks within the cluster two. Peaks in the clusters four and seven are located mostly within 50 kb from TSSs of repressed genes. Frequency of these peaks around induced genes is similar to the control. Interestingly, the peaks within cluster seven have increased frequency between 250-350 kb from induced genes.

5.3 Discussion

The concentration of NANOG determines self-renewal efficiency of ESCs (Chambers et al. 2003; Chambers et al. 2007). NANOG is considered to function through a direct activation of ESC sustaining factors mainly *Esrrb* (Festuccia et al. 2012) and through a direct repression of pro-differentiation factors including *Otx2* (Acampora et al. 2017). It has been shown that NANOG directly regulates $\sim$100 genes in ESCs (Festuccia et al. 2012). Therefore, we would expect that NANOG regulates expression of its gene targets by recognising cis-regulatory elements located close to these genes. However, NANOG binds in proximity of more than six thousand genes (Festuccia et al. 2012). This means that NANOG binding is not sufficient to explain NANOG

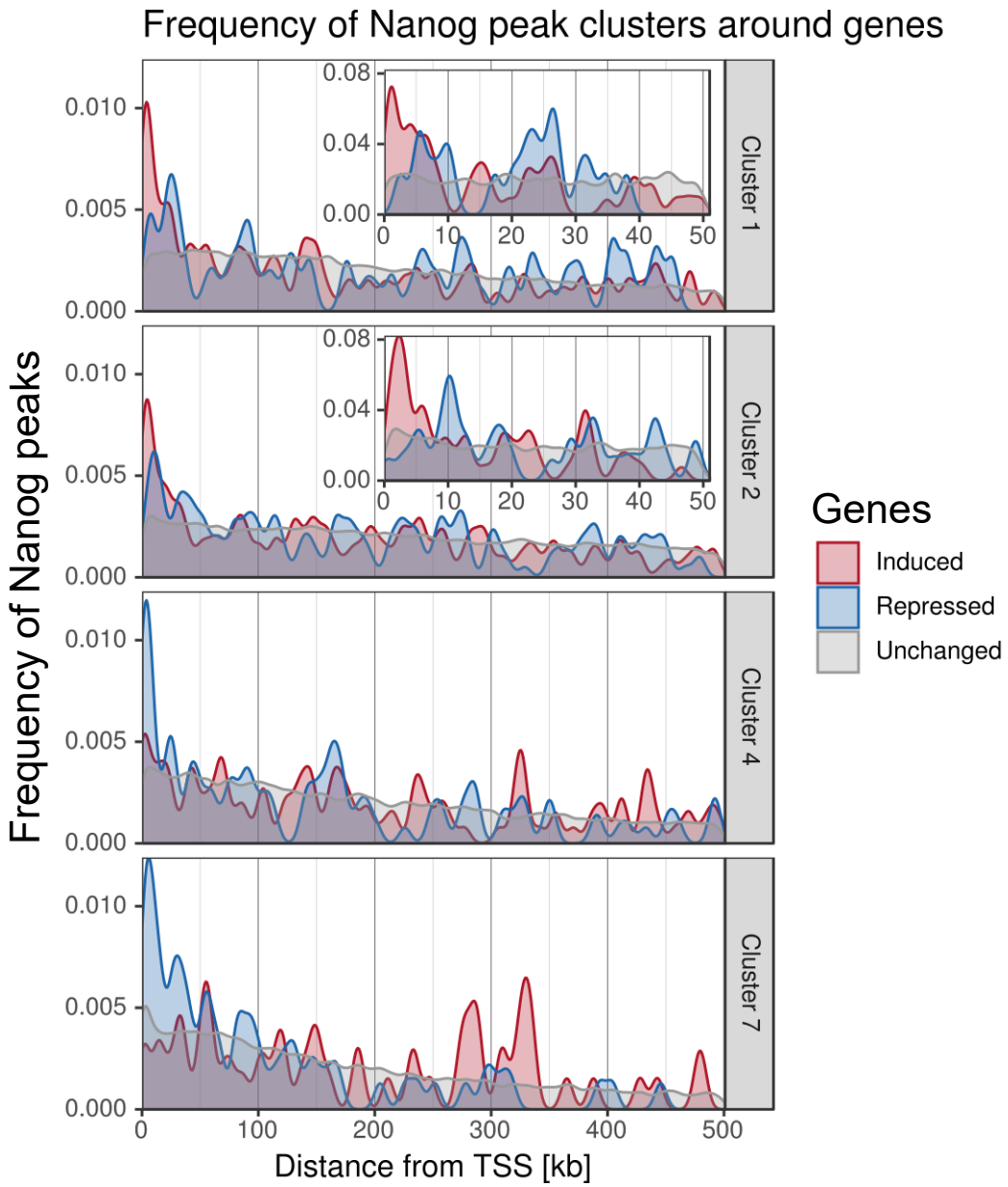


Figure 5.7: NANOG gene targets have distinct distribution patterns of NANOG peaks. The frequency of NANOG peaks from different clusters around genes was determined. NANOG peak clusters significantly associated with DEGs were analysed (see 5.2). Genes were grouped into three groups: induced by NANOG (red), repressed by NANOG induction (blue), or genes which are unchanged upon NANOG induction but still expressed in ESCs (grey). The zoomed in 50 kb region from the TSS is shown for clusters 1 and 2.

function. NANOG often co-localises with other transcription factors, including OCT4, SOX2 or RNA polymerase II (RNAPII). Therefore, combinatorial binding of NANOG and its protein partners could help us to explain why NANOG acts on a small proportion of genes and what distinguishes NANOG mediated gene activation and repression. This motivated me to explore the effects of NANOG on OCT4, SOX2 and RNAPII chromatin binding in ESCs.

I identified seven distinct clusters showing different combined patterns of OCT4, SOX2 and RNAPII chromatin binding after NANOG induction (Fig. 5.2). These clusters can be categorised into two main groups. The first group is characterised by a rapid increase of OCT4, SOX2 and RNAPII binding after NANOG induction (Fig. 5.2). The rapid response to NANOG supports the notion that NANOG influences binding of these factor directly. NANOG, OCT4 and SOX2 bind at these sites with a high intensity (Fig. 5.3). The responses of these TFs to NANOG seems to follow the same pattern reaching the highest values at 1 h after the induction, followed by a decrease of chromatin binding. This could be a result of an unknown cellular response to a sudden transport of NANOG protein into the nucleus after the tamoxifen treatment. These sites are often located between 3 to 10 kb from genes directly activated by NANOG and they significantly associate with these genes (Fig. 5.7 and 5.4). NANOG therefore might activate its gene targets by increasing OCT4, SOX2 and RNAPII chromatin binding at enhancers proximal to these genes. The exemplary case of OCT4, SOX2 and RNAPII recruitment by NANOG is the *Esrrb* locus where the majority of NANOG peaks belong to the cluster one (Fig. 5.5). MED1 and RNAPII have been shown to form condensates at the *Esrrb* locus (Cho et al. 2018) and undergo a phase separation together with OCT4, SOX2 and NANOG *in vitro* (Boija et al. 2018). Therefore, it is possible that NANOG nucleates formation of these large protein condensates, mediating gene activation consistent with the previously proposed in which concentration of coactivators at enhancers leads to gene activation (Hnisz et al. 2017; Boija et al. 2018). This could explain the observed high occupancy of NANOG, OCT4 and SOX2 at these sites (Fig. 5.3). The reduced frequency

of long-range interaction in *Nanog null* ESCs (Novo et al. 2018) indicates that formation of such condensates could stabilise promoter-enhancer contacts and promote gene expression. Phase separation mediated chromatin looping was also recently supported by a computational model (Shrinivas et al. 2019).

Interestingly, an increase of OCT4, SOX2 and RNAPII binding after NANOG induction is also observed relatively close to some genes repressed by NANOG (Fig. 5.4 and 5.7). For example, there are sites within the *Otx2* locus where NANOG induction decreases RNAPII binding (cluster four and seven) but also a sites where NANOG initially increases RNAPII binding (cluster two) (Fig. 5.6). This is reflected by a significant association of the cluster two with both repressed and activated genes (Fig. 5.4). Therefore, additional information are required to fully separate NANOG mediated gene activation and repression. A measurement of differential recruitment of chromatin coativators and corepressors and subsequent changes in the epigenetic environment after NANOG induction could provide the necessary information. In fact, NANOG was reported to interact with multiple coactivator and corepressor complexes in ESCs (Gagliardi et al. 2013). Moreover, functional dissection of these putative enhancers would provide a direct evidence that observed changes of RNAPII, OCT4 and SOX2 binding are functionally linked to gene activation or gene repression.

The second major group of NANOG binding sites consists of clusters four to seven and is marked by a weaker intensity of NANOG, OCT4 and SOX2 binding (Fig. 5.3). These binding of these TFs is weaker despite relative high chromatin accessibility (Fig. 5.3). This suggest that additional factors restrict NANOG, OCT4 and SOX2 binding at these sites. Differential motif analysis between the two major groups of NANOG binding sites could point to potential factors restricting binding of OCT4, SOX2 and NANOG at these sites. The second distinct pattern within the second major group is reduction of RNAPII binding shortly after NANOG induction. OCT4 and SOX2 binding is decreased by NANOG induction only in the cluster seven (Fig. 5.2). However, due to a low intensity of the TF binding and a small magnitude of the observed changes

(Fig. 5.2 and 5.3) it is necessary to further validate significance of these findings. Nevertheless, there is a clear correlation between RNAPII binding reduction (Fig. 5.2) and gene repression (Fig. 5.4). Therefore, displacement of RNAPII after NANOG induction could lead to gene repression. However, whether NANOG directly displaces RNAPII remains to be determined. RNAPII binding could be also decreased indirectly. For example, by recruitment of transcriptional repressors interacting with NANOG. NCOR1 or a polycomb group complex among other corepressor complexes are likely candidates as they were identified to interact with NANOG in ESCs (Gagliardi et al. 2013).

In addition, clusters four and seven have a higher RNAPII occupancy compared to majority of NANOG binding sites (Fig. 5.3). This could be explained by an increased proportion of promoter regions within these clusters (Fig. 5.3). Clusters four and seven are also often located 250-500 kb from the genes activated by NANOG (Fig. 5.7). This might indicate that NANOG redistributes OCT4, SOX2 and RNAPII between distal and proximal enhancers. Therefore, it would be interesting to determine the position of TAD boundaries relative to NANOG binding sites to determine if two main groups of NANOG binding sites are separated in different TADs.

Heurtier et al. has very recently reported a similar study of NANOG function in ESC self-renewal using a CRISPR-ON system which can increase the expression of endogenous *Nanog* after a chemical treatment. The authors, similarly to my results, observed that NANOG increases BRG1, OCT4, SOX2 and ESRRB binding at sites mainly associated with genes activated by NANOG which is in agreement with the results reported here. Moreover, Heurtier et al. proposed two distinct models of NANOG mediated gene repression. The first model proposes that NANOG blocks OCT4 and SOX2 binding and prevents gene activation. This model is based on observation of reduced OCT4 and SOX2 binding after NANOG induction at a sub-group of NANOG binding sites which associate with gene repression. These sites have been reported to have accessible chromatin configuration, measured by ATAC-seq, and increased occurrence within gene promoters. These

results agree with my analysis very well (Fig. 5.3). The second mechanism proposes that NANOG safeguards ESC self-renewal by sustaining H3K27me3 at gene promoters demonstrated by increased average levels of H3K27me3 at promoters of genes repressed by NANOG, mainly *Otx2*. Interestingly, shortly after NANOG induction OCT4, SOX2 and RNAPII binding at *Otx2* promoter increases (Fig. 5.6). This contradicts the first proposed model. This suggests that not only the response of TF network to NANOG but also the position of this event is important to dictate transcriptional response.

6 NANOG requires WR to efficiently direct the core regulatory factors and regulate gene expression

NANOG tryptophan repeat (WR) is essential for NANOG function in ESC self-renewal (Mullin et al. 2008). Substitution of tryptophan residues within the NANOG WR for alanines ($NANOG_{W10A}$), prevents NANOG homodimerisation (Mullin et al. 2017) and interaction with other proteins including RNA polymerase II (RNAPII) and SOX2 (Mullin et al. unpublished, Gagliardi et al. 2013). This suggests that the WR mediated protein-protein interactions are essential for NANOG function. However, the molecular mechanism underlying the role of the WR in NANOG mediated gene regulation is unclear.

To determine the role of the WR in NANOG function, *Nanog*$^{-/-}$ ES cell lines expressing fusion proteins of ER^{T2} with either wild-type NANOG ($NANOG_{WT}$), $NANOG_{W10A}$ or $NANOG_{N51A}$ variants ($NERT_{WT}$, $NERT_{W10A}$ and $NERT_{N51A}$ ESCs) were used (Zhang 2016). The $NANOG_{N51A}$ variant contains a classic point mutation within the homeodomain, preventing DNA binding (Noyes et al. 2008). Comparison of $NANOG_{WT}$, $NANOG_{W10A}$ and $NANOG_{N51A}$ variants allow to examine separate roles of the WR and the homeodomain in NANOG function. OCT4 chromatin binding before and after induction of $NANOG_{W10A}$ and $NANOG_{N51A}$ variants was measured using ChIP-seq (Fig. 5.1, method section 2.11.1). OCT4 ChIP-seq data were analysed together with previously generated ChIP-seq data sets of SOX2, RNAPII and NANOG variants in $NERT_{W10A}$ and $NERT_{N51A}$ ESCs (Fig. 5.1).

To determine effects of the W10A and N51A mutations on TF binding, variations between NERT clonal cell lines need to be assessed. To do this, ChIP-seq data sets generated before the tamoxifen treatment in these cell lines were used. At these conditions, all three cell lines are essentially *Nanog*$^{-/-}$ cells with NANOG expression restricted to the cytoplasm. It is therefore reasonable to expect that binding of OCT4, SOX2 and RNAPII should be equal in all

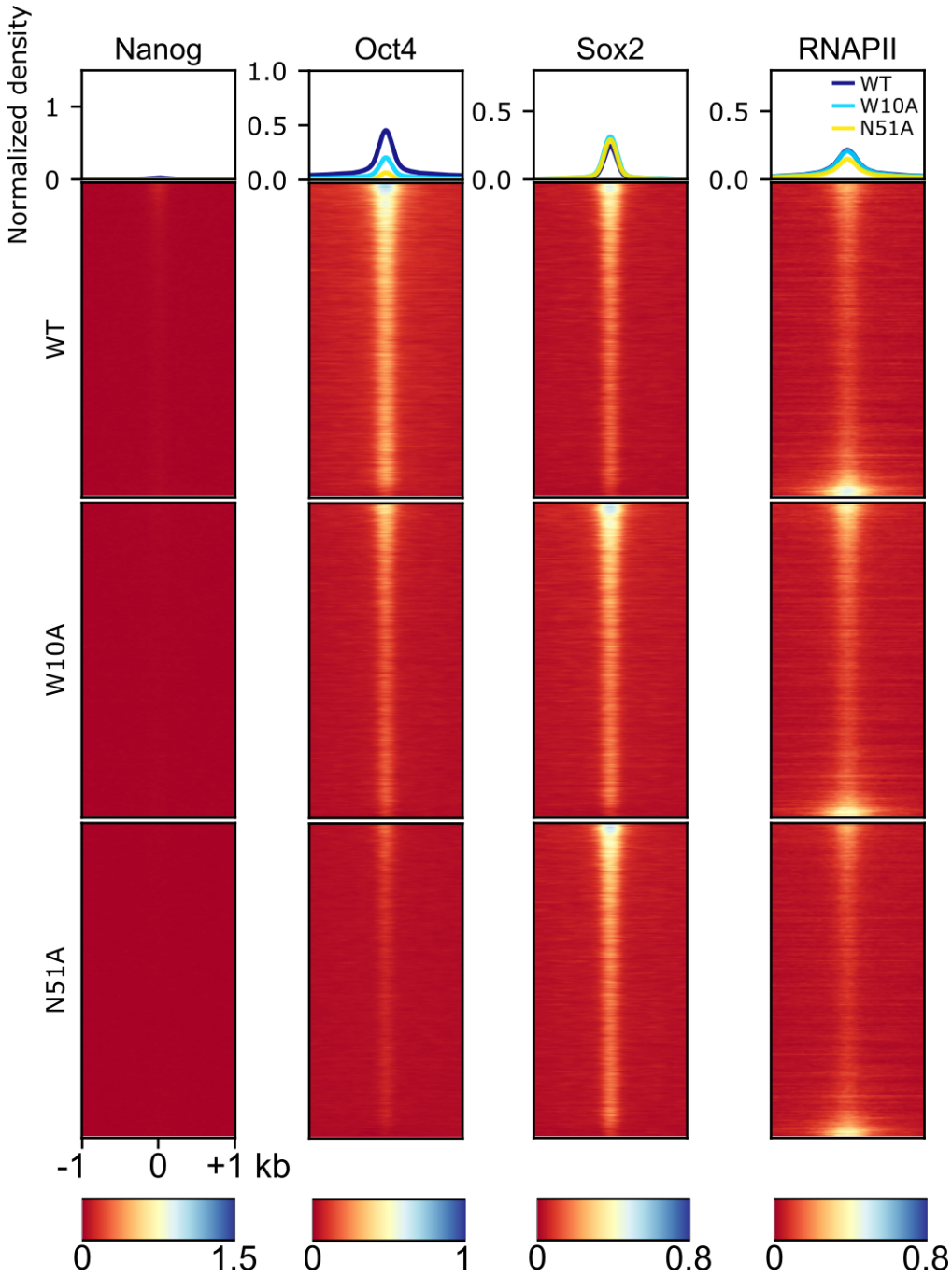


Figure 6.1: Clonal variation between NERT ES cell lines. Chromatin binding of OCT4, SOX2 and RNAPII at NANOG binding sites in $NERT_{WT}$, $NERT_{W10A}$, $NERT_{N51A}$ ESCs before NANOG induction. Each row represent a single NANOG binding site extended by 1 kb each direction. Rows were ordered by the average wild-type NANOG binding intensity after wild-type NANOG induction. Scales were set based on the intensities of the individual factors in the cells treated with tamoxifen.

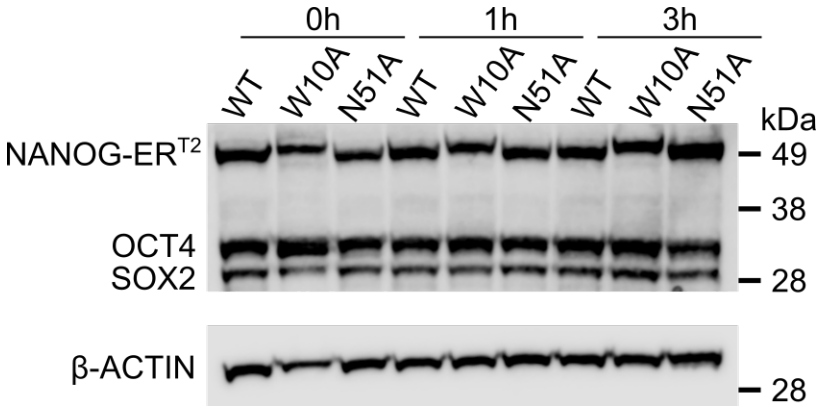


Figure 6.2: OCT4, SOX2 and NANOG levels in $NERT_{WT}$, $NERT_{W10A}$ and $NERT_{N51A}$ ESCs. NANOG variants were induced in $NERT_{WT}$, $NERT_{W10A}$ and $NERT_{N51A}$ ESCs for 0, 1 or 3 hours and OCT4, SOX2 and NANOG protein levels were assessed by immunoblot. β-ACTIN was used as a loading control.

samples. The signal of NANOG variants, OCT4, SOX2 and RNAPII before tamoxifen treatment at NANOG binding sites was compared (method section 2.13.2) (Fig. 6.1). As expected, NANOG binding was not detectable before tamoxifen treatment at majority of NANOG binding sites in all examined NERT cell lines (Fig. 6.1). However, low levels of NANOG binding was detected in the wild-type NERT cells before NANOG induction at sites where NANOG normally binds with the highest intensity. This might be due to an unanticipated transport of a small proportion of NANOG-ER^{T2} to the nucleus and a higher activity of wild-type NANOG compared to other NANOG variants. Binding of RNAPII and SOX2 before NANOG induction is fairly similar in all of the cell lines. However, OCT4 binding is on average almost 2-fold higher in the wild-type NERT cells compared to the $NERT_{W10A}$ cells (Fig. 6.1). The OCT4 signal is even lower in the $NERT_{N51A}$ ESCs at these conditions. As the OCT4 expression is similar in all three examined cell lines (Fig. 6.2), different OCT4 binding is possibly caused by a batch effect between the OCT4 ChIP samples.

6.1 $NANOG_{W10A}$ has impaired chromatin binding

A possibility that the W10A mutation impairs the binding of NANOG to chromatin could explain the observed defects of the $NANOG_{W10A}$ variant in the ES self-renewal (Mullin et al. 2017). Therefore, the average chromatin binding of wild-type, $NANOG_{W10A}$ and $NANOG_{N51A}$ variants at the different NANOG binding clusters after tamoxifen induction was compared (Fig. 6.3). Strikingly, $NANOG_{W10A}$ binding is hugely reduced at most of the NANOG binding sites, compared to the wild-type NANOG. The largest difference is in the cluster two with the average $NANOG_{W10A}$ binding intensity reduction by almost 10-fold (Fig. 6.3). The differences between $NANOG_{WT}$ and $NANOG_{W10A}$ are mainly at sites with a high intensity of NANOG signal. In the clusters five, six and seven, the differences are small. Interestingly, the average intensity of $NANOG_{W10A}$ is very similar in all the clusters. In addition, $NANOG_{W10A}$ binds stronger than $NANOG_{WT}$ at sites with vanishing levels of wild-type NANOG binding in the

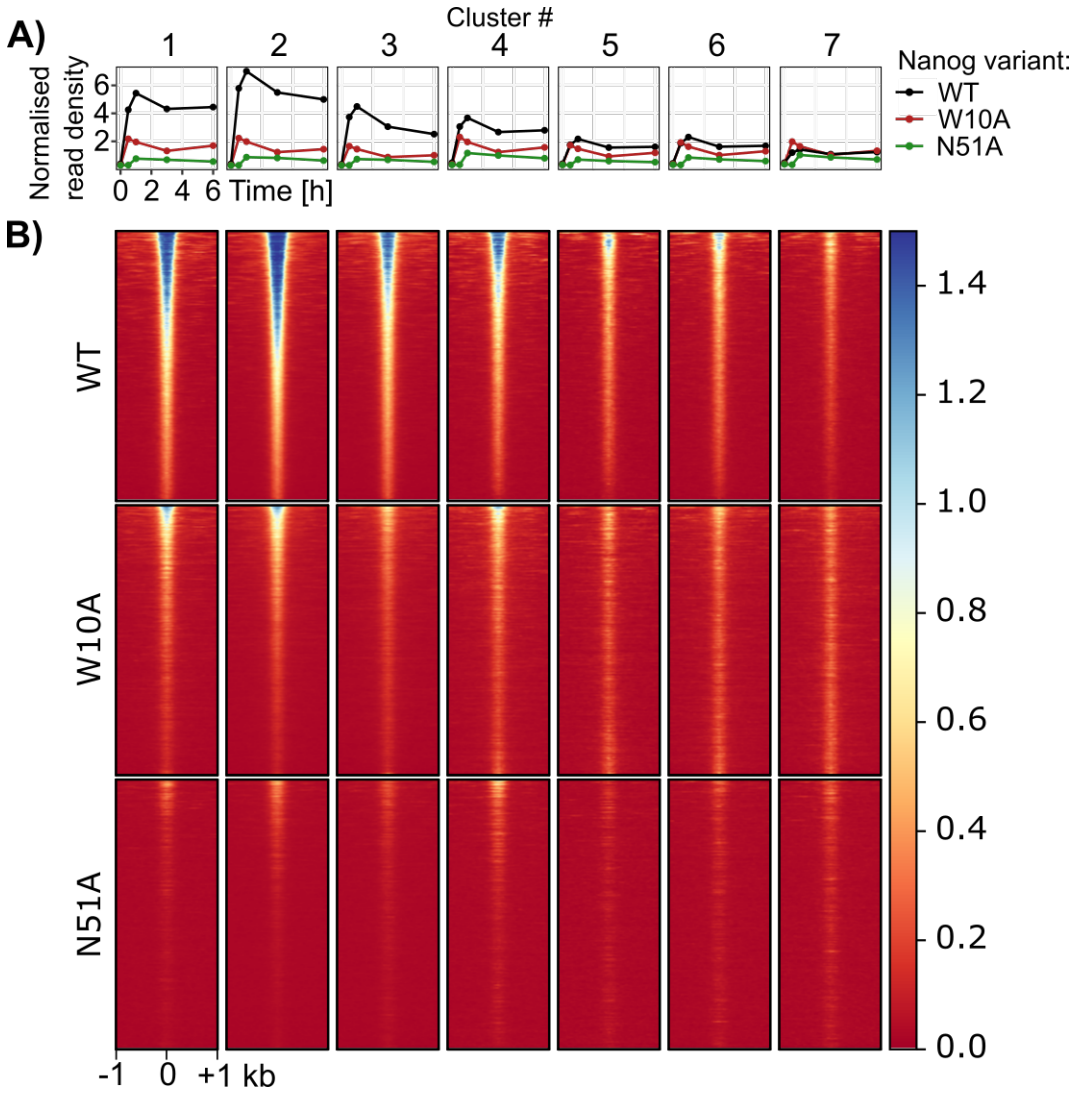


Figure 6.3: The WR is required for efficient NANOG chromatin binding. **A)** Average read density of $NANOG_{WT}$, $NANOG_{W10A}$ and $NANOG_{N51A}$ variants was measured at different time points after tamoxifen treatment of NERT ESCs. **B)** Combined signal of chromatin binding of individual NANOG variants after the induction at individual NANOG binding sites was visualised in form of a heatmap. Each row represents a single NANOG binding site extended by 1 kb each direction. Rows were ordered by the $NANOG_{WT}$ signal. Density plots represent the mean normalised signal of each NANOG mutant at the particular binding cluster.

clusters four, five and six. As expected, $NANOG_{N51A}$ variant, with the defect in DNA binding, is unable to bind most of the NANOG binding sites. However, $NANOG_{N51A}$ binds at the sites where wild-type NANOG binding has the highest intensity.

These results demonstrate that $NANOG_{W10A}$ has an impaired capacity to bind chromatin. This could be caused by a direct effect of the WR on the homeodomain or indirectly through protein-protein interactions. To determine if W10A mutation has a direct effect on NANOG capacity to bind DNA, affinity of the pure $NANOG_{WT}$ and $NANOG_{W10A}$ proteins to DNA was measured by electrophoretic mobility shift assay (EMSA) (method section 2.11.7). Surprisingly, the dissociation constant (K_d) of the $NANOG_{W10A}$ is almost 2-fold higher than the K_d of $NANOG_{WT}$ (Fig. 6.4). The measurements were consistent across three replicates. The bound fraction of the DNA reaches a plateau around 0.8 meaning that the full saturation was not reached (Fig. 6.4). Therefore, the accuracy of the measured apparent dissociation constants might be reduced (Altschuler et al. 2013).

6.2 $NANOG_{W10A}$ fails to increase OCT4, SOX2 and RNAPII binding at cis-regulatory elements

NANOG increases OCT4, SOX2 and RNAPII binding at thousands of sites grouped mainly in the identified clusters one, two and three. On the other hand, NANOG induction leads to a subtle decrease of RNAPII binding in the clusters four to seven. OCT4 and SOX2 binding decreases as well in the cluster seven (Fig. 5.2). To assess the roles of NANOG domains in this process, OCT4, SOX2 and RNAPII chromatin binding after W10A and $NANOG_{N51A}$ induction at identified NANOG binding sites was analysed (Fig. 6.5, 6.6 and 6.7).

Induction of the $NANOG_{W10A}$ fails to increase OCT4 binding at clusters one, two and three compared to wild-type NANOG (Fig. 6.5). At the cluster one, OCT4 signal increases gradually after $NANOG_{W10A}$ induction compared

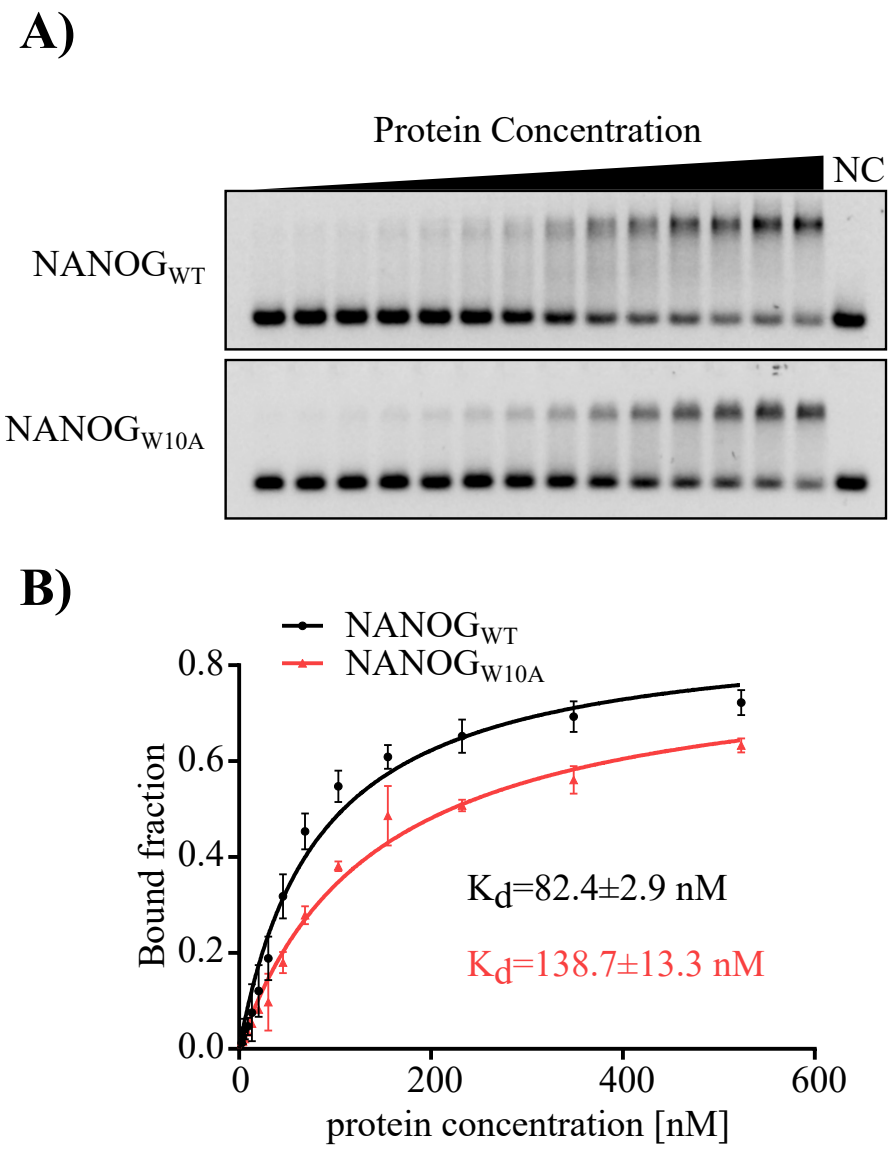


Figure 6.4: NANOG$_{W10A}$ has decreased affinity to DNA. A) Example of electrophoretic mobility shift assays of purified wild-type (NANOG$_{WT}$) and NANOG$_{W10A}$ proteins. Electrophoretic mobility of the DNA was measured at different concentrations of the NANOG$_{WT}$ and NANOG$_{W10A}$ proteins. The negative control (NC) contained no protein. **B)** Dissociation constant (K_d) for each NANOG mutant was determined by a non-linear fit to the measured data. n = 3, error bars represent standard deviation from the mean (points).

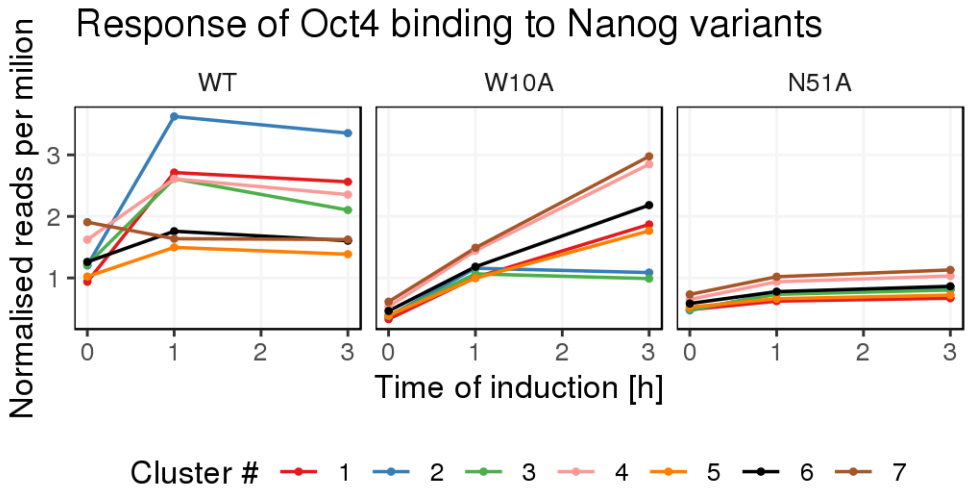


Figure 6.5: The WR is required for a specific OCT4 recruitment to NANOG binding sites. Average intensity of OCT4 ChIP-seq signal at NANOG binding clusters was measured after induction of wild-type, W10A and $NANOG_{N51A}$ mutants at 0, 1 and 3 hours. Values represent the median of normalised read counts at each cluster.

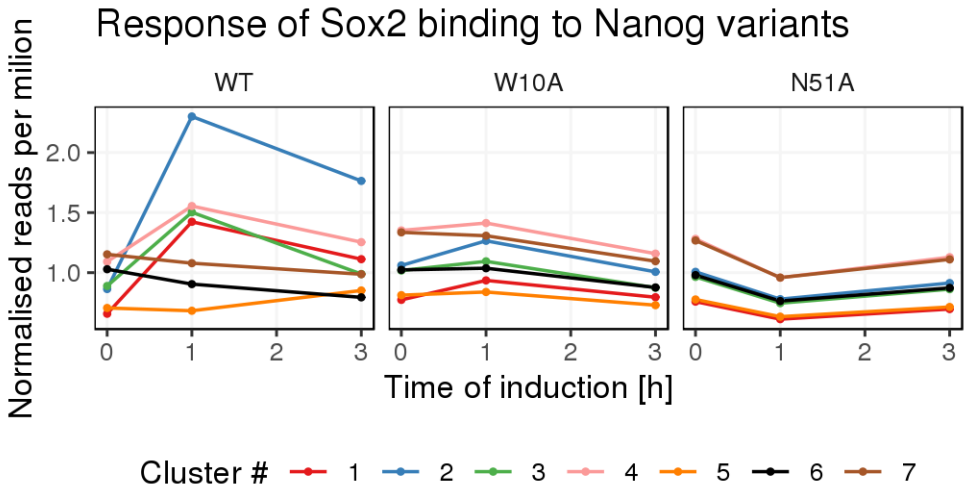


Figure 6.6: The NANOG WR is required for efficient SOX2 recruitment to the NANOG binding sites. Average intensity of SOX2 ChIP-seq signal at NANOG binding clusters was measured after induction of wild-type, W10A and $NANOG_{N51A}$ mutants at 0, 1 and 3 hours. Values represent the median of normalised read counts at each cluster.

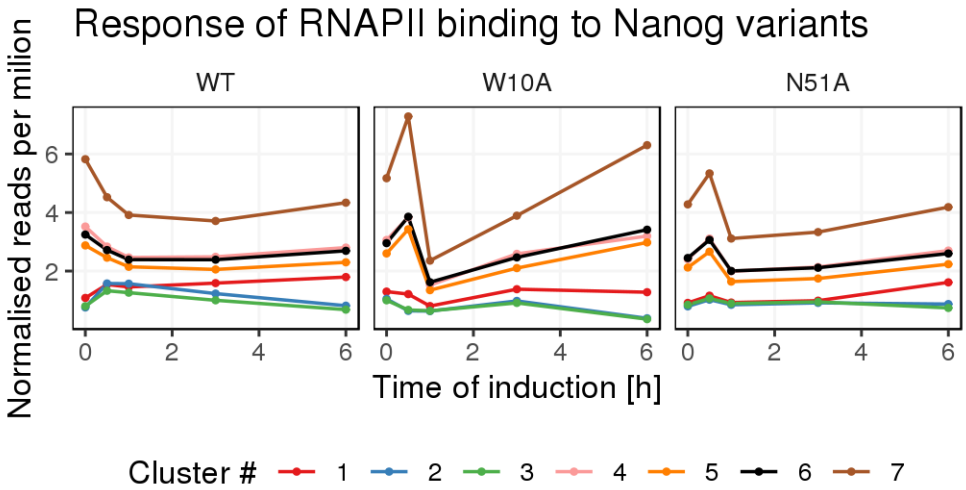


Figure 6.7: Mutation of the WR prevents efficient recruitment of RNAPII by NANOG to specific NANOG binding sites. Average intensity of RNAPII ChIP-seq signal at the NANOG binding clusters was measured after induction of wild-type, W10A and $NANOG_{N51A}$ mutants at three time points. Values represent the median of normalised read counts at each cluster.

to a sharp increase after wild-type NANOG induction. At clusters two and three, the increase of OCT4 binding after $NANOG_{W10A}$ induction is lower by 2 to 3.5-fold compared to wild-type NANOG induction and remains unchanged at the 3 hour time point. Interestingly, at clusters four to seven, $NANOG_{W10A}$ induction results in a higher OCT4 binding at 3 h compared to the wild-type induction. The most pronounced differences is at cluster seven. $NANOG_{N51A}$ induction has only a small effect on OCT4 binding at all examined clusters (Fig. 6.3). $NANOG_{W10A}$ induction do not alter SOX2 binding to same extent as wild-type NANOG induction (Fig. 6.6). The major differences are observed at clusters one and two where SOX2 binding is approximately 1.5 to 2-fold lower after $NANOG_{W10A}$ induction compared to induction of wild-type NANOG. Interestingly, SOX2 binding is decreased after 1 h of $NANOG_{N51A}$ induction at all clusters. The SOX2 binding is restored at 3 h time point. RNAPII binding is also influenced by $NANOG_{W10A}$ differently. At the clusters two and three, RNAPII binding is initially decreased by $NANOG_{W10A}$ induction compared to the increase caused by the wild-type NANOG induction (Fig. 6.7). $NANOG_{N51A}$ induction has a marginal effect on RNAPII binding at clusters two and three. At the cluster one, RNAPII binding is lower at 1 and 6 h time points after $NANOG_{W10A}$ induction compared to wild-type NANOG induction. $NANOG_{N51A}$ induction increased RNAPII binding at the cluster one similarly to wild-type NANOG but with delayed kinetics. At the clusters four to seven, W10A and $NANOG_{N51A}$ variants initially increase RNAPII binding after 30 min which is followed by a sharp decrease at 1 h. This is in contrast to a gradual decrease of RNAPII binding after wild-type NANOG induction (Fig. 6.7). At later time points, $NANOG_{W10A}$ induction increases RNAPII binding with the most noticeable effect at the cluster seven. $NANOG_{N51A}$ induction has a marginal effect on RNAPII binding at clusters four to seven after 1 h.

6.3 $NANOG_{W10A}$ fails to regulate its gene targets efficiently

Mutation of the NANOG WR, results in a lower chromatin binding and inability of NANOG to recruit OCT4, SOX2 and RNAPII to thousands of NANOG binding sites (Fig. 6.3, 6.5, 6.6 and 6.7). Does this influence the capacity of NANOG to regulate its gene targets as well? To answer this, the expression of previously identified NANOG gene targets after induction of wild-type and $NANOG_{W10A}$ proteins in the $NANOG_{WT}$ and $NANOG_{W10A}$ ESCs was compared. Expression tables and a list of differential expressed genes (DEGs) generated by Alison McGarvey was used.

The capacity of $NANOG_{W10A}$ to activate and repress genes is strongly reduced (Fig. 6.8). The differences between wild-type NANOG and $NANOG_{W10A}$ induction are clear at all the measured time points with a high statistical confidence. Not only the magnitude of the gene regulation but also the speed of the gene response to NANOG is decreased by the W10A mutation. There are evident defects of $NANOG_{W10A}$ to regulate its gene targets.

6.4 Discussion

The hallmark function of NANOG is to sustain LIF-independent self-renewal of ESCs (Chambers et al. 2003). This function is lost when NANOG loses the capacity to bind DNA or if the tryptophan repeat region (WR) is disrupted (Mullin et al. 2008; Mullin et al. 2017). The WR NANOG mutant (W10A) is unable to homodimerise (Mullin et al. 2017) and interact with other proteins including SOX2 and RNAPII (Mullin et al. unpublished). Comparison of gene regulatory potential of wild-type and $NANOG_{W10A}$ showed that the $NANOG_{W10A}$ mutant has a reduced capacity to regulate its gene targets (Fig. 6.8).

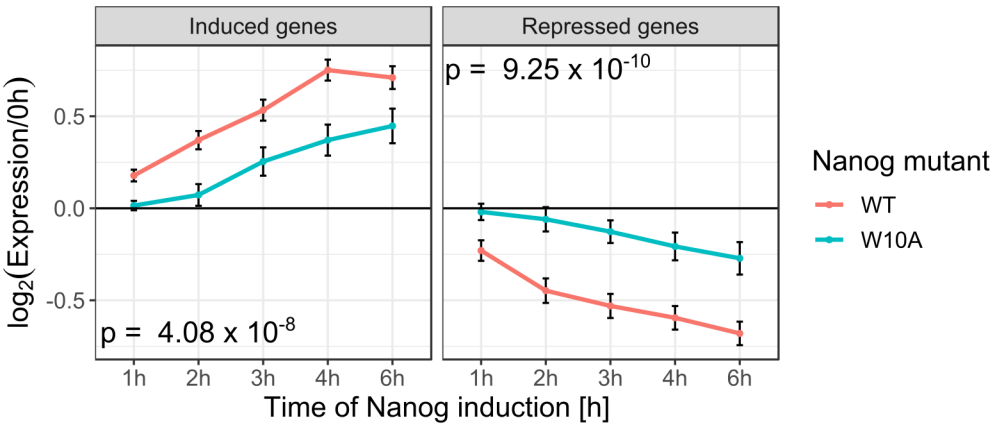


Figure 6.8: $NANOG_{W10A}$ fails to regulate its gene targets efficiently. Fold change expression of induced and repressed Nanog gene targets was measured after wild-type NANOG or $NANOG_{W10A}$ induction. The bars represent 95% confidence interval. The highest p-value among the different time points is reported.

To further reveal the role of the NANOG WR in gene regulation, chromatin binding of wild-type and $NANOG_{W10A}$ variants was compared. Surprisingly, $NANOG_{W10A}$ has largely reduced occupancy at most of the NANOG binding sites. At some sites, $NANOG_{W10A}$ binding is more than 10-fold lower (Fig. 6.3). The decreased affinity of the $NANOG_{W10A}$ to DNA *in vitro* supports the hypothesis that the WR is important for NANOG DNA binding (Fig. 6.4). However, the full binding saturation has not been reached which could decrease the accuracy of the measured apparent dissociation constant (Altschuler et al. 2013). The WR mediated NANOG homodimerisation could therefore stabilise NANOG-DNA binding. However, a direct experimental comparison is required to support this. The DNA binding of the SATB1 homeodomain protein has been also shown to be dependent on homeodomain independent homodimerisation (Purbey et al. 2008). Purbey et al. used EMSA experiments to demonstrate that high affinity binding of SATB1 to a palindromic DNA motif is dependent on homeodomain independent homodimerisation. To my knowledge, this is the only example of a homeodomain protein positively regulating its capacity to bind DNA by homodimerisation through a separate domain. However, it is unclear why NANOG homodimer binds a single NANOG motif with a higher affinity. It could be possible that there is a unknown sequence within the probe which would allow the binding of two NANOG molecules at the same time. This could be resolved by repeating the measurements with shorter DNA probes and different nucleotide composition. Alternatively, NANOG homodimerisation could allosterically alter the physical properties of the protein and stabilise the protein-DNA interaction. To answer this, a structural analysis of the W10A and wild-type NANOG proteins would be required.

Even though $NANOG_{W10A}$ has reduced chromatin binding, it still binds many NANOG binding sites. Interestingly, reduction of $NANOG_{W10A}$ chromatin binding at some NANOG sites is much higher than 2-fold reduction of DNA binding observed *in vitro* (Fig. 6.3, 6.4). This suggests that high affinity chromatin binding depends both on affinity to DNA and

additional protein-protein interactions. $NANOG_{W10A}$ induction, similarly to the $NANOG_{N51A}$ induction, fails to efficiently increase OCT4, SOX2 and RNAPII binding at the clusters one, two and three (Fig. 6.5, 6.6 and 6.7). The largest defect is at the sites in the cluster two. In the previous chapter, based on the observation of high occupancy of OCT4, SOX2 and RNAPII dependent on the presence of wild-type NANOG at these sites and the proposed model of the gene activation via phase separation (Boija et al. 2018; Hnisz et al. 2017), I proposed that NANOG might be required for a formation of protein condensates. The formation of the protein condensates is mediated through weak multi protein-protein interactions of a low-complexity domains (LCDs) (Boeynaems et al. 2018). NANOG, OCT4, SOX2 and RNAPII have been shown to form these protein condensates together with MED1, both in ESCs and *in vitro* (Boija et al. 2018; Sabari et al. 2018; Chong et al. 2018; Cho et al. 2018). The NANOG WR has also characteristics of a LCD. Therefore, it is possible that the W10A mutation prevents multi protein-protein interaction and results in a relatively weak NANOG, OCT4, SOX2 and RNAPII binding at these distal regulatory elements. Since the cluster one and two are significantly associated with the genes activated by NANOG (Fig. 2.13.5), the WR is likely to mediate formation of large regulatory complexes at these sites leading to gene activation. This hypothesis is consistent with previously proposed model in which TFs mediate gene activation through formation of protein condensates (Boija et al. 2018).

Moreover, in contrast to the wild-type NANOG induction, RNAPII binding is decreased approximately 2-fold at the 1 h time point at these clusters (Fig. 6.7). We can speculate that RNAPII recruitment by NANOG is redundant in the serum/LIF condition due to RNAPII recruitment by another factors, for example STAT3. It is known that *Nanog* is redundant for ESC self-renewal in the presence of LIF (Chambers et al. 2003) and it has been reported that NANOG is in fact redundant in OCT4, SOX2, ESRRB and BRG1 recruitment at some binding sites in the presence of LIF (Heurtier et al. 2019). Therefore it is possible that shortly after the $NANOG_{W10A}$ induction, $NANOG_{W10A}$ competes

with other TFs recruiting RNAPII. Since the $NANOG_{W10A}$ cannot interact with RNAPII (Mullin et al., unpublished), this results in a reduction of the RNAPII binding at these sites in the earliest time points. We can speculate that an unknown feedback process causes re-adjustment of the RNAPII and other TF binding in the later time points observed across the genome (Fig. 5.2 and 6.7).

Interestingly, the induction of DNA binding NANOG mutant ($NANOG_{N51A}$) has a small negative effect on SOX2 chromatin binding at the 1 h time point (Fig. 6.6). As NANOG directly interacts with SOX2 (Gagliardi et al. 2013), this could be caused by a competition between $NANOG_{N51A}$-SOX2 and SOX2-DNA binding leading to a decrease of the local SOX2 concentration at these genomic sites. However it is unknown why would $NANOG_{N51A}$-SOX2 complex bind chromatin weaker or less frequently than SOX2 alone and why is SOX2 binding restored at the 3 h time point (Fig. 6.6).

At the second major group of the NANOG binding sites (clusters four to seven), the differences between $NANOG_{WT}$, $NANOG_{W10A}$ and $NANOG_{N51A}$ induction on SOX2 and RNAPII binding are small (Fig. 6.6 and 6.7). Interestingly, $NANOG_{W10A}$ induction increases OCT4 occupancy to higher extent than the wild-type NANOG induction at 3 h. This is contradicting to the previously observed behaviours in the clusters two and three (Fig. 6.5). Therefore, it is possible that the clusters four to seven have a distinct functional roles to clusters one to three. However, the mechanisms underlying this abnormal OCT4 binding after the $NANOG_{W10A}$ induction at these clusters is unclear. Clusters four and seven are significantly associated with the genes repressed by NANOG (Fig. 5.4), therefore it could be possible that wild-type NANOG recruits a co-repressor via the WR which restricts OCT4 binding and activation of these genes. A possible candidate could be NCOR1 which directly interacts with NANOG through the WR (Mullin et al., unpublished). However, effect of NANOG on chromatin binding of corepressor complexes remains to be determined.

7 Concluding remarks and future directions

7.1 *Resf1* function in mouse development

Nanog plays an essential role in the establishment of the pre-implantation epiblast of a mouse embryo and it is important for efficient specification of the germ line later in the development. The hallmark function of NANOG capacity to sustain self-renewal of ESCs without LIF. This function is dependent on the tryptophan repeat (WR) within the C-terminal transactivation domain. Interestingly, the WR is also required for NANOG homodimerisation and interaction with a handful of other proteins. Therefore, I hypothesised that NANOG function is mediated through WR dependent protein-protein interactions. RESF1 is one of the NANOG protein partner interacting through the WR (Mullin et al. unpublished). RESF1 is an uncharacterised protein and its function in pluripotency and development is unknown. In this thesis, I studied the requirements of *Resf1* for ESC self-renewal and NANOG function.

I demonstrated that *Resf1* is not required for ESC and EpiSC self-renewal. Moreover, the capacity of *Resf1*$^{-/-}$ embryos to reach adulthood was recently reported (Dickinson et al. 2016). This demonstrates that *Resf1* is not essential for mouse embryo development. However, the *Resf1*$^{-/-}$ embryos appear in lower than Mendelian frequencies (Dickinson et al. 2016) indicating that *Resf1* has a supportive role during mouse development. Interestingly, ESCs lacking *Resf1* cultured in serum/LIF express lower amounts of the key pluripotency markers and display an increased propensity to differentiate (Fig. 3.7 3.8). Forced expression of *Resf1* has a small positive effect on the ESC self-renewal (Fig. 3.9). This suggests that RESF1 directly or indirectly supports the pluripotency GRN. In the following chapter, I showed that even though RESF1 is a nuclear protein interacting with NANOG in ESCs, it is not essential for NANOG function to sustain LIF-independent self-renewal (Fig.

3.11). Therefore, it is possible that RESF1-NANOG partnerships is functionally important in a different process.

NANOG is required for efficient specification of PGCs (Chambers et al. 2007; Zhang et al. 2018b) and NANOG overexpression bypasses the otherwise essential BMP signalling in the PGC induction (Murakami et al. 2016). Therefore, I tested *Resf1* requirements for specification of PGCLCs. The *Resf1*$^{-/-}$ cells display reduced capacity to activate expression of the key PGC markers and to form PGCLCs (Fig. 4.3 and 4.4). This could explain why *Resf1* knockout mice are infertile (Dickinson et al. 2016). Immunostaining of PGC markers, like DAZL, STELLA or BLIMP1 for example, in E7.25 *Resf1*$^{-/-}$ embryo could provide a direct evidence of *Resf1* requirements for early PGC specification. This would support the model that RESF1 acts together with NANOG in the PGC specification. However, *Nanog null* embryos, unlike *Resf1 null* embryos, are fertile (Zhang et al. 2018b; Dickinson et al. 2016). Therefore, RESF1 could be required in multiple steps of the germ line specification. RESF1 has been proposed to directly silence retrotransposons in ESCs (Fukuda et al. 2018). Inefficient silencing of retrotransposons leads to infertility (Molaro et al. 2014). Therefore, it is also possible that RESF1 safeguards proper germ line development by retrotransposon silencing.

I generated tools useful for further analysis of RESF1 function in pluripotency and PGC specification. The *Resf1*$^{-/-}$ ESCs (Fig. 3.5) could be used to generate inducible RESF1 ESCs. This would enable both the verification of RESF1 knockout phenotypes and determine if RESF1 regulates pluripotency genes and retrotransposable elements in ESCs directly. The regulatory potential of RESF1 in the PGC specification could be examined as well. The established *in vitro* PGCLC specification could serve as a functional readout allowing determination of the time window of RESF1 function. Rescue experiments with truncated versions of RESF1 could reveal functional domains of RESF1 which are currently unknown. This would be particularly interesting, since RESF1 has no homology to any functionally known protein domain. To determine if RESF1 is required for NANOG function in PGC specification,

NANOG can be overexpressed in PGC competent *Resf1*$^{-/-}$ EpiLCs and the activation of the key PGCLC markers could be evaluated.

Since there are no commercially available mouse RESF1 antibodies, the generated endogenously tagged *Resf1* ES cell lines (Rv5) (Fig. 3.2) are particularly useful. V5-tagged RESF1 can be immunoprecipiated and its interactions with DNA, RNA and proteins examined. In addition, Rv5 ESCs contain a single loxP site flanking the 3' end of the *Resf1* gene. This allows generation of conditional knockout *Resf1* cell line by insertion of a single loxP site into the 5' end of *Resf1* gene. Conditional knockout cell line could be further used to generate *Resf1*$^{flox/flox}$ mice which would allow assessment of *Resf1* function at specific developmental stage after a cross with a mouse expressing Cre under a lineage specific promoter.

In conclusion, I demonstrated that RESF1 is a nuclear protein interacting with NANOG and that RESF1 is not essential for NANOG function and ESC self-renewal. However, *Resf1* deletion causes decreased expression of the key pluripotency markers. Therefore it is possible that *Resf1* contributes to ESC self-renewal independently from NANOG. The observation of *Resf1* requirements for efficient PGCLC specification offer a possible explanation of mouse infertility caused by *Resf1* deletion.

7.2 The molecular model of NANOG function

NANOG binds tens of thousands sites within the ESC genome. Many of these sites are shared with other TFs including OCT4, SOX2 and RNAPII (Chen et al. 2008). However, the influence of NANOG on chromatin binding of these factors is unknown.

In this thesis, I determined seven clusters of NANOG binding sites with distinct response patterns of OCT4, SOX2 and RNAPII chromatin binding to NANOG induction. OCT4, SOX2 and RNAPII quickly increase chromatin binding at thousands of NANOG binding sites after NANOG induction (Fig. 5.2 and 5.3). The clusters one and two, displaying this behaviour, significantly associate with the genes activated by NANOG induction (Fig. 5.4). Therefore,

NANOG is likely to activate its gene targets by recruiting OCT4, SOX2 and RNAPII to cis-regulatory elements close to these sites (Fig. 7.1). The mutation of the NANOG WR or the NANOG homeodomain prevents NANOG to increase OCT4, SOX2 and RNAPII chromatin binding (Fig. 6.5 and 6.6 and 6.7) which can explain the reduced capacity of these NANOG mutants to regulate gene targets (Fig. 6.8). The NANOG WR contains ten repeats of five amino acids which classify this region as a low-complexity domain (LCD). Some LCDs can mediate formation of large protein condensates linked to gene regulation (Furlong et al. 2018; Hnisz et al. 2017). Therefore, it is possible that NANOG may act as a scaffold protein accumulating other transcription factors including OCT4, SOX2 and RNAPII via the WR. In fact, a strong chromatin binding of NANOG, OCT4 SOX2 and RNAPII at these sites (Fig. 5.3) and their capacity to form large protein condensates in ESCs (Boija et al. 2018; Sabari et al. 2018; Chong et al. 2018; Cho et al. 2018) supports this idea. Therefore, It would be interesting to determine the capacity of the WR to undergo a liquid-liquid phase separation *in vitro* and the capacity to form condensates on the ESC enhancers.

In addition, I demonstrated that the mutation of the NANOG WR decreases affinity of NANOG DNA binding *in vitro* (Fig. 2.11.7) and chromatin binding in ESCs (Fig. 6.3). This suggests that the WR is not only important for NANOG protein interactions but also for efficient DNA binding. This might be a result of a dual capacity of the WR to interact with itself and other proteins. NANOG homodimer could bind DNA more stably compare to a monomer similarly to SATB1 (Purbey et al. 2008). The induction of the DNA-binding deficient NANOG mutant ($NANOG_{N51A}$) has little effect on OCT4, SOX2 and RNAPII chromatin binding (Fig. 6.5, 6.6 and 6.7) demonstrating that the WR cannot support OCT4, SOX2 and RNAPII binding independently of DNA binding. This means that stabilisation of the transcription complex requires NANOG to bind DNA and be able to homodimerise and/or heterodimerise.

The second major group of NANOG binding sites (clusters four to seven) are defined by reduction of RNAPII chromatin binding shortly after NANOG

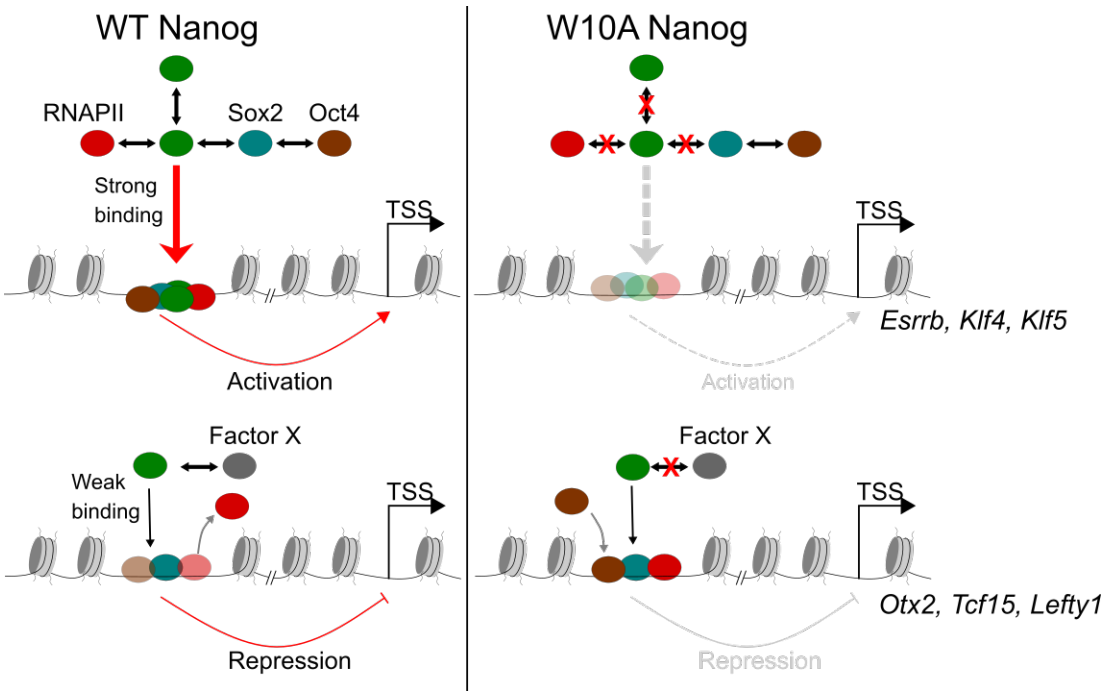


Figure 7.1: A model of NANOG function in gene regulation of ESC self-renewal. Wild-type NANOG, capable of homodimerisation and interaction with RNAPII and SOX2, recruits OCT4, SOX2 and RNAPII to cis-regulatory elements associated with activation of the NANOG gene targets. NANOG recruits unknown corepressor (Factor X) which restrict OCT4, SOX2 and NANOG binding, decreases RNAPII occupancy at sites associated with gene repressed by NANOG. Mutation of the NANOG tryptophan region (WR) for alanines (W10A) prevents NANOG homodimerisation and heterodimerisation. $NANOG_{W10A}$ is unable to efficiently bind chromatin, increase OCT4, SOX2 and RNAPII binding and activate NANOG gene targets. The $NANOG_{W10A}$ is also unable to recruit a putative corepressor (Factor X) which leads to increased OCT4 binding and inefficient gene repression.

induction (Fig. 5.2 and 5.3). The clusters four and seven are significantly associated with the genes downregulated by NANOG including *Otx2, Tcf15* and *Lefty1* (Fig. 5.4 and 7.1). Interestingly, even though these genomic regions are accessible (Fig. 5.3) the binding of NANOG, OCT4 and SOX2 is low compared to the NANOG binding sites from the group one (Fig. 5.3). This suggests that the binding of OCT4, SOX2 and NANOG is restricted by additional factor (Fig. 7.1). We can speculate that NANOG recruits a transcriptional corepressor which causes a reduction of RNAPII and other TFs occupancy leading to gene repression (Fig. 7.1). The small effect size could be caused by a redundant repression of the same genes by LIF signalling. This is supported by results of Heurtier et al. 2019 who reported that NANOG induction in the absence of LIF leads to an increase of polycomb repressive mark H3K27me3 at the TSSs of the genes repressed by NANOG whereas in the presence of LIF the effect is virtually not detectable. Interestingly, the induction of the NANOG WR mutant ($NANOG_{W10A}$) caused recruitment of OCT4 to these sites (Fig. 6.5). It might be possible that $NANOG_{W10A}$ is unable to recruit a corepressor complex to these sites while still binding the chromatin at the comparable levels to the wild-type NANOG. Therefore, it could be possible that $NANOG_{W10A}$ competes with other TFs which recruit the corepressor and allows OCT4 binding. However, it is not clear why the same pattern was not observed for SOX2 and RNAPII.

To further clarify the proposed model of NANOG function (Fig. 7.1), multiple additional experiments can be performed. The determination of the NANOG binding sites which are functional in ESCs would enable us to determine rules separating non-functional TF binding from the TF binding regulating gene expression. By focusing on the functional binding sites, we would be able to reduce noise in the observed data and propose more defined molecular model of NANOG function. The definition of functional NANOG binding sites is possible by combining multi-parallel report assays with assays enriching for NANOG binding sites similar to previously published approaches (Barakat et al. 2018; Muerdter et al. 2015). The reporter assays can also

validate the association between observed changes of TF binding and changes in the gene expression.

To further examine how NANOG homo and hetero-dimerisation influences binding of NANOG, OCT4, SOX2 and RNAPII to naked DNA or chromatin in ESCs, several experiments could be performed. Firstly, direct comparison of wild-type NANOG and $NANOG_{W10A}$ DNA binding by EMSA can determine if NANOG binds naked DNA as a dimer even if only a single homeodomain is interacting with the DNA. The analytical measurement of DNA affinity of NANOG mutants in complex with its partners would directly show that protein-protein-DNA interaction is important for stabilisation of this complex and strong DNA binding. A formation of large protein assemblies mediated by the NANOG WR could possibly be resolved by super-resolution microscopy. This would provide additional evidence of the role of the NANOG WR to nucleate formation of transcriptional complexes activating expression of the NANOG gene targets.

The mechanism of NANOG mediated gene repression can be further assessed by analysis of corepressor chromatin binding after NANOG induction. The likely candidates are NCOR1, polycomb complex and NuRD complex as they have been shown to interact with NANOG in ESCs (Gagliardi et al. 2013). The comparison of wild-type and $NANOG_{W10A}$ mutants will determine requirements of the WR for potential recruitment of these repressor complexes to the identified sites within clusters four and seven.

In summary, I demonstrated the capacity of NANOG to rapidly recruit OCT4, SOX2 and RNAPII to thousands of genomic sites mostly associated with gene activation. The NANOG DNA binding and the WR are essential for recruitment of these proteins. On the other hand, I determined a second group of NANOG binding sites which are significantly associated with gene repression. These sites display distinct characteristics with a noticeable reduction of RNAPII binding 30 min after NANOG induction. These results expand the understanding of the molecular processes underlying NANOG function and the roles of individual NANOG domains in these processes.

References

Acampora, Dario, Di Giovannantonio, Luca Giovanni, Garofalo, Arcomaria, Nigro, Vincenzo, Omodei, Daniela, Lombardi, Alessia, Zhang, Jingchao, Chambers, Ian and Simeone, Antonio (2017). "Functional antagonism between OTX2 and NANOG specifies a spectrum of heterogeneous identities in embryonic stem cells". In: *Stem cell reports* 9.5, pp. 1642–1659.

Altschuler, Sarah E, Lewis, Karen A and Wuttke, Deborah S (2013). "Practical strategies for the evaluation of high-affinity protein/nucleic acid interactions". In: *Journal of nucleic acids investigation* 4.1, p. 19.

Ambrosetti, Davide-Carlo, Basilico, Claudio and Dailey, Lisa (1997). "Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites." In: *Molecular and cellular biology* 17.11, pp. 6321–6329.

Ambrosetti, Davide-Carlo, Schöler, Hans R, Dailey, Lisa and Basilico, Claudio (2000). "Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4enhancer". In: *Journal of Biological Chemistry* 275.30, pp. 23387–23397.

Arman, Esther, Haffner-Krausz, Rebecca, Chen, Yali, Heath, John K and Lonai, Peter (1998). "Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development". In: *Proceedings of the National Academy of Sciences* 95.9, pp. 5082–5087.

Arnold, Cosmas D, Gerlach, Daniel, Stelzer, Christoph, Boryń, Łukasz M, Rath, Martina and Stark, Alexander (2013). "Genome-wide quantitative enhancer activity maps identified by STARR-seq". In: *Science* 339.6123, pp. 1074–1077.

Aubert, Jerôme, Dunstan, Hannah, Chambers, Ian and Smith, Austin (2002). "Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation". In: *Nature biotechnology* 20.12, p. 1240.

Avilion, Ariel A, Nicolis, Silvia K, Pevny, Larysa H, Perez, Lidia, Vivian, Nigel and Lovell-Badge, Robin (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function". In: *Genes & development* 17.1, pp. 126–140.

Banerji, Julian, Rusconi, Sandro and Schaffner, Walter (1981). "Expression of a β-globin gene is enhanced by remote SV40 DNA sequences". In: *Cell* 27.2, pp. 299–308.

Barakat, Tahsin Stefan, Halbritter, Florian, Zhang, Man, Rendeiro, Andre F, Perenthaler, Elena, Bock, Christoph and Chambers, Ian (2018). "Functional dissection of the enhancer repertoire in human embryonic stem cells". In: *Cell stem cell* 23.2, pp. 276–288.

Bedzhov, Ivan and Zernicka-Goetz, Magdalena (2014). “Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation”. In: *Cell* 156.5, pp. 1032–1044.

Benezra, Robert, Davis, Robert L, Lockshon, Daniel, Turner, David L and Weintraub, Harold (1990). “The protein Id: a negative regulator of helix-loop-helix DNA binding proteins”. In: *Cell* 61.1, pp. 49–59.

Berg, Debbie LC van den, Snoek, Tim, Mullin, Nick P, Yates, Adam, Bezstarosti, Karel, Demmers, Jeroen, Chambers, Ian and Poot, Raymond A (2010). “An Oct4-centered protein interaction network in embryonic stem cells”. In: *Cell stem cell* 6.4, pp. 369–381.

Bernardo, Andreia S, Faial, Tiago, Gardner, Lucy, Niakan, Kathy K, Ortmann, Daniel, Senner, Claire E, Callery, Elizabeth M, Trotter, Matthew W, Hemberger, Myriam, Smith, James C et al. (2011). “BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages”. In: *Cell stem cell* 9.2, pp. 144–155.

Bessonnard, Sylvain, De Mot, Laurane, Gonze, Didier, Barriol, Manon, Dennis, Cynthia, Goldbeter, Albert, Dupont, Geneviève and Chazaud, Claire (2014). “Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network”. In: *Development* 141.19, pp. 3637–3648.

Boeynaems, Steven, Alberti, Simon, Fawzi, Nicolas L, Mittag, Tanja, Polymenidou, Magdalini, Rousseau, Frederic, Schymkowitz, Joost, Shorter, James, Wolozin, Benjamin, Van Den Bosch, Ludo et al. (2018). “Protein phase separation: a new phase in cell biology”. In: *Trends in cell biology* 28.6, pp. 420–435.

Boija, Ann, Klein, Isaac A, Sabari, Benjamin R, Dall’Agnese, Alessandra, Coffey, Eliot L, Zamudio, Alicia V, Li, Charles H, Shrinivas, Krishna, Manteiga, John C, Hannett, Nancy M et al. (2018). “Transcription factors activate genes through the phase-separation capacity of their activation domains”. In: *Cell* 175.7, pp. 1842–1855.

Boroviak, Thorsten, Loos, Remco, Bertone, Paul, Smith, Austin and Nichols, Jennifer (2014). “The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification”. In: *Nature cell biology* 16.6, p. 513.

Boroviak, Thorsten, Loos, Remco, Lombard, Patrick, Okahara, Junko, Behr, Rüdiger, Sasaki, Erika, Nichols, Jennifer, Smith, Austin and Bertone, Paul (2015). “Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis”. In: *Developmental cell* 35.3, pp. 366–382.

Bowles, Josephine, Schepers, Goslik and Koopman, Peter (2000). “Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators”. In: *Developmental biology* 227.2, pp. 239–255.

Boyer, Laurie A, Lee, Tong Ihn, Cole, Megan F, Johnstone, Sarah E, Levine, Stuart S, Zucker, Jacob P, Guenther, Matthew G, Kumar, Roshan M,

Murray, Heather L, Jenner, Richard G et al. (2005). “Core transcriptional regulatory circuitry in human embryonic stem cells”. In: *cell* 122.6, pp. 947–956.

Boyle, Alan P, Davis, Sean, Shulha, Hennady P, Meltzer, Paul, Margulies, Elliott H, Weng, Zhiping, Furey, Terrence S and Crawford, Gregory E (2008). “High-resolution mapping and characterization of open chromatin across the genome”. In: *Cell* 132.2, pp. 311–322.

Bradley, Allan, Evans, Martin, Kaufman, Matthew H and Robertson, Elizabeth (1984). “Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines”. In: *Nature* 309.5965, p. 255.

Brehm, Alexander, Ohbo, Kazuyuki and Schöler, H (1997). “The carboxy-terminal transactivation domain of Oct-4 acquires cell specificity through the POU domain.” In: *Molecular and cellular biology* 17.1, pp. 154–162.

Brons, I Gabrielle M, Smithers, Lucy E, Trotter, Matthew WB, Rugg-Gunn, Peter, Sun, Bowen, Sousa Lopes, Susana M Chuva de, Howlett, Sarah K, Clarkson, Amanda, Ahrlund-Richter, Lars, Pedersen, Roger A et al. (2007). “Derivation of pluripotent epiblast stem cells from mammalian embryos”. In: *Nature* 448.7150, p. 191.

Brumbaugh, Justin, Russell, Jason D, Yu, Pengzhi, Westphall, Michael S, Coon, Joshua J and Thomson, James A (2014). “NANOG is multiply phosphorylated and directly modified by ERK2 and CDK1 in vitro”. In: *Stem cell reports* 2.1, pp. 18–25.

Buecker, Christa, Srinivasan, Rajini, Wu, Zhixiang, Calo, Eliezer, Acampora, Dario, Faial, Tiago, Simeone, Antonio, Tan, Minjia, Swigut, Tomasz and Wysocka, Joanna (2014). “Reorganization of enhancer patterns in transition from naive to primed pluripotency”. In: *Cell stem cell* 14.6, pp. 838–853.

Buehr, Mia, Meek, Stephen, Blair, Kate, Yang, Jian, Ure, Janice, Silva, Jose, McLay, Renee, Hall, John, Ying, Qi-Long and Smith, Austin (2008). “Capture of authentic embryonic stem cells from rat blastocysts”. In: *Cell* 135.7, pp. 1287–1298.

Burdon, Tom, Stracey, Craig, Chambers, Ian, Nichols, Jennifer and Smith, Austin (1999). “Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells”. In: *Developmental biology* 210.1, pp. 30–43.

Camus, Anne, Perea-Gomez, Aitana, Moreau, Anne and Collignon, Jérôme (2006). “Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo”. In: *Developmental biology* 295.2, pp. 743–755.

Carter, Ava C, Davis-Dusenbery, Brandi N, Koszka, Kathryn, Ichida, Justin K and Eggan, Kevin (2014). “Nanog-independent reprogramming to iPSCs with canonical factors”. In: *Stem cell reports* 2.2, pp. 119–126.

Chambers, Ian, Colby, Douglas, Robertson, Morag, Nichols, Jennifer, Lee, Sonia, Tweedie, Susan and Smith, Austin (2003). “Functional

expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells". In: *Cell* 113.5, pp. 643–655.

Chambers, Ian, Silva, Jose, Colby, Douglas, Nichols, Jennifer, Nijmeijer, Bianca, Robertson, Morag, Vrana, Jan, Jones, Ken, Grotewold, Lars and Smith, Austin (2007). "Nanog safeguards pluripotency and mediates germline development". In: *Nature* 450.7173, p. 1230.

Chambers, Ian and Tomlinson, Simon R (2009). "The transcriptional foundation of pluripotency". In: *Development* 136.14, pp. 2311–2322.

Chazaud, Claire, Yamanaka, Yojiro, Pawson, Tony and Rossant, Janet (2006). "Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway". In: *Developmental cell* 10.5, pp. 615–624.

Chen, Haixia, Guo, Renpeng, Zhang, Qian, Guo, Hongchao, Yang, Meng, Wu, Zhenfeng, Gao, Shan, Liu, Lin and Chen, Lingyi (2015). "Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells". In: *Proceedings of the National Academy of Sciences* 112.44, E5936–E5943.

Chen, Xi, Xu, Han, Yuan, Ping, Fang, Fang, Huss, Mikael, Vega, Vinsensius B, Wong, Eleanor, Orlov, Yuriy L, Zhang, Weiwei, Jiang, Jianming et al. (2008). "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells". In: *Cell* 133.6, pp. 1106–1117.

Cheng, Alec M, Saxton, Tracy M, Sakai, Ryuichi, Kulkarni, Sarang, Mbamalu, Geraldine, Vogel, Wolfgang, Tortorice, Christopher G, Cardiff, Robert D, Cross, James C, Muller, William J et al. (1998). "Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation". In: *Cell* 95.6, pp. 793–803.

Chew, Joon-Lin, Loh, Yuin-Han, Zhang, Wensheng, Chen, Xi, Tam, Wai-Leong, Yeap, Leng-Siew, Li, Pin, Ang, Yen-Sin, Lim, Bing, Robson, Paul et al. (2005). "Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells". In: *Molecular and cellular biology* 25.14, pp. 6031–6046.

Chickarmane, Vijay and Peterson, Carsten (2008). "A computational model for understanding stem cell, trophectoderm and endoderm lineage determination". In: *PLoS one* 3.10, e3478.

Cho, Won-Ki, Spille, Jan-Hendrik, Hecht, Micca, Lee, Choongman, Li, Charles, Grube, Valentin and Cisse, Ibrahim I (2018). "Mediator and RNA polymerase II clusters associate in transcription-dependent condensates". In: *Science* 361.6400, pp. 412–415.

Cho, Won-Ki, Jayanth, Namrata, English, Brian P, Inoue, Takuma, Andrews, J Owen, Conway, William, Grimm, Jonathan B, Spille, Jan-Hendrik, Lavis, Luke D, Lionnet, Timothee et al. (2016). "RNA Polymerase II cluster dynamics predict mRNA output in living cells". In: *Elife* 5, e13617.

Chong, Shasha, Dugast-Darzacq, Claire, Liu, Zhe, Dong, Peng, Dailey, Gina M, Cattoglio, Claudia, Heckert, Alec, Banala, Sambashiva, Lavis, Luke, Darzacq, Xavier et al. (2018). "Imaging dynamic and selective

low-complexity domain interactions that control gene transcription". In: *Science* 361.6400, eaar2555.

Chronis, Constantinos, Fiziev, Petko, Papp, Bernadett, Butz, Stefan, Bonora, Giancarlo, Sabri, Shan, Ernst, Jason and Plath, Kathrin (2017). "Cooperative binding of transcription factors orchestrates reprogramming". In: *Cell* 168.3, pp. 442–459.

Cinalli, Ryan M, Rangan, Prashanth and Lehmann, Ruth (2008). "Germ cells are forever". In: *Cell* 132.4, pp. 559–562.

Cisse, Ibrahim I, Izeddin, Ignacio, Causse, Sebastien Z, Boudarene, Lydia, Senecal, Adrien, Muresan, Leila, Dugast-Darzacq, Claire, Hajj, Bassam, Dahan, Maxime and Darzacq, Xavier (2013). "Real-time dynamics of RNA polymerase II clustering in live human cells". In: *Science* 341.6146, pp. 664–667.

Clift, Dean and Schuh, Melina (2013). "Restarting life: fertilization and the transition from meiosis to mitosis". In: *Nature reviews Molecular cell biology* 14.9, p. 549.

Cole, Megan F, Johnstone, Sarah E, Newman, Jamie J, Kagey, Michael H and Young, Richard A (2008). "Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells". In: *Genes & development* 22.6, pp. 746–755.

Corsinotti, Andrea, Wong, Frederick CK, Tatar, Tülin, Szczerbinska, Iwona, Halbritter, Florian, Colby, Douglas, Gogolok, Sabine, Pantier, Raphael, Liggat, Kirsten, Mirfazeli, Elham S et al. (2017). "Distinct SoxB1 networks are required for naïve and primed pluripotency". In: *Elife* 6, e27746.

DeBruine, Zachary J, Xu, HE and Melcher, Karsten (2017). "Assembly and architecture of the Wnt/β-catenin signalosome at the membrane". In: *British journal of pharmacology* 174.24, pp. 4564–4574.

Dickinson, Mary E, Flenniken, Ann M, Ji, Xiao, Teboul, Lydia, Wong, Michael D, White, Jacqueline K, Meehan, Terrence F, Weninger, Wolfgang J, Westerberg, Henrik, Adissu, Hibret et al. (2016). "High-throughput discovery of novel developmental phenotypes". In: *Nature* 537.7621, p. 508.

Do, Dang Vinh, Ueda, Jun, Messerschmidt, Daniel M, Lorthongpanich, Chanchao, Zhou, Yi, Feng, Bo, Guo, Guoji, Lin, Peiyu J, Hossain, Md Zakir, Zhang, Wenjun et al. (2013). "A genetic and developmental pathway from STAT3 to the OCT4–NANOG circuit is essential for maintenance of ICM lineages in vivo". In: *Genes & development* 27.12, pp. 1378–1390.

Dowen, Jill M, Fan, Zi Peng, Hnisz, Denes, Ren, Gang, Abraham, Brian J, Zhang, Lyndon N, Weintraub, Abraham S, Schuijers, Jurian, Lee, Tong Ihn, Zhao, Keji et al. (2014). "Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes". In: *Cell* 159.2, pp. 374–387.

Duval, D, Malaise, M, Reinhardt, B, Kedinger, C and Boeuf, H (2004). "A p38 inhibitor allows to dissociate differentiation and apoptotic processes

triggered upon LIF withdrawal in mouse embryonic stem cells". In: *Cell death and differentiation* 11.3, p. 331.

Eckersley-Maslin, Melanie A, Alda-Catalinas, Celia and Reik, Wolf (2018). "Dynamics of the epigenetic landscape during the maternal-to-zygotic transition". In: *Nature Reviews Molecular Cell Biology* 19.7, p. 436.

Ernst, Matthias, Oates, Andrew and Dunn, Ashley R (1996). "Gp130-mediated signal transduction in embryonic stem cells involves activation of Jak and Ras/mitogen-activated protein kinase pathways". In: *Journal of Biological Chemistry* 271.47, pp. 30136–30143.

Esch, Daniel, Vahokoski, Juha, Groves, Matthew R, Pogenberg, Vivian, Cojocaru, Vlad, Vom Bruch, Hermann, Han, Dong, Drexler, Hannes CA, Araúzo-Bravo, Marcos J, Ng, Calista KL et al. (2013). "A unique Oct4 interface is crucial for reprogramming to pluripotency". In: *Nature cell biology* 15.3, p. 295.

Evans, Martin J and Kaufman, Matthew H (1981). "Establishment in culture of pluripotential cells from mouse embryos". In: *Nature* 292.5819, p. 154.

Evsikov, AV, De Vries, WN, Peaston, AE, Radford, EE, Fancher, KS, Chen, FH, Blake, JA, Bult, CJ, Latham, KE, Solter, D et al. (2004). "Systems biology of the 2-cell mouse embryo". In: *Cytogenetic and genome research* 105.2-4, pp. 240–250.

Faunes, Fernando, Hayward, Penelope, Descalzo, Silvia Muñoz, Chatterjee, Sujash S, Balayo, Tina, Trott, Jamie, Christoforou, Andrew, Ferrer-Vaquer, Anna, Hadjantonakis, Anna-Katerina, Dasgupta, Ramanuj et al. (2013). "A membrane-associated β-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells". In: *Development* 140.6, pp. 1171–1183.

Festuccia, Nicola, Halbritter, Florian, Corsinotti, Andrea, Gagliardi, Alessia, Colby, Douglas, Tomlinson, Simon R and Chambers, Ian (2018). "Esrrb extinction triggers dismantling of naïve pluripotency and marks commitment to differentiation". In: *The EMBO journal* 37.21.

Festuccia, Nicola, Osorno, Rodrigo, Halbritter, Florian, Karwacki-Neisius, Violetta, Navarro, Pablo, Colby, Douglas, Wong, Frederick, Yates, Adam, Tomlinson, Simon R and Chambers, Ian (2012). "Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells". In: *Cell stem cell* 11.4, pp. 477–490.

Festuccia, Nicola, Osorno, Rodrigo, Wilson, Valerie and Chambers, Ian (2013). "The role of pluripotency gene regulatory network components in mediating transitions between pluripotent cell states". In: *Current Opinion in Genetics & Development* 23.5, pp. 504–511.

Frankenberg, Stephen, Gerbe, François, Bessonnard, Sylvain, Belville, Corinne, Pouchin, Pierre, Bardot, Olivier and Chazaud, Claire (2011). "Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling". In: *Developmental cell* 21.6, pp. 1005–1013.

Frum, Tristan and Ralston, Amy (2015). "Cell signaling and transcription factors regulating cell fate during formation of the mouse blastocyst". In: *Trends in Genetics* 31.7, pp. 402–410.

Fuda, Nicholas J, Ardehali, M Behfar and Lis, John T (2009). “Defining mechanisms that regulate RNA polymerase II transcription in vivo”. In: *Nature* 461.7261, p. 186.

Fukuda, Kei, Okuda, Akihiko, Yusa, Kosuke and Shinkai, Yoichi (2018). “A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells”. In: *Genome research*.

Furlong, Eileen EM and Levine, Michael (2018). “Developmental enhancers and chromosome topology”. In: *Science* 361.6409, pp. 1341–1345.

Gagliardi, Alessia, Mullin, Nicholas P, Tan, Zi Ying, Colby, Douglas, Kousa, Anastasia I, Halbritter, Florian, Weiss, Jason T, Felker, Anastasia, Bezstarosti, Karel, Favaro, Rebecca et al. (2013). “A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal”. In: *The EMBO journal* 32.16, pp. 2231–2247.

Galonska, Christina, Ziller, Michael J, Karnik, Rahul and Meissner, Alexander (2015). “Ground state conditions induce rapid reorganization of core pluripotency factor binding before global epigenetic reprogramming”. In: *Cell stem cell* 17.4, pp. 462–470.

Gao, Chan and Chen, Ye-Guang (2010). “Dishevelled: The hub of Wnt signaling”. In: *Cellular signalling* 22.5, pp. 717–727.

Ginsburg, MALKA, Snow, MH and McLAREN, ANNE (1990). “Primordial germ cells in the mouse embryo during gastrulation”. In: *Development* 110.2, pp. 521–528.

Greber, Boris, Wu, Guangming, Bernemann, Christof, Joo, Jin Young, Han, Dong Wook, Ko, Kinarm, Tapia, Natalia, Sabour, Davood, Sterneckert, Jared, Tesar, Paul et al. (2010). “Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells”. In: *Cell stem cell* 6.3, pp. 215–226.

Guibert, Sylvain, Forné, Thierry and Weber, Michael (2012). “Global profiling of DNA methylation erasure in mouse primordial germ cells”. In: *Genome research*.

Guo, Ge, Huang, Yue, Humphreys, Peter, Wang, Xiaozhong and Smith, Austin (2011). “A PiggyBac-based recessive screening method to identify pluripotency regulators”. In: *PloS one* 6.4, e18189.

Guo, Ge, Yang, Jian, Nichols, Jennifer, Hall, John Simon, Eyres, Isobel, Mansfield, William and Smith, Austin (2009). “Klf4 reverts developmentally programmed restriction of ground state pluripotency”. In: *Development* 136.7, pp. 1063–1069.

Habibi, Ehsan, Brinkman, Arie B, Arand, Julia, Kroeze, Leonie I, Kerstens, Hindrik HD, Matarese, Filomena, Lepikhov, Konstantin, Gut, Marta, Brun-Heath, Isabelle, Hubner, Nina C et al. (2013). “Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells”. In: *Cell stem cell* 13.3, pp. 360–369.

Hackett, Jamie A and Surani, M Azim (2014). “Regulatory principles of pluripotency: from the ground state up”. In: *Cell stem cell* 15.4, pp. 416–430.

Haeckel, Ernst (1868). *Natürliche Schöpfungsgeschichte: Gemeinverständliche wissenschaftliche Vorträge über die Entwickelungslehre im Allgemeinen und diejenige von Darwin, Göthe und Lamarck im Besonderen, über die Anwendung derselben auf den Ursprung des Menschen und andern damit zusammenhängende Gründfragen der Natur-Wissenschaft. Mit Tafeln, Holzschnitten, systematischen und genealogischen Tabellen*. Walter de Gruyter.

Hajkova, Petra, Erhardt, Sylvia, Lane, Natasha, Haaf, Thomas, El-Maarri, Osman, Reik, Wolf, Walter, Jörn and Surani, M Azim (2002). "Epigenetic reprogramming in mouse primordial germ cells". In: *Mechanisms of development* 117.1-2, pp. 15–23.

Hall, John, Guo, Ge, Wray, Jason, Eyres, Isobel, Nichols, Jennifer, Grotewold, Lars, Morfopoulou, Sofia, Humphreys, Peter, Mansfield, William, Walker, Rachael et al. (2009). "Oct4 and LIF/Stat3 additively induce Krüppel factors to sustain embryonic stem cell self-renewal". In: *Cell stem cell* 5.6, pp. 597–609.

Hamatani, Toshio, Carter, Mark G, Sharov, Alexei A and Ko, Minoru SH (2004). "Dynamics of global gene expression changes during mouse preimplantation development". In: *Developmental cell* 6.1, pp. 117–131.

Handyside, Alan H, O'Neill, Gerard T, Jones, Mary and Hooper, Martin L (1989). "Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line". In: *Roux's archives of developmental biology* 198.1, pp. 48–56.

Hayashi, Katsuhiko, Ogushi, Sugako, Kurimoto, Kazuki, Shimamoto, So, Ohta, Hiroshi and Saitou, Mitinori (2012). "Offspring from oocytes derived from in vitro primordial germ cell–like cells in mice". In: *Science*, p. 1222482.

Hayashi, Katsuhiko, Ohta, Hiroshi, Kurimoto, Kazuki, Aramaki, Shinya and Saitou, Mitinori (2011). "Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells". In: *Cell* 146.4, pp. 519–532. ISSN: 0092-8674.

Hayashi, Katsuhiko and Saitou, Mitinori (2013). "Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells". In: *Nature protocols* 8.8, p. 1513.

Hayashi, Katsuhiko, Sousa Lopes, Susana M Chuva de, Tang, Fuchou and Surani, M Azim (2008). "Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states". In: *Cell stem cell* 3.4, pp. 391–401.

Heurtier, Victor, Owens, Nick, Gonzalez, Inma, Mueller, Florian, Proux, Caroline, Mornico, Damien, Clerc, Philippe, Dubois, Agnes and Navarro, Pablo (2019). "The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells". In: *Nature communications* 10.1, p. 1109.

Hnisz, Denes, Shrinivas, Krishna, Young, Richard A, Chakraborty, Arup K and Sharp, Phillip A (2017). "A phase separation model for transcriptional control". In: *Cell* 169.1, pp. 13–23.

Hnisz, Denes, Weintraub, Abraham S, Day, Daniel S, Valton, Anne-Laure, Bak, Rasmus O, Li, Charles H, Goldmann, Johanna, Lajoie, Bryan R,

Fan, Zi Peng, Sigova, Alla A et al. (2016). “Activation of proto-oncogenes by disruption of chromosome neighborhoods”. In: *Science* 351.6280, pp. 1454–1458.

Huang, Yali, Osorno, Rodrigo, Tsakiridis, Anestis and Wilson, Valerie (2012). “In vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation”. In: *Cell reports* 2.6, pp. 1571–1578.

Hudson, Quanah J, Smith, Craig A and Sinclair, Andrew H (2005). “Conserved expression of a novel gene during gonadal development”. In: *Developmental dynamics: an official publication of the American Association of Anatomists* 233.3, pp. 1083–1090.

Institute, Broad (2018). *Picard Tools*. `http://broadinstitute.github.io/picard/`. Accessed: 2019-07-04; version 2.20.2.

Jackson, Dean A, Iborra, Francisco J, Manders, Erik MM and Cook, Peter R (1998). “Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei”. In: *Molecular biology of the cell* 9.6, pp. 1523–1536.

Jauch, Ralf, Choo, S, Ng, C and Kolatkar, P (2011). “Crystal Structure of the Dimeric Oct6 (Pou3fl) POU Domain Bound to Palindromic MORE DNA”. In: *Proteins: Structure Functions and Bioinformatics* 79.BNL-97886-2012-JA.

Jauch, Ralf, Ng, Calista Keow Leng, Saikatendu, Kumar Singh, Stevens, Raymond C and Kolatkar, Prasanna R (2008). “Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog”. In: *Journal of molecular biology* 376.3, pp. 758–770.

Ji, Xiong, Dadon, Daniel B, Powell, Benjamin E, Fan, Zi Peng, Borges-Rivera, Diego, Shachar, Sigal, Weintraub, Abraham S, Hnisz, Denes, Pegoraro, Gianluca, Lee, Tong Ihn et al. (2016). “3D chromosome regulatory landscape of human pluripotent cells”. In: *Cell stem cell* 18.2, pp. 262–275.

Jin, Jiali, Liu, Jian, Chen, Cong, Liu, Zhenping, Jiang, Cong, Chu, Hongshang, Pan, Weijuan, Wang, Xinbo, Zhang, Lingqiang, Li, Bin et al. (2016). “The deubiquitinase USP21 maintains the stemness of mouse embryonic stem cells via stabilization of Nanog”. In: *Nature communications* 7, p. 13594.

Johnson, David S, Mortazavi, Ali, Myers, Richard M and Wold, Barbara (2007). “Genome-wide mapping of in vivo protein-DNA interactions”. In: *Science* 316.5830, pp. 1497–1502.

Johnson, Martin H and Ziomek, Carol Ann (1981). “The foundation of two distinct cell lineages within the mouse morula”. In: *Cell* 24.1, pp. 71–80.

Joshi, Onkar, Wang, Shuang-Yin, Kuznetsova, Tatyana, Atlasi, Yaser, Peng, Tianran, Fabre, Pierre J, Habibi, Ehsan, Shaik, Jani, Saeed, Sadia, Handoko, Lusy et al. (2015). “Dynamic reorganization of extremely long-range promoter-promoter interactions between two states of pluripotency”. In: *Cell stem cell* 17.6, pp. 748–757.

Juven-Gershon, Tamar, Hsu, Jer-Yuan, Theisen, Joshua WM and Kadonaga, James T (2008). “The RNA polymerase II core promoter—the gateway to transcription”. In: *Current opinion in cell biology* 20.3, pp. 253–259.

Kamachi, Yusuke and Kondoh, Hisato (2013). “Sox proteins: regulators of cell fate specification and differentiation”. In: *Development* 140.20, pp. 4129–4144.

Kang, Minjung, Piliszek, Anna, Artus, Jérôme and Hadjantonakis, Anna-Katerina (2013). “FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse”. In: *Development* 140.2, pp. 267–279.

Karwacki-Neisius, Violetta, Göke, Jonathan, Osorno, Rodrigo, Halbritter, Florian, Ng, Jia Hui, Weiße, Andrea Y, Wong, Frederick CK, Gagliardi, Alessia, Mullin, Nicholas P, Festuccia, Nicola et al. (2013). “Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog”. In: *Cell stem cell* 12.5, pp. 531–545.

Keramari, Maria, Razavi, Janet, Ingman, Karen A, Patsch, Christoph, Edenhofer, Frank, Ward, Christopher M and Kimber, Susan J (2010). “Sox2 is essential for formation of trophectoderm in the preimplantation embryo”. In: *PloS one* 5.11, e13952.

Kim, Kun-Yong, Tanaka, Yoshiaki, Su, Juan, Cakir, Bilal, Xiang, Yangfei, Patterson, Benjamin, Ding, Junjun, Jung, Yong-Wook, Kim, Ji-Hyun, Hysolli, Eriona et al. (2018). “Uhrf1 regulates active transcriptional marks at bivalent domains in pluripotent stem cells through Setd1a”. In: *Nature communications* 9.1, p. 2583.

King, Hamish W and Klose, Robert J (2017). “The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells”. In: *Elife* 6, e22631.

Kinsella, Rhoda J, Kähäri, Andreas, Haider, Syed, Zamora, Jorge, Proctor, Glenn, Spudich, Giulietta, Almeida-King, Jeff, Staines, Daniel, Derwent, Paul, Kerhornou, Arnaud et al. (2011). “Ensembl BioMarts: a hub for data retrieval across taxonomic space”. In: *Database* 2011.

Kiyonari, Hiroshi, Kaneko, Mari, Abe, Shin-ichi and Aizawa, Shinichi (2010). “Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability”. In: *genesis* 48.5, pp. 317–327.

Klemm, Juli D and Pabo, Carl O (1996). “Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage.” In: *Genes & development* 10.1, pp. 27–36.

Klemm, Juli D, Rould, Mark A, Aurora, Rajeev, Herr, Winship and Pabo, Carl O (1994). “Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules”. In: *Cell* 77.1, pp. 21–32.

Kojima, Yoji, Kaufman-Francis, Keren, Studdert, Joshua B, Steiner, Kirsten A, Power, Melinda D, Loebel, David AF, Jones, Vanessa, Hor, Angelyn, Alencastro, Gustavo de, Logan, Grant J et al. (2014). “The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak”. In: *Cell stem cell* 14.1, pp. 107–120.

Krueger, Felix (2019). *Trim Galore*. (Visited on 04/07/2019).

Kunath, Tilo, Saba-El-Leil, Marc K, Almousailleakh, Marwa, Wray, Jason, Meloche, Sylvain and Smith, Austin (2007). "FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment". In: *Development* 134.16, pp. 2895–2902.

Kuroda, Takao, Tada, Masako, Kubota, Hiroshi, Kimura, Hironobu, Hatano, Shin-ya, Suemori, Hirofumi, Nakatsuji, Norio and Tada, Takashi (2005). "Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression". In: *Molecular and cellular biology* 25.6, pp. 2475–2485.

Laird, Diana J (2018). *How to lose your inheritance*.

Lane, Natasha, Dean, Wendy, Erhardt, Sylvia, Hajkova, Petra, Surani, Azim, Walter, Jörn and Reik, Wolf (2003). "Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse". In: *genesis* 35.2, pp. 88–93.

Langmead, Ben and Salzberg, Steven L (2012). "Fast gapped-read alignment with Bowtie 2". In: *Nature methods* 9.4, p. 357.

Latchman, David S (2001). "Transcription factors: bound to activate or repress". In: *Trends in biochemical sciences* 26.4, pp. 211–213.

Lawson, KA and Hage, WJ (1994). "Clonal analysis of the origin of primordial germ cells in the mouse". In: *Germline development* 165, pp. 68–84.

Lees-Murdock, DJ, De Felici, M and Walsh, CP (2003). "Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage". In: *Genomics* 82.2, pp. 230–237.

Leitch, Harry G, McEwen, Kirsten R, Turp, Aleksandra, Encheva, Vesela, Carroll, Tom, Grabole, Nils, Mansfield, William, Nashun, Buhe, Knezovich, Jaysen G, Smith, Austin et al. (2013). "Naive pluripotency is associated with global DNA hypomethylation". In: *Nature structural & molecular biology* 20.3, p. 311.

Lettice, Laura A, Williamson, Iain, Devenney, Paul S, Kilanowski, Fiona, Dorin, Julia and Hill, Robert E (2014). "Development of five digits is controlled by a bipartite long-range cis-regulator". In: *Development* 141.8, pp. 1715–1725.

Li, Heng, Handsaker, Bob, Wysoker, Alec, Fennell, Tim, Ruan, Jue, Homer, Nils, Marth, Gabor, Abecasis, Goncalo and Durbin, Richard (2009). "The sequence alignment/map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079.

Li, Ping, Tong, Chang, Mehrian-Shai, Ruty, Jia, Li, Wu, Nancy, Yan, Youzhen, Maxson, Robert E, Schulze, Eric N, Song, Houyan, Hsieh, Chih-Lin et al. (2008). "Germline competent embryonic stem cells derived from rat blastocysts". In: *Cell* 135.7, pp. 1299–1310.

Li, Zhongwei, Fei, Teng, Zhang, Jianping, Zhu, Gaoyang, Wang, Lu, Lu, Danyu, Chi, Xiaochun, Teng, Yan, Hou, Ning, Yang, Xiao et al. (2012). "BMP4 signaling acts via dual-specificity phosphatase 9 to control ERK activity in mouse embryonic stem cells". In: *Cell stem cell* 10.2, pp. 171–182.

Liao, Yang, Smyth, Gordon K and Shi, Wei (2019). “The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads”. In: *Nucleic acids research* 47.8, e47–e47.

Lim, Bomyi, Heist, Tyler, Levine, Michael and Fukaya, Takashi (2018). “Visualization of transvection in living Drosophila embryos”. In: *Molecular cell* 70.2, pp. 287–296.

Loh, Yuin-Han, Wu, Qiang, Chew, Joon-Lin, Vega, Vinsensius B, Zhang, Weiwei, Chen, Xi, Bourque, Guillaume, George, Joshy, Leong, Bernard, Liu, Jun et al. (2006). “The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells”. In: *Nature genetics* 38.4, p. 431.

Lu, Falong, Liu, Yuting, Inoue, Azusa, Suzuki, Tsukasa, Zhao, Keji and Zhang, Yi (2016). “Establishing chromatin regulatory landscape during mouse preimplantation development”. In: *Cell* 165.6, pp. 1375–1388.

Lyashenko, Natalia, Winter, Markus, Migliorini, Domenico, Biechele, Travis, Moon, Randall T and Hartmann, Christine (2011). “Differential requirement for the dual functions of β-catenin in embryonic stem cell self-renewal and germ layer formation”. In: *Nature cell biology* 13.7, p. 753.

Ma, Xinwei, Chen, Haixia and Chen, Lingyi (2016). “A dual role of Erk signaling in embryonic stem cells”. In: *Experimental hematology* 44.3, pp. 151–156.

Malaguti, Mattias, Nistor, Paul A, Blin, Guillaume, Pegg, Amy, Zhou, Xinzhi and Lowell, Sally (2013). “Bone morphogenic protein signalling suppresses differentiation of pluripotent cells by maintaining expression of E-Cadherin”. In: *Elife* 2, e01197.

Marks, Hendrik, Kalkan, Tüzer, Menafra, Roberta, Denissov, Sergey, Jones, Kenneth, Hofemeister, Helmut, Nichols, Jennifer, Kranz, Andrea, Stewart, A Francis, Smith, Austin et al. (2012). “The transcriptional and epigenomic foundations of ground state pluripotency”. In: *Cell* 149.3, pp. 590–604.

Martello, Graziano, Bertone, Paul and Smith, Austin (2013). “Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor”. In: *The EMBO journal* 32.19, pp. 2561–2574.

Martello, Graziano, Sugimoto, Toshimi, Diamanti, Evangelia, Joshi, Anagha, Hannah, Rebecca, Ohtsuka, Satoshi, Göttgens, Berthold, Niwa, Hitoshi and Smith, Austin (2012). “Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal”. In: *Cell stem cell* 11.4, pp. 491–504.

Martin, Gail R (1981). “Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells”. In: *Proceedings of the National Academy of Sciences* 78.12, pp. 7634–7638.

Masui, Shinji, Nakatake, Yuhki, Toyooka, Yayoi, Shimosato, Daisuke, Yagi, Rika, Takahashi, Kazue, Okochi, Hitoshi, Okuda, Akihiko, Matoba, Ryo, Sharov, Alexei A et al. (2007). “Pluripotency governed by Sox2 via

regulation of Oct3/4 expression in mouse embryonic stem cells". In: *Nature cell biology* 9.6, p. 625.

Matsuda, Takahiko, Nakamura, Takanori, Nakao, Kazuki, Arai, Takao, Katsuki, Motoya, Heike, Toshio and Yokota, Takashi (1999). "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells". In: *The EMBO journal* 18.15, pp. 4261–4269.

Meyenn, Ferdinand von, Iurlaro, Mario, Habibi, Ehsan, Liu, Ning Qing, Salehzadeh-Yazdi, Ali, Santos, Fátima, Petrini, Edoardo, Milagre, Inês, Yu, Miao, Xie, Zhenqing et al. (2016). "Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells". In: *Molecular cell* 62.6, pp. 848–861.

Mitsui, Kaoru, Tokuzawa, Yoshimi, Itoh, Hiroaki, Segawa, Kohichi, Murakami, Mirei, Takahashi, Kazutoshi, Maruyama, Masayoshi, Maeda, Mitsuyo and Yamanaka, Shinya (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells". In: *Cell* 113.5, pp. 631–642.

Miura, Taichi, Kume, Masahiko, Kawamura, Takeshi, Yamamoto, Kazuo, Hamakubo, Takao and Nishihara, Shoko (2018). "O-GlcNAc on PKCζ Inhibits the FGF4-PKCζ-MEK-ERK1/2 Pathway via Inhibition of PKCζ Phosphorylation in Mouse Embryonic Stem Cells". In: *Stem cell reports* 10.1, pp. 272–286.

Molaro, Antoine, Falciatori, Ilaria, Hodges, Emily, Aravin, Alexei A, Marran, Krista, Rafii, Shahin, McCombie, W Richard, Smith, Andrew D and Hannon, Gregory J (2014). "Two waves of de novo methylation during mouse germ cell development". In: *Genes & development* 28.14, pp. 1544–1549.

Moretto-Zita, Matteo, Jin, Hua, Shen, Zhouxin, Zhao, Tongbiao, Briggs, Steven P and Xu, Yang (2010). "Phosphorylation stabilizes Nanog by promoting its interaction with Pin1". In: *Proceedings of the National Academy of Sciences* 107.30, pp. 13312–13317.

Morgani, Sophie M, Canham, Maurice A, Nichols, Jennifer, Sharov, Alexei A, Migueles, Rosa Portero, Ko, Minoru SH and Brickman, Joshua M (2013). "Totipotent embryonic stem cells arise in ground-state culture conditions". In: *Cell reports* 3.6, pp. 1945–1957.

Morikawa, Masato, Koinuma, Daizo, Mizutani, Anna, Kawasaki, Natsumi, Holmborn, Katarina, Sundqvist, Anders, Tsutsumi, Shuichi, Watabe, Tetsuro, Aburatani, Hiroyuki, Heldin, Carl-Henrik et al. (2016). "BMP sustains embryonic stem cell self-renewal through distinct functions of different Krüppel-like factors". In: *Stem Cell Reports* 6.1, pp. 64–73.

Muerdter, Felix, Boryń, Łukasz M and Arnold, Cosmas D (2015). "STARR-seq—principles and applications". In: *Genomics* 106.3, pp. 145–150.

Mullen, Alan C, Orlando, David A, Newman, Jamie J, Lovén, Jakob, Kumar, Roshan M, Bilodeau, Steve, Reddy, Jessica, Guenther, Matthew G, DeKoter, Rodney P and Young, Richard A (2011). "Master transcription

factors determine cell-type-specific responses to TGF-β signaling". In: *Cell* 147.3, pp. 565–576.

Mullin, Nicholas P, Gagliardi, Alessia, Colby, Douglas, Hall-Ponsele, Elisa, Rowe, Arthur J, Chambers, Ian et al. (2017). "Distinct contributions of tryptophan residues within the dimerization domain to Nanog function". In: *Journal of molecular biology* 429.10, pp. 1544–1553.

Mullin, Nicholas and Chambers, Ian (2011). "The function of Nanog in pluripotency". In: *Nuclear Reprogramming and Stem Cells*. Springer, pp. 99–112.

Mullin, Nicholas, Yates, Adam, Rowe, Arthur J, Nijmeijer, Bianca, Colby, Douglas, Barlow, Paul N, Walkinshaw, Malcolm D and Chambers, Ian (2008). "The pluripotency rheostat Nanog functions as a dimer". In: *Biochemical Journal* 411.2, pp. 227–231.

Murakami, Kazuhiro, Günesdogan, Ufuk, Zylicz, Jan J, Tang, Walfred WC, Sengupta, Roopsha, Kobayashi, Toshihiro, Kim, Shinseog, Butler, Richard, Dietmann, Sabine and Surani, M Azim (2016). "NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers". In: *Nature* 529.7586, p. 403.

Nagy, Andras, Rossant, Janet, Nagy, Reka, Abramow-Newerly, Wanda and Roder, John C (1993). "Derivation of completely cell culture-derived mice from early-passage embryonic stem cells". In: *Proceedings of the National Academy of Sciences* 90.18, pp. 8424–8428.

Nichols, Jennifer, Jones, Kenneth, Phillips, Jenny M, Newland, Stephen A, Roode, Mila, Mansfield, William, Smith, Austin and Cooke, Anne (2009a). "Validated germline-competent embryonic stem cell lines from nonobese diabetic mice". In: *Nature medicine* 15.7, p. 814.

Nichols, Jennifer, Silva, Jose, Roode, Mila and Smith, Austin (2009b). "Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo". In: *Development* 136.19, pp. 3215–3222.

Nichols, Jennifer, Zevnik, Branko, Anastassiadis, Konstantinos, Niwa, Hitoshi, Klewe-Nebenius, Daniela, Chambers, Ian, Schöler, Hans and Smith, Austin (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4". In: *Cell* 95.3, pp. 379–391.

Nishimoto, Masazumi, Fukushima, Akiko, Okuda, Akihiko and Muramatsu, Masami (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". In: *Molecular and cellular biology* 19.8, pp. 5453–5465.

Nishioka, Noriyuki, Inoue, Ken-ichi, Adachi, Kenjiro, Kiyonari, Hiroshi, Ota, Mitsunori, Ralston, Amy, Yabuta, Norikazu, Hirahara, Shino, Stephenson, Robert O, Ogonuki, Narumi et al. (2009). "The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass". In: *Developmental cell* 16.3, pp. 398–410.

Niwa, Hitoshi, Burdon, Tom, Chambers, Ian and Smith, Austin (1998). “Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3”. In: *Genes & development* 12.13, pp. 2048–2060.

Niwa, Hitoshi, Miyazaki, Jun-ichi and Smith, Austin G (2000). “Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells”. In: *Nature genetics* 24.4, p. 372.

Niwa, Hitoshi, Ogawa, Kazuya, Shimosato, Daisuke and Adachi, Kenjiro (2009). “A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells”. In: *Nature* 460.7251, p. 118.

Novo, Clara Lopes, Javierre, Biola-Maria, Cairns, Jonathan, Segonds-Pichon, Anne, Wingett, Steven W, Freire-Pritchett, Paula, Furlan-Magaril, Mayra, Schoenfelder, Stefan, Fraser, Peter and Rugg-Gunn, Peter J (2018). “Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition”. In: *Cell reports* 22.10, pp. 2615–2627.

Noyes, Marcus B, Christensen, Ryan G, Wakabayashi, Atsuya, Stormo, Gary D, Brodsky, Michael H and Wolfe, Scot A (2008). “Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites”. In: *Cell* 133.7, pp. 1277–1289.

Nusse, Roel and Clevers, Hans (2017). “Wnt/β-catenin signaling, disease, and emerging therapeutic modalities”. In: *Cell* 169.6, pp. 985–999.

Ogawa, Kazuya, Nishinakamura, Ryuichi, Iwamatsu, Yuko, Shimosato, Daisuke and Niwa, Hitoshi (2006). “Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells”. In: *Biochemical and biophysical research communications* 343.1, pp. 159–166.

Ohinata, Yasuhide, Ohta, Hiroshi, Shigeta, Mayo, Yamanaka, Kaori, Wakayama, Teruhiko and Saitou, Mitinori (2009). “A signaling principle for the specification of the germ cell lineage in mice”. In: *Cell* 137.3, pp. 571–584.

Ohinata, Yasuhide, Payer, Bernhard, O’Carroll, Dónal, Ancelin, Katia, Ono, Yukiko, Sano, Mitsue, Barton, Sheila C, Obukhanych, Tetyana, Nussenzweig, Michel, Tarakhovsky, Alexander et al. (2005). “Blimp1 is a critical determinant of the germ cell lineage in mice”. In: *Nature* 436.7048, p. 207.

Okumura-Nakanishi, Sayaka, Saito, Motoki, Niwa, Hitoshi and Ishikawa, Fuyuki (2005). “Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells”. In: *Journal of Biological Chemistry* 280.7, pp. 5307–5317.

Ornitz, David M and Itoh, Nobuyuki (2015). “The fibroblast growth factor signaling pathway”. In: *Wiley Interdisciplinary Reviews: Developmental Biology* 4.3, pp. 215–266.

Oshimori, Naoki and Fuchs, Elaine (2012). “The harmonies played by TGF-β in stem cell biology”. In: *Cell stem cell* 11.6, pp. 751–764.

Osorno, Rodrigo, Tsakiridis, Anestis, Wong, Frederick, Cambray, Noemí, Economou, Constantinos, Wilkie, Ronald, Blin, Guillaume, Scotting, Paul J, Chambers, Ian and Wilson, Valerie (2012). “The developmental

dismantling of pluripotency is reversed by ectopic Oct4 expression". In: *Development* 139.13, pp. 2288–2298.

Ouyang, Zhengqing, Zhou, Qing and Wong, Wing Hung (2009). "ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells". In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21521–21526.

Pantier, Raphaël P (2018). "Molecular interactions of TET proteins in pluripotent cells". Doctoral dissertation. The University of Edinburgh.

Peaston, Anne E, Evsikov, Alexei V, Graber, Joel H, De Vries, Wilhelmine N, Holbrook, Andrea E, Solter, Davor and Knowles, Barbara B (2004). "Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos". In: *Developmental cell* 7.4, pp. 597–606.

Pevny, Larysa H and Lovell-Badge, Robin (1997). "Sox genes find their feet". In: *Current opinion in genetics & development* 7.3, pp. 338–344.

Piliszek, Anna, Grabarek, Joanna B, Frankenberg, Stephen R and Plusa, Berenika (2016). "Cell fate in animal and human blastocysts and the determination of viability". In: *MHR: Basic science of reproductive medicine* 22.10, pp. 681–690.

Plusa, Berenika, Piliszek, Anna, Frankenberg, Stephen, Artus, Jérôme and Hadjantonakis, Anna-Katerina (2008). "Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst". In: *Development* 135.18, pp. 3081–3091.

Purbey, Prabhat Kumar, Singh, Sunita, Kumar, P Pavan, Mehta, Sameet, Ganesh, KN, Mitra, Debashis and Galande, Sanjeev (2008). "PDZ domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1". In: *Nucleic acids research* 36.7, pp. 2107–2122.

Quinlan, Aaron R and Hall, Ira M (2010). "BEDTools: a flexible suite of utilities for comparing genomic features". In: *Bioinformatics* 26.6, pp. 841–842.

Ramalho-Santos, Miguel and Willenbring, Holger (2007). "On the origin of the term "stem cell"". In: *Cell stem cell* 1.1, pp. 35–38.

Ramírez, Fidel, Ryan, Devon P, Grüning, Björn, Bhardwaj, Vivek, Kilpert, Fabian, Richter, Andreas S, Heyne, Steffen, Dündar, Friederike and Manke, Thomas (2016). "deepTools2: a next generation web server for deep-sequencing data analysis". In: *Nucleic acids research* 44.W1, W160–W165.

Ran, F Ann, Hsu, Patrick D, Wright, Jason, Agarwala, Vineeta, Scott, David A and Zhang, Feng (2013). "Genome engineering using the CRISPR-Cas9 system". In: *Nature protocols* 8.11, p. 2281.

Rao, Tata Purushothama and Kühl, Michael (2010). "An updated overview on Wnt signaling pathways: a prelude for more". In: *Circulation research* 106.12, pp. 1798–1806.

Rappolee, Daniel A, Basilico, Claudio, Patel, Yogesh and Werb, Zena (1994). "Expression and function of FGF-4 in peri-implantation development in mouse embryos". In: *Development* 120.8, pp. 2259–2269.

Reményi, Attila, Tomilin, Alexey, Pohl, Ehmke, Lins, Katharina, Philippsen, Ansgar, Reinbold, Rolland, Schöler, Hans R and Wilmanns, Matthias (2001). “Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping”. In: *Molecular cell* 8.3, pp. 569–580.

Respuela, Patricia, Nikolić, Miloš, Tan, Minjia, Frommolt, Peter, Zhao, Yingming, Wysocka, Joanna and Rada-Iglesias, Alvaro (2016). “Foxd3 promotes exit from naive pluripotency through enhancer decommissioning and inhibits germline specification”. In: *Cell stem cell* 18.1, pp. 118–133.

Robertson, Gordon, Hirst, Martin, Bainbridge, Matthew, Bilenky, Misha, Zhao, Yongjun, Zeng, Thomas, Euskirchen, Ghia, Bernier, Bridget, Varhol, Richard, Delaney, Allen et al. (2007). “Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing”. In: *Nature methods* 4.8, p. 651.

Rodda, David J, Chew, Joon-Lin, Lim, Leng-Hiong, Loh, Yuin-Han, Wang, Bei, Ng, Huck-Hui and Robson, Paul (2005). “Transcriptional regulation of nanog by OCT4 and SOX2”. In: *Journal of Biological Chemistry* 280.26, pp. 24731–24737.

Ross, Sarah and Hill, Caroline S (2008). “How the Smads regulate transcription”. In: *The international journal of biochemistry & cell biology* 40.3, pp. 383–408.

Rossant, Janet and Tam, Patrick PL (2017). “New insights into early human development: lessons for stem cell derivation and differentiation”. In: *Cell Stem Cell* 20.1, pp. 18–28.

Routledge, Daniel and Scholpp, Steffen (2019). “Mechanisms of intercellular Wnt transport”. In: *Development* 146.10, dev176073.

Sabari, Benjamin R, Dall’Agnese, Alessandra, Boija, Ann, Klein, Isaac A, Coffey, Eliot L, Shrinivas, Krishna, Abraham, Brian J, Hannett, Nancy M, Zamudio, Alicia V, Manteiga, John C et al. (2018). “Coactivator condensation at super-enhancers links phase separation and gene control”. In: *Science* 361.6400, eaar3958.

Saitou, Mitinori, Barton, Sheila C and Surani, M Azim (2002). “A molecular programme for the specification of germ cell fate in mice”. In: *Nature* 418.6895, p. 293.

Sato, Masatake, Kimura, Tohru, Kurokawa, Ken, Fujita, Yukiko, Abe, Koichiro, Masuhara, Masaaki, Yasunaga, Teruo, Ryo, Akihide, Yamamoto, Mikio and Nakano, Toru (2002). “Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells”. In: *Mechanisms of development* 113.1, pp. 91–94.

Sato, Noboru, Meijer, Laurent, Skaltsounis, Leandros, Greengard, Paul and Brivanlou, Ali H (2004). “Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor”. In: *Nature medicine* 10.1, p. 55.

Schaefer, Kristina N and Peifer, Mark (2019). “Wnt/Beta-catenin signaling regulation and a role for biomolecular condensates”. In: *Developmental cell* 48.4, pp. 429–444.

Schrode, Nadine, Saiz, Néstor, Di Talia, Stefano and Hadjantonakis, Anna-Katerina (2014). “GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst”. In: *Developmental cell* 29.4, pp. 454–467.

Schröter, Christian, Rué, Pau, Mackenzie, Jonathan Peter and Arias, Alfonso Martinez (2015). “FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells”. In: *Development* 142.24, pp. 4205–4216.

Schulz, Katharine N and Harrison, Melissa M (2018). “Mechanisms regulating zygotic genome activation”. In: *Nature Reviews Genetics*, p. 1.

Schwarz, Benjamin A, Bar-Nur, Ori, Silva, José CR and Hochedlinger, Konrad (2014). “Nanog is dispensable for the generation of induced pluripotent stem cells”. In: *Current Biology* 24.3, pp. 347–350.

Seisenberger, Stefanie, Andrews, Simon, Krueger, Felix, Arand, Julia, Walter, Jörn, Santos, Fátima, Popp, Christian, Thienpont, Bernard, Dean, Wendy and Reik, Wolf (2012). “The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells”. In: *Molecular cell* 48.6, pp. 849–862.

Seki, Yoshiyuki, Yamaji, Masashi, Yabuta, Yukihiro, Sano, Mitsue, Shigeta, Mayo, Matsui, Yasuhisa, Saga, Yumiko, Tachibana, Makoto, Shinkai, Yoichi and Saitou, Mitinori (2007). “Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice”. In: *Development* 134.14, pp. 2627–2638.

Shi, Yigong and Massagué, Joan (2003). “Mechanisms of TGF-β signaling from cell membrane to the nucleus”. In: *cell* 113.6, pp. 685–700.

Shrinivas, Krishna, Sabari, Benjamin R, Coffey, Eliot L, Klein, Isaac A, Boija, Ann, Zamudio, Alicia V, Schuijers, Jurian, Hannett, Nancy M, Sharp, Phillip A, Young, Richard A et al. (2019). “Enhancer features that drive formation of transcriptional condensates”. In: *Molecular Cell* 75.3, pp. 549–561.

Silva, Jose, Nichols, Jennifer, Theunissen, Thorold W, Guo, Ge, Oosten, Anouk L van, Barrandon, Ornella, Wray, Jason, Yamanaka, Shinya, Chambers, Ian and Smith, Austin (2009). “Nanog is the gateway to the pluripotent ground state”. In: *Cell* 138.4, pp. 722–737.

Simon, Claire S, Hadjantonakis, Anna-Katerina and Schröter, Christian (2018). “Making lineage decisions with biological noise: Lessons from the early mouse embryo”. In: *Wiley Interdisciplinary Reviews: Developmental Biology* 7.4, e319.

Smallwood, Sébastien A, Tomizawa, Shin-ichi, Krueger, Felix, Ruf, Nico, Carli, Natasha, Segonds-Pichon, Anne, Sato, Shun, Hata, Kenichiro, Andrews, Simon R and Kelsey, Gavin (2011). “Dynamic CpG island methylation landscape in oocytes and preimplantation embryos”. In: *Nature genetics* 43.8, p. 811.

Smith, Austin (2017). “Formative pluripotency: the executive phase in a developmental continuum”. In: *Development* 144.3, pp. 365–373. ISSN: 0950-1991.

Smith, Austin G, Heath, John K, Donaldson, Deborah D, Wong, Gordon G, Moreau, J, Stahl, Mark and Rogers, David (1988). “Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides”. In: *Nature* 336.6200, p. 688.

Smith, Zachary D, Chan, Michelle M, Mikkelsen, Tarjei S, Gu, Hongcang, Gnirke, Andreas, Regev, Aviv and Meissner, Alexander (2012). “A unique regulatory phase of DNA methylation in the early mammalian embryo”. In: *Nature* 484.7394, p. 339.

Stavridis, Marios P, Lunn, J Simon, Collins, Barry J and Storey, Kate G (2007). “A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification”. In: *Development* 134.16, pp. 2889–2894.

Stock, K Julie, Brookes, Emily and Pombo, Ana (2019). *Phospho-sensitive chromatin immunoprecipitation of RNA Polymerase II*. URL: https://www.epigenesys.eu/images/stories/protocols/pdf/20111121162410_p48.pdf (visited on 26/06/2019).

Sun, Liang Tso, Yamaguchi, Shinpei, Hirano, Kunio, Ichisaka, Tomoko, Kuroda, Takao and Tada, Takashi (2014). “Nanog co-regulated by Nodal/Smad2 and Oct4 is required for pluripotency in developing mouse epiblast”. In: *Developmental biology* 392.2, pp. 182–192.

Tai, Chih-I and Ying, Qi-Long (2013). “Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state”. In: *J Cell Sci* 126.5, pp. 1093–1098.

Tee, Wee-Wei, Shen, Steven S, Oksuz, Ozgur, Narendra, Varun and Reinberg, Danny (2014). “Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs”. In: *Cell* 156.4, pp. 678–690.

Ten Berge, Derk, Kurek, Dorota, Blauwkamp, Tim, Koole, Wouter, Maas, Alex, Eroglu, Elif, Siu, Ronald K and Nusse, Roel (2011). “Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells”. In: *Nature cell biology* 13.9, p. 1070.

Tesar, Paul J, Chenoweth, Josh G, Brook, Frances A, Davies, Timothy J, Evans, Edward P, Mack, David L, Gardner, Richard L and McKay, Ronald DG (2007). “New cell lines from mouse epiblast share defining features with human embryonic stem cells”. In: *Nature* 448.7150, p. 196.

Tomioka, Mizuho, Nishimoto, Masazumi, Miyagi, Satoru, Katayanagi, Tomoko, Fukui, Nobutaka, Niwa, Hitoshi, Muramatsu, Masami and Okuda, Akihiko (2002). “Identification of Sox-2 regulatory region which is under the control of Oct-3/4–Sox-2 complex”. In: *Nucleic acids research* 30.14, pp. 3202–3213.

Toyooka, Yayoi, Shimosato, Daisuke, Murakami, Kazuhiro, Takahashi, Kadue and Niwa, Hitoshi (2008). “Identification and characterization of subpopulations in undifferentiated ES cell culture”. In: *Development* 135.5, pp. 909–918.

Vallier, Ludovic, Reynolds, Daniel and Pedersen, Roger A (2004). “Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway”. In: *Developmental biology* 275.2, pp. 403–421.

Vallier, Ludovic, Mendjan, Sasha, Brown, Stephanie, Chng, Zhenzhi, Teo, Adrian, Smithers, Lucy E, Trotter, Matthew WB, Cho, Candy H-H, Martinez, Amelie, Rugg-Gunn, Peter et al. (2009). “Activin/Nodal signalling maintains pluripotency by controlling Nanog expression”. In: *Development* 136.8, pp. 1339–1349.

Vincent, Stéphane D, Dunn, N Ray, Sciammas, Roger, Shapiro-Shalef, Miriam, Davis, Mark M, Calame, Kathryn, Bikoff, Elizabeth K and Robertson, Elizabeth J (2005). “The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse”. In: *Development* 132.6, pp. 1315–1325.

Weber, Susanne, Eckert, Dawid, Nettersheim, Daniel, Gillis, Ad JM, Schäfer, Sabine, Kuckenberg, Peter, Ehlermann, Julia, Werling, Uwe, Biermann, Katharina, Looijenga, Leendert HJ et al. (2010). “Critical function of AP-2gamma/TCFAP2C in mouse embryonic germ cell maintenance”. In: *Biology of reproduction* 82.1, pp. 214–223.

Whyte, Warren A, Orlando, David A, Hnisz, Denes, Abraham, Brian J, Lin, Charles Y, Kagey, Michael H, Rahl, Peter B, Lee, Tong Ihn and Young, Richard A (2013). “Master transcription factors and mediator establish super-enhancers at key cell identity genes”. In: *Cell* 153.2, pp. 307–319.

Wickham, Hadley (2016). *ggplot2: elegant graphics for data analysis*. Springer.

Wicklow, Eryn, Blij, Stephanie, Frum, Tristan, Hirate, Yoshikazu, Lang, Richard A, Sasaki, Hiroshi and Ralston, Amy (2014). “HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst”. In: *PLoS genetics* 10.10, e1004618.

Wilder, Phillip J, Kelly, David, Brigman, Kristen, Peterson, Cynthia L, Nowling, Tamara, Gao, Qing-Sheng, McComb, Rodney D, Capecchi, Mario R and Rizzino, Angie (1997). “Inactivation of the FGF-4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny”. In: *Developmental biology* 192.2, pp. 614–629.

Williams, R Lindsay, Hilton, Douglas J, Pease, Shirley, Willson, Tracy A, Stewart, Colin L, Gearing, David P, Wagner, Erwin F, Metcalf, Donald, Nicola, Nicos A and Gough, Nicholas M (1988). “Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells”. In: *Nature* 336.6200, p. 684.

Wong, Frederick CK, Chambers, Ian and Mullin, Nicholas P (2016). “Sox2-dependent regulation of pluripotent stem cells”. In: *Sox2*. Elsevier, pp. 163–185.

Wray, Jason, Kalkan, Tüzer, Gomez-Lopez, Sandra, Eckardt, Dominik, Cook, Andrew, Kemler, Rolf and Smith, Austin (2011). “Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency

network and increases embryonic stem cell resistance to differentiation". In: *Nature cell biology* 13.7, p. 838.

Wray, Jason, Kalkan, Tuzer and Smith, Austin G (2010). *The ground state of pluripotency*.

Wu, Jingyi, Huang, Bo, Chen, He, Yin, Qiangzong, Liu, Yang, Xiang, Yunlong, Zhang, Bingjie, Liu, Bofeng, Wang, Qiujun, Xia, Weikun et al. (2016). "The landscape of accessible chromatin in mammalian preimplantation embryos". In: *Nature* 534.7609, p. 652.

Yabuta, Yukihiro, Kurimoto, Kazuki, Ohinata, Yasuhide, Seki, Yoshiyuki and Saitou, Mitinori (2006). "Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling". In: *Biology of reproduction* 75.5, pp. 705–716.

Yamaji, Masashi, Seki, Yoshiyuki, Kurimoto, Kazuki, Yabuta, Yukihiro, Yuasa, Mihoko, Shigeta, Mayo, Yamanaka, Kaori, Ohinata, Yasuhide and Saitou, Mitinori (2008). "Critical function of Prdm14 for the establishment of the germ cell lineage in mice". In: *Nature genetics* 40.8, p. 1016.

Yamaji, Masashi, Ueda, Jun, Hayashi, Katsuhiko, Ohta, Hiroshi, Yabuta, Yukihiro, Kurimoto, Kazuki, Nakato, Ryuichiro, Yamada, Yasuhiro, Shirahige, Katsuhiko and Saitou, Mitinori (2013). "PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells". In: *Cell stem cell* 12.3, pp. 368–382.

Yamanaka, Yojiro, Lanner, Fredrik and Rossant, Janet (2010). "FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst". In: *Development* 137.5, pp. 715–724.

Yates, Adam and Chambers, I (2005). *The homeodomain protein Nanog and pluripotency in mouse embryonic stem cells*.

Ye, Shoudong, Li, Ping, Tong, Chang and Ying, Qi-Long (2013). "Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1". In: *The EMBO journal* 32.19, pp. 2548–2560.

Yeo, Jia-Chi, Jiang, Jianming, Tan, Zi-Ying, Yim, Guo-Rong, Ng, Jia-Hui, Göke, Jonathan, Kraus, Petra, Liang, Hongqing, Gonzales, Kevin Andrew Uy, Chong, Han-Chung et al. (2014). "Klf2 is an essential factor that sustains ground state pluripotency". In: *Cell stem cell* 14.6, pp. 864–872.

Yi, Fei, Pereira, Laura, Hoffman, Jackson A, Shy, Brian R, Yuen, Courtney M, Liu, David R and Merrill, Bradley J (2011). "Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal". In: *Nature cell biology* 13.7, p. 762.

Yi, Fei, Pereira, Laura and Merrill, Bradley James (2008). "Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal". In: *Stem cells* 26.8, pp. 1951–1960.

Ying, Qi-Long, Nichols, Jennifer, Chambers, Ian and Smith, Austin (2003). "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3". In: *Cell* 115.3, pp. 281–292.

Ying, Qi-Long, Wray, Jason, Nichols, Jennifer, Batlle-Morera, Laura, Doble, Bradley, Woodgett, James, Cohen, Philip and Smith, Austin (2008). “The ground state of embryonic stem cell self-renewal”. In: *Nature* 453.7194, p. 519.

Ying, Ying and Zhao, Guang-Quan (2001). “Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse”. In: *Developmental biology* 232.2, pp. 484–492.

Yoshida, Kanji, Chambers, Ian, Nichols, Jennifer, Smith, Austin, Saito, Mikiyoshi, Yasukawa, Kiyoshi, Shoyab, Mohammed, Taga, Tetsuya and Kishimoto, Tadamitsu (1994). “Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways”. In: *Mechanisms of development* 45.2, pp. 163–171.

Young, Richard A (2011). “Control of the embryonic stem cell state”. In: *Cell* 144.6, pp. 940–954.

Yuan, Huabing, Corbi, Nicoletta, Basilico, Claudio and Dailey, Lisa (1995). “Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3.” In: *Genes & development* 9.21, pp. 2635–2645.

Zhang, Jingchao (2016). “Functional analysis of the role of the Nanog Tryptophan Repeat in ES cells”. Doctoral dissertation. The University of Edinburgh.

Zhang, Jingchao, Zhang, Man, Acampora, Dario, Vojtek, Matúš, Yuan, Detian, Simeone, Antonio and Chambers, Ian (2018a). “OTX2 restricts entry to the mouse germline”. In: *Nature* 562.7728, p. 595.

Zhang, Man, Leitch, Harry G, Tang, Walfred WC, Festuccia, Nicola, Hall-Ponsele, Elisa, Nichols, Jennifer, Surani, M Azim, Smith, Austin and Chambers, Ian (2018b). “Esrrb complementation rescues development of Nanog-Null germ cells”. In: *Cell reports* 22.2, pp. 332–339.

Zhang, Yong, Liu, Tao, Meyer, Clifford A, Eeckhoute, Jérôme, Johnson, David S, Bernstein, Bradley E, Nusbaum, Chad, Myers, Richard M, Brown, Myles, Li, Wei et al. (2008). “Model-based analysis of ChIP-Seq (MACS)”. In: *Genome biology* 9.9, R137.

Zhao, Feng-Qi (2013). “Octamer-binding transcription factors: genomics and functions”. In: *Frontiers in bioscience (Landmark edition)* 18, p. 1051.

Zhu, Lihua J, Gazin, Claude, Lawson, Nathan D, Pagès, Hervé, Lin, Simon M, Lapointe, David S and Green, Michael R (2010). “ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data”. In: *BMC bioinformatics* 11.1, p. 237.