



MONASH University

*Dynamics of transcription factor targeting and chromatin states
during reprogramming*

Sara Alaei Shehni (BSc, MSc)

Department of Anatomy and Developmental Biology
Monash University, Australia

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in February 2016

Notice

Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Table of Contents:

Summary	vii
Declaration	viii
List of PhD publications:	x
Acknowledgment:	xii
Abbreviations.....	xiv
CHAPTER 1	1
Introduction	1
1.1. History of pluripotent cells.....	1
1.2. Induced Pluripotent Stem Cells (iPSCs):.....	5
1.2.1. Required transcription factors for reprogramming:	5
1.2.1.1. Oct4:	6
1.2.1.2. Sox2:.....	7
1.2.1.3. Klf4:.....	7
1.2.1.4. C-Myc:.....	8
1.2.2. Sourcing cells for reprogramming:.....	9
1.3. Different methods to generate iPSC:	10
1.4. Applications of iPSCs:	11
1.4.1. Understanding reprogramming and developmental mechanisms.....	11
1.4.2. Cell therapy:	11
1.4.3. Drug screening and disease modelling:	12
1.5. Unveiling the reprogramming process:	15
1.5.1. Defining of intermediates during mouse embryonic fibroblast (MEF) reprogramming	15
1.5.2. Targets of transcription factors during reprogramming:	18
1.5.3. Initiation of the reprogramming process:	19
1.6. Chromatin and epigenetic changes in pluripotent cells:	21
1.6.1. Chromatin structure:	21
1.7. Epigenetic changes during reprogramming:	22
1.7.1. Histones variants and modifications during reprogramming.....	24
1.7.1.1. Histone variants	24
1.7.1.2. Histone modifications:.....	24
1.7.1.3. Nucleosome shifting and chromatin remodelling.....	29
1.7.1.4. Targeting the epigenetic machinery to increase reprogramming efficiency:	30
1.8. Protein complexes during reprogramming:	30

1.8.1. Polycomb group (PcG) proteins:.....	31
1.8.2. The Tip60-p400 complex (known as NuA4 HAT):.....	31
1.8.3. The Nucleosome Remodeling and Deacetylase (NuRD) complex:.....	32
1.9. Conclusion:	33
CHAPTER 2.....	35
Material & Methods	35
2.1. Materials and Suppliers:.....	35
2.1.1. Cell culture reagents:	35
2.1.2. Antibodies:	36
2.1.3. Disposables:	37
2.1.4. Equipment:	37
2.2. Cell Culture:.....	38
2.2.1. Media preparation	38
2.2.2. Mouse genotyping:.....	39
2.2.2.1. Polymerase chain reactions (PCR):.....	39
2.2.2.2. DNA gel electrophoresis:.....	40
2.2.3. Reprogramming of MEFs into induced pluripotent stem cells:	40
2.2.3.1. Generation of mouse embryonic fibroblasts	40
2.3. FACS isolation of reprogramming intermediates	42
2.3.1. Antibody labelling of cells:.....	42
2.3.2. Preparation of compensation tubes:	43
2.3.2. Fluorescent Activated Cell Sorting (FACS):	44
2.4. Magnetic-activated cell sorting (MACS):.....	44
2.4.1. Auto MACS Buffers:	45
2.4.2. Labelling with AutoMACS antibody:.....	45
2.5. Virus production:	45
2.5.1. Concentration of primary viral supernatants.....	46
2.5.2. Virus infection	46
2.6. AP staining.....	46
2.7. Western Blot:	47
2.7.1. Western Blot buffers:	47
2.7.2. Western Blot procedure:	47
2.8. Multiple ligation PCR amplification:.....	48
2.8.1. MLPA analysis:	49
2.9. Assay for Transposase Accessible Chromatin (ATAC):	49
2.9.1. Library QC:.....	50

2.9.2. ATAC analysis:	50
2.10. Chromatin immunoprecipitation (ChIP).....	51
2.10.1. ChIP buffers:	52
2.10.2. Quantitative Real Time PCR:	53
2.10.2.1. Fluidigm ChIP:	53
CHAPTER 3	55
Reprogramming factor levels and their impact on reprogramming kinetics and efficiency.55	
3.1. Introduction	55
3.2. Results:	60
3.2.1. Morphological changes and changes in cell surface marker expression (OKSM mouse model).....	60
3.2.1.1: Cell Morphology	60
3.2.1.2. Changes in cell surface marker expression during reprogramming	62
3.2.1.3. Rescue of refractory cells with addition of more transcription factors:	64
3.2.2. Generation of an improved mouse model with higher levels of OKSM expression	66
3.2.2.1. The amount of Oct4 and Sox2 expression in M2 versus M3 mouse model	68
3.2.2.2. M3rtTA followed similar kinetics as M2rtTA	70
3.2.2.3. Higher reprogramming efficiency in M3rtTA compared to M2rtTA.....	73
3.2.2.4. Assessment of the minimum time requirement for reprogramming cells from both models to become independent of doxycycline:.....	75
3.2.3. Variation in the stoichiometry of reprogramming factors leads to changes in reprogramming	77
3.2.3.1. Cell morphology	80
3.2.3.2. The ratio/stoichiometry of reprogramming factors in the three different mouse models	82
3.2.3.3. Different reprogramming efficiencies in OKSM, OSKM and OSKM/OSKM mouse models	84
3.2.3.4. Different reprogramming kinetics were found for the three mouse models.....	86
3.3. Conclusion:.....	90
CHAPTER 4	91
Mapping of Oct4 and Sox2 during reprogramming.....	91
4.1. Introduction	91
4.2. Chromatin Immunoprecipitation (ChIP) Optimisation:	93
4.2.1. Optimization of sonication cycles:	94
4.2.2. Optimization of the number of cells:.....	96
4.2.3. Optimisation of crosslinking time:	98
4.2.4. Optimisation of the amount of antibody:.....	100
4.2.5. Optimising the type of capture beads:	102

4.2.6. Conclusion of the optimisation of the ChIP protocol:.....	104
4.3. The dynamics of transcription factor targeting during cellular reprogramming	105
4.3.1. Oct4 and Sox2 bind to different targets in different patterns during reprogramming	105
4.3.2. Purification of reprogramming intermediates with AutoMACS.....	108
4.3.3. Roadmap of Oct4 and Sox2 binding during reprogramming in the Ssea1+ population	110
4.4. Conclusion:	129
CHAPTER 5.....	131
Chromatin accessibility during reprogramming.....	131
5.1. Introduction.....	131
5.2. Results.....	134
5.2.1. The chromatin state in DNase I hypersensitive (DHS) region of reprogramming intermediates	134
5.2.2. Detection of DHS site using a modified Multiplex Ligation-dependent Probe Amplification (MLPA)	136
5.2.3 Identification of open region chromatin in refractory and reprogrammed cells during reprogramming.....	145
5.2.3.1. Changes in chromatin accessibility in refractory and reprogrammed cells on a genome wide scale in OSM targets during reprogramming.....	145
5.2.3.2. Is gene activity related to chromatin accessibility?.....	150
5.2.3.3. The binding sites for Oct4 and Sox2 are located in regions of open chromatin in iPSC	156
5.2.3.4. The correlation between Oct4 and Sox2 binding sites and transcriptional activity based on chromatin accessibility.....	158
5.3. Conclusion:	166
CHAPTER 6.....	167
General discussion and future directions.....	167
6.1. Different reprogrammable mouse models.....	167
6.2. The effect of chromatin dynamics in transcription during reprogramming	170
6.3. Future direction.....	177
Appendix: List of primers, probes and target genes.....	178
References:.....	186

Summary

Induced pluripotent stem cells (iPSCs) have been produced from a number of mouse and human somatic cell types, upon the enforced expression of the transcription factors (TFs) Oct4, Klf4, Sox2, and c-Myc (OKSM). Reprogramming leads to inactivation of the somatic cell program and activation of the embryonic stem (ES) cell-specific program for self-renewal and pluripotency. Cells undergoing reprogramming initially downregulate the fibroblast-associated marker Thy1, then activate the Ssea1 marker. However only Ssea1+ cells can form iPSCs and Thy1+ and Thy1- cells are refractory cells. I firstly find why some cells are refractory and how the refractory cells can be rescued. Next an efficient reprogramming mouse model generates. Then both the spatial and temporal binding properties of these transcription factors during reprogramming investigates using Chromatin Immunoprecipitation, followed by qPCR which showed different binding patterns during reprogramming. I also investigates the compaction and protection of chromatin by DNase-MLPA and ATAC- seq across the different cell populations. Finally we studied the correlation among the TF binding, chromatin changes, nucleosomes shifting and transcriptional activity. The results of this study will further our understanding of transcriptional regulation by OKSM and how gene regulatory networks transpire during cellular reprogramming.

Declaration

In accordance with Monash University Doctorate Regulation 17.2 Doctor of Philosophy and Research Master's regulations the following declarations are made:

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

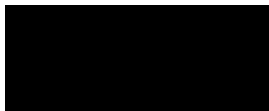
This thesis includes 1 review paper, 1 method paper, 1 book chapter and 1 peer review paper. The original paper published in a peer reviewed journal and 0 unpublished publications. The core theme of the thesis is mapping of transcription factors and chromatin state during reprogramming. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Department of Anatomy and Developmental Biology and the Australian Regenerative Medicine Institute under the supervision of Associate Professor Jose Polo.

In the case of chapter 1, 2 and 3 my contribution to the work involved the following:

Thesis chapter	Publication title	Publication status	Nature and extent of candidate's contribution
Chapter 1	"Epigenetic Changes During Reprogramming"	Published	First author

Chapter 2	“Cell surface marker mediated purification of iPS cell intermediates from a reprogrammable mouse model”	Published	Co-first author
	“Isolation of reprogramming intermediates during generation of induced pluripotent stem cells from mouse embryonic fibroblasts”	Published	Middle author
Chapter 3	“A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells”	Published	Middle author
	“An improved reprogrammable mouse model harbouring the reverse tetracycline-controlled transcriptional transactivator 3”	In press	First author

Signed:



Date: 10/2/2016

List of PhD publications:

Sara Alaei et al. Dynamic of transcription factor targeting and chromatin state during reprogramming. **Manuscript in preparation.**

Sara Alaei et al. An improved reprogrammable mouse model harbouring the reverse tetracycline-controlled transcriptional transactivator 3. **Stem Cell Research. In press.**

Christian M. Nefzger, **Sara Alaei**, Jose M. Polo. Isolation of reprogramming intermediates during generation of induced pluripotent stem cells from mouse embryonic fibroblasts, **Methods Mol Biol**, (2015); 1330:205-18

Phong Dang Nguyen*, Georgina Elizabeth Hollway*, Carmen Sonntag, Lee Miles, Tom Edward Hall, Silke Berger, Kristine Joy Fernandez, David Baruch Gurevich, Nicholas James Cole, **Sara Alaei**, Mirana Ramialison, Robert Lyndsay Sutherland, Jose Maria Polo, Graham John Lieschke and Peter David Currie. Induction of haematopoietic stem cells by somite-derived endothelial cells is controlled by *meox1*. **Nature**, (2014). Aug 21; 512(7514):314-8. *equal contribution.

Sara Alaei, Anja Knaupp, Jose Polo, Epigenetic changes during reprogramming, Mini-review, **ASBMB**, (2014). 45(1), 17-19.

Sara Alaei*, Christian M. Nefzger*, Anja Knaupp, Melissa Holmes, Jose Polo. Cell Surface Markers Mediated Purification Of IPS Cell Intermediates From A Reprogramming Mouse Model. **Journal of Visualized Experiments (JoVE)**, (2014), Sep 6; (91). *equal contribution.

Jose M. Polo*, Endre Anderssen*, Ryan M. Walsh, Benjamin A. Schwarz, Christian M. Nefzger, Sue Mei Lim, Marti Borkent, Effie Apostolou, **Sara Alaei**, Jennifer Cloutier, Ori Bar-Nur, Sihem Cheloufi, Matthias Stadtfeld, Maria Eugenia Figueroa, Daisy Robinton, Sridaran Natesan, Ari Melnick, Jinfang Zhu, Sridhar Ramaswamy, and Konrad Hochedlinger . A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell*, **(2012)**. 151, 1617-1632. *equal contribution.

Acknowledgments:

The PhD has been an amazing experience which was quite demanding but also filled with days of happiness: happiness because of the hilarious, friendly and educational environment in our lab. First I would like to express my thanks to my main supervisor A/Prof Jose Polo for his helpful mentorship during my PhD. Jose, working in your lab was a great experience that helped me to further my knowledge of epigenetics and made me a better researcher. I am glad that I passed these four years of my PhD journey with you and under your supervision. I will never forget all of your support. I would like to thank Dr Christian Nefzger, Dr Anja Knaupp, Dr Sue Mei Lim, Dr Melissa Holmes, Dr Stefan White and Dr Fernando Rossello in our lab. Anja and Christian, my lovely German friends, I will never forget those days that you both tried hard to make me happier and shared with me your great experience in science and life. Sue Mei, my lovely friend, I will never forget your support when I started in the lab. We were three in the lab, you, me and Christian. If our lab is friendly now, it is because of you and Christian who pioneered this environment. Stefan, you are a really nice supervisor, and I really appreciate you for always motivating me regardless of the results. With one positive sentence that you told me, I could work several months nonstop. I would like to thank Dr Fernando Rossello for helping me in bioinformatics analysis and also my present and future PhD colleagues Jaber Firas, Ethen Liu, Suzan de Boer and Joseph Chen for their support. I would also like to thank Dr Trevor Wilson and Dr Vivien Vasic in MIMR and sequencing facility and also Dr Shalima Nair and Prof. Suzan Clark from the Garvan institute. I would like to thank Dr Kathy Davidson and Mathew Tiedemann and Nathalia Yen Tan for helping me with proofreading.

I would like to express my thanks to all my friends in Iran who always support me emotionally despite the distance between us. Thanks to Sina, Parastoo, Shaghayegh, Azam and Fatemeh Zahra. Thank to another group of friends in Melbourne; Roshanak, Vahid,

Hadi, Salman, Saeedeh, Farnaz, Shaghayegh, Kiana, Tahmineh and Saeed for their friendship and their support. Also thanks to my engineer friends at Monash University: Mahsa, Morteza, Negin, Fatemeh, Hossein, Esmaeil, Sajjad – thank you for being a part of my PhD life.

Arezo, you are an amazing friend I am so lucky that I have you in my life, I will never forget your caring attitude whenever I need support emotionally. Fahimeh, you are a gift for me and I am so thankful for having you in my life. Mitra, you are fantastic friend and I am so glad that I work in the same building with you.

Special thanks to my wonderful family who always support me. Many thanks to my lovely parents. I cannot express my feeling of how much I love you. Mum, you are my best friend and my best teacher not only in high school but also throughout my whole life. Dad, you are the hero in my life and thanks for all your support. I am so proud that I have such a thoughtful and supportive, understanding and lovely family. I would like to dedicate this thesis to my lovely parents. Thanks to my lovely sisters and brothers: Sahar, Mozhdeh, Sajjad and Milad I love you so much. God gave me such an amazing present, having you beside me.

I thank God who gave me the box full of gifts like you. I love you all.

Abbreviations

AP	Alkaline Phosphatase
ALS	Amyotrophic lateral sclerosis
BMP	Bone Morphogenetic Protein
ChIP	Chromatin immunoprecipitation
DBD	DNA binding domain
DHS	DNaseI Hypersensitive Site
Dox	Doxycycline
ESCs	Embryonic Stem Cells
EG	Embryonic germ cell
EMT	Epithelial to Mesenchymal Transition
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9me3	Histone 3 lysine 9 trimethylation
H3K4me1	Histone 3 lysine 4 monomethylation
H3K27me3	Histone 3 lysine 27 trimethylation
HDACs	Histone deacetylases
hESC	Human Embryonic Stem Cell
HMG	High Mobility Group
ICM	Inner Cell Mass
iPSCs	Induced Pluripotent Stem Cells
MEFs	Mouse Embryonic Fibroblasts
MET	Mesenchymal to Epithelial Transition
OKSM	Oct4, Klf4, Sox2, c-Myc
PouHD	Pou Homeodomain
PouS	Pou Specific domain

PSC	Pluripotent stem cell
rtTA	Reverse tetracycline-dependent Transactivator
TetOP	Tetracycline inducible promoter
Tet	Tetracycline
TFs	Transcription Factors

CHAPTER 1

Introduction

1.1. History of pluripotent cells

The generation of mammalian cells commences with a unicellular totipotent zygote, which gives rise to the blastocyst. Within the blastocyst, cells from the inner cell mass (ICM) contribute to the development of the whole organism. Embryonic stem cells (ESCs) are obtained by *in vitro* culture from the ICM cells, and are able to generate cells of all three embryonic germ layers, confirming their pluripotent state (Evans and Kaufman 1981). They are stem cells with unique properties of pluripotency and self-renewal ability. These cells provide a novel source of cells to replace those that are lost or impaired as a result of disease in the future and hence harbour great therapeutic potential (Cowan, Klimanskaya et al. 2004).

Pluripotency and self-renewal are two important properties of ESCs that make them useful in understanding mechanisms of various diseases, to produce drugs and potentially treat patients suffering from conditions such as diabetes, Parkinson's disease and spinal cord injuries. However, widespread use of these cells has been hampered due to particular methods of derivation of human ESC lines. For example, some extraction methods cause destruction of an embryo and causing ethical controversies while also potentially leading to immune compatibility issues in patients (Takahashi and Yamanaka 2006). Generation of pluripotent stem cell-like cells has been achieved experimentally using different techniques such as cell-cell fusion, nuclear transfer and transcription factor mediated reprogramming (Hochedlinger and Jaenisch 2002, Takahashi and Yamanaka 2006, Tsubouchi, Soza-Ried

et al. 2013). These approaches may potentially resolve the controversial issues. Pluripotent stem cells (PSCs) can be generated from unipotent cells or multipotent cells by overcoming epigenetic barriers (Matsui, Zsebo et al. 1992, Hochedlinger and Jaenisch 2002, Eggan, Baldwin et al. 2004). Reprogramming of somatic cells into iPS cells had previously been established based on seminal work outlining three events. The first event was the demonstration that the genetic information is not lost during development which was shown by nuclear transfer. This led to the generation of tadpoles from unfertilised eggs by nuclear transfer from intestinal cells of adult frogs (Gurdon 1962) and after 40 years of research ends with the birth of the first cloned mammal, Dolly, via cloning of epithelial cells (Wilmut, Schnieke et al. 1999). The second event was the discovery of transcription factors as primary determinants of cell fate. For example, formation of legs instead of antenna by manipulating a transcription factor in *Drosophila*, and the use of MyoD (a transcription factor in myocytes) to convert fibroblasts into myocytes (Schneuwly, Klemenz et al. 1987). The third important stream was the discovery of the required factors in ESCs to maintain pluripotency (Tada, Takahama et al. 2001). Therefore, generation of mouse ESC culture conditions and long-term maintenance of mouse ESCs were crucial for iPSCs establishment. The combination of these three research streams and identification of transcription factors led to the generation of iPSCs (Takahashi and Yamanaka 2006).

Briefly, there are three different strategies to generate pluripotent stem cells (PSC). The first strategy is using nuclear transfer in which the somatic nuclei can regain pluripotency upon injection into oocytes (Dominko, Mitalipova et al. 1999, Gurdon and Melton 2008) (Figure 1A). The second strategy is cell-cell fusion between embryonic cells and somatic cells (Miller and Ruddle 1976, Tada, Takahama et al. 2001) (Figure 1B) and the third strategy is the generation of induced pluripotent stem cells (iPSCs) by transcription factor mediated reprogramming. In 2006, Takahashi and Yamanaka generated iPSCs from

mouse embryonic fibroblasts and adult fibroblasts by retrovirus-mediated transfection of four transcription factors; Oct4, Klf4, Sox2 and c-Myc (OKSM) (Gurdon 1960, Takahashi and Yamanaka 2006) (Figure 1C).

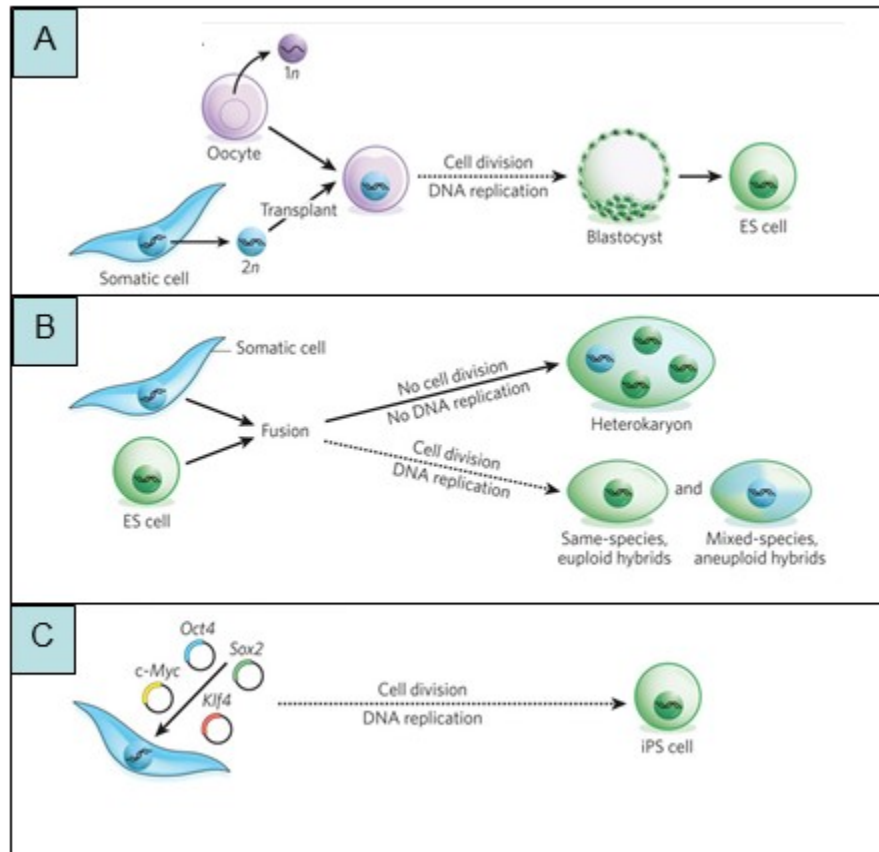


Figure 1. Three different approaches to pluripotent stem cell generation. A) Nuclear transfer **B)** Cell fusion **C)** Transcription-factor mediated reprogramming. (Modified from Yamanaka et al (Yamanaka and Blau 2010)).

1.2. Induced Pluripotent Stem Cells (iPSCs):

1.2.1. Required transcription factors for reprogramming:

Transcription factors (TFs) bind to specific DNA sequences and either activate or suppress transcription. TFs control gene expression by recruiting different proteins which modify chromatin that facilitates or abrogates the TF binding by RNase Pol II (Young 2011). In order to reprogram somatic cells, Takahashi and Yamanaka initially used a cohort of 24 TFs, which included the core pluripotency transcription factors Oct4, Sox2, Klf4, c-Myc, known to interact with coactivators that bind to RNA polymerase II. However, c-Myc does not play a crucial role in RNA polymerase II recruitment, but binds to sequences at core promoter sites (Takahashi and Yamanaka 2006, Young 2011).

With expression of other pluripotency transcription factors such as Sall4, Rex1, Dax1 and Tcl1, the efficiency of reprogramming and the number of iPSC colonies generated increases (Maherali and Hochedlinger 2008, Orkin and Hochedlinger 2011). Depending on the relative expression levels of endogenous factors in each cell type, the number of transcription factors required for reprogramming differs. For example, the transcription factor required for neural stem cell reprogramming is only Oct4 because of endogenous expression of Sox2, Klf4 and c-Myc, whereas the factors needed for melanocyte reprogramming are Oct4, Klf4, c-Myc due to expression of endogenous Sox2 (Maherali, Sridharan et al. 2007, Maherali and Hochedlinger 2008, Hochedlinger and Plath 2009).

Oct4, Sox2 and Nanog bind to target gene promoters which play a dual role, including the repression of the genes which are necessary for differentiation, and also the activation of genes which are necessary for maintenance of pluripotency (Plath and Lowry 2011).

1.2.1.1. Oct4:

Oct4 (POU domain, class 5, homeobox 1 (Pou5f1)) is an essential mediator of ESCs in order to maintain their self-renewal and pluripotent properties (Nichols, Zevnik et al. 1998). It plays an important role during embryonic development, with its expression restricted to Embryonic Stem (ES) and Embryonic germ (EG) cells. It positively regulates expression of genes that are involved in the activation of pluripotency and represses genes involved in differentiation (Nichols, Zevnik et al. 1998). Oct4 dosage is important: reduced expression of Oct4 leads to trophoectoderm development, whereas enhanced expression is associated with primitive endoderm differentiation (Kim, Chu et al. 2008). Furthermore, Oct4 has been shown to promote tumour growth in a dose-dependent manner (Gidekel, Pizov et al. 2003, Campbell, Perez-Iratxeta et al. 2007). Pou proteins have a bipartite DNA-binding domain with two subdomains, a Pou-specific domain (PouS), and a homeodomain (PouHD), connected by a variable linker region. Binding of the N-terminal tail of PouHD is specific to the minor groove of DNA, while the major groove of DNA harbours the binding site for the globular regions of PouS and PouHD. The minor groove of DNA also binds to the high mobility group (HMG) domain of the Sox2 protein (Werner, Huth et al. 1995). Oct4 and Sox2 bind to a canonical motif which is a composite motif formed by combination of individual Oct4 and Sox2 binding sites. Oct4 recognizes and binds to an octamer with the consensus sequence ATGC(A/T)AAT, whereas Sox2 binds the consensus sequence C(T/A)TTGTT. Regulatory regions of key pluripotency genes such as *Pou5f1*, *Nanog* and *Utf1* etc. include a canonical motif which Oct4 and Sox2 are able to bind separately (Aksoy, Jauch et al. 2013); however, transcription of some of the genes is associated with binding of both factors onto the canonical motif. The PouS helix a1 and HMG helix a3 involve in an interaction between Oct4 and Sox2, which leads to fibroblast reprogramming.

1.2.1.2. Sox2:

Sox2 (SRY- related HMG-box family member 2) is another transcription factor that is required for self-renewal ability of undifferentiated PSCs (González, Boué et al. 2011).

Sox2 also plays an important role in neural induction and maintenance of neural progenitors. Sox2 interacts with Oct4 and together they bind to DNA motifs located in the enhancer or promoter regions of genes to regulate activation or repression of these genes (van den Berg, Snoek et al. 2010). In ES cells, they activate pluripotency-associated genes while repressing differentiation -associated genes. Many studies showed the same target genes enriched by Oct4 and Sox2 (Loh, Wu et al. 2006, Kim, Chu et al. 2008).

Sox2 is one of the four factors required for reprogramming; however, the high level of endogenous Sox2 expression in some cells, such as neural stem cells, allows for omission of the Sox2 transgene when reprogramming neural stem cells (Eminli, Utikal et al. 2008).

1.2.1.3. Klf4:

Klf4 is a member of the Kruppel-like family of zinc finger transcription factors which are involved in cell proliferation, differentiation and survival by acting as transcriptional activator and repressor. Loss-of-function experiments with RNA interference (siRNA) demonstrate that Klf4 regulates expression of pluripotency associated genes along with two other family members Klf2 and Klf5. Klf4 interacts with core ES cell maintenance transcription factors to activate *Lefty 1* together with Oct4 and Sox2. Three C-terminal Klf4 zinc fingers facilitate interaction with both factors and are necessary to activate transcription of the target gene *Nanog*.

This binding is essential for the collective function of the ES cell maintenance network. It is able to activate and repress the transcription of different genes depending on the cellular state (Vangapandu and Ai 2009).

Oct4 and Sox2 each bind non-competitively to the same Klf4 binding site. This result was confirmed by an experiment which showed that increasing the amount of Sox2 did not change the interaction between Klf4 and Oct4 (Wei, Yang et al. 2009). A complex including Oct4, Sox2 and Klf4 activates downstream targets which are needed for reprogramming. Interestingly, gain-of-function experiments showed that Klf4 alone plays a role in activating the *Nanog* promoter in 293T cells, whereas Oct4 or Sox2 alone are unable to activate *Nanog* promoter. These results suggest that in the reprogramming complex, Klf4 acts as transactivator (Wei, Yang et al. 2009).

It is known that low expression of Klf4 generates “paused” iPSCs, which remain at a distinct intermediate stages during reprogramming, whereas a lowered expression of the other three factors, Oct4, Sox2 and c-Myc do not generate iPSCs. At the early stage of iPSC generation, Klf4 together with c-Myc repress expression somatic cell-specific lineage genes; however, at the late stage of reprogramming, Klf4 cooperates with Oct4 and Sox2 to enhance pluripotency gene expression (Polo, Anderssen et al. 2012). Therefore, the low level of Klf4 expression may be enough for somatic cell associated gene repression, but is not sufficient for pluripotency-associated gene activation. Furthermore, by recruiting chromatin cohesin, which can control transcription, Klf4 is able to influence chromosomal interactions with the Oct4 locus (Wei, Gao et al. 2013, Nishimura, Kato et al. 2014). These chromosomal interactions are only observed in reprogramming population including Ssea1⁺ cells during reprogramming and not in refractory population (Ssea1⁻ cells), suggesting that Klf4 may be playing a role during the transition from Ssea1⁻ cells to Ssea1⁺ cells by mediating chromosomal interaction (Nishimura, Kato et al. 2014).

1.2.1.4. C-Myc:

C-Myc is a transcription factor involved in cell cycle progression and cell proliferation through regulation of miRNA (Thompson, Challoner et al. 1985, Lin, Jackson et al. 2009).

It plays a critical role in oncogenesis and tumor suppression (González, Boué et al. 2011). C-Myc is required in the early stages of reprogramming. In addition, it is required for silencing of genes that promote differentiation. C-Myc expression increases histone acetylation and methylation (H3K4me3), which is often associated with active transcription and open chromatin. It also controls proliferation through regulation of miRNA and recruits multiple complexes in chromatin modification and structure, such as histone acetyltransferases (GCN5, p300), chromatin-remodelling complexes, histone deacetylases (HDACs) and histone demethylases (Lin, Jackson et al. 2009, Orkin and Hochedlinger 2011). Studies show that differences in c-Myc levels affect the efficiency and kinetics of reprogramming (Knoepfler 2008, Singh and Dalton 2009). Cells with high endogenous levels of c-Myc expression do not require exogenous c-Myc for reprogramming. Because c-Myc is able to maintain cells in a proliferative state, these cells also respond more efficiently to the other reprogramming factors (Kim, Chu et al. 2008, Singh and Dalton 2009). For example, the reprogramming efficiency in keratinocytes is higher than in fibroblasts due to the higher levels of c-Myc expression in keratinocytes. These differences in the level of c-Myc expression in various cell types show exogenous c-Myc plays an essential role in reprogramming efficiency.

1.2.2. Sourcing cells for reprogramming:

Multiple cell types such as hematopoietic, muscle and adipose cells can generate iPSC; however, timing and efficiency of reprogramming varies significantly depending on the cell type. For example, reprogramming of mouse embryonic fibroblasts (MEFs) takes 8-12 days, compared with granulocytes that take 20-25 days. In general, the reprogramming process is more efficient and faster in less differentiated cells or progenitor cells compared with terminally differentiated cells. For example, haematopoietic progenitors generate iPS cells up to 300 times more efficiently than differentiated B and T cells (Eminli, Foudi et al.

2009). In addition, keratinocytes reprogram more efficiently than fibroblasts (González, Boué et al. 2011). Probable reasons for this are the lack of a mesenchymal to epithelial transition (MET) stage, a difference in cell cycle status, higher endogenous expression of c-Myc, and endogenous expression of Klf4 in keratinocytes versus fibroblasts (Maherali, Ahfeldt et al. 2008).

1.3. Different methods to generate iPSC:

Different methods of iPSC generation rely on different approaches to deliver transgenes into target cells. Originally OKSM were delivered using Moloney Murine Leukemia Virus (MMLV)-derived retroviruses (Takahashi and Yamanaka 2006, Wernig, Meissner et al. 2007). Silencing of transgenes plays a critical role in this delivery system, as only the cells that express endogenous pluripotency genes are able to form iPSCs, and incomplete silencing leads to partially reprogrammed cells. The efficiency of this delivery system ranges from 0.1% and 0.01% for mouse and human fibroblasts respectively. In addition to a low reprogramming efficiency, the risk of insertional mutagenesis is another drawback of this retroviral delivery system. Lentiviral vectors have been used with more efficiency compared to MMLV-derived retroviruses (Sommer, Stadtfeld et al. 2009). Moreover, Tetracycline (Tet)-inducible reprogramming lentiviruses provide a controllable system for the expression of transcription factors that are widely used (Yao, Sukonnik et al. 2004, Hockemeyer, Soldner et al. 2008, González, Boué et al. 2011). This system has several limitations associated with it such as viral transgene silencing, heterogeneous expression, random locations of viral integrations, and screening of mice offspring with multiple viral transgenes (Wernig, Lengner et al. 2008, Stadtfeld, Maherali et al. 2009).

Following injection of clonal primary iPSCs into a mouse blastocyst, doxycycline (dox), which is tetracycline analogue, inducible iPSC chimeras can produce secondary MEFs

(Wernig, Lengner et al. 2008, Stadtfeld, Maherali et al. 2009). A major breakthrough in the reprogramming field was the introduction of a secondary system that relied on a dox-inducible polycystronic cassette harbouring the four factors OKSM in the 3' untranslated region of the Collagen locus (*Col1 α 1*) (Stadtfeld, Maherali et al. 2009) and a reverse tetracycline-dependent transactivator (M2-rtTA) targeted to Rosa26. Upon the addition of dox, the rtTA is transcribed and binds to Tetracycline inducible promoter (TetOP) followed by the transcription of the four factors OKSM.

The efficiency of iPSC colony formation in this system is increased ten-fold in comparison to the lentiviral systems as this system allows temporal control of OKSM expression in mouse models, producing isogenic cells that are mostly all conducive to reprogramming (Stadtfeld, Maherali et al. 2009).

1.4. Applications of iPSCs:

1.4.1. Understanding reprogramming and developmental mechanisms

Research into the molecular mechanisms of pluripotency factors and epigenetic modification during development has been assisted by iPSC generation. With the observation of higher reprogramming efficiency in less differentiated cells compared to more differentiated cells, we postulated that the change in epigenetic memory required in progenitor cells is much less than in terminally differentiated cells. The identification of specific phenotypes and subsequent purification of subpopulations of reprogramming “intermediates” allowed researchers to study the role of transcription factors and chromatin modifications during the reprogramming process (Stadtfeld and Hochedlinger 2010).

1.4.2. Cell therapy:

The patient-specific iPS-derived tissues can be used without the need of immunosuppressive drugs because the source of the cells is genetically identical to their

host (Kimbrel and Lanza 2015). iPSCs can be generated from various types of the cells, depending on the patient's needs. The Riken institute in Japan is focused on the treatment of Age-related Macular Degeneration (AMD) with iPSC-derived retinal pigment epithelium (RPE) and the first patient was treated in September 2014 (Reardon and Cyranoski 2014). However because the iPSCs derived from another patient carried genetic mutation, they paused research. Moreover a recent study showed human embryonic stem cell (hESC)-derived mature cardiomyocytes could regenerate contractile cardiac muscle in non-human primate hearts (Chong, Yang et al. 2014). In addition to using iPSC derived cardiac cells in cell therapy, they can be used to understand the pathogenesis of disease.

Consequently, iPSCs have the potential to become a platform for personalised medicine by allowing a patient's own cells to become a source of therapeutic tissue (Amabile and Meissner 2009).

1.4.3. Drug screening and disease modelling:

Cell and animal models are used to understand the complexity of different disease processes and to investigate drug effects on patients. iPSCs can be used in the early stages of the drug discovery process to help identify new drugs and to evaluate their safety and efficacy in these specialized cell types. These preclinical results could then be used to advance compounds through the next stages of drug development (Ebert and Svendsen 2010).

Several disease models have been established by reprogramming of somatic cells from patients with various diseases such as spinal muscular atrophy and diabetes followed by differentiation of these cells to disease related cell types (Colman and Dreesen 2009, Zhang, Lian et al. 2011, Mekhoubad, Bock et al. 2012). iPSCs can be used to derive cell types that display a disease phenotype. For example, a diabetic cardiomyopathy model has been generated using human induced pluripotent cells (hESC) (Drawnel, Boccardo et al. 2014).

Smooth muscle atrophy (SMA) and familial dysautonomia (FD) were the first neurodegenerative diseases modelled using human iPSCs. The SMA model was generated using iPSCs from a child with a mutation in SMA type 1 (Ebert, Yu et al. 2009). Amyotrophic Lateral Sclerosis (ALS) is another neurodegenerative disease that is characterized by muscular atrophy and degeneration of motor neurons. Clinical trials were unsuccessful in ALS mouse models, however ALS models based on iPSCs generated from two different familial forms have been successful in that motor neurons were generated from the patients (Dimos, Rodolfa et al. 2008). In this approach researchers can study early pathophysiology and also provide cells for drug discovery (Ebert and Svendsen 2010, Zhang, Lian et al. 2011) (Figure 1).

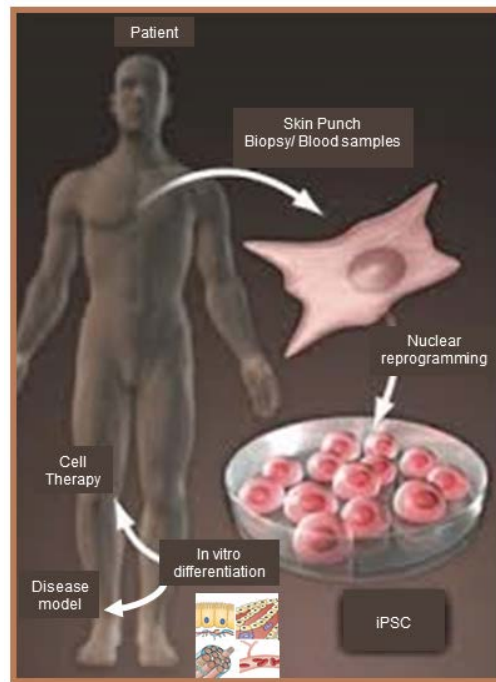


Figure 2. Promise of iPSCs. IPSC applications include cell therapy and disease modelling.

(Modified from <https://www.ucsf.edu/news/2013/04/104816/will-cell-therapy-become-%E2%80%9Cthird-pillar%E2%80%9D-medicine>)

1.5. Unveiling the reprogramming process:

1.5.1. Defining of intermediates during mouse embryonic fibroblast (MEF) reprogramming

Reprogramming takes place with a transition from MEFs to iPSCs, which is accompanied by molecular changes in each step. Factor-induced reprogramming is a gradual process with the formation of defined intermediate cell populations (Stadtfeld, Maherali et al. 2008). In 2012, we showed that this process occurred in two stages based on key molecular changes (Polo, Anderssen et al. 2012). A small number of cells are able to progress from one stage to another, with just 5% of the cells forming iPSCs (Plath and Lowry 2011). During fibroblast reprogramming, cells have to pass through a MET stage. This is accompanied by a loss of mesenchymal properties and the gain of some epithelial features. Bone Morphogenetic Protein (BMP) signalling has major effects on the induction of MET from MEF during reprogramming (Li, Liang et al. 2010, Samavarchi-Tehrani, Golipour et al. 2010). MET happens in the early stages of reprogramming until day 5 (Samavarchi-Tehrani, Golipour et al. 2010). These events are followed by upregulation of alkaline phosphatase (AP) and *Ssea1* before the expression of the pluripotency genes *Nanog* or *Oct4*. The requirement for exogenous factors during reprogramming is indispensable during the first 10 days (Maherali, Sridharan et al. 2007, Stadtfeld, Maherali et al. 2008). The late stage of reprogramming is accompanied by reactivation of endogenous *Oct4* and *Sox2* (Stadtfeld, Maherali et al. 2008).

Intermediates during reprogramming have been defined based on surface markers such as *Thy1* and *Ssea1* (Stadtfeld, Maherali et al. 2008, Polo, Anderssen et al. 2012). *Thy1* (CD90) is a GPI-anchored glycoprotein of the immunoglobulin superfamily, which is a known fibroblast marker. *Ssea1* is a glycosphingolipid which is expressed by blastomeres of late-

cleavage and morula-stage embryos. Ssea1 is a prominent marker of embryonic stem cells and its expression is downregulated upon differentiation of mouse embryonic cells (Draper and Andrews 2006, Wright and Andrews 2009). On the first day of reprogramming, MEFs (in the absence of doxycycline) maintain a Thy1⁺ and Ssea1⁻ phenotype. Three days after dox addition, downregulation of the fibroblast associated marker Thy1 occurs in one third of the cells, while no Ssea1 expression is detected (Stadtfeld, Maherali et al. 2008). The percentage of Ssea1⁺ cells marginally increases to 5 and 10% on day 5 and 7 respectively and by day 9, upregulation of an Oct4-GFP reporter is observed, which indicates the cells express Oct4 from the endogenous locus and have thus completed the reprogramming process. Overall, there is a transition of cells from Thy1⁺ to Thy1⁻ and Thy1⁻ to Ssea1⁺ cells. However, only the Ssea1⁺ population has the potential to give rise to iPSC with activation of Oct4-GFP at the late days of reprogramming (Figure 3) (Polo, Anderssen et al. 2012). The population of cells that fail to downregulate Thy1 remain Thy1⁺ throughout the reprogramming process and are known to be refractory to reprogramming.

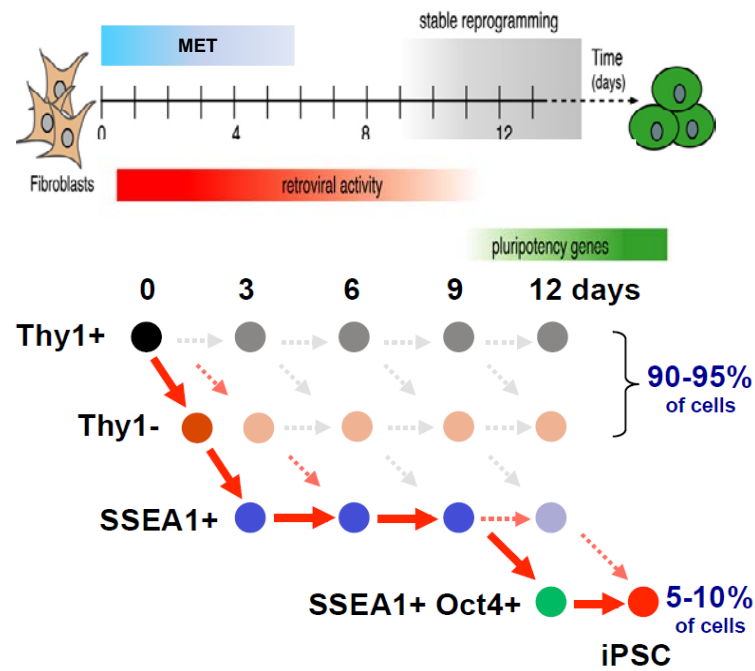


Figure 3. Transition of MEF to iPSC with different intermediates and different stages. Ssea1+ populations can form iPSCs with expression of Oct4 at the later stages of reprogramming and pass through the mesenchymal to epithelial transition stage (MET). (Modified from (Stadtfield, Maherali et al. 2008, Polo, Anderssen et al. 2012)).

1.5.2. Targets of transcription factors during reprogramming:

Identification of transcriptional targets of Oct4 in ESCs has shown a partnership between Oct4, Sox2, and Nanog, and this corporation forms the frame of the core transcriptional network which maintain pluripotency through coordination of a series of feedback and feed-forward cycles (Boyer, Lee et al. 2005). In ESCs, Oct4 often binds in partnership with Sox2, which binds to a neighbouring Sox element (Loh, Wu et al. 2006). The presence of several transcription factors binding sites and active histone marks in the interactome of Oct4 suggests the role of chromatin looping and cooperative regulation of transcription factors in maintenance of the pluripotent state (Jerabek, Merino et al. 2014). Further studies demonstrate the functional role of chromatin loops in gene activation in pluripotent cells. Moreover, Klf4 plays an important role in chromatin rearrangements around the Pou5f1 locus (Jerabek, Merino et al. 2014).

Studies on the DNA binding profile of Klf4 by genome-wide chromatin immunoprecipitation and microarray analysis (ChIP-ChIP) showed overlap of Oct4, Sox2 and Klf4 on promoters of pluripotent associated genes, implicating transcriptional synergy among these three factors (Wei, Yang et al. 2009).

C-Myc, another reprogramming factor, is not required for reprogramming, but its expression with Oct4, Sox2 and Klf4 increases the efficiency and accelerated the kinetics of the early steps of reprogramming (Soufi, Donahue et al. 2012). C-Myc is able to bind to 3000 promoters, including some pluripotency factors (Kim, Chu et al. 2008); however, c-Myc also engages distinct transcriptional networks from the core pluripotency networks. During reprogramming, c-Myc targets are activated prior to expression of target genes of the core factors (Oct4, Sox2, and Klf4). Targets of the c-Myc network are activated in early phase of reprogramming, and it possibly acts through facilitating chromatin accessibility. The c-Myc module is highly represented in ES-associated signatures that have been widely

used in assessing the relatedness of cancer and embryonic cells (Orkin and Hochedlinger 2011). Indeed, c-Myc targets reveal upregulation during the first days of reprogramming without change until the end of the reprogramming process (Polo, Anderssen et al. 2012). Overall, c-Myc, alone or in combination with other factors, plays an important factor in inducing early gene expression (Kim, Chu et al. 2008, Sridharan, Tchieu et al. 2009).

Four or more transcription factors are able to bind to approximately 800 gene promoters. The targets of Oct4, Sox2, Klf4, Nanog, Dax1 are similar, whereas the target of c-Myc is completely different (Kim, Chu et al. 2008, Soufi, Donahue et al. 2012).

The comparison between binding of OKSM in iPSC and ES cells showed that in the early reprogramming steps, four factors can find their targets and the genes co-targeted by four factors was approximately 10 times greater compared with co-targeted genes in ES cells. In 48 hours, co-binding site of three factors (O, K, S) to c-Myc were more than those in ES cells. The majority of the genes change expression in early stage of reprogramming (Soufi, Donahue et al. 2012). The silent genes associated with differentiation are clearly targeted by OKSM at early stage of reprogramming whereas this change is not observed in ES cells. It is still unknown where and when reprogramming factors find their targets at different time points.

1.5.3. Initiation of the reprogramming process:

Reprogramming is regulated through the binding of OKS to their targets at distal enhancer and promoter regions (Soufi, Donahue et al. 2012). But how these transcription factors control reprogramming remains largely unknown, though presumably it involves collaboration of many factors, including various cofactors which bind to OKSM as well as chromatin structure (ie. DNA methylation and histone variation). The binding of OKS during reprogramming is a crucial event for later steps to promote reprogramming by

modification in chromatin accessibility and consequently transcriptional activity of pluripotent genes.

A low level of transgenic Oct4 is able to generate iPS cells with aberrant methylation as well as defects in functional assays of pluripotency, such as tumourigenicity in xenografts and low levels of tetraploid complementation. These defects could be rescued by increasing the amount of exogenous Oct4 protein (Radziskeuskaya and Silva 2014) and we showed that refractory cells with low level of exogenous Oct4 can be rescued by increasing OKSM copy numbers (Polo, Anderssen et al. 2012). Therefore, we postulate that high amounts of Oct4 in the early days of reprogramming may facilitate opening of the chromatin and consequently lead to activation of early reprogramming genes (Radziskeuskaya and Silva 2014). Oct4, Sox2 and Klf4 can access closed chromatin at their targets without recruitment of other factors at the early stage of reprogramming, but cofactors are required for later stages. Zaret and his colleagues showed cooperation between OKSM at the first days of reprogramming at different targets. Many of the OKSM targets are related to regulation of MET (Soufi, Donahue et al. 2012). Oct4 also plays a role in induction of MET and chromatin compaction in somatic cells related regions. Therefore, Oct4 down-regulates Snail which is the regulator of the Epithelial to Mesenchymal transition (EMT) via repression of Tgfb β 3 and TgfbR3 and together with Sox2 repress the expression of EMT regulator Zeb2 and facilitate MET via activation of mir200 miRNA family (Radziskeuskaya and Silva 2014). A map of transcription factors binding sites and their targets at different time points as well as changes in chromatin modification will greatly improve our knowledge of reprogramming process.

1.6. Chromatin and epigenetic changes in pluripotent cells:

1.6.1. Chromatin structure:

Chromatin was discovered by Walther Flemming in 1882 through the development of novel histological staining methods that enabled the observation of a unique fibrous structure in the nucleus (Flemming 1882). Chromatin is the combination of DNA and proteins found in the nucleus of eukaryotic cells. Genomic DNA is wrapped in a chromatin structure and the structural units that make up chromatin are defined as nucleosomes. The nucleosome is the basic structural unit of chromatin, consisting of 147 bp of DNA compacted with a four-histone protein core. The chromatin is wrapped in repeated units of nucleosomes. The connection between these nucleosomes is around 20-50 bp of unwrapped linker DNA (Olins and Olins 2003).

Fifty years after the discovery of chromatin, Emil Heitz introduced the notion that chromatin can be characterised by two different states, euchromatin and heterochromatin (Heitz 1928). Heterochromatin is highly condensed DNA, which is associated with repressed genes (Quina, Buschbeck et al. 2006). In heterochromatin, DNA is inaccessible to nucleases and it is methylated in the dinucleotide cytosine-phosphoguanine (CpG), therefore this state of chromatin can protect chromosome ends. Heterochromatin forms usually on satellite centromeric and pericentromeric repeats and histone H3 on lysine 9 and histone H4 on lysine 20 are tri-methylated (Lachner, O'Carroll et al. 2001).

Euchromatin is often associated with actively transcribed genes. This part is accessible to nucleases with unmethylated CpG islands and hyperacetylated N-terminal lysine residues (Quina, Buschbeck et al. 2006). Active genes are also associated with methylation in lysine 4, lysine 79 histone H3 residues and phosphorylation in Ser10 histone H3. Moreover

transcriptional activation is associated with methylation of Arg2, Arg 17, Arg 26 of H3 and there is a cross-talk between histone methylation and acetylation (Bauer, Daujat et al. 2002).

1.7. Epigenetic changes during reprogramming:

Epigenetic changes are associated with alterations in gene regulation without altering the DNA sequence, and these changes are governed by chromatin modifications. Gene activity is ultimately regulated by the state of the chromatin, which is the result of a combination of nucleosome dynamics and positioning, as well as DNA methylation, histone variants and histone modifications (Figure 4). This encompasses the epigenetic regulation of gene transcription.

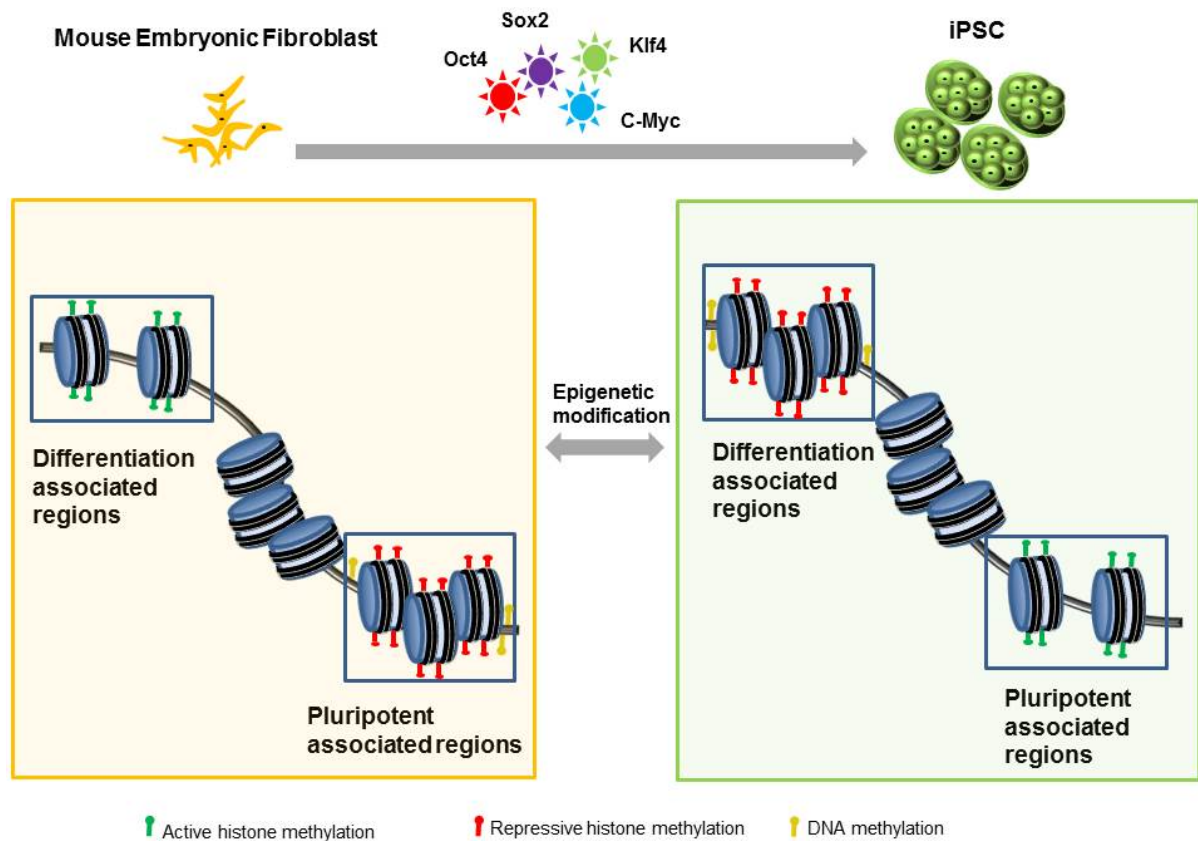


Figure 4. The epigenome of somatic cells and ES cells/iPSCs is different. Fibroblasts (differentiated cells) shown on the left exhibit a relaxed chromatin pattern in regions associated with differentiation (or fibroblast identity) and these regions are marked with active markers such as H3K4me3 (green). In contrast, the regions containing pluripotency genes are more condensed, with repressive histone marks such as H3K9me3, H3K27me3 (red) and DNA methylation (yellow). These patterns are reversed in pluripotent cells (shown at right). (Alaei-Shehni, Knaupp et al. 2014)

1.7.1. Histones variants and modifications during reprogramming

1.7.1.1. Histone variants

H2AZ, a variant of the nucleosome core histone H2A, is one of the chromatin regulators enriched in the promoters of developmentally prominent genes and plays an essential role in ES cell differentiation (Creyghton, Markoulaki et al. 2008). H2A can be replaced by H2AZ by the ATP-dependent chromatin remodelling complex SWR1 (Mizuguchi, Shen et al. 2004). H2AZ and the Polycomb group complex (PcG) occupy the promoters of pluripotency associated genes, and depletion of H2AZ leads to an increase in the pluripotent state of the cells (Creyghton, Markoulaki et al. 2008). Histone H3 variants are also chromatin regulators that play a role in chromatin dynamics. The variant H3.3 is involved in maintenance of non-pluripotent genes (Filipescu, Szenker et al. 2013) and the expression of pluripotent genes in *Xenopus* oocytes occurs after nuclear transfer (Ng and Gurdon 2008). The K4 residue of H3.3 has been shown to play a role in maintaining the epigenetic memory by being modified in transcriptionally active genes (Ng and Gurdon 2008)

1.7.1.2. Histone modifications:

Different post-translational modifications of histones have a notable effect on transcription. Three of the most common post-translational histone modifications occurring in chromatin are methylation, acetylation and phosphorylation (Quina, Buschbeck et al. 2006, Kouzarides 2007). Several lysines in the so-called histone tails are the main target for post-translational modification. Methylation of histones occurs by lysine methyltransferase such as G9a that can modify one single lysine in a single histone, which can then lead to activation or repression of transcription depending on the histone and the site. Methylation of histone H3 lysine 4 (H3K4), histone H3 lysine 36 (H3K36) and histone H3 lysine 79 (H3K79) is associated with transcriptional activation, whereas methylation of H3K9,

H3K27, H4K20 is related to repression of transcription (Quina, Buschbeck et al. 2006). Demethylation of lysine by histone demethylase LSD1 in H3K4, for example, represses transcription whereas demethylation of H3K9 by demethylase LSD1 and a complex of androgen receptors, activates transcription (Shi, Lan et al. 2004). In ES cells, promoters with CpG rich regions at housekeeping genes are usually enriched in H3K4me3 and are generally active, whereas developmentally associated genes are enriched for both H3K4me3 and H3K27me3, so called bivalent domains, and are usually untranscribed (Meissner, Mikkelsen et al. 2008). As mentioned above, during reprogramming, the epigenome is reset and becomes similar to that of ES cells (Maherali, Sridharan et al. 2007). The methylation pattern of H3K27 and H3K4 tri-methylation, for example, is similar in ES and iPSCs, whereas fibroblasts have a different histone methylation pattern (Figure 4) (Maherali, Sridharan et al. 2007). Moreover, investigation of epigenetic changes in intermediate cell populations during reprogramming has shown that changes in the histone methylation pattern of H3K4 and H3K27 follow two waves (Figure 5) (Polo, Anderssen et al. 2012). The genes change their H3K4me3 and H3K27me3 status first in the early stage, followed by a period of minor changes and concludes with a second wave of changes in histone modification (Polo, Anderssen et al. 2012). A closer look at chromatin changes at enhancers and promoters in the early stage of reprogramming reveals that H3K4me2 appears in promoters which are associated with pluripotent genes in the late stage of reprogramming (Koche, Smith et al. 2011). Histone acetylation is a covalent modification and is primarily associated with transcriptional activation. Three main families of acetyltransferases are involved in this modification: GNAT, MYST, and CBP/P300 (Kouzarides 2007). Repressive chromatin remodelling factors such as PcG and the nucleosome remodelling and deacetylase complex (NuRD) induce histone deacetylation and trimethylation of H3K27 (H3K27me3) in promoters and lead to suppression of

differentiation pathways in pluripotent cells (Boyer, Lee et al. 2005, Kaji, Caballero et al. 2006).

OKSM targets are categorised into three classes based on chromatin accessibility and gene activation. The first group of targets are located in open region chromatin in somatic cells and they are able to bind OKSM immediately and led to down regulation of somatic cell associated genes and MET which is related to first wave of reprogramming (Polo, Anderssen et al. 2012, Gifford, Ziller et al. 2013, Zhu, Adli et al. 2013).

The second class of OKSM targets include distal regulatory elements which need additional chromatin remodelling for gene activity (Soufi, Donahue et al. 2012). Some of these elements contain the H3K4 me1 mark and are called permissive elements that bind to transcription factors before their promoter region and prior to gene expression. Binding of transcription factors to distal elements of the target genes leads to crosstalk with its promoter and subsequently a poised chromatin state (Taberlay, Kelly et al. 2011, Soufi, Donahue et al. 2012). Some other distal elements include DNase-I-resistant loci that Sall4 belong to this group. Binding of OKS to these targets facilitates binding of c-Myc. Therefore OKS are called pioneer factors that are able to bind to close chromatin and allow recruitment of chromatin remodellers and other TFs (Soufi, Donahue et al. 2012).

Heterochromatic regions which are characterized by H3K9 me3 are associated with core pluripotency genes such as Nanog and Sox2. The third group of OKSM target genes located in the heterochromatic regions and OKSM are not able to bind these regions immediately and need broad chromatin remodelling for gene activation. These H3K9 enriched regions are associated with repression of lineage specific genes in differentiation. In addition, bivalent genes are marked by both H3K4 methylation and H3K27 methylation in iPSCs

and ES cells. These genes are inactive in iPSCs and ES cells but poised for gene activation during differentiation or development (Wen et al. 2009).

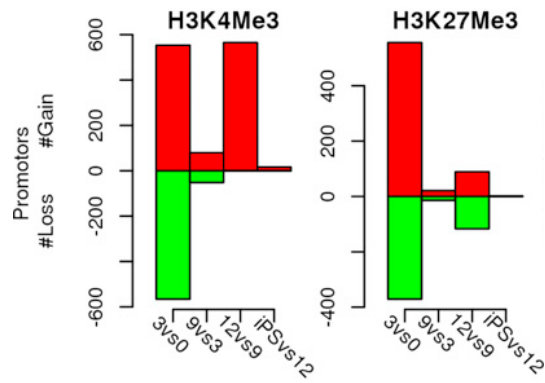


Figure 5. Enrichment for H3K4me3 or H3K27me3 at promoters of differentially expressed genes in progressing intermediates (Polo, Anderssen et al. 2012). The changes in H3K4 me3 and H3K27 me3 (day 3 versus day 0 (3 vs 0), day 9 versus day 3 (9 vs 3) and day 12 versus day 9 (12 vs 9) and iPS versus day 12 (iPS vs 12)).

1.7.1.3. Nucleosome shifting and chromatin remodelling

Nucleosome dynamics and positioning can modulate DNA accessibility and therefore affect the binding of transcription factors to their target sites (Barrero, Boué et al. 2010). Transcription factors recruit different chromatin regulators to promoter regions to modify and mobilise nucleosomes, in order to increase their access to DNA sequences.

It has, for instance, been shown that the presence of nucleosomes in the Oct4 binding sites in the pluripotency associated gene *Nanog* prevent Oct4 accessibility to the target in differentiated cells and consequently lead to gene repression even if the DNA is demethylated (You, Kelly et al. 2011). Recent work has shown that Oct4, Sox2 and Klf4 function as “pioneer” factors at enhancers of genes that promote reprogramming (Beato and Eisefeld 1997). The ability of these pioneer transcription factors to bind to compacted chromatin enables them to induce a cascade of changes in chromatin, which occurs during the later stages of reprogramming. Nucleosome positions can be changed by chromatin remodellers and transcription factors (Beato and Eisefeld 1997, Clapier and Cairns 2009). The effect of transcription factors is related to their competition with the nucleosome to access their target (Clapier and Cairns 2009, Segal and Widom 2009). This competition depends on the affinity of nucleosomes and transcription factors for the DNA sequence and also depends on their concentrations (Polach and Widom 1996, Miller and Widom 2003). Repositioning of nucleosomes by transcription factors occurs with the assistance of ATP-dependent chromatin remodelling factors. ATP-dependent chromatin remodelling complexes affect the interaction between DNA and histones and sequentially allows accessibility of transcription factors to their targets. The BAF complex, for example, is involved in regulating pluripotency gene expression and facilitates the binding of Oct4 to its target promoters (Singhal, Graumann et al. 2010). The addition of Brg1 and Baf155, which are components of the BAF complex, to the reprogramming cocktail (OKSM)

increases demethylation of the promoters of Oct4, Nanog, and Rex1 which is followed by an increase in reprogramming efficiency (Singhal, Graumann et al. 2010)

1.7.1.4. Targeting the epigenetic machinery to increase reprogramming efficiency:

Generation of iPSCs from somatic cells is theoretically simple, but reprogramming efficiency is low and varies from 0.01% to 5% depending on the reprogramming system and the cell of origin (Hochedlinger and Plath 2009). However, reprogramming efficiency can be significantly increased by targeting the epigenetic machinery. DNA methyltransferase inhibitors such as 5-aza-cytidine (AZA) have a positive effect on reprogramming. It has been shown that AZA leads to reactivation of endogenous Oct4 expression in partially reprogrammed cells and enhances their ability to form ES cell-like colonies (Meissner, Mikkelsen et al. 2008, Mikkelsen, Hanna et al. 2008). Histone deacetylase inhibitors such as valproic acid (VPA), trichostatin A (TSA), and suberoylanilide hydroxamic acid (SAHA) increase reprogramming efficiencies in mouse and human fibroblasts with the transcription factors Oct4, Sox2 and Klf4 (Huangfu, Maehr et al. 2008). Moreover, knockdown of the NuRD complex member of the methyl-CpG binding domain family of proteins, Mbd3, by short-hairpin RNA has a positive effect on the reprogramming of somatic cells (Luo, Ling et al. 2013). Depletion of Mbd3 increases reversion of epiblast stem cells, primordial germ cells and somatic cells, such as fibroblast, haematopoietic and neural cells, to a pluripotent state with efficiency approaching 100% (Brumbaugh and Hochedlinger 2013, Rais, Zviran et al. 2013).

1.8. Protein complexes during reprogramming:

Multiple protein complexes that catalyse covalent modifications of histones have important roles in ES cells and iPSC, including proteins involved in histone methylation, acetylation, and ubiquitylation (Niwa 2007). Chromatin factors and protein complexes including

SWI/SNF, Tip60-p400, NuRD, Polycomb group protein (PcG), HMTs, HDMs, CHD, DNMTs are essential to maintain pluripotency in embryonic stem cells (Orkin and Hochedlinger 2011).

1.8.1. Polycomb group (PcG) proteins:

Polycomb group (PcG) proteins function as silencers of differentiation pathways in pluripotent cells. Polycomb-mediated repression is essential for maintenance of pluripotency (Taberlay, Kelly et al. 2011). PcG proteins act in two different complexes known as PRC1 and PRC2. PRC2 consists of the four core proteins EED, Suz12, Ezh2, and RbAp46/48 and PRC1 is composed of the core subunits Ring1A and 1B. PRC1 catalyses monoubiquitination of histone H2A at lysine 119. The PRC2 via SET domain of Ezh2 or the related protein Ezh1, catalyses di- and tri-methylation of histone H3 at lysine 27 (Taberlay, Kelly et al. 2011, Young 2011). H3K27me3 also binds to EED and activates the complex. PRC2 and PRC1 components bind numerous differentiation-related genes that are silenced in ESCs (Margueron and Reinberg 2011). Ezh2 accompanied with c-Myc facilitate MET by silencing TGF- β signalling in human iPSC generation (Rao, Dhele et al. 2015). Increased expression of PRC2 was observed during mouse reprogramming, on the other hand knocking down of PRC components led to decrease in mouse and human iPSC generation. Therefore, the chromatin regulators such as PcG components impact reprogramming (Rao, Dhele et al. 2015).

1.8.2. The Tip60-p400 complex (known as NuA4 HAT):

The Tip60-p400 complex is a c-Myc-associated complex (Fazzio, Huff et al. 2008). Tip60-p400 is also essential for maintaining the ESC state. The multisubunit of the Tip60-p400 complex has two chromatin regulatory functions. Tip60 serves as a protein acetyltransferase and p400, a member of the Swi2/Snf2 family, acts to exchange histones H2AZ-H2B within nucleosomes (Gaspar-Maia, Alajem et al. 2010). siRNA-mediated

inhibition of components of the Tip60-p400 complex leads to differentiation of ESCs (Fazio, Huff et al. 2008). Consistent with gene expression profiling, it has been suggested that Tip60-p400 and Nanog lie within a common pathway, as Nanog depletion leads to less p400 binding at targets (Orkin and Hochedlinger 2011).

1.8.3. The Nucleosome Remodeling and Deacetylase (NuRD) complex:

Chromatin remodeling complexes (CRCs) are able to hydrolyze ATP and utilize the energy to disrupt nucleosome-DNA interaction and consequently lead to accessibility of co-regulator proteins to associated regions and allow them to influence on transcription.

NuRD complexes are chromatin remodelling factors that utilise nucleosome remodelling and histone deacetylase activities to create repressive chromatin structures (Hu and Wade 2012). NuRD is one of four major families of CRCs that contains the conserved subunit members. (Nair, Li et al. 2013). NuRD complex contains seven polypeptides including the ATPase Mi-2, the histone deacetylases HDAC1 and HDAC2, metastasis-associated proteins Mta1, Mta2 and methyl binding domain proteins Mbd2 and Mbd3. NuRD is able to bind to methylated DNA to form heterochromatin and eventually can silence transcription. Mbd3/NuRD directly affect maintenance of ESC lineage commitment via regulation of pluripotency associated gene expression. Metastatic tumor antigen 1 (MTA1) is the core subunit of the NuRD complex, but the role of MTA1/NuRD in gene transcription is still unclear. A recent study showed that MTA1 methylation by G9a methyltransferase and demethylation by LSD1 leads to changes in nucleosome positioning and transcriptional outcome (Shi, Lan et al. 2004, Li, Pakala et al. 2012). Methylation of MTA1 is required for formation of NuRD repressor complex; however, the demethylation of MTA1 can identify the bivalent histone H3K4-AcK9 mark. A study on MTA1 binding to DNA showed that MTA1 does not bind to DNA, but it interacts with DNA through core histones, preferentially histone H3 (Nair, Li et al. 2013).

1.9. Conclusion:

Despite significant advances in our understanding of the mechanisms of reprogramming, it is still not clear how the TFs that initiate the reprogramming process find and modulate their target genes during this process. In this study we will address these mechanistic questions to determine the TF dynamics that orchestrate reprogramming and the subsequent generation of iPS cells. We expect that this work will contribute significantly to both basic science and clinical outcomes and applications.

From the basic science point of view we aim to answer two key questions regarding the mechanism behind nuclear reprogramming and TF targeting in general: 1) How do transcription factors find their targets and modulate gene expression? 2) How does the interaction between chromatin and TFs influence accessibility and nucleosome position? Answering these questions are not only important for the understanding of somatic cell reprogramming, but also to understand and model any process involving cell identity such as differentiation, direct cell conversion and development.

From a clinical perspective, we will directly address two relevant questions: 1) Why do some cells fail to reprogram, despite expressing OKSM? 2) How can we improve the reprogramming process to make it safer and more efficient, to better enable the therapeutic application of iPS cells? Addressing these questions will undoubtedly aid in making the process more efficient, which in turn allows the generation of iPS cells from fewer cell numbers. iPS technology is poised to enter broad spectrum clinical trials, however there are still concerns about administration techniques of iPSCs into patients and the potential for genomic mutations. Previous studies have increased the understanding of iPSC tumorigenic potential and offered possible solutions to decrease this risk (Utikal, Polo et al. 2009), while also presenting methods to monitor the reprogramming process and select the

best cell of origin for downstream lineage-specific differentiation (Polo, Anderssen et al. 2012).

In summary, this study aims to tackle fundamental mechanistic questions regarding the reprogramming process, while at the same time shedding light onto the basic principle of chromatin targeting and gene regulation by transcription factors. This will provide a better understanding of the molecular and cellular requirements of direct lineage conversion and aid in efforts to apply iPS cell technology for therapeutic purposes in the future.

CHAPTER 2

Material & Methods

2.1. Materials and Suppliers:

2.1.1. Cell culture reagents:

- β -Mercaptoethanol (Life Technologies, Melbourne, Australia, Cat.No. 21985-023)
- Sodium pyruvate (Life Technologies, Cat. No. 11360-070)
- GlutaMAX TM (Life Technologies, Cat.No. 35050-070)
- MEM Non-Essential Amino Acids (NEAA) (Life Technologies, Cat.No. 11140-050)
- Embryonic Stem Cell qualified Foetal Bovine Serum (FBS) (LifeTechnologies, Cat.No. 10439024)
- Trypsin-EDTA (0.25%, Life Technologies, Cat.No. 25200-056)
- DMEM High Glucose (Life Technologies, Cat.No. 11960-044)
- Advanced Dulbecco's Modified Eagle Medium (Invitrogen, Australia)
- DPBS (Dulbecco s phosphate buffer saline) (Life Technologies, Cat.No. 14190-144)
- DMSO (Dimethyl Sulfoxide) (Sigma-Aldrich, Sydney, Australia, Cat.No. D8418)
- KnockOut DMEM (Life Technologies, Cat.No. 10829-018)
- Leukemia Inhibitory Factor (LIF) (Millipore, Cat.No. ESG1107)
- Doxycyclin hyclate (Sigma-Aldrich, Cat.No. 33429-100MG-R)
- Penicillin/streptomycin (Life Technologies, Cat.No. 15140)

- Gelatine from porcine skin (Sigma-Aldrich, Cat.No. G1890)
- Propidium Iodide solution (PI) (Sigma-Aldrich, Cat.No. P4864-10ML)
- Irradiated MEFs (Life Technologies, Cat.No. S1520-100)
- SYBR[®] Green I nucleic acid gel stain (Sigma- Aldrich. Cat. No. S9430)
- Polybrene- (Millipore, Cat.No. TR-1003-G)
- Formaldehyde solution (Sigma- Aldrich, Cat.No. 252549)
- Glycine (Merck Millipore, Cat.No. 104201)
- NuPAGE[™] Novex[™] 4-12% Bis-Tris Protein Gels (Novex, Cat.No. NP0336BOX)
- Precision protein [™] Dual color standards (BioRad, Cat.No. 1610374)

2.1.2. Antibodies:

- Anti-mouse CD90.2 (Thy1.2) Pacific Blue (eBioscience, San Diego, CA, USA, Cat.No. 48-0902-82)
- Anti-Human/Mouse SSEA-1 Biotin (eBioscience, Cat.No. 13-8813)
- Streptavidin PE-Cy 7(eBioscience, Cat.No., 25-4317-82)
- Anti- Mouse CD326(Epcam) Fitc (eBioscience, 48-5791-82)
- Anti Ssea1 microbeads (Milteyi Biotec, Cat.No. 130-094-530)
- Anti- Oct4 antibody- ChIP Grade (abcam, Cat.No. ab19857)
- Anti Sox2 antibody – ChIP Grade (abcam, Cat.No. ab59776)
- Mouse Klf4 antibody (R&D systems, Cat.No. AF3158)

2.1.3. Disposables:

- 75-cm² Tissue culture flasks (Corning/BD Bioscience, Melbourne, Australia, Cat.No. 430641)
- 6 well tissue culture plates (Corning/BD Bioscience, Cat.No. 3516)
- 15-ml centrifuge tubes (Corning/BD Bioscience, Cat.No. 430791)
- 50-ml centrifuge tubes (Corning/BD Bioscience, Cat.No. 430829)
- Disposable surgical blades (Swann Morton, Sheffield, England, Cat.No. 0201)
- Cryogenic vials (Corning/BD Bioscience, Cat No. 430487)
- Cell strainers (70 µm Nylon, BD Bioscience, Melbourne, Australia, Cat. No. 352340)

2.1.4. Equipment:

- Haemocytometer (Boeco, Hamburg, Germany)
- Bench top Centrifuge Thermo, Melbourne, Australia)
- Inverted tissue-culture microscope with phase contrast microscope (Moticam 2300 or similar)
- Class II biosafety cabinet with aspirator for tissue culture (Biocabinet or similar)
- “Mister Frosty” Freezing Container (NALGENE or similar)
- Incubator (Thermo or similar at 5% CO₂ under normoxic or low oxygen(5%) conditions)

- Water bath (Ratek or similar)
- Nitrogen tank (CHART or similar)
- Autoclave (GETINGE or similar)
- Fluorescence activated Cell sorter with 405nm, 488nm, 560nm excitation laser
(Influx, BD Biosciences)
- T100 TM Thermal cycler (Bio RAD, Sydney, Australia)
- Odyssey Scanner (LI-COR Biosciences - Millennium Science)
- AutoMACS pro separator (Miltenyi Biotec)
- Magnetic rack (Invitrogen)

2.2. Cell Culture:

2.2.1. Media preparation

iPS cell medium: Knockout DMEM containing 15% FBS (vol/vol), 1% GlutaMAX (vol/vol), 1% Non-essential amino acids (vol/vol), 0.1% β -Mercaptoethanol (vol/vol), 1000 Units/ml LIF, 1% Pen/Strep (vol/vol) To make 500 ml of medium, add 75ml FBS. 5ml NEAA, 5 ml GlutaMAX, 5ml penicillin/streptomycin, 500 μ l β -Mercaptoethanol, 50 μ l LIF to Knockout DMEM media and mix thoroughly.

MEF (Mouse Embryonic Fibroblast) medium: DMEM containing 10% FBS (vol/vol), 1% GlutaMAX, 1% Non-essential amino acids, 1% Sodium pyruvate, 0.1% β -Mercaptoethanol. To make 500ml media, mix 50ml FBS, 5ml GlutaMAX, 5ml NEAA, 5ml Sodium pyruvate, 500 μ l β -Mercaptoethanol in DMEM media.

Freezing medium: 10% DMSO (vol/vol) and 90% FBS (vol/ vol)

Gelatine solution: Add 0.1% gelatine (wt/vol) in milli-Q water. Autoclave to properly dissolve gelatine.

Gelatine coating of T75 flasks/10cm dishes/6 well-plates: Add 3ml 0.1% gelatine solution to a T75 flask /10cm dish and 500µl to each well of a 6 well-plate to cover their surface. Incubate at room temperature (RT) for 20 min. Aspirate gelatine before use.

FACS labelling buffer: PBS containing 1% FBS (vol/vol). To make 10 ml, mix 100 µl FBS into 10 ml PBS.

2.2.2. Mouse genotyping:

For genotyping, 25 ul reaction mix was used containing DNA template which extracted from embryos head (1ng), Taq DNA polymerase recombinant (Life Technologies, Cat No. 10342-020); Forward and reverse primer (5µM); Mgcl₂ (2.5 mM). Polymerase chain reaction (PCR) was performed in a Dyad PCR machine (Bio-Rad) (see PCR conditions for rtTA, Oct 4, OKSM).

2.2.2.1. Polymerase chain reactions (PCR):

PCR conditions for m2rtTA PCR:

94°C for 3min, 35 cycle of (94°C for 30 sec, 65°C for 1 min, 72°C for 1 min), then 72°C for 2 min, hold at 12°C. Following PCR kit was used: Taq DNA polymerase (Life Technologies, Cat No. 10342-020); primer concentration: 5µM; optimal MgCl₂ concentration: 3.5mM

Product size is 650bp for wild type allele and 340bp for mutant allele.

PCR conditions for Collagen-StemCa/OKSM PCR:

95°C for 1 min, 2 cycles of (94°C for 30 sec, 70°C for 45 sec), 6 cycles of (94°C for 30 sec, 68°C for 45 sec) and 30 cycles of (94°C for 20 sec, 60°C for 1 min) followed by 72°C for 5 min and hold at 12 °C. PCR kits used: Taq DNA polymerase recombinant (Life Technologies, Cat No. 10342-020); primer concentration: 5µM; optimal MgCl₂ concentration: 2.5 mM

Product size is 331bp for wild type allele and 551bp for mutant allele.

2.2.2.2. DNA gel electrophoresis:

DNA fragments were visualized on 1-3% agarose containing gel red Nucleic Acid gel stain.

2.2.3. Reprogramming of MEFs into induced pluripotent stem cells:

The two transgenic mouse strains (OKSM, M2rtTA) or (OKSM, M3 rtTA) can be obtained as homozygous (homo) founder animals from the Jackson Institute. Reprogrammable mice can be generated by crossing animals from the OKSM strain with animals from the M2rtTA strain. MEFs were generated from mice at embryonic day (E) 13.5.

2.2.3.1. Generation of mouse embryonic fibroblasts

Perform a timed mating between a male or female of the OKSM strain with a member of the opposite gender of the m2rtTA strain. Harvest embryos at embryonic day 13.5 by culling the mother and removing the uterine horn. Please note that it is essential to spray the dead female's abdomen generously with 80% ethanol solution before removing the horn to avoid microbial contamination. Transfer the uterine horn into a 10cm tissue culture dish with 10ml of sterile Phosphate buffered saline (PBS).

Using a dissection microscope in a class-I tissue culture hood employ a pair of surgical scissors to cut the uterine horn into pieces containing one embryo.

Use fine dissection forceps to carefully remove the uterine envelope and the extra embryonic membranes surrounding each embryo. Transfer each embryo to a separate 10cm dish with 10 ml of PBS. Commence by removing each embryo's head, feet, legs, tail and internal organs (heart, liver, intestine etc.) using the forceps. At the end the experimenter should be left with a degutted torso. Transfer the embryo's head to an Eppendorf tube and freeze so it can be used to extract DNA for genotyping should there be any doubt about the embryos.

Transfer the embryo's torso into an empty gelatine coated 10cm plate and use two surgical blades to mince the embryo for 2 minutes. Add 200ul of trypsin/EDTA solution on top of the minced embryo and incubates for 3-5 minutes at room temperature. Continue mincing for an additional 2 minutes and add 10ml of MEF media to the plate. Use a 10ml serological pipette mix the plate's contents and transfer to a tissue culture incubator (both normoxic and hypoxic incubators can be used, refer to discussion for more information).

After 24-48 hours, the plate should be densely covered with MEFs. At this stage, either further propagate the cells, use them for reprogramming, or freeze them down. To freeze the cells remove the culture media, rinse the dish with 10ml PBS to remove traces of media and overlay with 3ml Trypsin/EDTA solution. Incubate for 3-5 minutes at 37°C, quench by adding 5ml of MEF media and transfer to a 15ml centrifugation tube. Spin down and resuspend the pellet in 3ml freeze media. Transfer to 3 cryovials and freeze down in a Mr Frosty container.

For successful reprogramming experiments low passage (P) MEFs (P0-P2) are essential. MEFs were either used right after derivation or alternatively cryopreserved MEFs were thawed. If cryopreserved MEFs were used, thaw and allow cells to recover for 24-48 hours before reprogramming. For reprogramming we need to seed MEFs at densities between 2.5×10^3 cells/cm² on gelatine coated tissue culture plastic in doxycycline (2μg/ml)

containing iPSC media (12ml of media for a T75 flask/ 3ml of media for a well in a 6-well plate).

Reprogramming begins with the addition of doxycycline (2ug/ml) supplemented iPS cell media to the cells. The media needs to be changed every other day. At different time points the cells were harvested. After day 12, switch the cells to iPSC media without doxycycline and culture for another 4-7 days (media changes daily or every other day with double the media volume). Aberrant iPSC colonies that have not become independent of induced OKSM expression will disappear during this period. At day 16-20, the iPS cells can be collected either by picking individual colonies or by passaging whole cultures.

2.3. FACS isolation of reprogramming intermediates

Reprogramming intermediates were FACS extracted from day 0, day 3, day 6, day 9, day 12, and day 16 reprogramming cultures as well as from established iPSC lines using cell surface markers Thy-1, Ssea-1, and Epcam.

2.3.1. Antibody labelling of cells:

For each cell pellet, 200µl of labelling buffer (PBS + 0.1% FBS) supplemented with antibodies (anti-Thy-1 pacific blue 1:400, anti-Ssea1 biotin 1:400 and anti-Epcam Fitc 1:400) was prepared. The labelling media with antibodies was added to the cell pellets and incubated on ice for 10 min. The tubes were gently tapped to get cells back into suspension and incubated for another 10 min on ice. Then, the cell pellets were washed with 10ml of cold PBS and spun down at 200g for 5min at 4°C. While cells were spinning down the labelling media with secondary antibodies was prepared. For each tube to be stained, 200µl of labelling media were supplemented with Streptavidin-PeCy7 (1:200) (The volume of labelling media was changed depends according to cell numbers). The pellets were

resuspended in 200µl of labelling media supplemented with Streptavidin-PeCy7. The cells were incubated for 10 min on ice. The tube was gently tapped to resuspend cells and incubated for another 10 min on ice. The cells were washed with 10ml cold PBS and spun down at 200g for 5min at 4°C. Then, the cell pellets were resuspended in labelling media supplemented with propidium iodide (PI) (2ug/ml). Cells were resuspended in PI solution at approximately 1×10^7 cells/ ml. The cell suspension was passed through a 70µm strainer to remove clumps and transferred to an appropriate FACS sample tube. The cells were kept on ice until FACS was performed.

2.3.2. Preparation of compensation tubes:

Colour compensation had to be performed for the three channels used to detect Thy1 Ssea1 and Epcam. In addition, some unlabelled cells were required to set voltages and to determine gates for sorting.

1. Resuspend iPS cell pellet in 800µl of labelling buffer.
2. Remove 200µl of the iPS cell suspension (“unlabelled cells”), pass through a 70µm strainer and supplement with propidium iodine (PI) to a final concentration of 2µg/ml and transfer into a FACS tube.
3. Split the remaining 600µl of the iPS cell suspension into three 15ml tubes.
4. To the first tube add 0.5µl of the anti-Thy1 Pacific Blue antibody (Pacific Blue compensation control), to the second tube add 0.5µl of the anti-Epcam-Fitc antibody (Fitc compensation tube), to the third tube add 0.5µl of the anti Ssea-1-Biotin antibody (Pe-Cy7 compensation control)
5. Mix tubes by tapping gently and incubate on ice for 10min.
6. Gently tap the tube to resuspend cells and incubate for another 10min on ice.
7. Add 10ml of cold PBS to each tube and centrifuge at 200g for 5min at 4°C.

8. Resuspend the pellets of all three tubes each in 200µl of FACS labelling media.
9. Supplement the cell suspensions of the Pacific Blue and the Fitc compensation controls with PI to 2µg/ml, pass through a 70µm strainer and transfer to an appropriate FACS tube. Keep on ice.
10. Add 1µl of Streptavidin-PeCy7 conjugate to the suspension of cells that have been labelled with the anti-Ssea-1-biotin antibody. Mix by tapping tube gently.
11. Incubate tube on ice for 10min.
12. Gently tap the tube to resuspend cells and incubate for another 10min on ice.
13. Add 10ml of PBS to each tube and centrifuge at 200g for 5min at 4°C.

Resuspend the pellet in 200µl FACS labelling buffer supplemented with PI (2ug/ml) and pass through a 70µm strainer, transfer to a FACS tube and keep on ice until needed.

2.3.2. Fluorescent Activated Cell Sorting (FACS):

Due to the cell size of MEFs and some subpopulations of reprogramming cells, a cell sorter with a 100µm nozzle was used. The compensation tubes were used to set the appropriate voltages for forward and side scatter and the required fluorescent channels (PB, Fitc/GFP, PI, and Pe-Cy7) and to perform colour compensation. When running the samples, cell debris, aggregates and dead cells were excluded. For days 3, 6 and 9 the reprogramming cultures were subfractioned into three populations: Ssea1-/Thy1.2+ cells, Ssea1-/Thy1.2- cells (both populations are refractory to reprogramming) and the reprogramming intermediates the Ssea1+/Thy1.2+ cells. In order to set the gates unlabelled cells were used. (Nefzger, Alaei et al. 2014).

2.4. Magnetic-activated cell sorting (MACS):

For large scale of sorting of intermediates, AutoMACS was used.

2.4.1. Auto MACS Buffers:

Running buffer:

Running buffer containing Dulbecco's Phosphate-Buffered Saline (DPBS) 1X, 0.5 % Bovine Serum Albumin (BSA), 2 mM EDTA (Sigma Aldrich).

Rinsing buffer:

Dulbecco's phosphate-buffered Saline (DPBS) 1X, 2mM EDTA.

2.4.2. Labelling with AutoMACS antibody:

The cells were counted with the cell counter (Invitrogen) and then 80µl of running buffer was added per 10^7 cells and cells were passed through a 70µm cell strainer. Then, 20µl of anti-Ssea1 microbeads (Miltenyi Biotec) were added to 10^7 cells. The cells were incubated for 15 min on ice and then washed with 10 ml PBS and centrifuged at 1500 rpm for 3 min. The PBS was removed and 400µl of running buffer was added. Then the Automacs separator was set up and the positive fraction was captured through the column.

2.5. Virus production:

One day prior to transfection, 293T or 293FT cells were cultured into T75 culture vessels in the absence of antibiotics. Immediately before transfection, the serum containing media was changed to un-supplemented, serum free Advanced Dulbecco's Modified Eagle Medium (7.5ml) (Invitrogen, Australia). Preparation of the lipofectamine-DNA complexes for one T75 vessel went as follows: In 1.5ml of OptiMEM medium (Invitrogen, Australia) transfer vector (6,750ng), packaging plasmids psPax2 (4,200ng) and pMD2G (2,700ng) were combined with 15µl of Invitrogen's proprietary "plus" component. 5 min later 30µl of Lipofectamine LTX (Invitrogen, Australia) was added and the solution mixed,

complexes were allowed to form (at room temperature) for 30 min and then added to the T75 vessel. Five hours after transfection the serum free media was exchanged with supplemented Advanced Dulbecco's Modified Eagle's Medium media containing 2% FBS (Invitrogen, Australia), non-essential amino acids (Sigma-Aldrich, Australia), 2mM L-Glutamine (Invitrogen, Australia) and 30 μ M cholesterol (Sigma-Aldrich, Australia). Viral supernatants were harvested after 24 and 36 hours and frozen down at -70°C.

2.5.1. Concentration of primary viral supernatants

10ml primary viral supernatants were filtered through a 0.45 μ m PVDF membrane (Millipore, Massachusetts) and subsequently centrifuged on 2.5ml of a 20% sucrose cushion prepared in TNF buffer (50mM Tris-HCl; 100mM NaCl; 0.5mM EDTA at pH 7.4) (all reagents from Sigma-Aldrich, Australia) for 3 hours at 50,000xg at 4°C. After centrifugation pellets were resuspended in 100 μ l of ES maintenance media. To ensure proper resuspension the pellet was pipetted up and down at least 20 times. Viral concentrates were stored in aliquots at -70°C until used.

2.5.2. Virus infection

The cells were infected with 500 μ l ES media containing polybrene diluted 1:1600 and 100 μ l of LV Myc and LV OKSM viruses. Then the infected cells were spun down in 1900 rpm for one hour at RT and they were incubated overnight at 37°C. The media was changed to ESC medium with doxycycline.

2.6. AP staining

The Vector Red alkaline phosphatase substrates (Cat. No. SK 5100) were used. The cells were washed with PBS and the substrate working solution mix containing 100 mM Tris-HCl (pH 8.2-8.5 buffer). To make the substrate working solution, 2 drops of reagent 1 were added to 5 ml of 100 mM Tris-HCl and mixed well and then 2 drops of reagent 2

and then 2 drops of reagent 3 were added and mixed. With 30 min incubation in the dark, the AP stained colonies appeared and were counted.

2.7. Western Blot:

2.7.1. Western Blot buffers:

10% SDS

10x Western transfer buffer (10 x WTB): 60.5 g Tris/L and 150.14 g Glycine/L.

1x WTB: 200ml 5xWTB (250mM Tris, 2M Glycine), 200ml Methanol, 500ul 20% SDS
make the volume up to 1L with dH₂O

Tris-buffered Saline (TBS): 2.42g/L Tris 11.69 g/L NaCl, pH 7.4

TBST: 0.5% (v/v) Tween 20 in TBS

Blocking buffer: 5% (w/v) skim milk powder in TBS

2.7.2. Western Blot procedure:

The total protein concentration was determined in the samples using the Qubit protein assay. 10-30ug of total protein was loaded to start with. The same amount of protein was calculated to load per well to be able to compare between samples. The samples were diluted with 4x denaturing and reducing loading dye and incubated at 100 °C for 10 min or incubated. The gel cassette was assembled and filled up with 1x MOPS SDS running buffer. The samples and 5ul of pre-stained protein ladder were loaded. The gel was run at 200V constant until the dye-front was at the bottom of the gel. The tops of the wells were removed with a spatula and the Western Blot sandwich was set up. The tank was filled with cold 1x WTB. The gel was run at 400mA for 2 hours and then the membrane was incubated in 50ml blocking buffer (PBS + 5% BSA) for 1h on the shaker at RT (or

overnight in the cold room). The membrane was washed 1 x 5min in PBST (PBS + 0.05% Tween20), then the membrane was incubated in primary antibody (Oct4, Sox2) solution (1:1000 diluted with PBST + 5% BSA) in the cold room overnight. The membrane was washed three times with PBST and was then incubated in secondary antibody solution (1:5000 diluted with PBST + 5% BSA) for 2h at room temperature wrapped in foil. After incubation, the antibody solution was removed and the membrane was washed three times with PBST wrapped in foil. The samples were analyzed on an Odyssey scanner.

2.8. Multiple ligation PCR amplification:

A modified Multiplex Ligation-dependent Probe Amplification (MLPA) method was used to identify genomic regulatory regions. This method is a multiplexed PCR method originally developed for detecting deletion and duplication events in genomic DNA. It was adapted to allow the analysis of up to 50 different genomic regions in one PCR reaction to identify DNase I hypersensitivity. It is based on the binding of two-half probes after annealing to a complementary sequence of interest. Each MLPA probe consists of two oligonucleotides, the left probe and the right probe. These were designed based on DNase I hypersensitive sites, and each of them have different lengths. The same sequences were added to the 5' and 3' end of left and right side of the probe respectively. For the 5' end, the sequence of up to 50 probes can be used in one PCR reaction. The advantage of the same forward and reverse sequence at the end of each probe is it allows the use of a single primer pair in the amplification of all probes in a PCR reaction. These primers have a fluorescent label at the 5' end. In this process, DNA is denatured and the probe mix is added to the DNA sample to allow hybridisation to the target. The left and right side of the probes are ligated to complete one probe. During PCR, only the ligated probes can be amplified.

In this method, the cell nuclei were extracted using the NE-PER nuclear and cytoplasmic extraction kit (Invitrogen) according to the manufacturer's specifications. The nuclei were washed with DNase I buffer, and then treated with the required amount of DNase I enzyme. The reactions were incubated for 20 min at room temperature and stopped by adding nuclei lysis buffer and incubated at 55°C for 45 min. The digested DNA was recovered and cleaned with Highpure PCR purification tubes (Qiagen). The DNA was then denatured at 98°C for 5 min and allowed to cool to room temperature. MLPA probes and MLPA buffer were mixed and incubated for 1 min at 95°C followed by 16 hrs at 60°C. Then the samples were analysed using capillary electrophoresis.

2.8.1. MLPA analysis:

Peak data was normalized to the average of all positive probes in the samples with all other reactions being normalized to negative probes.

2.9. Assay for Transposase Accessible Chromatin (ATAC):

50,000 cells were sorted by FACS and centrifuged at 2400 rpm for 5 min at 4°C, then the cells were washed with 50µl PBS. Cells were resuspended in 50µl of cold lysis buffer containing Tris-HCl (10mM), NaCl (10mM), MgCl₂ (3mM) and then centrifuged at 2400rpm in 4°C for 5 min and the supernatant was discarded. 25µl of 2X TD Buffer (Nextera DNA sample prep kit, Cat. No. FC-121-1031), 2.5µl of Tn5 Transposase (Nextera DNA sample prep kit, Cat. No. FC-121-1031) and 22.5µl of nuclease-free water were added to the cell pellet and the sample was incubated for 30 min at 37°C. Sample was immediately purified using the Qiagen Min-Elute reaction kit (Qiagen). Transposed purified DNA was eluted in 10µl EB buffer (10mM Tris buffer, pH 8) and was stored at -20°C. The next step was PCR amplification for making a library with index PCR primers. The reaction contained 10µl of the transposed DNA, 10µl of nuclease-free water,

25µM Ad1_noMX (Customized Nextera PCR Primer 1), 2.5µl (1.25µM in 50µl Rx), 25µM Ad2.1 (Customized Nextera PCR Primer 2), 2.5µl (1.25µM in 50µl Rx), 25µl of NEBNext High-Fidelity 2x PCR Master Mix. The PCR conditions were: 72 °C – 5 min, 98 °C – 30 sec, (98 °C – 10 sec, 63 °C – 30 sec (repeat 9 x)), 72 °C – 1 min, Hold at 4°C. The amplified libraries were purified using the Qiagen Min-Elute PCR clean up kit (Qiagen Cat. No. 28004). The sample was eluted in 20µl Elution Buffer and the quality was checked using KAPA quantification kit.

2.9.1. Library QC:

The quality of samples was checked using KAPA quantification kit. 10nM of the samples was used for KAPA. The samples were diluted in the buffer containing 10mM Tris and 0.05% tween 20. Library Quantification Primer Pre-mix (10X), contained the following primers:

Primer Name	Primer Sequence
Primer 1	5'-AAT GAT ACG GCG ACC ACC GA-3'
Primer 2	5'-CAA GCA GAA GAC GGC ATA CGA-3'

2.9.2. ATAC analysis:

The ATAC-Seq dataset was processed as described in Mo et al (PMID: 26087164) (Mo, Mukamel et al. 2015). In brief, ATAC-Seq sequencing reads (paired-end with 100nt read length) were adaptor-trimmed and quality filtered using Cutadapt (DOI: <http://dx.doi.org/10.14806/ej.17.1.200>). Surviving pairs were aligned to the complete mouse genome using Bowtie2 (PMID: 22388286). Duplicate sequencing reads were removed using the MarkDuplicates from Picard (<http://broadinstitute.github.io/picard>)

and sub-nucleosomal fragments (those minor to 100bp) were extracted using a combination of BEDtools' bamToBed (PMID: 20110278) and a custom script.

The analysed sample was normalized to a total of 10^7 uniquely mapped sequencing reads and a read density matrix of a 10kb region surrounding the TSS of 15k genes associated with the GO term cell cycle (GO:007049) was calculated using HOMER (PMID: 20513432). Exploratory analyses were performed by creating a custom track in UCSC's genome browser (PMID: 12045153).

2.10. Chromatin immunoprecipitation (ChIP)

Cells were fixed in 1% formaldehyde for 10 min, quenched with glycine and washed 3 times with PBS. Cells were then resuspended in lysis buffer and sonicated for 6 cycles: 20 sec on, 30 sec off, in a Bioruptor (Diagenode, Philadelphia, PA) to shear the chromatin to an average length of 600bp. Supernatants were precleared using protein-A agarose beads (Diagenode) and 10% input was removed and put aside. Immunoprecipitations were performed using polyclonal antibodies to Oct4 (abcam) and Sox2 (abcam). DNA-protein complexes were pulled-down using pre-blocked protein A agarose beads and washed. The chromatin was washed with dilution buffer, low-salt washing buffer and high-salt washing water respectively. DNA was recovered by overnight incubation at 65°C to reverse the cross-linking and purified using QIAquick PCR purification columns (Qiagen,). Enrichment of the different target genes was detected by quantitative real time PCR (qPCR).

2.10.1. ChIP buffers:

Lysis buffer. For 10ml.

Ingredient	Volume	Final concentration (cc)
20% SDS	500µl	1%
0.5M EDTA pH 8	200µl	10mM
2M Tris-HCl pH 8.1	250µl	50mM
Water	9.05ml	
Roche protease inhibitor	Add before use	

Dilution buffer. For 50ml

Ingredient	Volume	Final cc
5M NaCl	1.65ml	165mM
20% SDS	25µl	0.01%
Triton X-100	550µl	1.1%
0.5M EDTA pH 8.0	120µl	1.2mM
2M Tris HCl pH8.0	417µl	16.7mM
Water	47.24ml	

Low salt washing buffer. For 50ml

Ingredient	Volume	Final cc
5M NaCl	1.5ml	150mM
5% Na deoxycholate	5ml	0.5%

20% SDS	250µl	0.1%
20% Nonidet P-40	2.5ml	1%
0.5M EDTA pH 8.0	100µl	1mM
2M Tris HCl pH8.0	1.25ml	50mM

High salt washing buffer. For 50ml

Ingredient	Volume	Final cc
5M NaCl	5ml	500mM
5% Na deoxycholate	5ml	0.5%
20% SDS	250µl	0.1%
20% Nonidet P-40	2.5ml	1%
0.5M EDTA pH 8.0	100µl	1mM
2M Tris HCl pH8.0	1.25ml	50mM
Water	35.9ml	

2.10.2. Quantitative Real Time PCR:

The qPCR was performed by ABI 7500 for ChIP optimizing conditions. The qPCR condition was 94°C for 10 min, 40 cycle of (94°C for 30 sec, 60°C for 1 min), then 1 cycle (95°C for 15 sec, 60 °C for 1 min, 95 °C for 15 sec).

2.10.2.1. Fluidigm ChIP:

Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) was performed using the Eva Green qPCR Master Mix reagent (Fluidigm) in Ssea1+ cells. Triplicate samples for each primer pair were analysed on a BioMark instrument (Fluidigm 24 x 192 chip). The limit of detection was set to 40 cycles. Data normalized with negative targets.

Required equipment for Fluidigm

BioMark HD System

IFC Controller MX (for the 48.48 Dynamic Array IFC) or HX (for the 96.96 Dynamic Array IFC)

Standard 96-well Thermal Cycler

Software requirements for analysis

Fluidigm® Real-Time PCR Analysis Software v.3.0.2 or higher and BioMark HD Data Collection Software v.3.1.2 or higher is required for this protocol.

All the primers and probes were listed in Appendix part.

CHAPTER 3

Reprogramming factor levels and their impact on reprogramming kinetics and efficiency

3.1. Introduction

As mentioned in Chapter 1, there are different methods to generate iPSC cells such as using retroviruses and lentiviruses. However due to some disadvantages of retroviral and lentiviral systems, such as viral transgene silencing, heterogeneous expression, and random viral integrations, the development of a secondary system was a milestone in the reprogramming field (Stadtfield, Maherali et al. 2010) which allows temporal control of TFs and generation of isogenic cells for reprogramming. This mouse model expresses M2 reverse tetracycline transactivator (rtTA) from the Rosa 26 locus, and a doxycycline-inducible polycistronic cassette encoding Oct4, Sox2, Klf4 and c-Myc (OKSM) in the 3' untranslated region of the *Col1 α 1* locus (Stadtfield, Maherali et al. 2010) (Figure 1). Using this secondary reprogramming mouse model, iPSCs are generated from somatic cells upon ectopic expression of OKSM. This leads to the more efficient generation of iPSC cells from somatic cells, compared to viral systems (Sommer, Stadtfield et al. 2009, González, Boué et al. 2011). Reprogramming is initiated through doxycycline (dox) addition, and there is temporal control of the expression of the exogenous OKSM by dox addition or dox withdrawal. Expression of endogenous Oct4 at the end of reprogramming indicates that these cells are independent of exogenous factors. In order to monitor expression of endogenous Oct4, this mouse strain also carries an Oct4-GFP reporter under control of the endogenous Oct4 promoter.

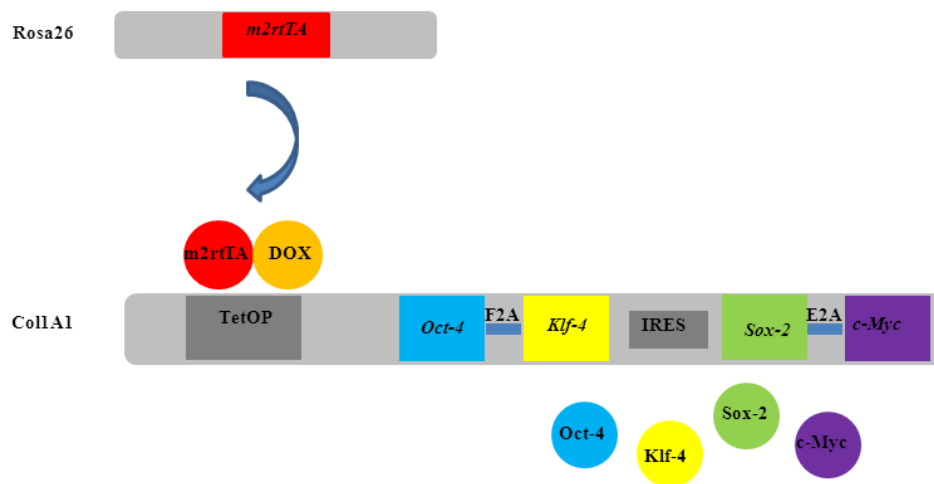


Figure 1. Schematic of the Rosa26 M2rtTA-Coll1a1 OKSM mouse model. The OKSM cassette (Oct4, Sox2, Klf4, c-Myc) is under control of the Tet O promoter (doxycycline inducible promoter). Following doxycycline addition, the M2rtTA can activate the Tet O promoter and the four transcription factors can be transcribed (Nefzger, Alaei et al. 2014).

Reprogramming for mouse fibroblasts takes 12 days and is associated with changes in cell morphology, surface marker expression, epigenetic state and transcriptome (Maherali, Sridharan et al. 2007, Polo, Anderssen et al. 2012). Intermediate populations that arise during the reprogramming process have been defined, based on surface marker expression such as Thy1, which is a fibroblast identity marker, and Ssea1, a pluripotency-associated marker (Stadtfeld, Maherali et al. 2008). During reprogramming, there are defined transitions for successfully reprogramming cells from a Thy1⁺ state to a Thy1⁻ state followed by activation of Ssea1 (Stadtfeld, Maherali et al. 2008, Polo, Anderssen et al. 2012). We demonstrated previously that cells unable to downregulate Thy1 are refractory to the reprogramming process and are not able to give rise to iPSCs. Only cells that are able to reactivate Ssea1 will fully progress through the reprogramming process (Polo, Anderssen et al. 2012).

We previously investigated why Thy1⁺ cells are not able to generate iPSC and we hypothesized the lack of reprogramming seen in refractory cells might be due to a lack of exogenous factors. To do this, the cells were sorted at different time points (day 3, 6, 9, 12) using Thy1 and Ssea1 markers, and the amount of Oct4 protein and OKSM cassette (StemCCA) was measured in these populations (Thy1⁺ and Ssea1⁺ populations). We showed there were higher levels of the unprocessed protein from the OKSM cassette (StemCCA) (previously explained by Carey et al (Carey, Markoulaki et al. 2011)) in Ssea1⁺ cells compared to Thy1⁺ cells (Polo, Anderssen et al. 2012) (Figure 2). This result suggested that the amount of OKSM protein plays an important role in directing the Thy1⁺ population towards becoming Ssea1⁺ cells.

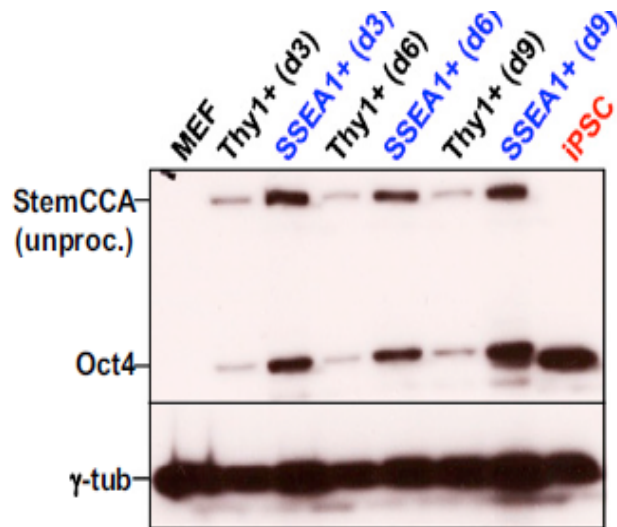


Figure 2. Densitometric quantification of Western Blot analysis. The cells were sorted at different time points and the amount of Oct4 and StemCCA cassette expression was analysed. High levels of StemCCA cassette and Oct4 protein expression in Ssea1+ cells was observed compared to Thy1+ cells at day 3, 6, 9 (Polo, Anderssen et al. 2012).

In the first part of this chapter, based on the different amounts of exogenous proteins in refractory and pluripotent cells, we tested whether supplementation with addition of OKSM might be able to reprogram more cells into iPSC. To address this, we examined the importance of acquiring more of the exogenous OKSM to rescue refractory cells. Subsequently, we aimed to generate a more efficient mouse model based on our findings regarding varying levels of protein expression in the refractory and pluripotent population. Then the dynamic changes in kinetics and efficiency of the improved version of the mouse model versus M2rtTA mouse model was investigated.

In the second part of this chapter, the importance of the ratios of the different transcription factors on the efficiency and kinetics of reprogramming was investigated. We hypothesized that different amounts and ratios of the transcription factors have an effect on the reprogramming efficiency. This study was performed with three different mouse models: OKSM which express reprogramming factors in order of Oct4, Klf4, Sox2 and c-Myc (Stadtfield, Maherali et al. 2010), OSKM which reflect expression of Oct4, Sox2, Klf4 and c-Myc (Carey, Markoulaki et al. 2009) and a cross between OKSM and OSKM.

3.2. Results:

3.2.1. Morphological changes and changes in cell surface marker expression (OKSM mouse model)

3.2.1.1: Cell Morphology

First, the morphological changes following dox induction were characterized. First the cells were seeded in 6-well plate in reprogramming conditions. The cell images were taken at different time points (day 3, 6, 9, 12) using bright field phase contrast microscopy. Our result confirmed previous data and showed that at day 3 the size of reprogramming cells decreases and around day 6, early colony-like patches started to emerge (Figure 3). These cell patches continued to grow in size over time (Figure 3). Established iPSC cultures possess characteristically dome-shaped colonies, and should be mostly devoid of differentiated cells. Occasionally, additional passaging of iPS cultures may be required to remove undifferentiated/partially reprogrammed cells (Nefzger, Alaei et al. 2014). Therefore from a morphological point of view, from MEF to iPSC the cells decrease in size and then form dome-shape colonies.

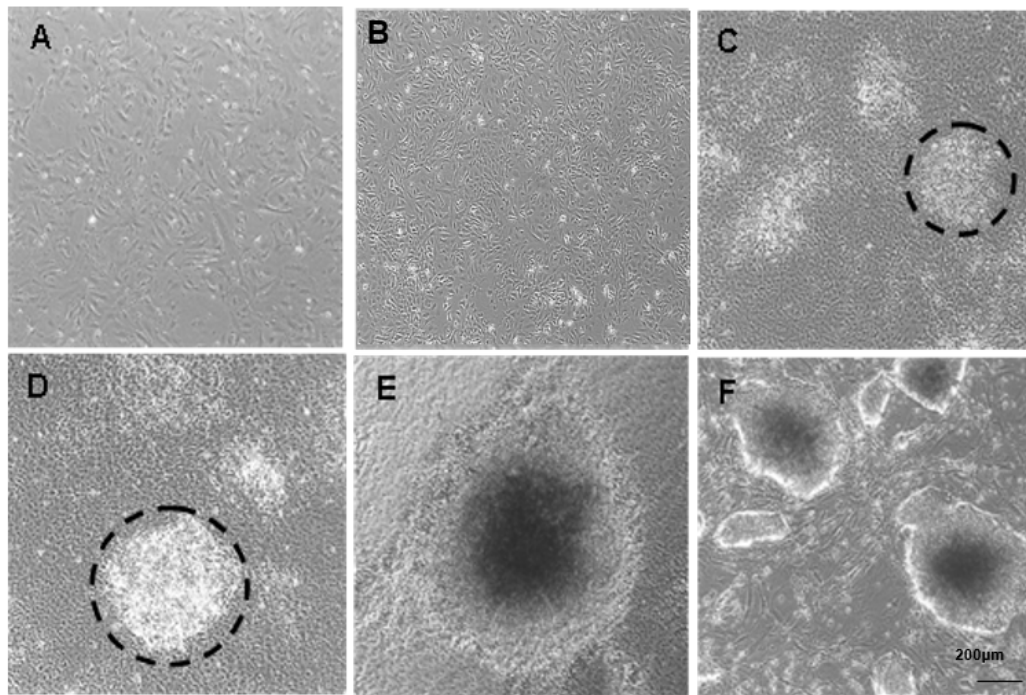
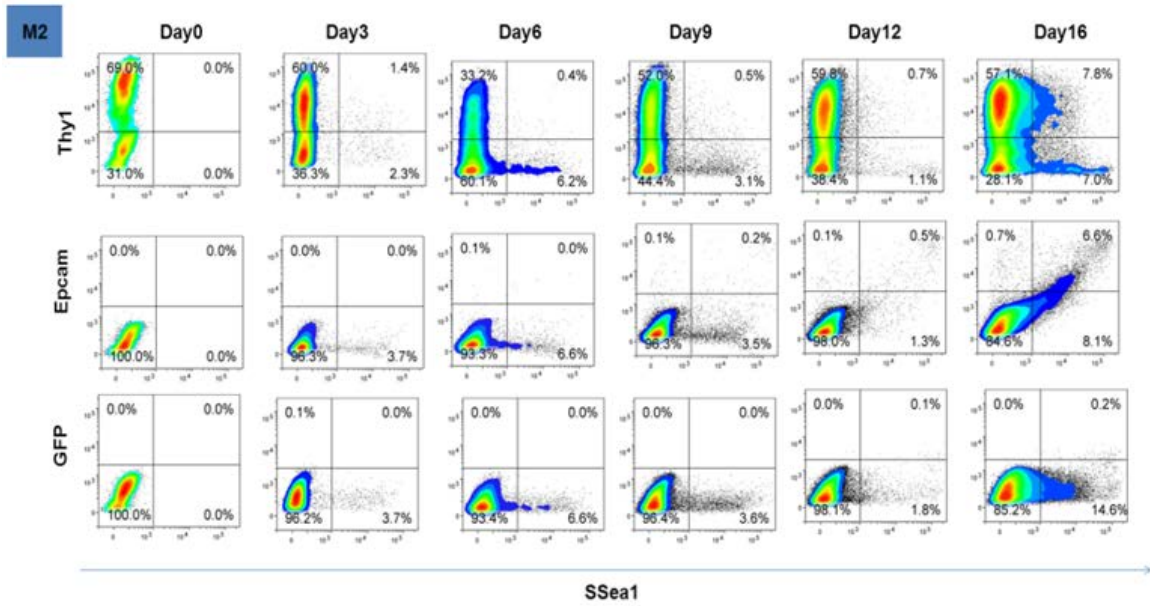


Figure 3. Cell morphology during reprogramming. MEFs with their distinct appearance (A) undergo morphological changes during reprogramming. Day 3 cultures (B) showed a decrease in cell size, however, on days 6 and 9 (C and D, respectively) patches of cells become apparent (indicated by black circles) which form proper iPS cell colonies by day 12 (E). Panel F depicts an established iPS cell culture that has been re-plated onto a feeder layer of irradiated MEFs. Scale bar: 200 μm (Nefzger, Alaei et al. 2014).

3.2.1.2. Changes in cell surface marker expression during reprogramming

The expression of a set of previously published (Stadtfeld, Maherali et al. 2008, Polo, Anderssen et al. 2012) cell surface markers, such as Thy1, Ssea1, Oct4-GFP, and Epcam, during reprogramming were investigated (Figure 4A). Previous studies have shown upregulation of Ssea1 and downregulation of Thy1 during reprogramming (Stadtfeld, Maherali et al. 2008, Polo, Anderssen et al. 2012). Here, we characterized the kinetics of our M2rtTA-OKSM mouse model in three different biological replicates. The percentages of Thy1⁺ population that were able to shift from a Thy1⁺ to an Ssea1⁺ state were then determined. The reprogramming kinetics showed the downregulation of fibroblast marker Thy1 from day 3 onward, with the pluripotent marker Ssea1 upregulated from the same time point onwards. The percentage of Ssea1⁺ cells on day 3 is about $1.8\% \pm 0.6$ (n=3) and this increases at day 6 ($7.2\% \pm 2.8$). Epcam⁺/Ssea1⁺ cells appeared on day 9, being $0.13\% \pm 0.05$ of the total population. The percentage of Epcam/Ssea1 double positive cells increased slightly on day 12 to $0.2\% \pm 0.25$, and on day 16 to $3.8\% \pm 3.8$. Oct4-GFP expression was present in $0.03\% \pm 0.05$ of the population on day 12 and increased to $0.2\% \pm 0.3$ on day 16 (Figure 4).

A



B

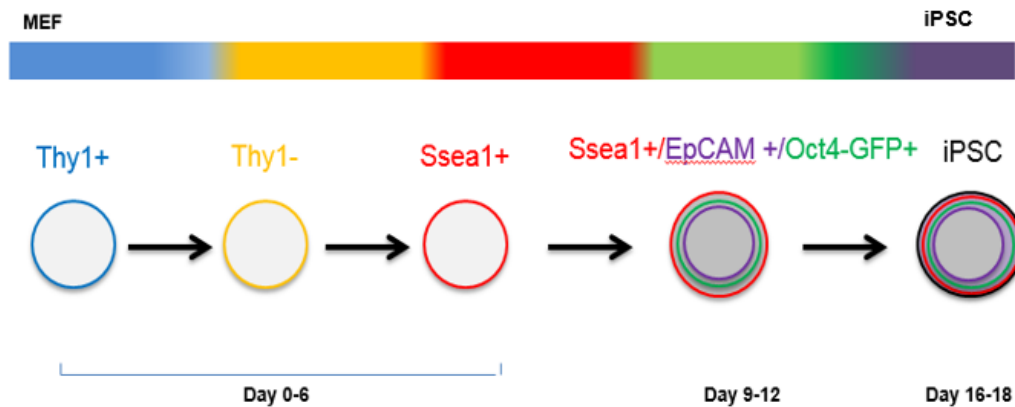


Figure 4. Representative reprogramming kinetics for the M2rtTA mouse model for Ssea1, Thy1, Epcam, and GFP at different time points. A. Representative FACS plots in M2rtTA mouse model. The reprogramming starts with downregulation of Thy1 on day 3 and up regulation of Ssea1, an early marker of reprogramming. Ssea1 was expressed on day 3, with continued expression until the end of reprogramming. Epcam was expressed from day 9 onward and the GFP reporter gene was expressed on days 12/16 (n=3). **B.** Schematic expression of reprogramming surface markers during reprogramming.

3.2.1.3. Rescue of refractory cells with addition of more transcription factors:

As mentioned previously, a lower amount of exogenous factors at the protein level in Thy1⁺ cells was observed when compared to Ssea1⁺ cells (Figure 2). In order to study the potential of each intermediate to form iPSC, the cells were harvested at different time points (day 3, 6, 9 and 12) and fluorescence-activated cell sorting (FACS) was performed to separate the different cell populations (Thy1⁺, Thy1⁻, Ssea1⁺) at the corresponding time points (Figure 6). Then the sorted cells were cultured in reprogramming conditions and the analysis of the reprogramming efficiency was performed by counting the number of colonies after alkaline phosphatase (pluripotency marker) staining (Figure 5C). These studies confirm that the Thy1⁺ cells were not able to give rise to iPSCs. Therefore, in an attempt to rescue these cells, Thy1⁺ cells at different stages of reprogramming (days 3, 6, 9, and 12) were infected with lentiviruses expressing additional copies of OKSM in one group, and with c-Myc viruses in another group. These two groups were compared with a control group where only dox was added (Figure 5B). Interestingly, Thy1⁺ cells with extra copies of OKSM were able to form iPS colonies, but not the c-Myc infected cells or the untreated control cells. Hence the addition of more exogenous c-Myc is not sufficient to allow colonies to form from refractory cells. This demonstrates that overexpression of the four factors can rescue refractory cells and suggests that the inability to sustain OKSM protein expression in Thy1⁺ cells on or after day 3 broadly contributed to their failure to form iPSCs (Figure 5 A).

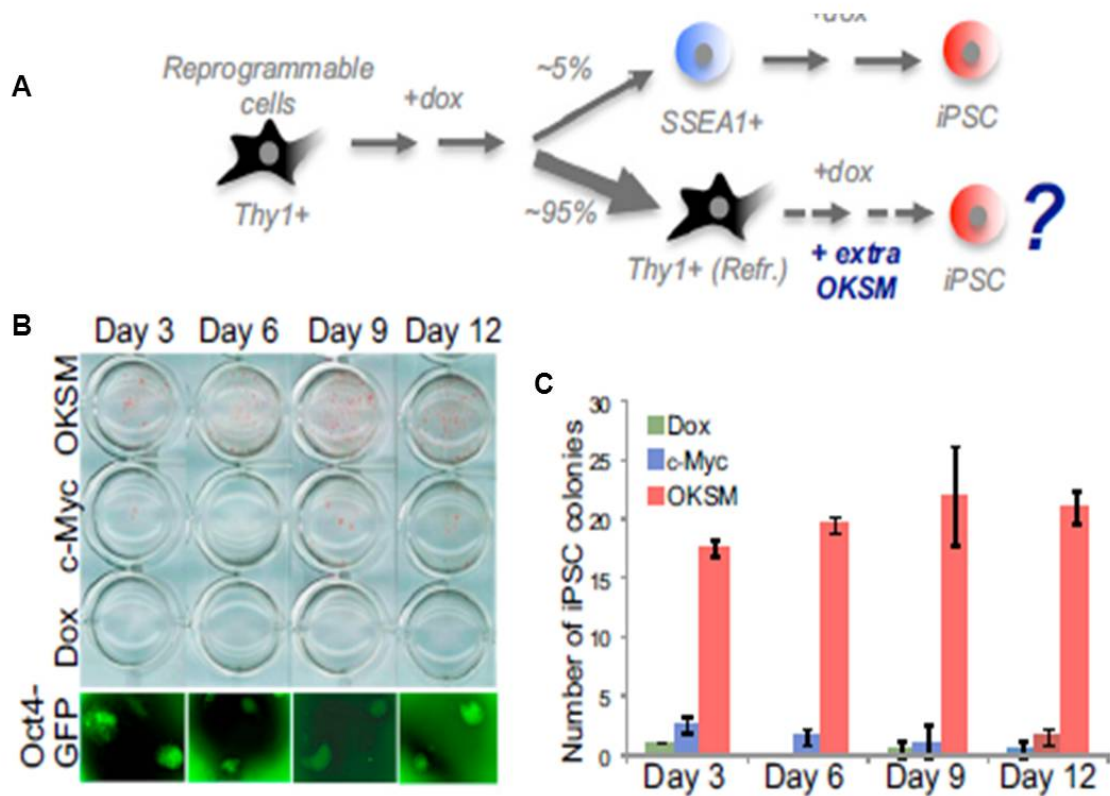


Figure 5. Rescue of *Thy1*⁺ cells with infection of OKSM virus at different time points. A. Schematic figure: *Thy1*⁺ cells are able to generate *iPSC* with the addition of extra OKSM. **B.** Transfection of *Thy1*⁺ cells with OKSM and c-Myc. **C.** Number of *iPSC* colonies in the *Thy1*⁺ population after infection with OKSM and c-Myc during reprogramming (Polo, Anderssen et al. 2012).

3.2.2. Generation of an improved mouse model with higher levels of OKSM expression

Recently a genetically modified mouse strain was generated with strong expression of the M3rtTA gene, which is a more potent version of the M2rtTA gene, providing consistent induction of tetracycline-responsive element (TRE)-controlled transgene. Briefly, the tet inducible system consists of two essential components: a tetracycline-responsive element (TRE), and a reverse tet-transactivator protein rtTA. In this system M3rtTA leads to a high level of transgene expression (Dow, Nasr et al. 2014). Therefore we hypothesised that the M3rtTA reprogramming mouse might generate high levels of OKSM expression. For the purposes of this investigation, the M3rtTA mouse was crossed with the OKSM reprogramming mouse to increase the level of OKSM expression. This provides an alternative mouse model for investigating the mechanism underlying somatic cell reprogramming with generation of high amounts of intermediates.

To generate heterozygous MEF for three loci (OKSM, Oct4-GFP reporter and M3rtTA), a cross between mice homozygous for Oct4 and heterozygous (het) for M3rtTA, and mice homozygous (homo) for OKSM, was performed. This resulted in half of the offspring carrying a het/het/het genotype for all three loci (Figure 6).

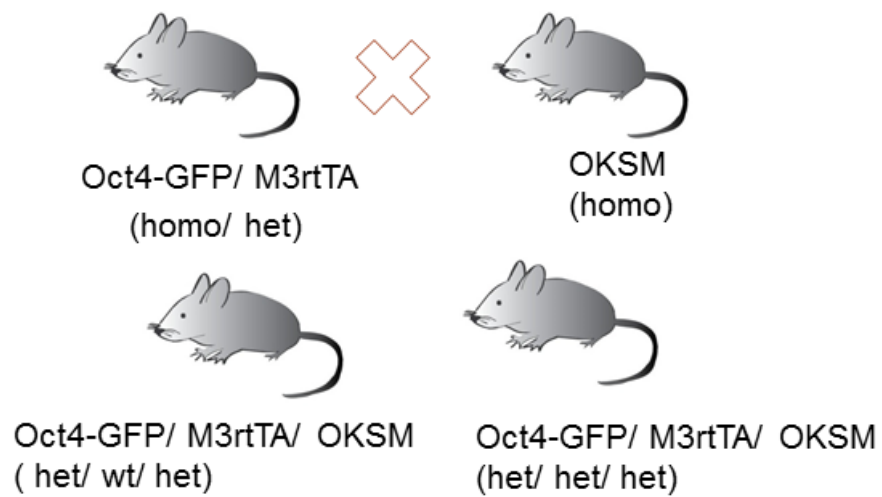


Figure 6. Schematic depicting generation of the Oct4-GFP/M3rtTA/OKSM mouse. A cross between Oct4-GFP/ M3rtTA mouse and OKSM mouse results in the generation of het/het/het offspring.

3.2.2.1. The amount of Oct4 and Sox2 expression in M2 versus M3 mouse model

Based on our finding that a high copy number of OKSM can lead to Thy1⁺ cells forming iPSC cells, we hypothesized that a secondary mouse model with higher levels of reprogramming factor expression might be able to reprogram more efficiently. First, to find out which is the more efficient mouse model for the OKSM system, we measured the level of the four factors at the protein level in the M2rtTA and M3rtTA mouse models. To achieve this, the cells of these two different mouse models were cultured and reprogramming was started by dox addition. The cells were sorted on day 6 to isolate Thy1⁺ cells and Ssea1⁺ cells in the M2rtTA and M3rtTA mouse models, and the amount of Oct4 and Sox2 protein was determined by Western Blot. High amounts of Oct4 and Sox2 protein were observed in the Ssea1⁺ population on day 6 in cells containing M3rtTA, whereas there was less Oct4 and Sox2 protein in the Ssea1⁺ population in cells carrying M2rtTA (Figure 7A). The level of Oct4 and Sox2 in Ssea1⁺ cells in the M3rtTA model on day 6 was approximately 3 times more compared to Ssea1⁺ cells in the M2rtTA model on day 6. However, the amount of Oct4 and Sox2 protein was similar in the Thy1⁺ population in M2rtTA and M3rtTA mouse strains on day 6 (Figure 7B, C). Since overexpression of Oct4 and Sox2 was observed in the M3rtTA mouse, the kinetics and efficiency of reprogramming in the M3rtTA model was then examined, in order to better understand the difference between the M2rtTA and M3rtTA mouse models.

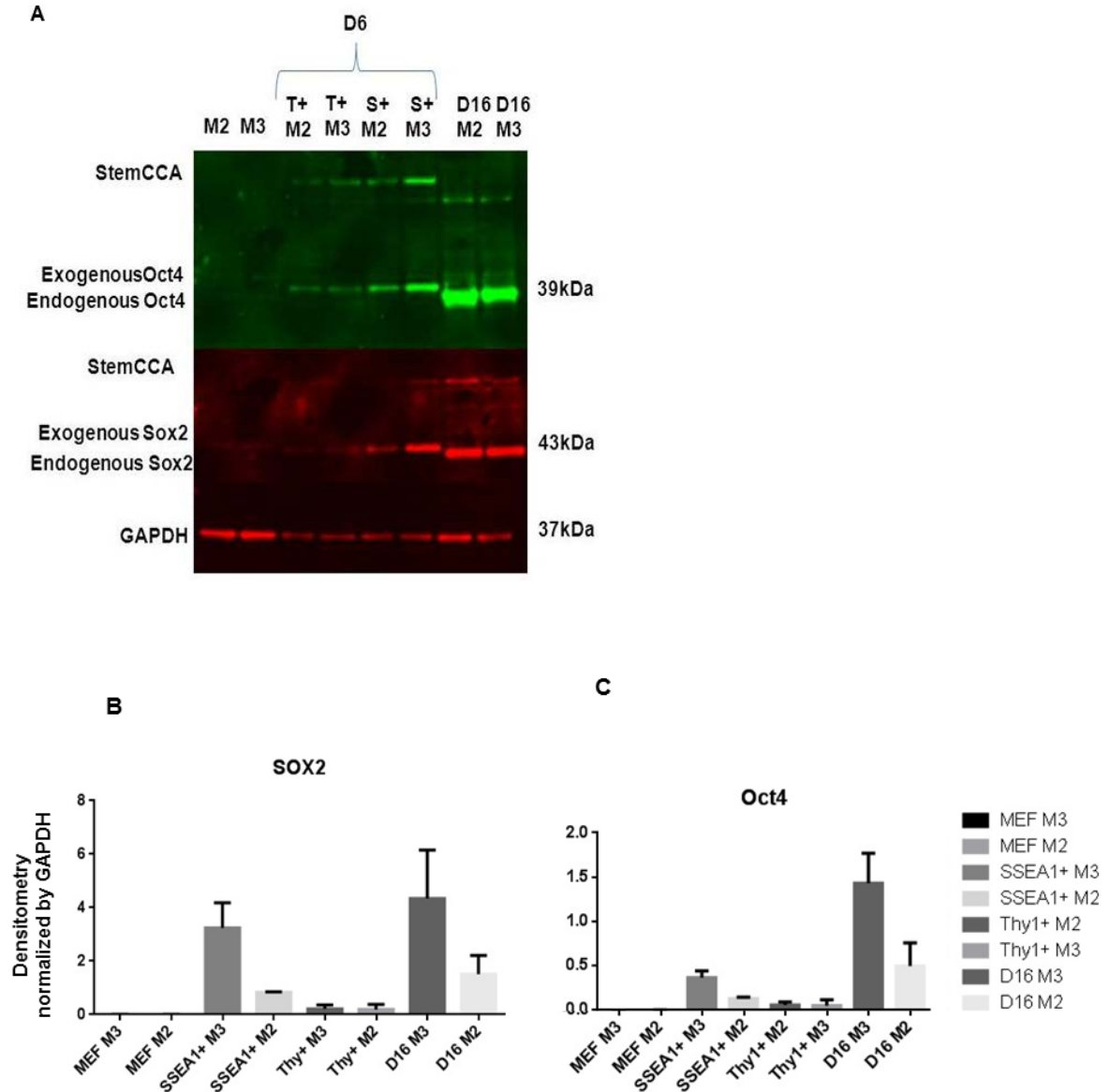


Figure 7. Quantification of the amount of Oct4 and Sox2 in MEFs, on day 6 and day 16 of reprogramming. **A)** Western Blot for Oct4 and Sox2 and GAPDH for the indicated cell populations. The Thy1⁺ and Ssea1⁺ cells from both M2rtTA and M3rtTA were sorted on day 6 and Western Blot was performed for MEFs in both models (was shown as M2, M3) and Thy1⁺ and Ssea1⁺ cells on day 6 for both M2 and M3rtTA (was shown as T+ M2, T+ M3, S+ M2 and S+ M3) and for whole population of iPSC at day 16 for both models. The higher molecular weight band for exogenous Oct4 compared with endogenous Oct4 (iPSC) reflects unprocessed protein originating from the polycistronic construct. **B)** Sox2 densitometric quantification of Western Blot analysis **C)** Oct4 densitometric quantification of Western Blot analysis (n=2).

3.2.2.2. M3rtTA followed similar kinetics as M2rtTA

To further understand the subtle differences between these two models, the kinetics of reprogramming were investigated and a time course experiment was performed. To do this, MEFs of both models were seeded at the same density, cultured in ES media and after 24 hours reprogramming was started with dox addition. Then the cells were collected at different time points (day 3, 6, 9, 12) and FACS performed based on Ssea1, Thy1, Epcam and GFP expression. Analysis of surface marker expression showed that reprogrammable MEFs containing M2rtTA or M3rtTA followed the same kinetics (Figure 8), but importantly a higher number of cells were able to generate Ssea1⁺ cells in the M3rtTA mouse model, compared to M2rtTA.

The percentage of Ssea1⁺ cells in the M3rtTA model was significantly increased from day 3 onwards from $8.3\% \pm 0.5$ to $31.8\% \pm 7.3$ on day 9 (Figure 8) whereas in M2rtTA the percentage of Ssea1⁺ cells changed from $1.8\% \pm 0.6$ on day 3 to $3.3\% \pm 0.9$ on day 9. Epcam⁺/Ssea1⁺ cells appeared at day 6 in the M3rtTA model compared to M2rtTA where they appeared later on day 9. The percentage of Ssea1⁺/Epcam⁺ for M3rtTA was $0.5\% \pm 0.2$, $4.6\% \pm 1.45$, $7.3\% \pm 2.82$ and $8.9\% \pm 0.6$ on days 6, 9, 12 and 16 respectively. The percentage of Ssea1⁺/Epcam⁺ for M2rtTA increased from $0.13\% \pm 0.05$ on day 9 to 0.5% and $3.8\% \pm 3.8$ on day 12 and day 16 respectively. GFP expression was observed in $0.3\% \pm 0.1$ of M3rtTA cells on the last day of reprogramming, similar to M2rtTA cells. These results showed that both mouse strains followed similar reprogramming kinetics by downregulation of Thy1 and upregulation of Ssea1 during reprogramming and expression of Epcam at later stage of reprogramming. However in the M3rtTA mouse model a higher number of the Thy1⁺ cells are able to generate the Ssea1⁺ population, and consequently form iPSC, compared to M2 (Figure 8, 9).

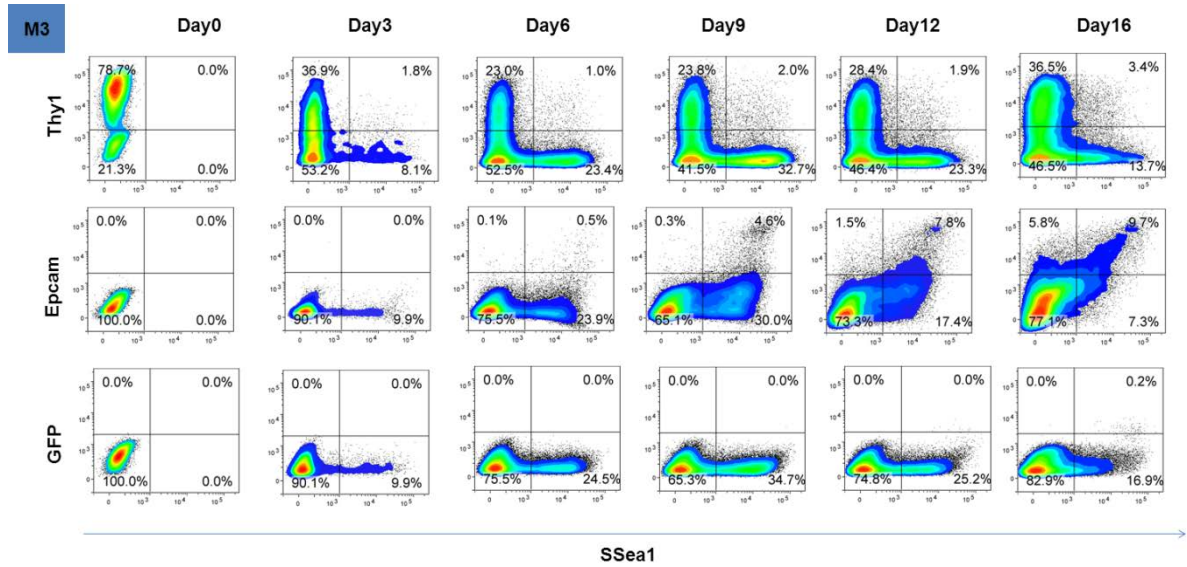


Figure 8. Representative FACS plots for M3rtTA. The expression of surface markers such as Thy1, Ssea1, Epcam, and Oct4-GFP during reprogramming in the M3rtTA mouse model. Our results show downregulation of Thy1 on day 3 and upregulation of Ssea1 followed by expression of Epcam and GFP reporter from day 9 and day 12/16 respectively (n=3).

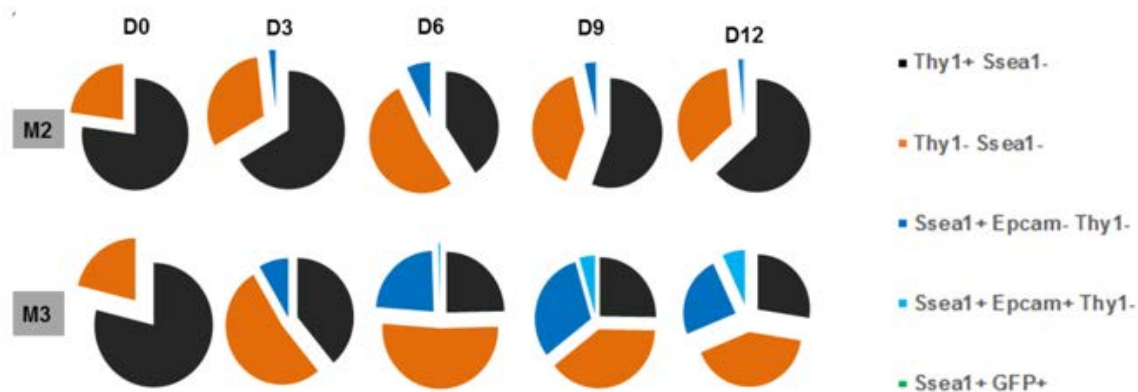


Figure 9. Quantification and comparison of the different cell populations during reprogramming in the M2rtTA and M3rtTA mouse models. A higher number of Thy1+ cells moved to the Ssea1+ state in the M3rtTA mouse model.

3.2.2.3. Higher reprogramming efficiency in M3rtTA compared to M2rtTA

Based on the changes in kinetics that showed high percentage of Ssea1+ cells in M3rtTA compared to M2rtTA, we hypothesized the reprogramming efficiency of M3rtTA is higher than M2rtTA MEF. To address this, 20,000 cells of M2rtTA and M3rtTA were seeded in 6-well plates, and alkaline phosphatase activity was measured after 15 days of dox exposure and 4 days of dox withdrawal. The results of reprogramming efficiency for OKSM/M3rtTA MEFs showed a significantly higher number of colonies generated when compared to the OKSM/M2-rtTA strain (Figure 10A). Quantification of colonies for the M2rtTA mouse showed 8 ± 2.8 colonies compared to 110 ± 24.5 colonies for M3rtTA (Figure 10B). These results demonstrate the higher reprogramming efficiency of the M3rtTA mouse model compared to M2rtTA.

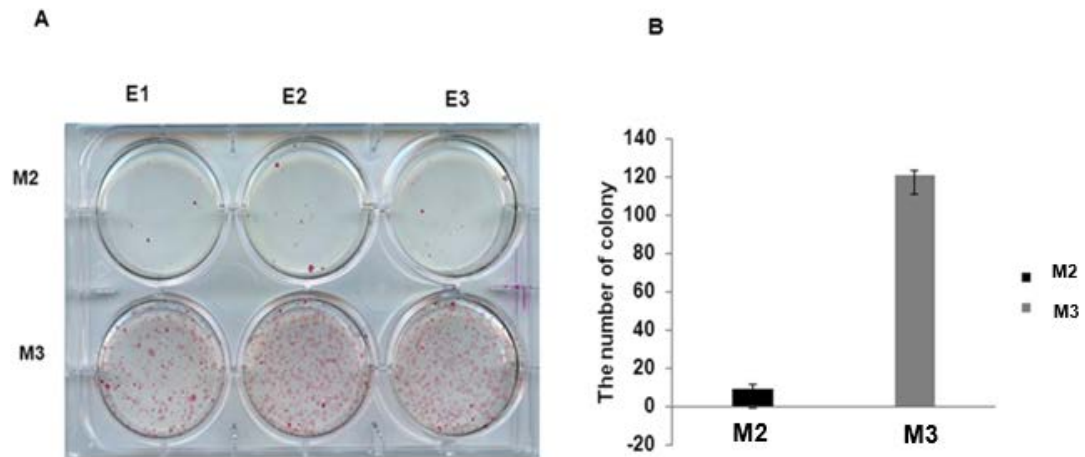


Figure 10. Measuring reprogramming efficiency with alkaline phosphatase (AP) assay in M2rtTA (M2) and M3rtTA (M3) mouse models in three different embryos (E1, E2, E3). A. Alkaline phosphatase positive colonies in M2rtTA vs. M3rtTA. **B.** Quantification of AP positive colonies in M2rtTA and M3rtTA cells (n=3). The number of colonies in M3rtTA was 12-fold more than M2rtTA.

3.2.2.4. Assessment of the minimum time requirement for reprogramming cells from both models to become independent of doxycycline:

To assess the difference in time required for reprogramming cells to become independent of doxycycline in M3rtTA and M2rtTA, 20,000 cells were cultured into 6-well plates and reprogramming was started with dox addition. Then at individual time points (day 3, 6, 9 and 12), the media was changed to ES media without doxycycline. The ‘point of no return’ of each time point was investigated by counting the alkaline phosphatase-positive colonies 5 days later after dox withdrawal. The result showed that dox addition until day 9 was essential in both M2rtTA and M3rtTA. After day 9 they were both able to generate colonies independent of dox, however the number of colonies was much higher in M3rtTA compared to M2rtTA (Figure 11). Therefore, despite of high number of iPSC colonies and high percentage of Ssea1+ cells in M3rtTA vs M2rtTA, both mouse models (M2rtTA and M3rtTA) become independent of dox at the same time.

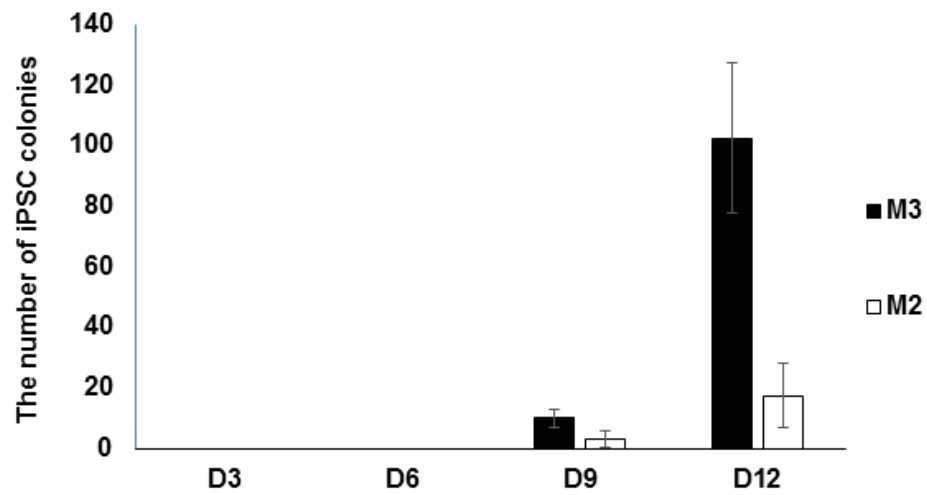


Figure 11. Determination of the ‘point of no return’ during reprogramming. For both M3rtTA (M3) and M2rtTA (M2) mouse models, dox was withdrawn either on day 3, 6, 9 or 12 and colonies counted (AP staining) on day 16.

3.2.3. Variation in the stoichiometry of reprogramming factors leads to changes in reprogramming

The first part of this chapter demonstrated that the amount of four factor expression is important for reprogramming efficiency, but the question remained whether the ratio of reprogramming factors also plays an important role. We will examine and compare in detail which mouse model is the most efficient for reprogramming, and determine how the kinetics vary between different reprogramming systems. To address this issue, we isolated MEFs from embryos of three different mouse models (OKSM, OSKM, a hybrid of OKSM/OSKM) with the same genotype. The mouse genotype was homo-homo for both loci (Collagen1a with four factor cassette(s) and Rosa26 locus with M2rtTA).

OKSM mouse:

As described previously, the OKSM polycistronic cassette is located in the Collagen1a locus and an internal ribosomal entry site (IRES) is located between Klf4 and Sox2. Oct4 and Klf4 are connected by a self-cleaving peptide F2A (derived from the foot and mouth disease virus) and E2A (from equine rhinitis A virus) is placed between Sox2 and c-Myc and this cassette is under control of the Tet O promoter. Additionally, an optimized form of the reverse tetracycline controlled transactivator (M2rtTA), a B-globin intron, PolyA signal and a PGK-puromycine cassette, are all inserted at the Rosa 26 locus. Upon dox addition, the ubiquitously expressed M2rtTA binds to the Tet O promoter, initiating the transcription of the four factors OKSM (Stadtfield, Maherali et al. 2010) (Figure 1).

OSKM mouse:

As with the OKSM model, this mouse also carries the M2rtTA gene in the Rosa 26 locus and a copy of the four transcription factors (Oct4, Sox2, Klf4, c-Myc, in that order) in an expression cassette, called 4F2A (Carey, Markoulaki et al. 2009). However there are

differences in OKSM and OSKM mouse models. In the OSKM model, each factor is separated by three 2A self-cleavage peptides. P2A is located between Oct4 and Sox2, T2A is located between Sox2 and Klf4, and E2A separates Klf4 and c-Myc. The 4F2A cassette is also located in the Colla locus (Figure 12).

There is an abundance of literature regarding the M2rtTA OKSM and OSKM systems (Stadtfield, Maherali et al. 2008, Wernig, Lengner et al. 2008, Carey, Markoulaki et al. 2009, Stadtfield, Maherali et al. 2010, Polo, Anderssen et al. 2012), but little is known about the comparison between different mouse systems in the context of morphology, reprogramming factor ratio and reprogramming efficiencies.

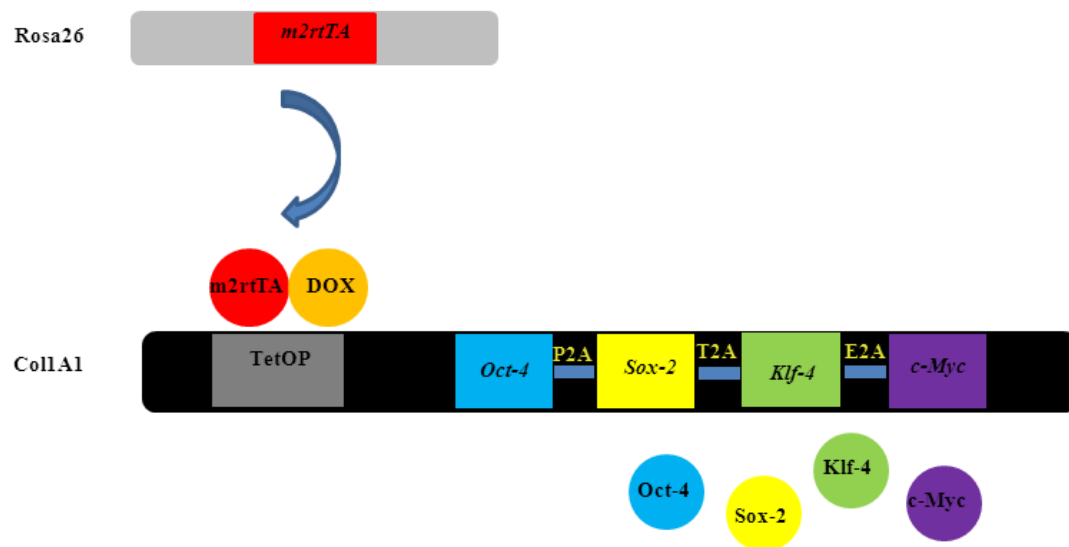


Figure 12. Schematic representation of the OSKM mouse model. The difference in OSKM and OKSM (Figure 1) is the order of the transcription factors in the cassette. In the OSKM mouse model the order is Oct4, Sox2, Klf4 and c-Myc. When rtTA bind to the Tet O promoter, all four transcription factors can be transcribed.

3.2.3.1. Cell morphology

To observe the differences among the three mouse models (OKSM, OSKM, OSKM/OKSM) in terms of cell morphology, the cells were seeded in 6-well plate and images were taken using bright field phase contrast microscopy. The cell morphology shows the difference between the three groups during reprogramming. The OSKM cells start to decrease in size at a later time point compared to OKSM and OSKM/OKSM mouse model (Figure 13). The early colony-like patches were only formed on day 9 in the OSKM model, whereas this started earlier (on day 6) in both the OKSM and OSKM/OKSM models. Moreover, the final iPSC colonies were bigger in the OKSM model compared to the OSKM mouse model, however the size of OSKM/OKSM iPSC colonies was in between the size of OSKM and OKSM model colonies (Figure 13).

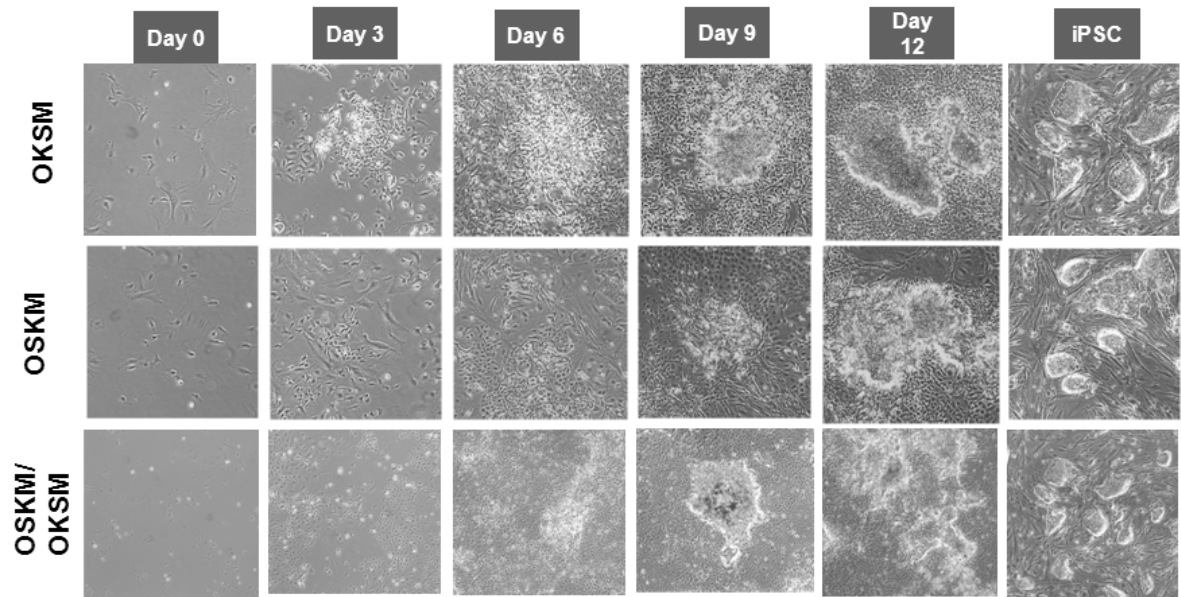


Figure 13. Morphological changes in OKSM, OSKM, and OKSM/OSKM mouse models during reprogramming. The bright field phase contrast microscopy in three different mouse models at different time courses (Day 0, 3, 6, 9, 12 and iPSC). Cells have started to decrease in size and form patches around day 6 in OKSM cells, however patch formation occurs on day 9 for OSKM and OSKM/OKSM cells. OKSM iPSC colonies were bigger than colonies derived from OSKM cells.

3.2.3.2. The ratio/stoichiometry of reprogramming factors in the three different mouse models

As overexpression of reprogramming factors led to different efficiency in M2rtTA and M3rtTA mouse models (OKSM cassette), we hypothesized that the changes in the stoichiometry of reprogramming factors might lead to changes in reprogramming efficiency in the three mouse models. Therefore in order to determine the relative amount of the different transcription factors in each mouse model, Western Blot analysis was performed for Oct4, Klf4, Sox2 and c-Myc on day 3. As shown in figure 14, the amount of Klf4 for the OKSM mouse model was significantly lower compared to other transcription factors in this mouse model. Comparatively, the amount of Sox2 and c-Myc was less than the amount of Oct4 and Klf4 in the OKSM model. The OKSM/OSKM mouse model showed approximately equal expression ratios of four transcription factors (Figure 14). This result showed that there are different ratios of the four transcription factors in the three different mouse models (OKSM, OKSM/OSKM, OSKM).

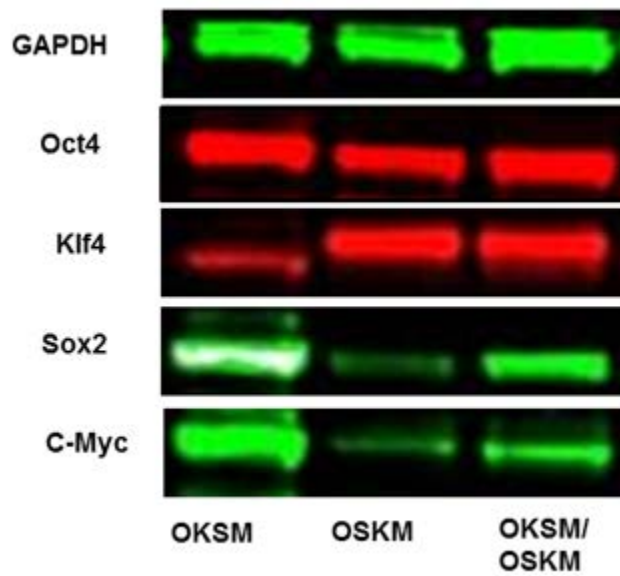


Figure 14. Western Blot analysis for Oct4, Sox2, Klf4, c- Myc and GAPDH for the OKSM, OSKM and OKSM/OSKM mouse models at day3 of reprogramming. In three different mouse models, the cells were collected on day 3 and Western Blot was performed for four transcription factors (Oct4, Sox2, Klf4 and c-Myc). Different ratios of the four transcription factors were observed in the three different mouse models.

3.2.3.3. Different reprogramming efficiencies in OKSM, OSKM and OSKM/OSKM mouse models

In order to determine the efficiency and kinetics of reprogramming, we examined the same genotype for three different mouse models (homo rtTA-homo in reprogramming cassette). The same number of cells (20,000 cells) from each model were seeded into a 6-well plate. After 15 days dox treatment and then 3 days dox withdrawal, these cells were then stained for alkaline phosphatase (AP) activity. The three mouse systems showed different reprogramming efficiencies. The number of AP-stained colonies in the OKSM models was higher compared to OSKM only and the hybrid of OSKM/OKSM model, and it was 6 times more compared to the OSKM model and 2 times more compared to a hybrid of OSKM/OKSM. Morphologically, the size of the AP stained colonies was also bigger in the OKSM model (Figure 15).

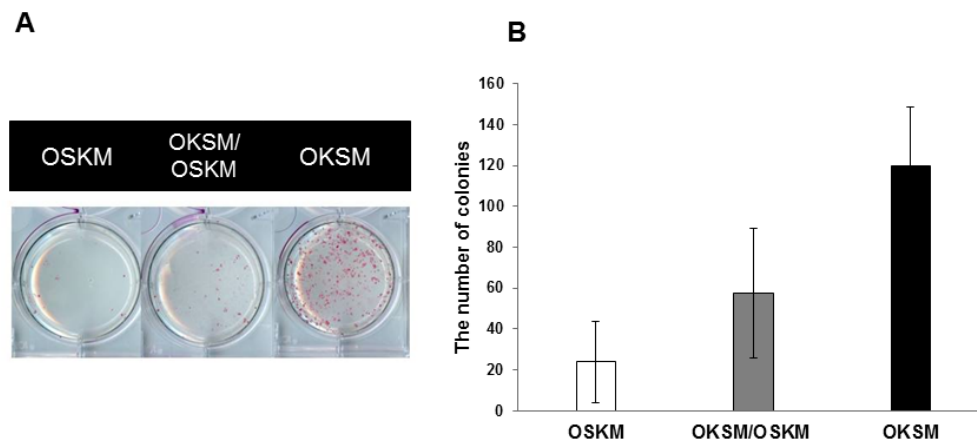


Figure 15. Reprogramming efficiency in three different mouse models. After the reprogramming process cultures from the three models were submitted to (A) AP staining followed by (B) quantification of AP positive colonies. The highest number of alkaline phosphatase stained colonies was observed in OKSM cells.

3.2.3.4. Different reprogramming kinetics were found for the three mouse models

To understand the differences in kinetics in the three mouse models, time course reprogramming experiments were performed. We observed that during reprogramming the percentage of Ssea1⁺ cells in the OKSM model was much higher than for either OSKM/OKSM or OSKM mouse models (see figure 16, representative FACS blots). The percentage of Ssea1⁺/Epcam⁻ population in hybrid of OKSM/OSKM showed a similar pattern however with low expression of Ssea1 compared to the OKSM model. It showed $4.2\% \pm 2.1$ on day 3, $17.3\% \pm 2.2$ on day 6 and $2.1\% \pm 0.8$ of cells were Ssea1⁺/Epcam⁻ on day 9. This population was highly reduced in the OSKM mouse strain with $0.2\% \pm 0.1$, $0.5\% \pm 0.2$ and $0.6\% \pm 0.4$ on day 3, 6, 9 respectively. We also examined the changes in Epcam expression in three mouse strains. The expression of Epcam⁺/Ssea1⁺ cells in the OKSM mouse model showed $1.3\% \pm 0.2$, $7.4\% \pm 1.2$ and $30.5\% \pm 5.2$ positive cells on days 3, 6 and 9 respectively, however in the OKSM/OSKM mouse model the numbers were found to be $0.2\% \pm 0.2$ on day 3, $10.2\% \pm 3.2$ on day 6 and $26.4\% \pm 4.5$ on day 9. In the OSKM mouse model, the Epcam⁺/Ssea1⁺ cells can be detected on day 6 at $0.3\% \pm 0.1$ increasing to $0.3\% \pm 0.2$ by day 9 (Figure 16). Interestingly we also observed that the order of surface marker expression was different in these three mouse models. Epcam as one of the late pluripotency markers was expressed prior to Ssea1 in both the OSKM and the OKSM/OSKM models, however this order was reversed in the OKSM mouse model with Ssea1 being expressed prior to Epcam (Figure 16).

On day 16 the percentage of GFP⁺/Ssea1⁺ was increased in the OKSM model with $2.2\% \pm 0.3$ compared to only $0.5\% \pm 0.3$ and $0.7\% \pm 0.1$ in OKSM/OSKM and OSKM models respectively. Overall these results showed different kinetics on the cell surface during reprogramming in three different mouse systems (Figure 17).

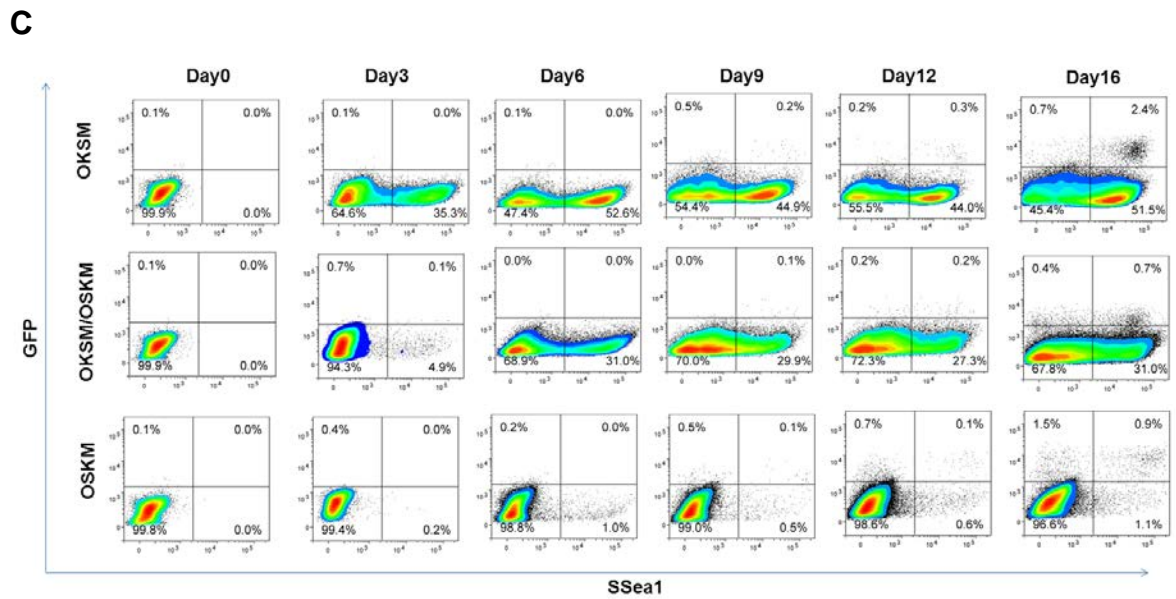
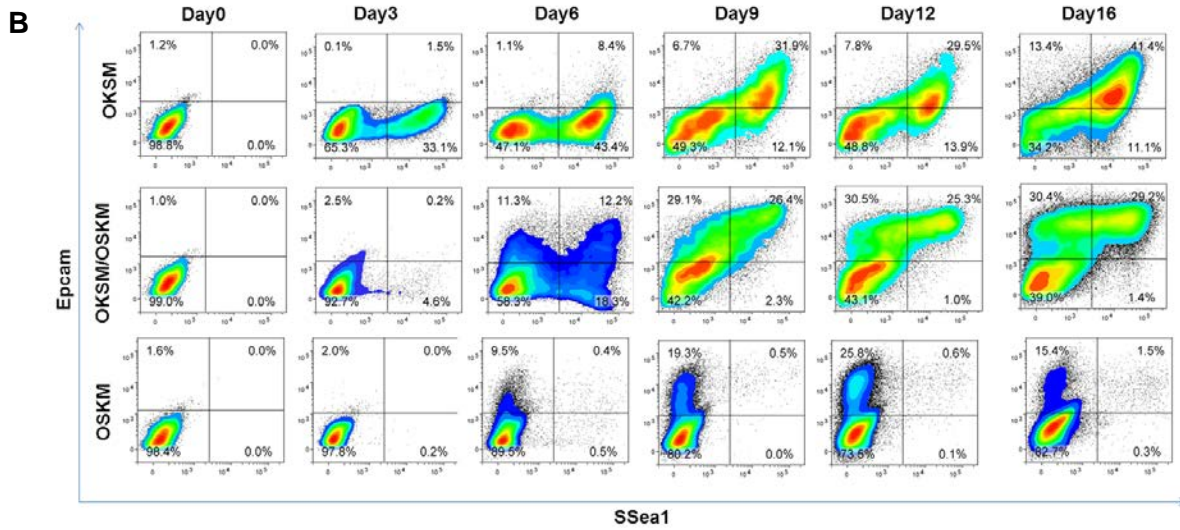
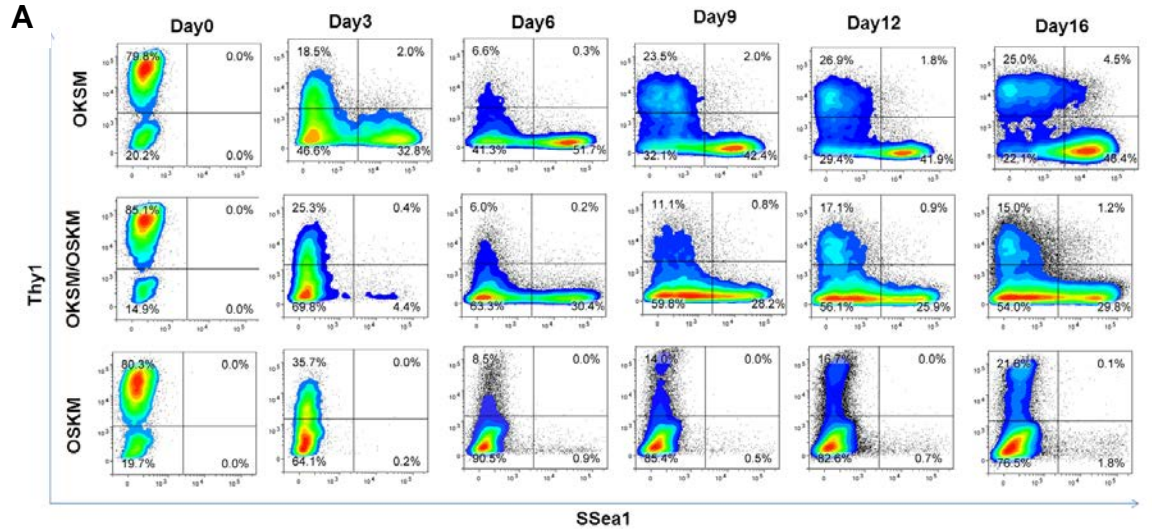


Figure 16. Representative FACS plot for OKSM, OKSM/OSKM and OSKM mouse models during reprogramming. Expression of Thy1 (A), Ssea1 (B), Epcam (C), GFP (D) at different time points in three different mouse models.

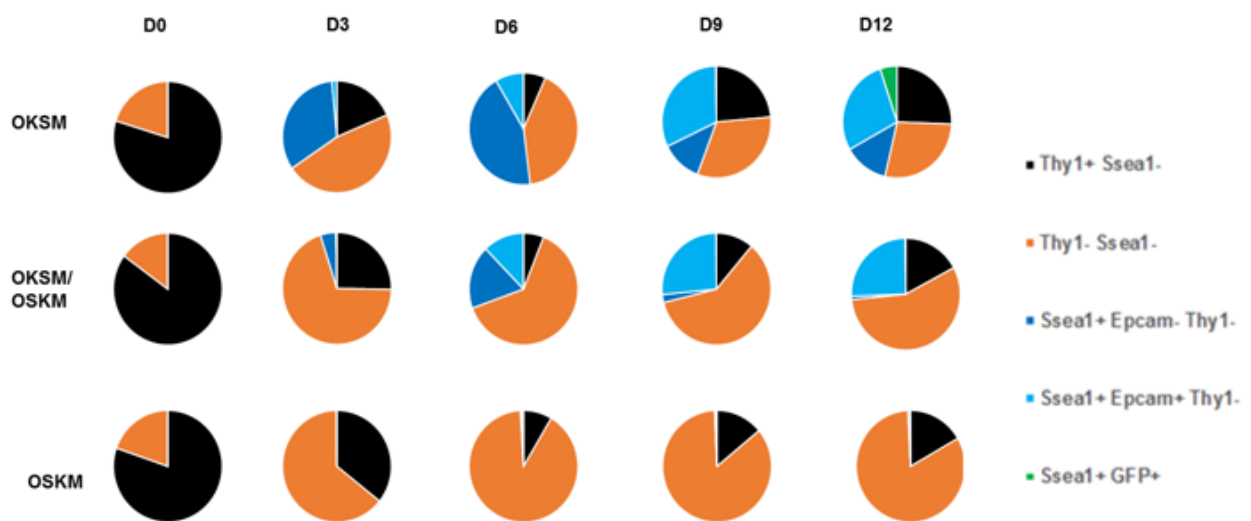


Figure 17. Quantification and comparison between the different cell populations during reprogramming in the different mouse models (OKSM, OKSM/OSKM, OSKM). The quantification was performed based on the percentage of each population in FACS plots.

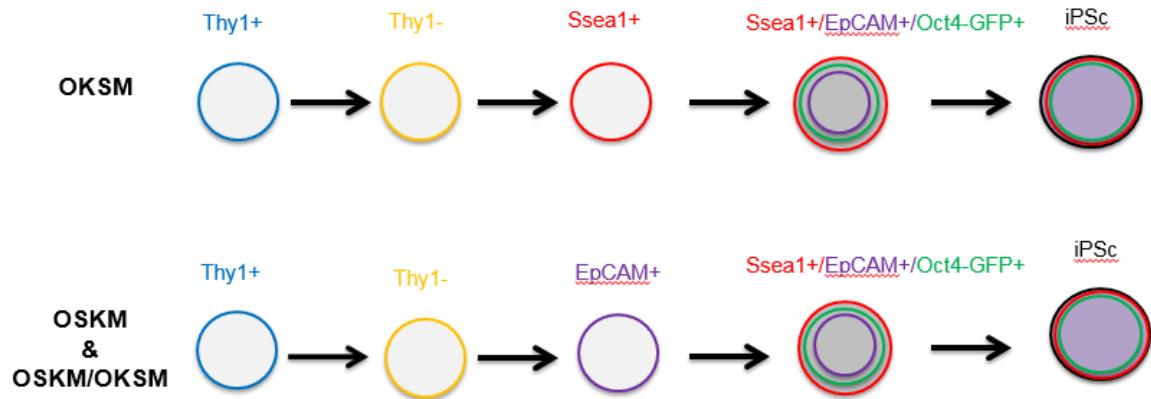


Figure 18. Epcam expression during reprogramming in three mouse models. Epcam + populations appear in OSKM and OKSM/OSKM prior to Ssea1+ cells. (Modified from (Nefzger, Alaei et al. 2014))

3.3. Conclusion:

Overall, the results presented in this chapter demonstrate that the amounts and the ratio of the transcription factors have an effect on the efficiency of iPS cell generation. With viral transduction we expressed more OKSM factors in refractory Thy1⁺ cells of the OKSM/M2rtTA mouse model, which allowed these cells to form iPS cells. Based on this idea that over expression of four transcription factors rescued the refractory cells, we generated an improved version of the OKSM mouse model where we substitute the M2rtTA with the more active M3rtTA, resulting in the expression of higher levels of Oct4 and Sox2 and higher reprogramming efficiency. Nevertheless, the increase in OKSM expression did not change reprogramming kinetics, and both MEFs of M2rtTA and M3rtTA mouse strains followed the same reprogramming pathway. Our investigation of three different mouse strains (OKSM, OSKM/OKSM and OSKM, in all cases with M2rtTA) showed the effects of different reprogramming factor ratios. The results suggested that not only does the difference in the amount of reprogramming factors observed in M2rtTA and M3rtTA models change reprogramming efficiency, but they also show that the changes in the ratio of reprogramming factors in the three different mouse models also has an effect on reprogramming efficiency. Furthermore, the stoichiometry of transcription factors had a significant impact on reprogramming kinetics. The exact relation between each transcription factor expression and reprogramming efficiency is still unknown. Altogether the changes in the ratio of transcription factors effects on expression of surface markers sequence and consequently reprogramming efficiency. It is hoped that this new reprogramming mouse will provide a higher number of reprogramming intermediates in order to study various molecular aspects of the reprogramming system, such as transcription factor binding sites and DNA methylation.

CHAPTER 4

Mapping of Oct4 and Sox2 during reprogramming

4.1. Introduction

The pluripotent stem cell state is under the control of a transcriptional circuitry that includes the reprogramming factors OKSM. The conversion from somatic cells to pluripotent cells takes two weeks, and it is still unclear how the four TFs govern reprogramming and bind to their targets. During embryonic development some TFs bind to heterochromatin, and direct the accessibility of other TFs to their targets (Cirillo, Lin et al. 2002). In 2012, the Zaret laboratory had studied the binding of four factors during 48 hours (hrs) in human reprogramming (Soufi, Donahue et al. 2012). In this study, Oct4, Sox2 and Klf4 bound to closed chromatin and c-Myc played a facilitator role to help Oct4, Sox2, and Klf4 bind to chromatin. The combination of four factors bound to the distal elements, however combination of Klf4 and c-Myc bind to promoter regions. After 48 hrs of reprogramming induction, most OSKM co-bound complexes moved from the distal regions closer to the TSS, with the exception of bound Sox2 that remained distal. It was found that the four factors bound to related enhancer sites 48 hrs after induction of reprogramming (Soufi, Donahue et al. 2012). Analysis during 48 hrs of reprogramming also indicated that Oct4, Sox2, Klf4 and c-Myc occupy the human genome in different patterns when compared to the late stage of reprogramming of mouse fibroblast (pre-iPSC), iPSC and ES cells. A major difference during the early phases of OKSM binding and pre-iPSC cells is that Oct4, Sox2 and Klf4 bind together with c-Myc to the distal regions of heterochromatin; however in

pre-iPS and ES cells, c-Myc is involved in a different network of target genes. Moreover, the binding of each TF separately 48 hrs of reprogramming showed that c-Myc bind to open chromatin, whereas other factors alone bind to closed chromatin. Therefore, apart from c-Myc, the other transcription factors including Oct4, Sox2 and Klf4 bind to one side of the DNA helix, and the other side of DNA helix is left free to bind to histone octamers or other chromatin complexes (Soufi, Donahue et al. 2012). Soufi *et al*, also demonstrated that the minority of c-Myc and Klf4 binding sites were enriched for active histone marks such as H3K4me2, H3K4me3, H3K27ac and H3K9ac.

Overall, the results from Soufi and colleagues showed that c-Myc does not play a role in the occupancy of Oct4, Sox2 and Klf4 to the OSK only sites. However, it did increase occupancy of Oct4, Sox2 and Klf4 in OSKM binding sites, indicating supportive binding of c-Myc with OSK. Promoters of active genes in fibroblasts were bound by c-Myc and Klf4, and gene expression changed over 48 hrs. In the first two days of reprogramming, the open histone modifications had minimal effects on the binding of the four factors to distal elements (Soufi, Donahue et al. 2012).

We had previously studied gene expression and histone modifications on day 3, 6, 9 and 12 of reprogramming, and analysis of H3K4me3 and H3K27me3 showed two waves of changes which were consistent with mRNA transcriptional phases (Polo, Anderssen et al. 2012). We also showed the binding pattern of Oct4 and Sox2 at three different targets during reprogramming. The results demonstrated that Fut9, Fucosyltransferase 9, was enriched for Oct4 and Sox2 during the early days of reprogramming, whereas Nanog and Lefty1 (both late pluripotent markers) were enriched for Oct4 alone during the first days of reprogramming, and these targets were occupied by both Oct4 and Sox2 from day 12 onwards (Polo, Anderssen et al. 2012). Although these studies investigated changes in the binding of reprogramming factors during reprogramming specially in early days of

reprogramming (Polo, Anderssen et al. 2012, Soufi, Donahue et al. 2012), it is still unclear what changes in TF binding happen during two weeks of reprogramming, and which transcription factors affect pluripotency-associated gene expression during reprogramming.

In this chapter, to further understanding of how TFs govern reprogramming, the binding patterns of two transcription factors to their targets were investigated: Oct4 and Sox2. First, ChIP for Oct4 and Sox2 will be optimized by testing a range of variables, including sonication cycle, amount of antibody, the type of precipitation beads, and time of crosslinking to obtain maximum enrichment of Oct4 and Sox2 targets. Second, the reprogramming population was purified by AutoMACS to enrich for Ssea1+ and capture the maximum number of reprogramming intermediates. Third, the binding sites of Oct4 and Sox2 to some of their targets will be examined, and the results will be correlated with gene expression data during reprogramming. This chapter will therefore shed light on dynamics of Oct4 and Sox2 binding during entire process of reprogramming.

4.2. Chromatin Immunoprecipitation (ChIP) Optimisation:

The ChIP was optimised in order to capture maximum enrichment of target genes by Oct4 and Sox2. During ChIP, transcription factors are crosslinked to DNA using formaldehyde, and the corresponding binding sites in the DNA are captured by ChIP using a specific antibody against the transcription factor of interest (Park 2009). Many factors in the ChIP protocol, including the time of formaldehyde fixation, the number of sonication cycles, the salt concentration in the wash buffers, the number of washes and the amount and lot number of the antibody used, can have a major impact on the amount of DNA and the purity of the DNA recovered. Background binding of DNA to the beads or unspecific binding of the antibody can often impair the ChIP results. Therefore, the ChIP protocol was optimised in several steps, including the number of cells and sonication cycles, the type of beads used

(e.g. Agarose or magnetic A/G Dynabeads), and the amount and lot numbers of Sox2 and Oct4 antibody.

4.2.1. Optimization of sonication cycles:

The first step in the ChIP optimisation was the number of sonication cycles, which is different from cell to cell. The cells needed to lyse and then DNA needed to shear in small fragments. The fragments must not be too short or too long. A long fragment could be associated with several transcription factors binding site and it is not specific to TF of interest. On the other hand, short fragments less than 100bp are not suitable for making a sequencing library.

To optimise this, 5 million cells were cross-linked and lysed with lysis buffer. DNA was then sheared using the Bioruptor-Diagenode with either 2, 5, 10, or 20 cycles. The sonicated DNA was then checked on an agarose gel to determine the approximate size of the DNA fragments (Figure 6). The number of sonication cycles between 5 and 10 cycles gave a fragment size between 300-500 bp which is the optimum size (Figure 6).

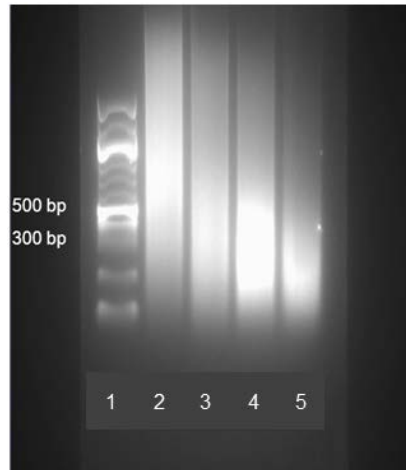


Figure 6. Optimisation of the number of sonication cycles. Different sonication cycles were tested. Lane 1 shows the ladder (0.1- 10kb), lane 2 (2 cycles), lane 3 (5 cycles), lane 4 (10 cycles) and lane 5 (20 cycles). Chromatin samples were separated on a 2% agarose gel and 5 and 10 sonication cycles provided the optimum fragment size for ChIP (between 300 and 500bp).

4.2.2. Optimization of the number of cells:

A limiting cell dilution experiment (1, 2.5, 5, 10 million iPSC cells) was performed in order to determine the minimum number of cells required to do ChIP. To test this, the cells were crosslinked and sonicated then ChIP was performed in iPS cells according to Polo laboratory standard protocol. The significant difference in enrichment between IgG-ChIP and Oct4/Sox2 ChIP was observed, therefore in this study we ignored the result of IgG-ChIP and only focused on positive and negative targets which is more accurate. This showed that the minimum number of cells required to successfully perform ChIP was 5 million cells. The Rif1 and Nanog studied as positive targets and Myog as a negative target (Figure 5).

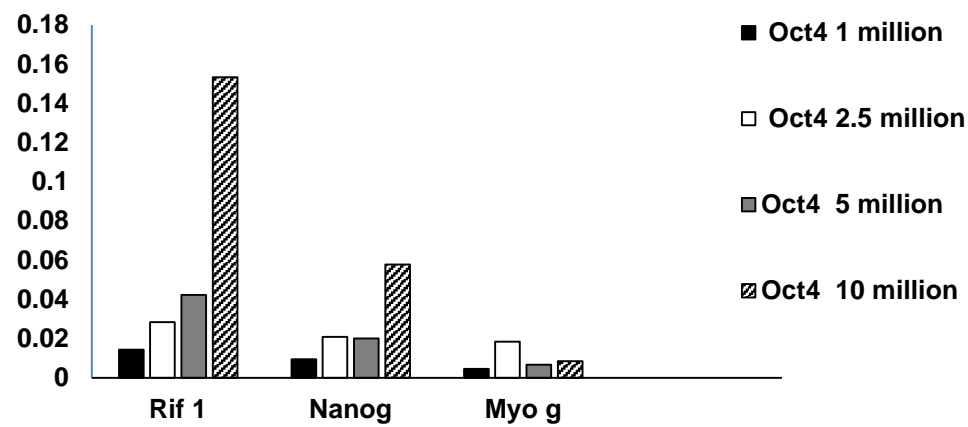


Figure 5. The number of cells required for ChIP in iPSC. Representative result of different numbers of iPSC (1, 2.5, 5 and 10 million cells) were tested. The minimum number of cells with reasonable enrichment for positive targets (Rif 1 and Nanog) was observed in 5 million cells.

4.2.3. Optimisation of crosslinking time:

The next step was optimizing the time taken for formaldehyde-crosslinking. The time of crosslinking is critical because excessive crosslinking could mask the antigen epitope and consequently adversely affect the binding capacity. Conversely, a low incubation time may cause low crosslinking of the transcription factor to the DNA. The time required for crosslinking was tested by crosslinking iPSCs with formaldehyde for 5 and 10 min using 5 and 10 sonication cycles, and then performing an Oct4 ChIP (Figure 7). The results showed there was no significant difference between positive and negative targets in 5 min crosslinking and 5 sonication cycles and the optimum conditions of protein-DNA crosslinking with formaldehyde were observed within 10 min incubation and 5 sonication cycles.

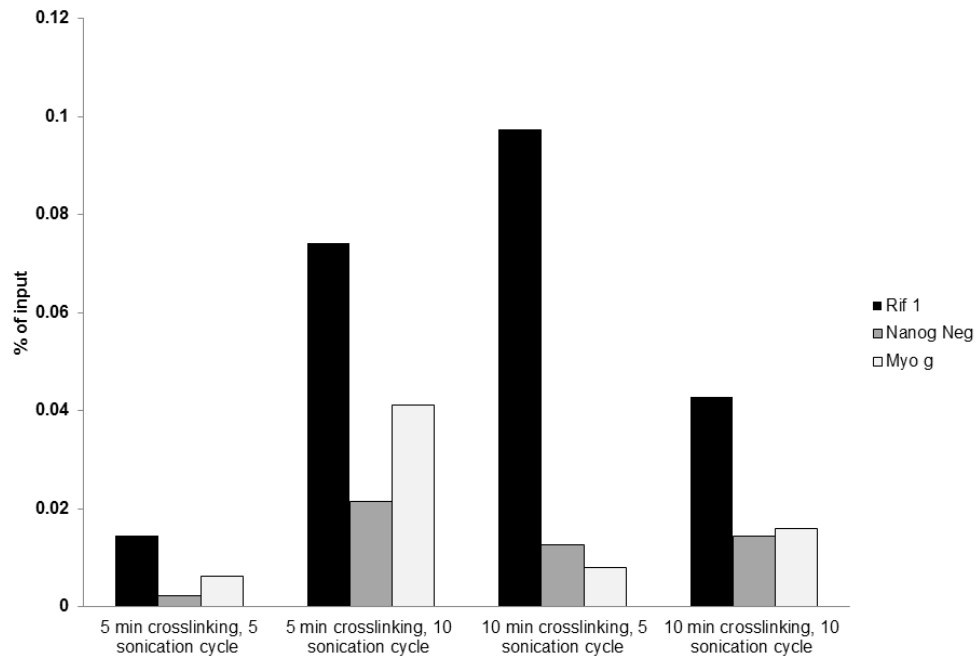


Figure 7. Representative result of different crosslinking times with two different sonication cycles. Crosslinking with formaldehyde was performed in 5 and 10 min incubations, followed by two different sonication cycles (5 and 10 sonication cycles). Highest enrichment of a positive target (Rif1) was observed in 10 min crosslinking followed by 5 sonication cycles. MyoG and Nanog-neg were used as negative targets.

4.2.4. Optimisation of the amount of antibody:

Next, the amount of antibody for 5 million cells was tested. Different titres of Sox2 (2.5, 5 and 10ug antibody per ChIP) were tested and the qPCR results showed high enrichment by Sox2 in positive targets such as Rif1 and Dido1 with 2.5ug antibody (Figure 8). The enrichment of positive targets (Rif1 and Dido1) compared to negative targets (MyoG and Nanog neg) showed less background (compared with negative targets) using 2.5µg antibody. The high enrichment was also observed using 5ug antibody but the background was also higher.

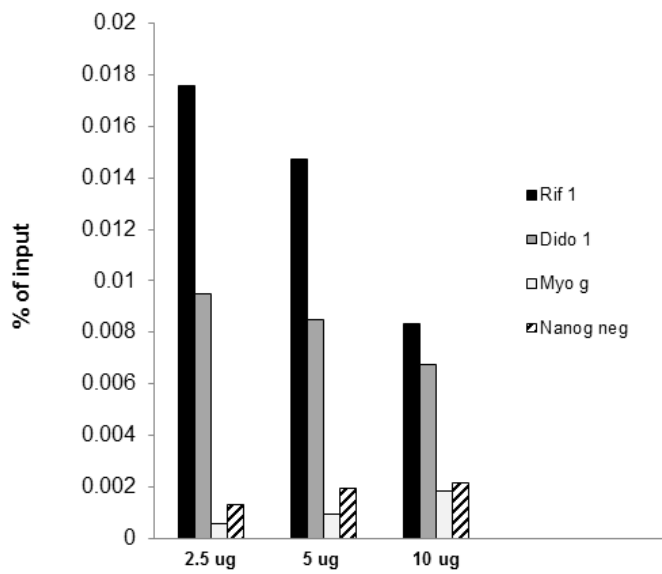


Figure 8. Representative experiment, to determine the optimum amount of ChIP antibody.

Different amounts of anti-Sox2 antibody (2.5, 5 and 10 μ g) were tested using 5 million iPSCs.

Highest enrichment was observed with 2.5ug antibody, which also showed less background. Dido1

and Rif1 were used as positive targets, and MyoG and Nanog neg were used as negative targets.

4.2.5. Optimising the type of capture beads:

The next optimisation step was to compare different types of capture beads and their effect on the amount of Oct4 and Sox2 enrichment. Protein A/G agarose beads (Diagenode) and Dynabeads-Protein G (Invitrogen) were tested. Agarose beads have a high binding capacity because of their porous structure and the high surface area compared to magnetic Dynabeads. However, this may potentially cause high background. Five million cells were used and after lysis and sonication, Dynabeads and agarose beads were used in Oct4 and Sox2 ChIPs. High enrichment was observed in positive target genes such as Rif1 and Dido1 with agarose beads compared to magnetic Dynabeads (Figure 10). To reduce the background, the washing time with the different buffer (dilution buffer, low binding buffer and high binding buffer (see Chapter 2)) was increased from 10 to 20 min.

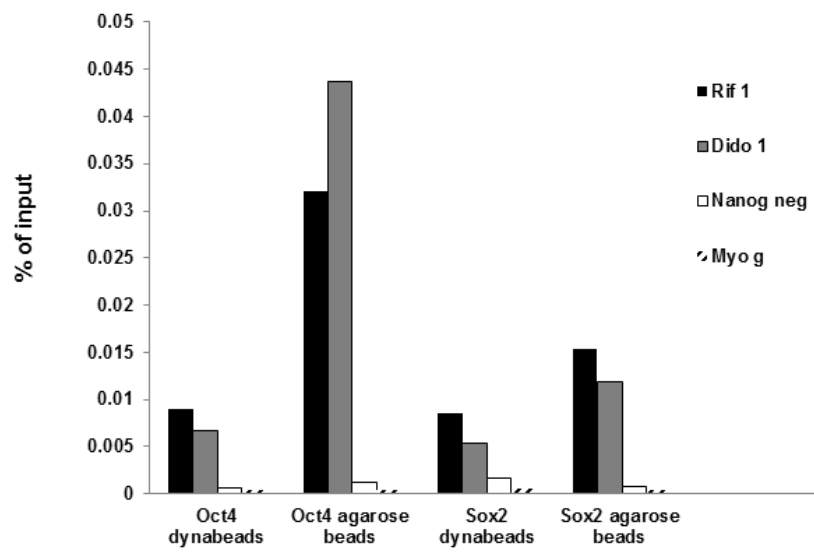


Figure 10. Testing different types of chromatin immunoprecipitation beads. Using Rif1 and Dido1 as positive targets, the highest enrichment was observed with agarose beads. Nanog- neg and MyoG were used as negative targets.

4.2.6. Conclusion of the optimisation of the ChIP protocol:

The result of all optimisation steps showed that using 5 million cells crosslinked with formaldehyde for 10 min, 5 sonication cycles, 2.5ug of antibody (batch specific), using agarose beads to capture the desired fragments, the Oct4 and Sox2 ChIP showed the highest enrichment with the least background (Table 1). Although ChIP in 10 million cells showed higher enrichment, it was difficult to capture 10 million intermediate cells during reprogramming. These conditions were used for performing ChIP in Oct 4 and Sox2 targets during reprogramming.

Table 1: Optimal conditions for ChIP

Number of cells	Sonication cycles	Amount of Antibody	Type of capture beads	Time of formaldehyde crosslinking
5 million	5 cycles	2.5ug	Agarose beads	10 min

4.3. The dynamics of transcription factor targeting during cellular reprogramming

4.3.1. Oct4 and Sox2 bind to different targets in different patterns during reprogramming

In order to assess how the reprogramming factors mediate reprogramming, it is important to understand how they bind to the genome. Due to low expression of the Ssea1 marker during the early stages of reprogramming, bulk populations of cells undergoing reprogramming, containing a heterogeneous population of Thy1+, Thy1-, Ssea1+ cells, were investigated. These cells were analysed to see whether there were any differences in Oct4 and Sox2 binding at different time points during reprogramming. Also, based on our previous study, gene expression during reprogramming occurred in two waves (Polo, Anderssen et al. 2012). The first wave of reprogramming (occurring between MEF and day 6) is associated with genes that directed cell proliferation and metabolism, whilst the second wave of reprogramming (between day 6 and day 12) is related to pluripotency-associated genes (Polo, Anderssen et al. 2012). Therefore, the first and second wave of reprogramming were investigated in terms of the binding patterns to determine whether there are any differences in these two waves in terms of TF binding.

The bulk populations of cells were crosslinked and sonicated and ChIP was performed based on optimal conditions. Analysis of Oct4 and Sox2 binding to the three targets Rif1, Klf9 and Dido1 showed that Oct4 and Sox2 have similar binding pattern in these targets, with high enrichment during the first wave of reprogramming, followed by low enrichment during the second wave of reprogramming (Figure 11). These three targets also showed high enrichment by Oct4 and Sox2 in iPSCs. Upon doxycycline induction, Oct4 and Sox2 are expressed in both the reprogramming cells (e.g. the Ssea1+ population) and the

refractory cells (e.g. the Thy1+ population). However, it is not clear at this stage whether these two transcription factors bind to their targets in both populations and consequently the DNA enriched by ChIP from the bulk population is an average of those two populations, or whether the transcription factors only bind to their targets in reprogramming cells.

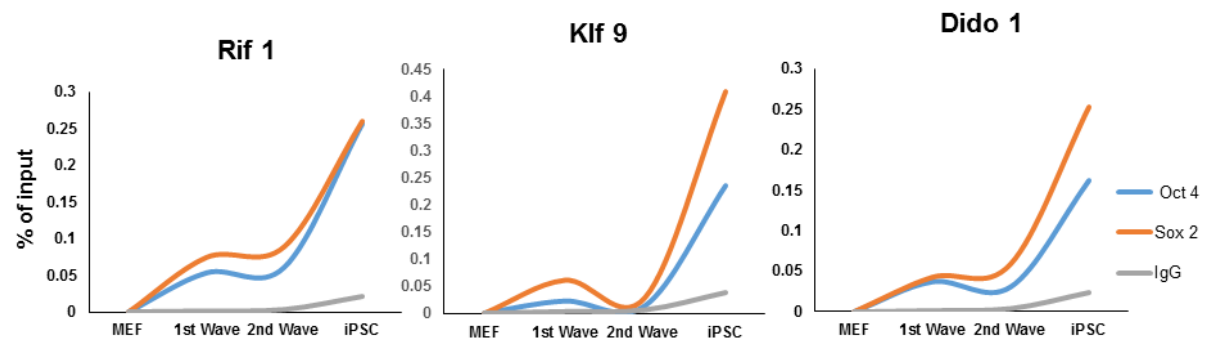


Figure 11. Oct4 and Sox2 binding to Rif1, Klf9 and Dido1 in MEF, during the first wave of reprogramming (between day 0 and day 6) and the second wave of reprogramming (between day 6 and day 12), as well as in iPSC. The highest enrichment was observed in iPSC for Oct4 and Sox2.

4.3.2. Purification of reprogramming intermediates with AutoMACS

The reprogramming intermediate cells with expression of Ssea1 marker during reprogramming needed to be sorted. To capture high numbers of reprogramming intermediates in a time efficient manner, we used magnetic isolation via magnetic cell isolation (AutoMACS - Milteny Biotec) instead of fluorescence-activated cell sorting (FACS) to purify Ssea-1+ intermediate cells. There is a clear time advantage in using AutoMACS compared to FACS, however, the Ssea-1+ cell populations' purity was reduced to 70-80% rather than >95%, which, depending on the assay to be used in the future, might pose a problem. Therefore, for our purpose, using AutoMACS was time efficient and cheaper. For ChIP, 80-90% purity is sufficient because of the high number of cells (5-10 million cells) used. The cells were labelled with anti-Ssea1 (CD15) magnetic beads for sorting Ssea1+ cells by MACS at different time points, and the purity of post-sort MACS was validated with FACS (Figure 12). The result showed that the percentage of Ssea1+ population before MACS in four time courses (day 3, 6, 9, 12) was 3.63, 9.83, 15.5 and 19.5% respectively. The isolation of positive fraction (Ssea1+ population) showed 54.3, 73.6, 74.6 and 77.7% after MACS (Figure 12). Therefore with AutoMACS, the reprogramming cells were successfully isolated and used for ChIP.

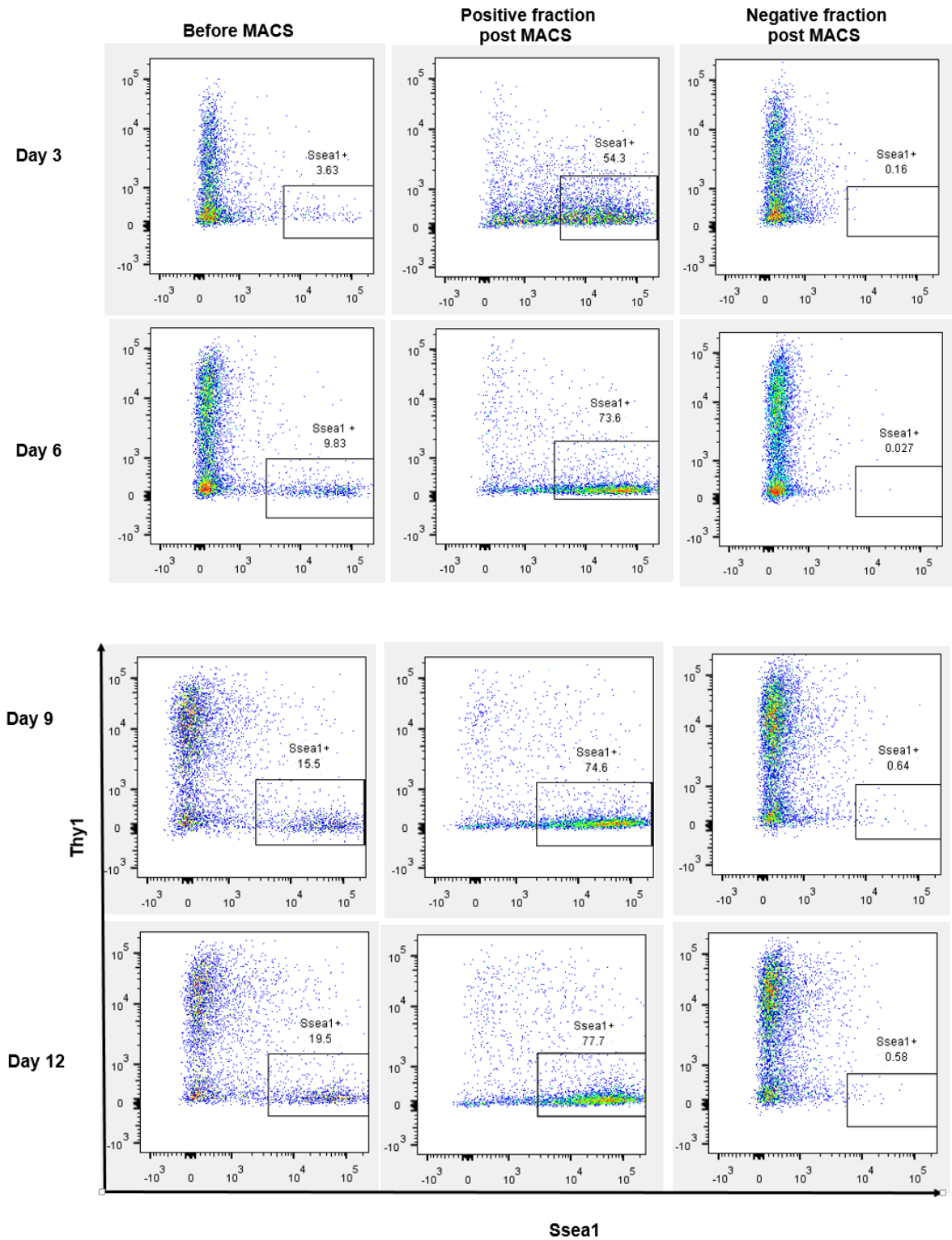


Figure 12. Isolation of reprogramming intermediates during reprogramming by AutoMACS.

The percentage of Ssea1 before MACS and after MACS in positive fraction and negative fraction in days 3, 6, 9, 12.

4.3.3. Roadmap of Oct4 and Sox2 binding during reprogramming in the Ssea1+ population

To dissect in more detail how Oct4 and Sox2 bind to their targets in reprogramming cells (Ssea1+ cells) and consequently to understand whether the transcription factors only bind to their targets in reprogramming cells or there are any differences between ChIP in reprogramming cells and bulk population. The reprogramming cells were enriched by MACS at different time points during reprogramming, and ChIP for Oct4 and Sox2 was performed in Ssea1+ cells. Despite all optimisation steps it was impossible to sort enough Ssea1+ cells during reprogramming to successfully perform ChIP-sequencing for Oct4 and Sox2. As an alternative, a Biomark Real-time quantitative PCR (Fluidigm) ChIP was set up which allowed analysis of Oct4 and Sox2 binding to 70 genes previously identified as targets (Kim, Chu et al. 2008) using minimal amounts of DNA. Oct4 and Sox2 ChIP results were provided by Fluidigm for 40 target genes, but for other target genes it failed. To assess when the Oct4 and Sox2 targets recruit their TFs, and whether there was any similarity in binding in different targets between these two waves of reprogramming, ChIP for Oct4 and Sox2 was performed in MACS isolated intermediate cells.

Interestingly, Sox2 targets showed four different patterns during the first and second waves of reprogramming. Some targets, such as Klf2, Sall1, Slc20a and Tdgf1, were enriched for Sox2 at the first wave of reprogramming, with less change until the end of reprogramming (Figure 13A). Other targets, such as Id3, Rest, Mybl2, and Sox13, showed another pattern of enrichment with little enrichment during the first wave of reprogramming, followed by high enrichment at the end of reprogramming (Figure 13B). Targets such as Dido1 and Nodal were enriched during the first wave of reprogramming, followed by slight enrichment at the end of reprogramming and in iPSCs (Figure 13C). The fourth binding pattern showed high enrichment for Ssbp3, Rif1 and Cbx7 during the first wave and in

iPSCs, but little enrichment was observed at the second wave of reprogramming (Figure 13D).

Oct4 targets also showed four binding patterns, with high enrichment at the first and second wave of reprogramming in different targets (Figure 14). High enrichment of Oct4 in the first wave of reprogramming and in iPSC was observed in several genes, such as Cbx1, Phc1, Rest, Nodal, Rif1, Pou5f1, Lefty1, Klf4 (Figure 14A), however some other target genes, such as Sox2, Usp44, Nkx2.2, Rp13b, Rybp, Tcf7 and Cdx2, Arhgap2, Dido1 and Cbx7, only showed high enrichment at the second wave of reprogramming (Figure 14B). Kdelc1 and Tcl were enriched for Oct4 only in the first wave of reprogramming (Figure 14C). Oct4 was recruited to Hox13, Utf1, Mybl2, Sox13 and Zfp148 at the second wave of reprogramming and in iPSC (Figure 14D). Therefore, different patterns of Oct4 and Sox2 binding were observed in different target genes, suggesting different temporal activity of each target gene at different time points.

Of all investigated target genes, ten common Oct4 and Sox2 target genes were studied in both Oct4 and Sox2 binding sites. Most of these target genes were highly enriched at day 3 by both factors, although some of the target genes showed different patterns during reprogramming (Figure 15).

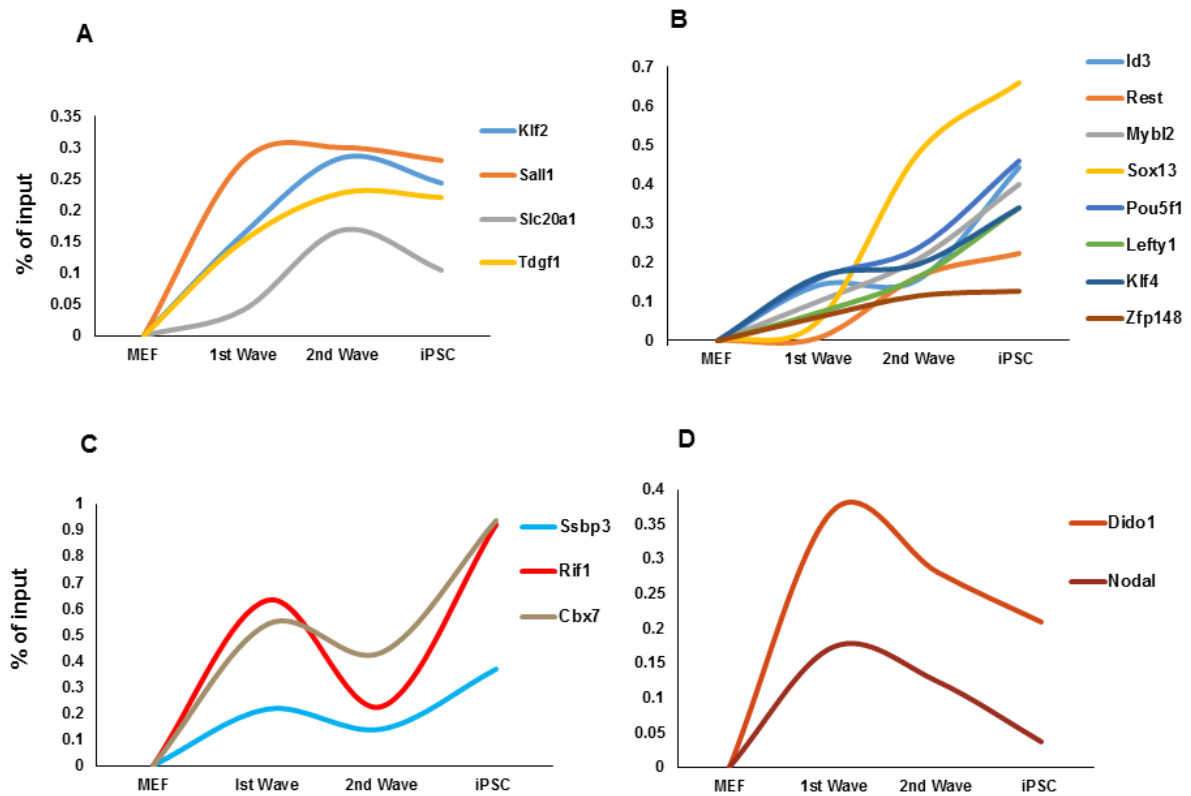


Figure 13. Four patterns of Sox2 binding during reprogramming. A. The first pattern with high enrichment at the first wave of reprogramming with no change until the end, including Klf2, Sall1 Slc20a1 and Tdgf1. B. The second pattern, with high enrichment at the second wave of reprogramming, was observed in Id3, Rest, Mybl2, Sox13, Pou5f1, Lefty1, Klf4 and Zfp148. C. The third pattern, with high enrichment at the first wave and low enrichment at the second wave of reprogramming was observed in Ssbp3, Rif1 and Cbx7. D. The fourth pattern of reprogramming shows high enrichment at the first wave of reprogramming, followed by a gradual decrease in binding in second wave and iPSC stages.

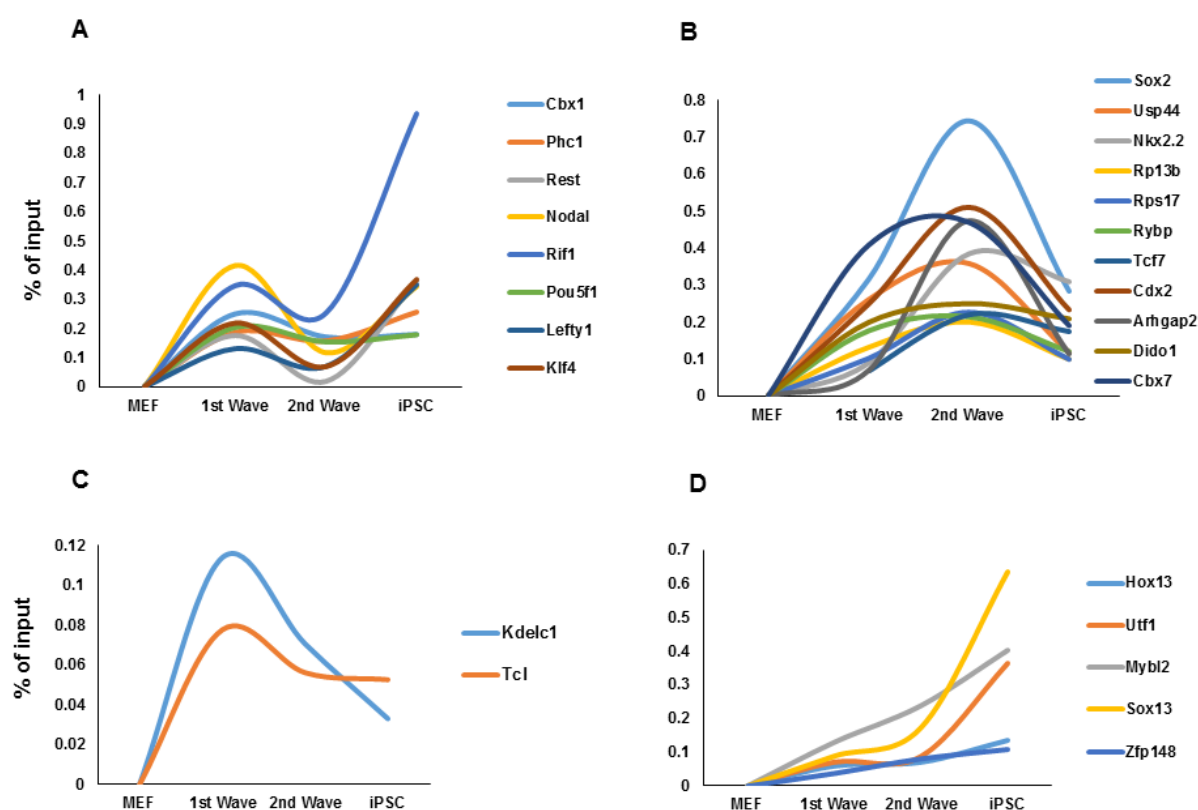


Figure 14. Four different patterns of Oct4 binding to its targets during reprogramming. A.

The first pattern showed high enrichment at the first wave, followed by a little enrichment at the second wave of reprogramming, and then high enrichment in iPSCs, and was observed for Cbx1, Phc1, Rest, Nodal, Rif1, Pou5f1, Lefty1 and Klf4. B. The second pattern showed high enrichment at the second wave and low enrichment in first wave and iPSC, and was observed in Sox2, Usp44, Nkx2.2, Rp13b, Rybp, Tcf7, Cdx2, Arhgap2, Dido1 and Cbx7. C. Only two target genes (Kdelc1 and Tcf) were observed with the third binding pattern, with high enrichment at the first wave and low enrichment at the second wave and in iPSCs. D. The fourth Oct4 binding pattern showed a gradual enrichment for its targets during reprogramming, and was observed for Hox13, Utf1, Mybl2, Sox13 and Zfp148.

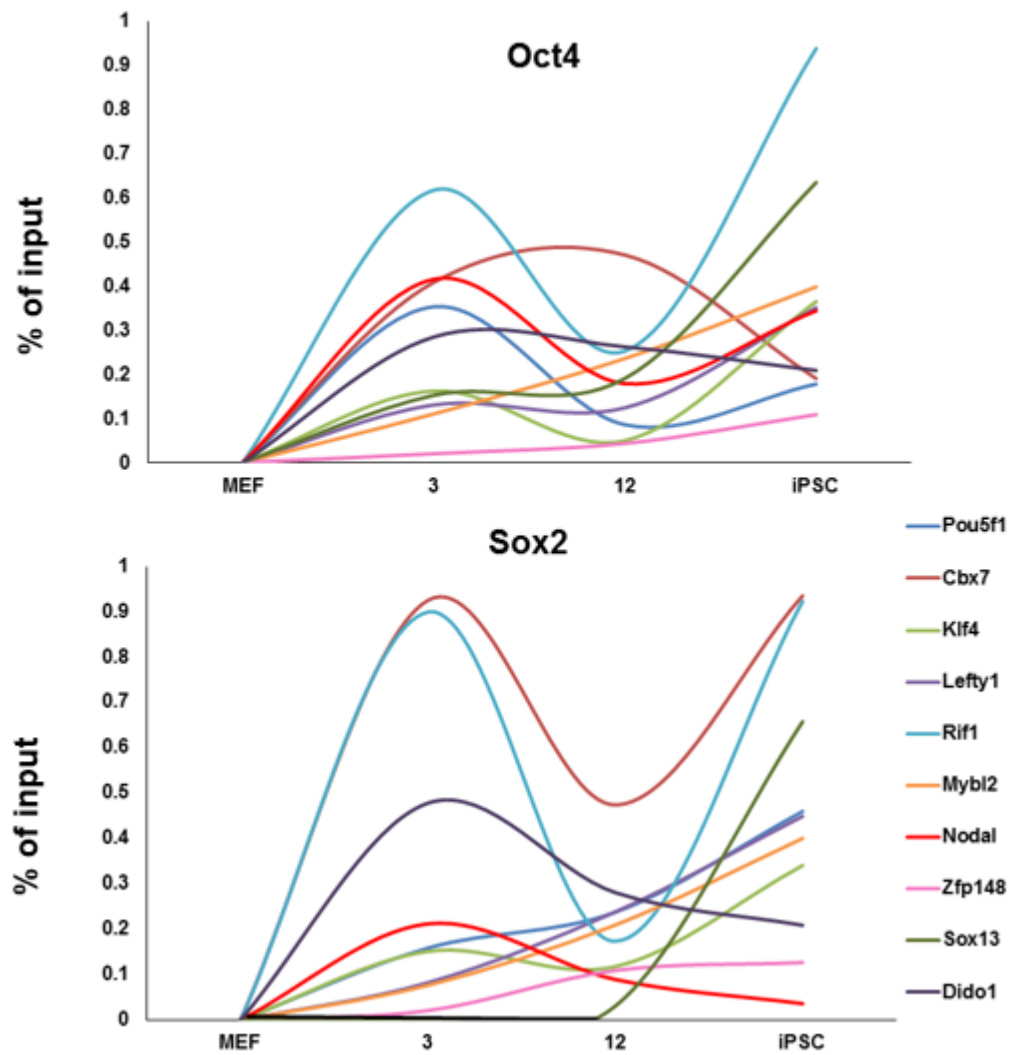


Figure 15. Sox2 and Oct4 binding to different targets during reprogramming. Most of the targets showed highest enrichment by Oct4 and Sox2 at day 3 and in iPSC. The top figure showed enrichment of 10 target genes by Oct4 and the bottom one showed enrichment by Sox2.

As we previously showed that some target genes upregulated during first and second wave of reprogramming and some other target genes downregulated during reprogramming (Polo, Anderssen et al. 2012), here the differences in the binding pattern of Oct4 and Sox2 to these target genes was investigated in terms of similarity in gene expression. Two important pluripotent associated genes, Lefty1 and Pou5f1, showed upregulation at the end of reprogramming (Figure 16 C). In contrast, Mybl2, Rif1, Cbx7 and Nanog showed another gene expression pattern, with gradual upregulation till the end of reprogramming. Similar gene expression patterns were observed for Zfp148 and Dido1, with upregulation at first wave of reprogramming followed by down regulation at the end of reprogramming. Sox13 and Klf4 showed two different gene expression patterns. To examine the binding pattern of Oct4 and Sox2 at these genes, we analysed them based on their gene expression pattern.

As stated above, Lefty1 and Pou5f1 are pluripotency genes and are upregulated at late days of reprogramming. However, despite similar expression patterns, the two genes showed different Oct4 and Sox2 enrichment during reprogramming. Both Oct4 and Sox2 bound to Lefty1 promoter (1kb upstream of the Transcriptional Start Site (TSS)), at the early days of reprogramming with increased Sox2 enrichment at the end of reprogramming. Pou5f1 showed a similar pattern of enrichment for Sox2, however, the enrichment for Oct4 showed high enrichment at day3 with no change until end of reprogramming, as was the case for Sox2. This result suggested that both transcription factors might act as transcriptional activators for Lefty1 and Pou5f1 during reprogramming and in iPSC.

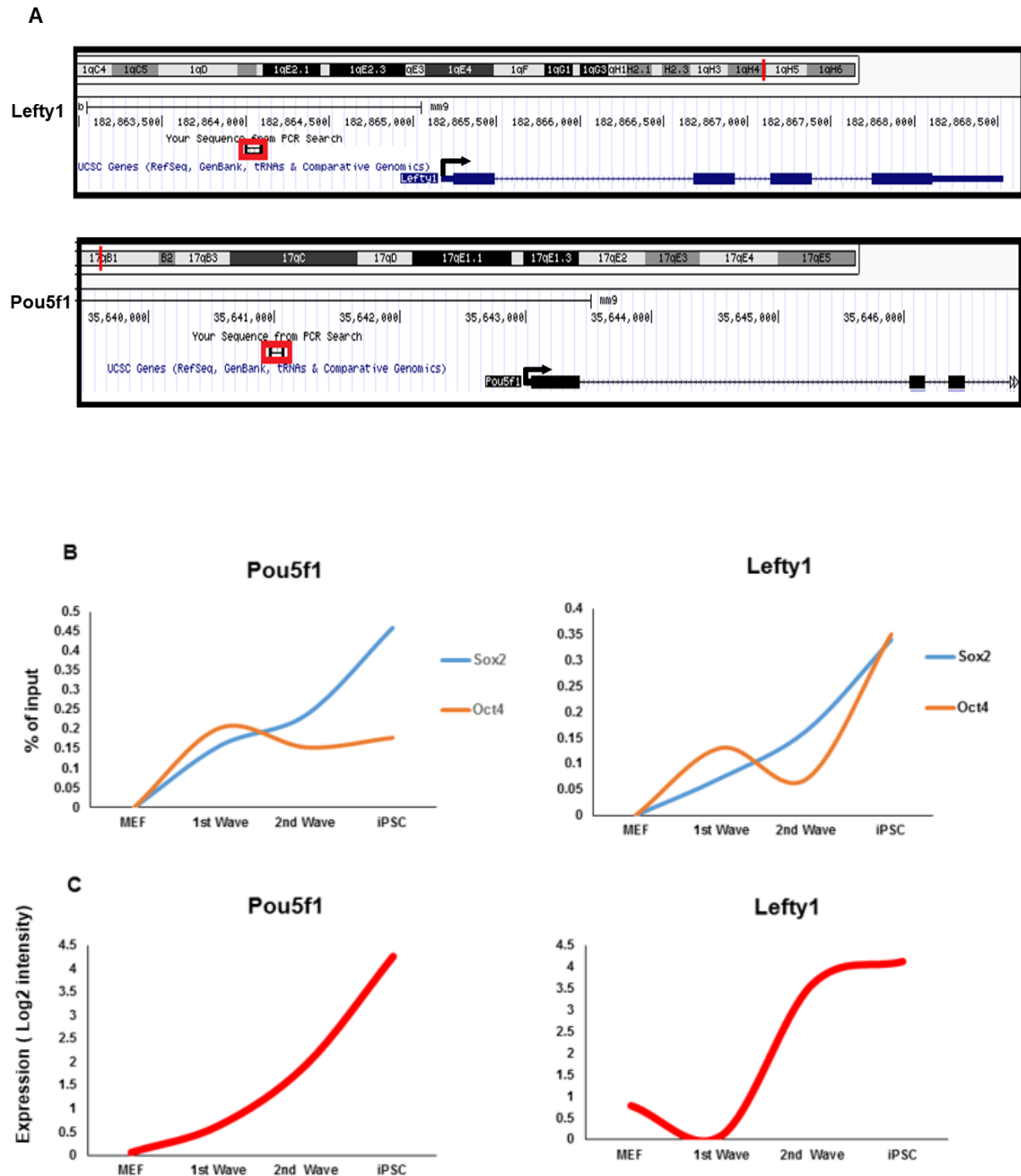


Figure 16. Lefty1 and Pou5f1 activity during reprogramming. **A.** Screen capture of the Lefty1 and Pou5f1 gene track from the UCSC Genome Browser. The red boxes show binding sites of Oct4 and Sox2 in Lefty1 and Pou5f1 promoter regions. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression levels during reprogramming.

Mybl2 (Myb-related protein B-2), Rif 1 (Rap 1-interacting factor 1 homolog), Cbx7 and Nodal followed a similar transcriptional pattern with gradual upregulation until the end of reprogramming (Figure 17C). Mybl2 has an important role in inner cell mass (ICM) formation and ES cell generation (Tanaka, Patestos et al. 1999). Upregulation of the Mybl2 gene was observed in the late G1 stage and leads to progression into the S phase of the cell cycle (Lorvellec, Dumon et al. 2010). Therefore, this data suggest that it plays a role in first and second wave of reprogramming due to its effects on proliferation and ES cell generation. The ChIP data showed Oct4 and Sox2 enrichment at Mybl2 binding site, which was consistent with gradual expression of Mybl2 during reprogramming. These results therefore suggested that, during reprogramming, Oct4 and Sox2 bound to the Mybl2 promoter gradually in a similar pattern to activate gene expression (Figure 17 A, B, C).

Rif 1 (Rap 1-interacting factor 1 homolog) is expressed at the early days of reprogramming, with high expression during the late stages of reprogramming (Figure 17C). This gene is also involved in cell cycling and is a telomere-associated protein (Cornacchia, Dileep et al. 2012). As it is related to the cell cycle based on our data (Polo, Anderssen et al. 2012), it was not surprising that it was affected during the first wave of reprogramming. The promoter occupancy analysed by ChIP showed that Oct4 and Sox2 binding to Rif1 followed a similar pattern during reprogramming, with high enrichment during the first wave of reprogramming and in iPSCs, whereas it showed low enrichment for both in second wave of reprogramming (Figure 17B).

Oct4 and Sox2 occupy the Cbx7 promoter during the first wave of reprogramming, followed by low enrichment for Oct4 at the second reprogramming wave. Similar signals were observed in both Oct4 and Sox2. In iPSC, there was low enrichment by Sox2 and high enrichment by Oct4 (Figure 18B).

Nodal is a member of TGF- β superfamily and plays a role as an important regulator in embryonic stem cell fate (Strizzi, Postovit et al. 2009). It is also a late stage pluripotency marker that is gradually expressed at the end of reprogramming. It was enriched in this study at day 3 for Oct4 and Sox2, whereas low enrichment was observed at day 12. Low binding of Sox2 at the promoters of Nodal was observed in iPSC, however interestingly Oct4 showed high enrichment in iPSC (Figure 18).

Therefore, although gene activity for Mybl2, Rif1, Cbx7 and Nodal showed a similar pattern, the binding patterns of Oct4 and Sox2 was different. Thus, the other factors might affect gene expression.

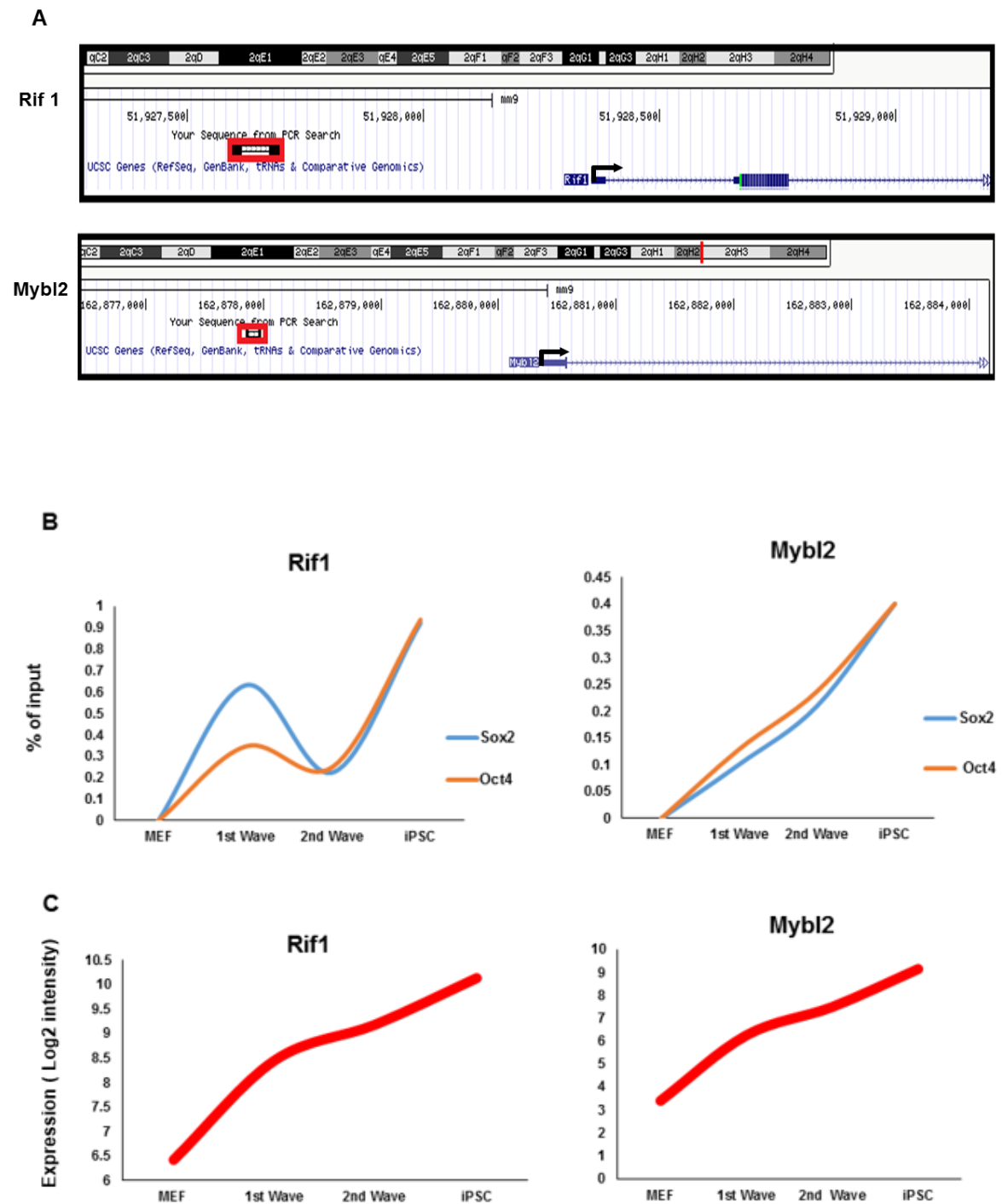


Figure 17. Rif1 and Mybl2 activity during reprogramming. A. Screen capture of the Rif1 and Mybl2 gene track from the UCSC Genome Browser. The red boxes show binding sites of Oct4 and Sox2 in Rif1 and Mybl2 promoter regions. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression during reprogramming.

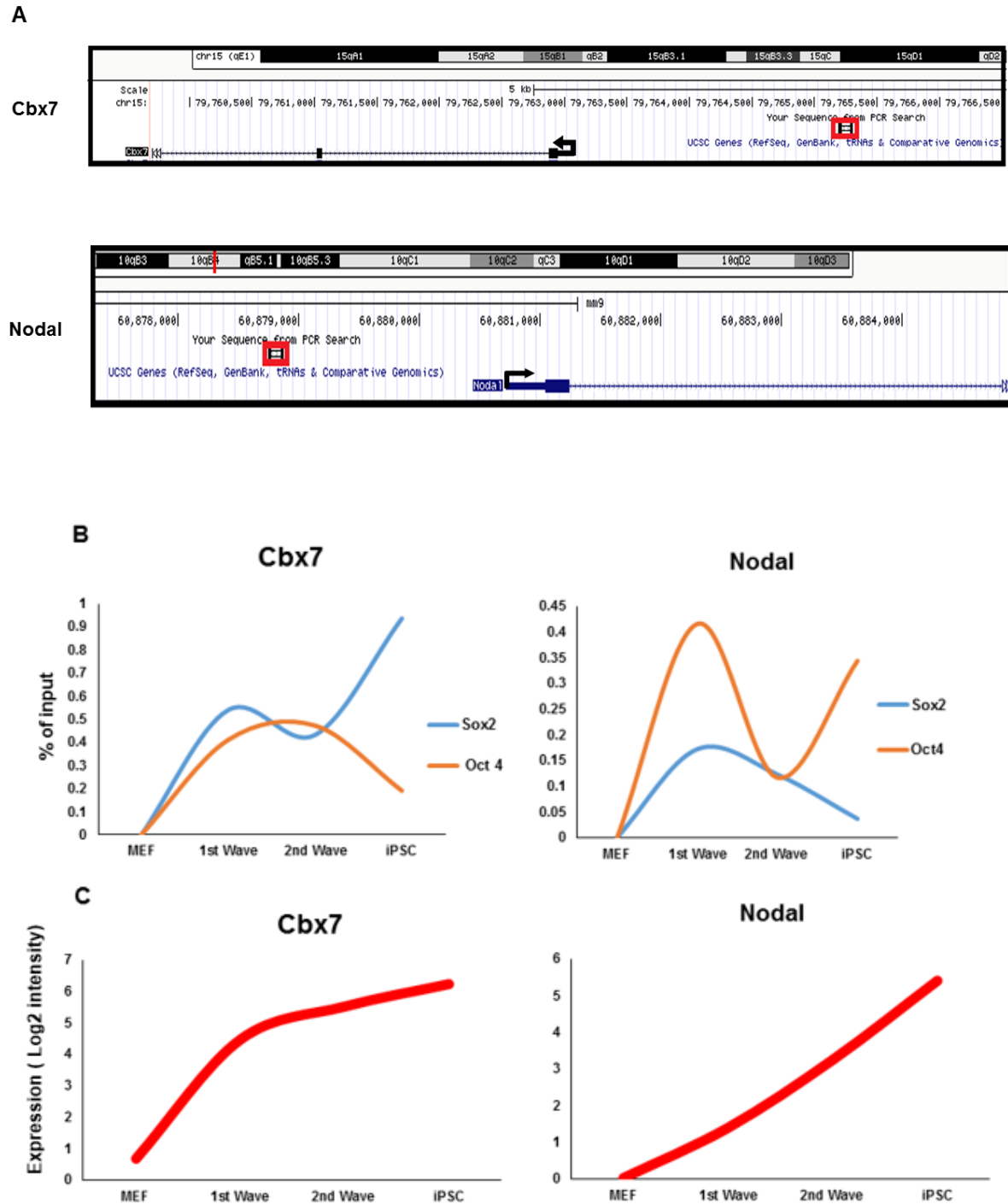


Figure 18. Cbx7 and Nodal activity during reprogramming. **A.** Screen capture of the Cbx7 and Nodal gene track from the UCSC Genome Browser. The red boxes show binding site of Oct4 and Sox2 in Cbx7 and Nodal promoter region. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression during reprogramming.

Another gene expression pattern was analysed for Sox13, with up-regulation during the first phase of reprogramming. There was further upregulation during the second phase, but down regulation was observed in iPSCs (Figure 18). Sox13 is also expressed in all three embryonic cell lineages: ectoderm, mesoderm and endoderm, suggesting that this gene plays a role in both developmental and differentiation processes (Wang, Ristevski et al. 2006). Oct4 and Sox2 binding, as determined by ChIP, showed an opposite pattern compared to the gene expression data (Figure 19). For Sox2, no binding was observed in Sox13 until day 3, and low binding was shown by Oct4 during first wave of reprogramming. The highest enrichment for both Oct4 and Sox2 was observed in iPSCs. These results suggest that Oct4 and Sox2 form part of a repressor complex in Sox13, which suppresses Sox13 expression in iPSCs. Recent data showed a point mutation in Sox13 did not make a difference in iPSC generation (Aksoy, Jauch et al. 2013), suggesting that expression of this gene does not have a pivotal role in iPSC generation. Our results also showed low expression of this gene in iPSCs (Figure 19C).

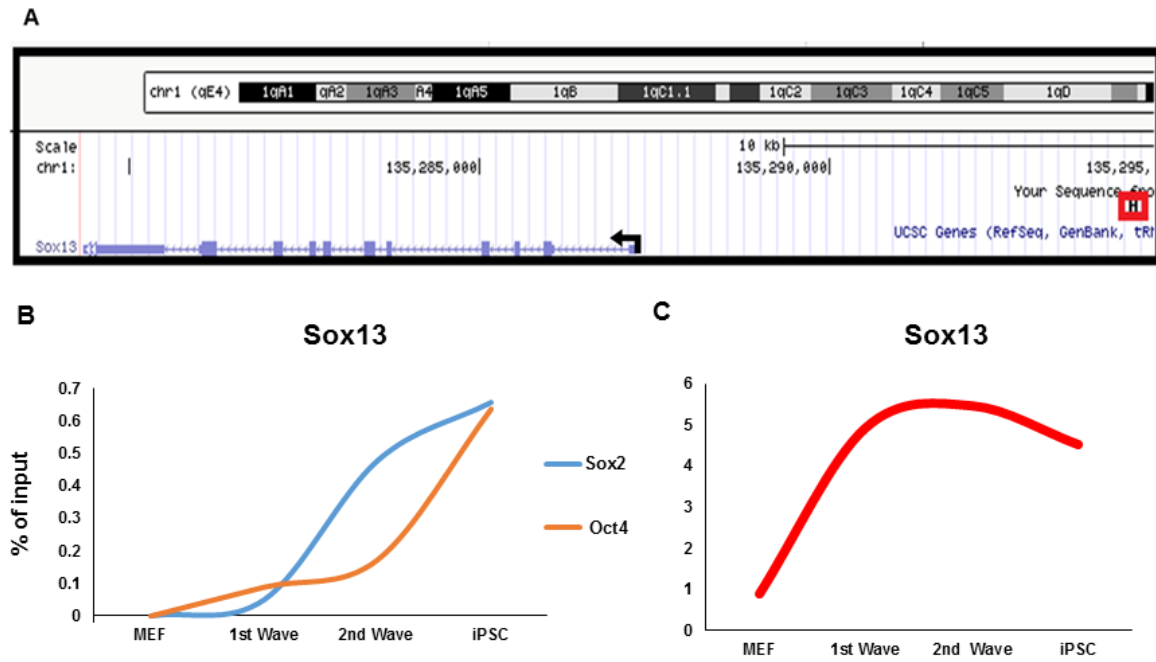


Figure 19. Sox13 activity during reprogramming. **A.** Screen capture of the Sox13 gene track from the UCSC Genome Browser. The red inset shows binding site of Oct4 and Sox2 in Sox13 promoter region. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression during reprogramming, Y axis shows Log2 intensity.

Klf4 plays an important role in chromatin rearrangements around the Pou5f1 locus (Jerabek, Merino et al. 2014) and it is the only target gene down-regulated at day 3, and upregulated in iPSC (Figure 20). We observed high enrichment for Oct4 and Sox2 during the first phase of reprogramming but there was a decrease in binding on day 12 followed by the highest enrichment at the end of reprogramming. Despite some Oct4 and Sox2 binding during the first wave of reprogramming (Figure 20B), there is no Klf4 expression (Figure 20C). This may be due to insufficient binding, or because other factors from the activator complex are missing. High enrichment of Oct4 and Sox2 was observed in iPSC, which is consistent with the Klf4 expression in iPSC (Figure 20).

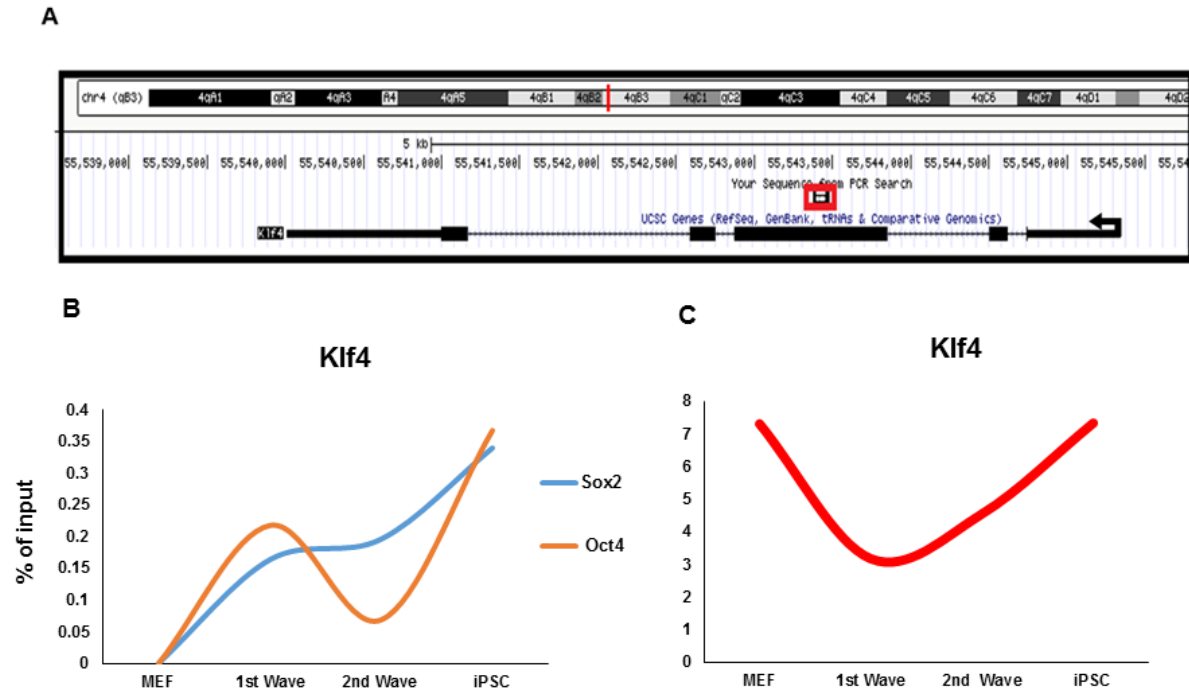


Figure 20. Klf4 activity during reprogramming. **A.** Screen capture of the Klf4 gene track from the UCSC Genome Browser. The red box shows binding sites of Oct4 and Sox2 in Klf4 promoter region. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression during reprogramming, Y axis showed Log2 intensity.

Zfp148 (Zinc finger protein 148) is required for development of fetal germ cells and Sox2 positively effect of Zfp148 expression. The transcription factor Zfp148 also interacts with p53 as a tumour suppressor and it is therefore associated with regulation of cell cycling (Campolo, Gori et al. 2013). Dido1 (Death-inducer obliterator 1) is one of the target genes that, when depleted in mouse ES cells, leads to differentiation. Both of these genes (Zfp148 and Dido1) showed no significant expression changes during reprogramming, however, they showed a different binding pattern by Oct4 and Sox2 during ChIP. The gene enrichment by ChIP against Oct4 and Sox2 showed an increase in Oct4 and Sox2 binding at first wave of reprogramming, with a gradual increase at the end of reprogramming and in iPSCs (Figure 21).

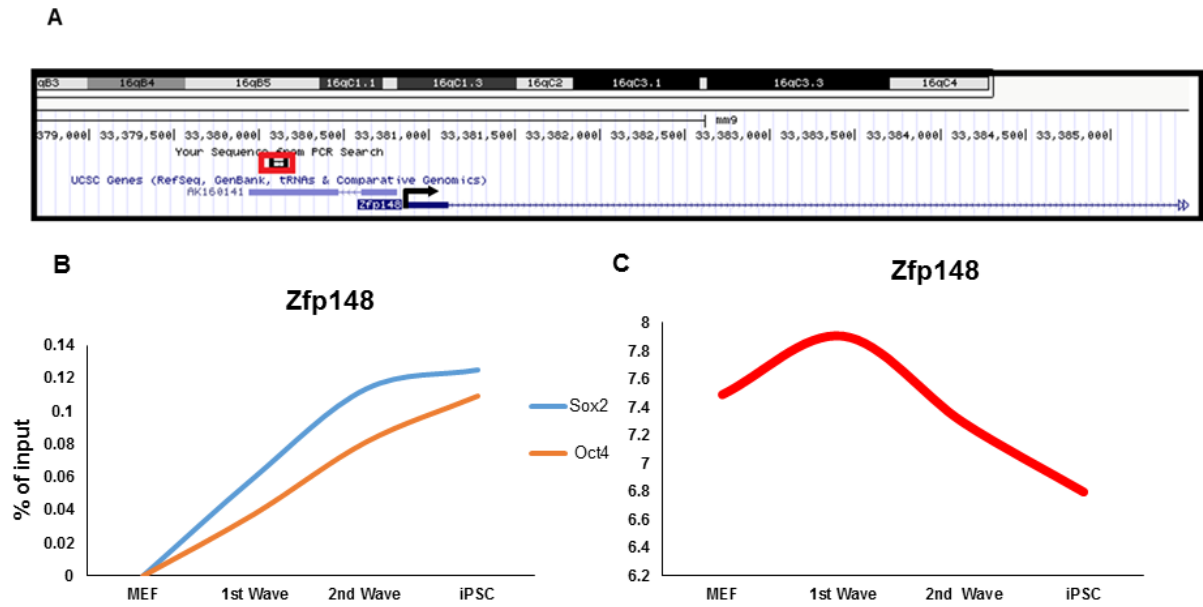


Figure 21. Zfp148 activity during reprogramming. A. Screen capture of the Zfp148 gene track from the UCSC Genome Browser. The red box shows binding sites of Oct4 and Sox2 in Zfp148 promoter region. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression during reprogramming, Y axis showed Log2 intensity.

Nanog and Oct4 are able to regulate Dido1 expression, and Dido1 was also able to regulate Oct4 and Nanog expression (Liu, Kim et al. 2014). Therefore, ectopic expression of Dido1 led to inhibition of differentiation. Dido 1 expression changed during the first and second waves of reprogramming, with no significant change in iPSC. The result for Oct4 and Sox2 ChIP showed that Dido1 was enriched for Oct4 and Sox2 during the first wave, and binding remained consistent at the end of reprogramming (Figure 22).

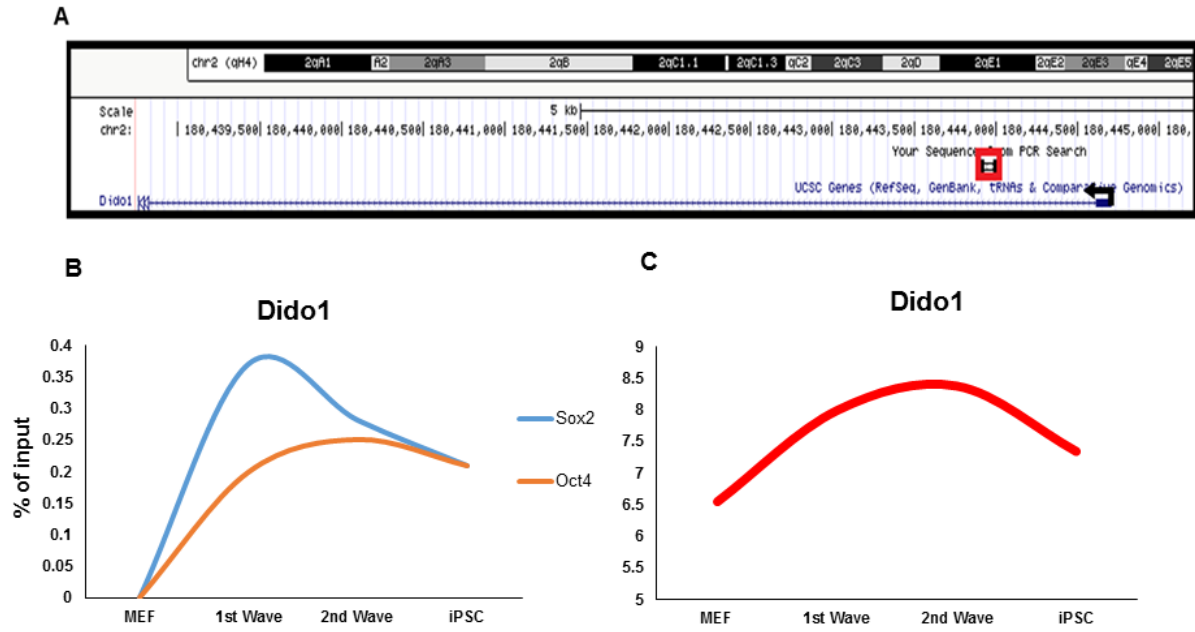


Figure 22. Dido1 activity during reprogramming. **A.** Screen capture of the Dido1 gene track from the UCSC Genome Browser. The red inset shows binding site of Oct4 and Sox2 in Dido1 promoter region. **B.** Dynamic changes in Oct4 and Sox2 binding during reprogramming. **C.** Gene expression during reprogramming, Y axis shows Log2 intensity.

All investigated target genes showed that, despite of similarity in gene activity, Oct4 and Sox2 bind differently to their targets, suggesting a temporal role for each transcription factor in activation of pluripotency-associated genes during reprogramming.

4.4. Conclusion:

Transcription is regulated by transcription factors promoting or impairing transcriptional initiation and elongation (Li, Carey et al. 2007). Transcription factors can act as both activators and repressors, depending on the developmental status of the cell. TFs mediate transcription by competing with activators and repressors for binding sites, or due to changes in the chromatin state (Kadonaga 2004). Both Oct4 and Sox2 require other transcription factors to direct the reprogramming process and express the associated target genes.

In this chapter, some of the targets of Oct4 and Sox2 during cellular reprogramming were investigated. ChIP was used to map the DNA interaction of Oct4 and Sox2 during the first and second waves of reprogramming. The data showed that Oct4 and Sox2 bound differently to their different targets during the first and second waves of reprogramming. Most of the target genes, such as *Rif1*, *Nodal*, *Pou5f1*, *Dido1*, and *Klf4*, were enriched during the first wave of reprogramming and in iPSCs, which might be due to changes in chromatin state at early days of reprogramming. This research showed that the binding of Oct4 and Sox2 to the promoter region is not always consistent with expression of the target gene. It is also showed that some target genes such as *Mybl2*, Oct4 and Sox2 are bound at the same time to the promoter region during reprogramming, and it seems the binding of both factors together is needed for gene expression. However for other candidates, such as *Nodal* and *Cbx7*, Oct4 and Sox2 are bound to the promoter of these targets individually. They bind to promoter regions together at the first wave of reprogramming, however in

iPSCs the gene expression might not rely on the binding of both factors. Thus, these results suggest that the binding of only one of these transcription factors is not sufficient for expression of some target genes, indicating that the gene expression might require additional factors that might not be expressed at this stage. On the other hand, as mentioned in Chapter 3, the amount and ratio of transcription factors affect the efficiency of reprogramming. Therefore, low expression of some of the target genes might be because additional binding factors are required to activate those genes, or it may be due to the lack of transcription factors or the ratio of different transcription factors at their promoter regions. Only Oct4 and Sox2 binding was investigated in this chapter, however several studies have shown that the role of Klf4 accompanied with Oct4 and Sox2 in activation of some target genes at the end of reprogramming and co-bound of Klf4 and c-Myc repress somatic cell genes (Wei, Gao et al. 2013, Nishimura, Kato et al. 2014). The result described in this chapter suggested that investigating the chromatin state and protein complexes during reprogramming will help us understand what impact TF binding has on their targets, and how this affects gene expression. Therefore in the next chapter the chromatin states during reprogramming will be investigated, and the effect on gene expression and Oct4 and Sox2 binding will be examined.

CHAPTER 5

Chromatin accessibility during reprogramming

5.1. Introduction

Chromatin with repeated nucleosome units has important roles in biological processes such as gene expression and DNA replication. Gene expression is regulated by the binding of DNA binding proteins, such as transcription factors and RNA polymerases, to the DNA. DNA packaging plays an important role in gene regulation. The binding of regulatory factors to nucleosome-free regions, which are often located at promoters and enhancers, leads to transcriptional activation. During reprogramming, the chromatin undergoes conformational changes in the region of the transcribed genes. This reconfiguration leads to an open chromatin structure and subsequent accessibility of transcription factors to regulate gene expression (Cirillo, Lin et al. 2002). Chromatin in pluripotent stem cells is relatively open compared with chromatin in somatic cells and its overall structure is less condensed. Moreover the ratio between relaxed euchromatin and compacted heterochromatin is higher than in differentiating cells (Niwa 2007). Heterochromatin is packed with nucleosomes, and therefore most TFs are unable to access to their regulatory regions, although several pioneer reprogramming factors have been shown to access closed and/or open chromatin. For example, FoxA can bind to its regulatory DNA on closed chromatin through the winged-helix DNA binding domain (DBD), which acts like linker histone. Reprogramming factors can also access closed chromatin (Soufi, Donahue et al. 2012, Soufi, Garcia et al. 2015) at 48 hrs post-induction of reprogramming. In these studies, co-binding of factors Oct4/c-Myc and Sox2/c-Myc was observed in nucleosome-enriched chromatin. This differs for Klf4/c-Myc co-binding, which is detected more frequently at

the nucleosome-free regions. They showed that initially Oct4, Sox2, Klf4 and c-Myc bind closed chromatin, and partial motifs were detected on nucleosomes. However, c-Myc alone is not able to bind nucleosomes, but only in combination with other factors (Soufi, Donahue et al. 2012). Despite the recent study on 48 hrs of reprogramming, it is still unclear what changes in chromatin state occur during the two weeks of reprogramming. It is also not obvious if inaccessibility to OKSM targets in Thy1+ cells underlies unsuccessful reprogramming.

We hypothesized that chromatin becomes more accessible during reprogramming, allowing more transcription factors to bind to their targets. To test this, the regulatory and accessible regions in refractory cells (Thy1+) and reprogrammed cells (Ssea1+ cells) were examined using DNase I-MLPA and ATAC-seq. Further studies of nucleosome positioning were then performed in Ssea1+ cells on a genome-wide scale, by ATAC-seq, which is an efficient and rapid method requiring low cell numbers (50,000 cells).

DNase I hypersensitive sites are markers of regulatory DNA and are short regions of chromatin that play functional roles within genetic regulatory active elements, including promoters, enhancers, silencer, insulators, and locus control regions. Either bound transcription factors or nucleosomes protect the accessible regions from digestion. There are different methods to identify accessible chromatin by enzymatic reactions such as DNase-MLPA, MNase-seq, DNase-seq and ATAC-seq. In this chapter, DNase I hypersensitive sites will be examined by DNase I-Multiple Ligation Probe amplification (DNase-MLPA) which is a low-cost, PCR-based technique, and then accessible chromatin will be investigated further by ATAC-seq. An advantage of the ATAC assay compared to DNase-seq and FAIRE-seq is the low number of the cells required: 50,000 cells compared to 1-50 million cells (FAIRE protocol) or 50 million cells (DNase-seq). Additionally, these protocols contain a number of steps and are time consuming (Buenrostro, Giresi et al. 2013).

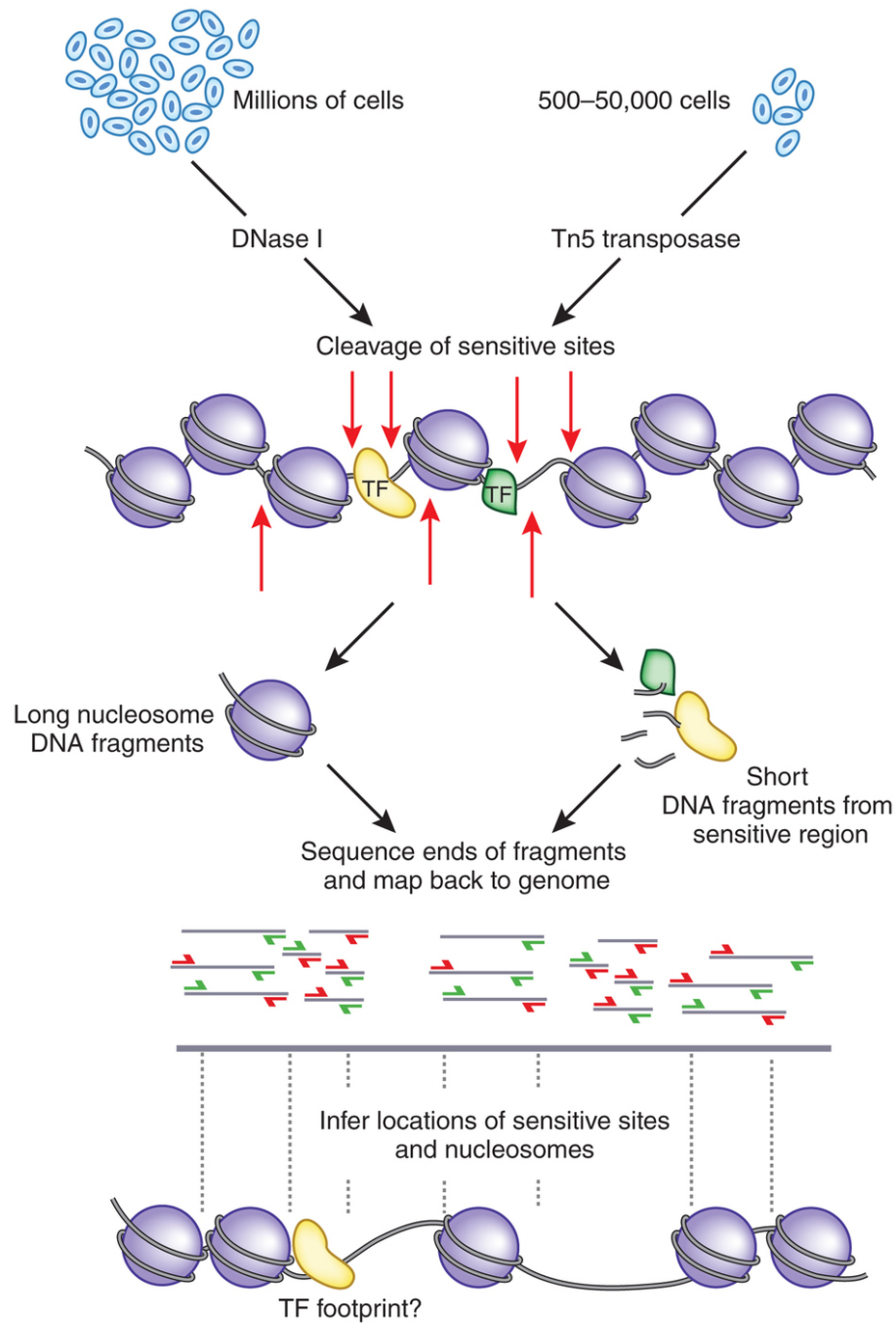


Figure 1. Comparison between DNaseI- seq and ATAC-seq in identification of open region chromatin. Using millions of cells in DNase-seq and only 50,000 cells in ATAC-seq (Buenrostro, Giresi et al. 2013).

5.2. Results

5.2.1. The chromatin state in DNase I hypersensitive (DHS) region of reprogramming intermediates

To confirm that the chromatin state changes during reprogramming, 1 million cells (MEF, day 3, day 9 or iPSC) were sorted and treated with 1 unit DNase I enzyme. DNA was run on an agarose gel and purified using a PCR purification kit (Roche). The DHS sites are exposed to DNase I enzyme in open regions chromatin, and the digestion of these regions was greater than compacted regions. The chromatin at day 3 was more accessible compare to MEF and at day 9 DNA showed more digestion in agarose gel compare to day 3, which means the chromatin is more open (Figure 2). This demonstrated that in general, chromatin becomes more accessible in the transition from a differentiated cell to a pluripotent cell.

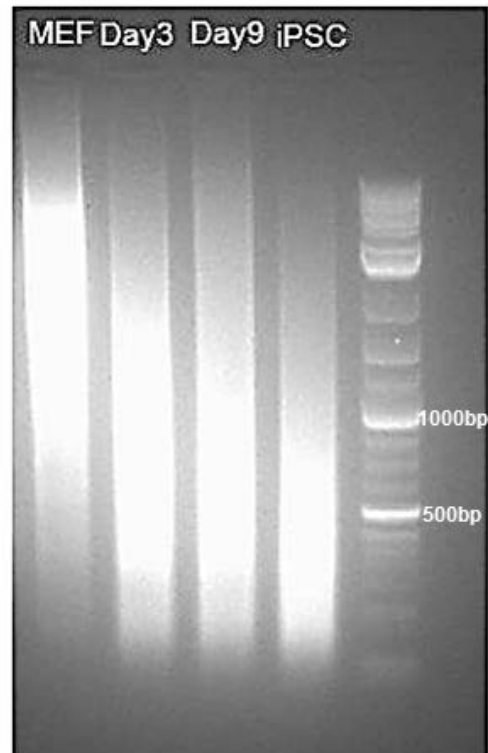


Figure 2. DNase I digestion during reprogramming in MEF, day 3, day 9 and iPSC. The DNA showed progressively more digestion from MEF to iPSC.

5.2.2. Detection of DHS site using a modified Multiplex Ligation-dependent Probe

Amplification (MLPA)

During reprogramming, refractory cells (Thy1+ and Thy1- cells) are not able to form iPSC whereas the reprogramming population (Ssea1+ cells) does generate iPSC colonies. To gain further insight into chromatin changes during reprogramming and observe how chromatin changes at different time points during reprogramming from MEF to iPSC, DNase I-MLPA was used. Different probes were designed, based on DNase I hypersensitive sites of known targets of OKSM (Appendix 2). MEF, iPSC and three different populations (Thy1+, Thy1- and Ssea1+) were sorted at different time points. Nuclei were extracted from one million cells, and then were exposed to DNase I. However, the amount of DNase I needed to be optimized to avoid overdigestion of the DNA. Different enzyme concentrations were examined (0 units (U), 1U, 3U) and the optimum amount of enzyme based on the fragment size was determined in reprogramming intermediates (Figure 3). The optimum fragment size was between 200- 500bp which has previously shown the optimum size for analysing DNase I hypersensitive sites using DNaseI-MLPA (Ohnesorg, Eggers et al. 2009, Ohnesorg, Eggers et al. 2012). The result showed that 1U of enzyme was the optimum amount. Next, DNase I-MLPA were tested in MEF and iPSC using several probes of different lengths. Following capillary electrophoresis, each probe was examined. The result showed a reduction in peak heights for the Sox2, Pou5f1 and Zfp42 in iPSC compared to MEF, suggesting that the chromatin is more accessible in iPSC (Figure 4).

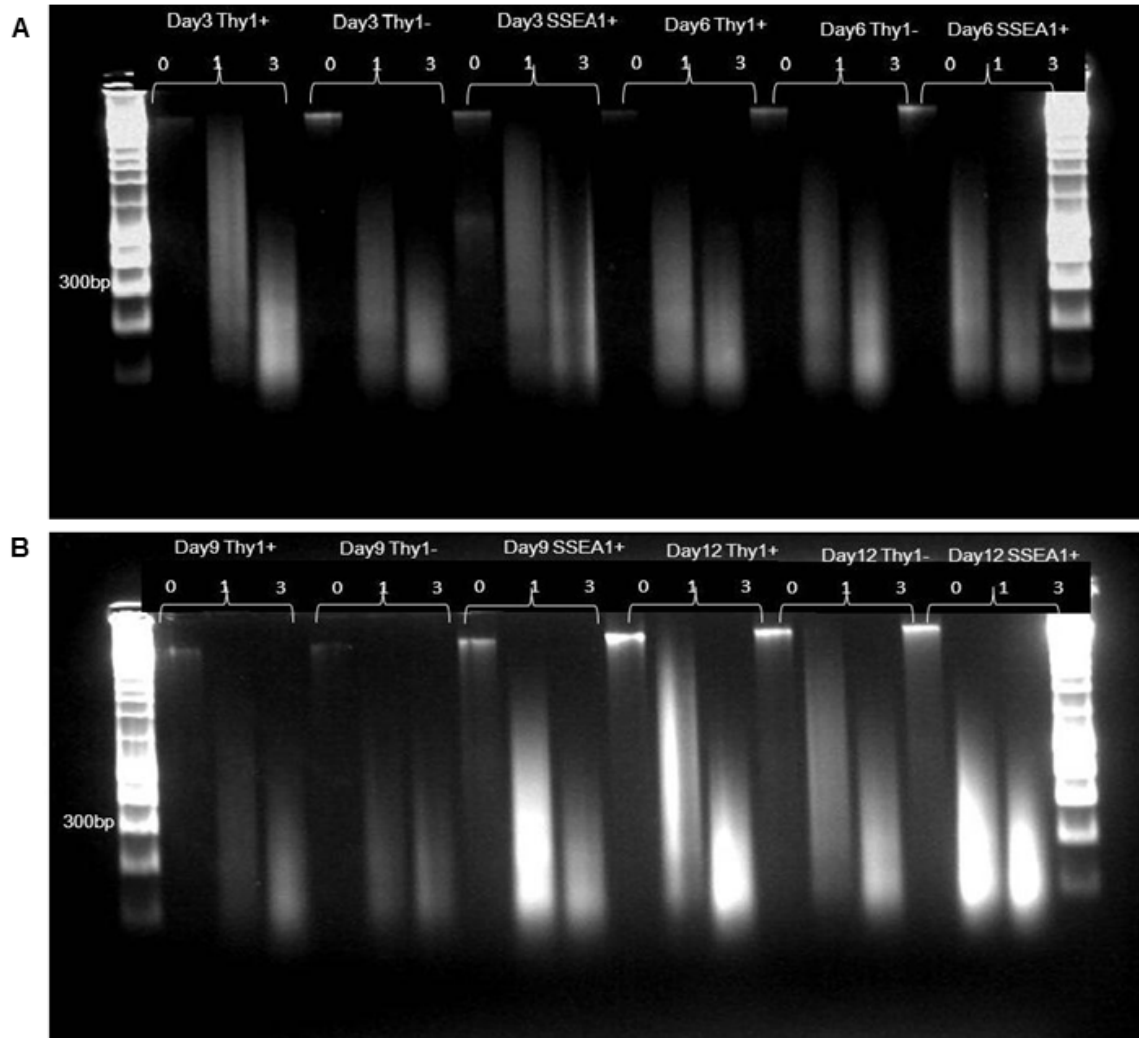


Figure 3. DNase I digestion of intermediates at different time points (day 3, 6, 9, 12). The agarose gel after DNase I treatment showed that one unit enzyme results in DNA fragments of the optimum size for DNase I-MLPA, and three units of enzyme cause over digested DNA. **A.** DNase I digestion at day 3 and day 6 with 0 unit, 1 and 3 units enzyme in three populations (Thy1+, Thy1-, Ssea1+) **B.** DNase I digestion at day 9 and day 12 with 0, 1 and 3 units enzyme in three different populations.

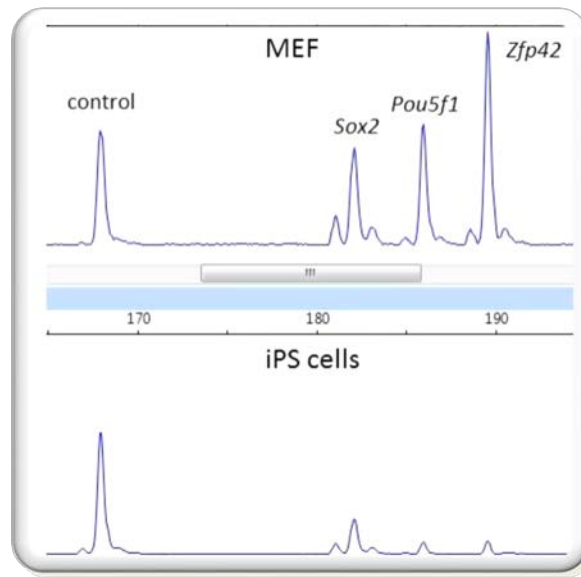


Figure 4. Screen capture of target gene peaks in iPS cells and MEF. Each peak corresponds to a specific genomic locus. The peaks for Sox2, Pou5f1 and Zfp42 in iPS cells are significantly reduced compared to MEF peaks, indicating that these genomic loci have been digested by DNase I as they are in regions of open chromatin.

To examine how the chromatin changes during reprogramming, DNase I hypersensitive sites in Ssea1⁺ and Thy1⁺ intermediates were then investigated by DNase I-MLPA. The result of DNase I digestion in Ssea1⁺ cells showed that the chromatin starts to become accessible for some candidates of the known target genes by day 3, with no further change until day 9. It showed more accessibility at day 12 and in iPSC (Figure 5), suggesting that chromatin becomes more open and accessible for regulatory proteins in the Ssea1⁺ population. The chromatin compaction for refractory cells (Thy1⁺) was similar to what was observed in the Ssea1⁺ population in the early days of reprogramming, and all targets are accessible for different protein complexes. However, in the Ssea1⁺ population the chromatin stayed open until the end of reprogramming and in iPSCs, it showed more relaxation during reprogramming, whereas in the Thy1⁺ population after early days of reprogramming the chromatin started to be compacted and consequently most of the targets were protected and inaccessible for DNase I digestion (Figure 6). Some target genes, such as Lefty1, Rif1 and Nanog, showed less compaction during reprogramming in the Ssea1⁺ population, whereas the chromatin showed more compaction at the end of reprogramming in the Thy1⁺ population (Figure 8). In contrast, Thy1, a fibroblast gene, showed the opposite pattern of chromatin accessibility in Ssea1⁺ and Thy1⁺ cells. There was low accessibility from day 3 onwards in the Ssea1⁺ population, and less compaction in the Thy1⁺ population at later time points (Figure 7). Overall, the data demonstrated that chromatin changes in refractory and reprogrammed cells was similar in early wave of reprogramming, however major changes occurred during the second wave of reprogramming. Furthermore, the chromatin compaction results were consistent with the gene activity profile. The relaxed state of the chromatin corresponds to an increase in gene expression. For example, for Lefty1, Nanog and Rif1 the chromatin becomes more relaxed and transcription of these genes increases (Figure 9).

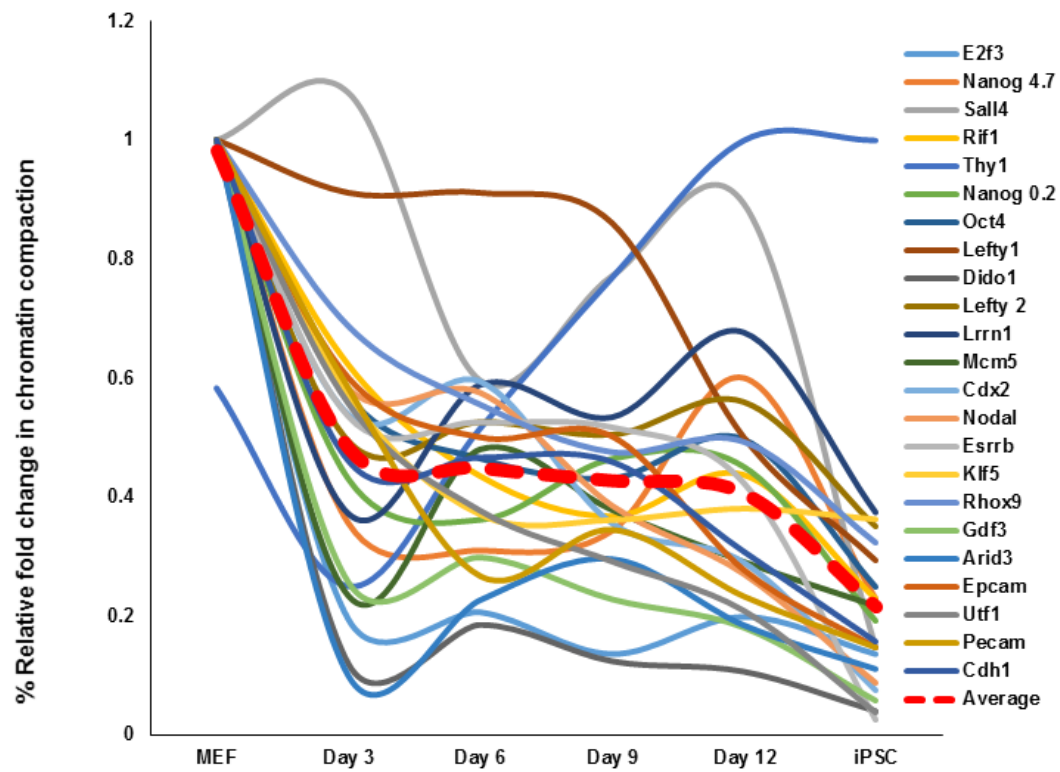


Figure 5. The average changes in Ssea1+ population during reprogramming. The dotted line shows the average change of chromatin compaction. The chromatin becomes increasingly more open at the end of reprogramming.

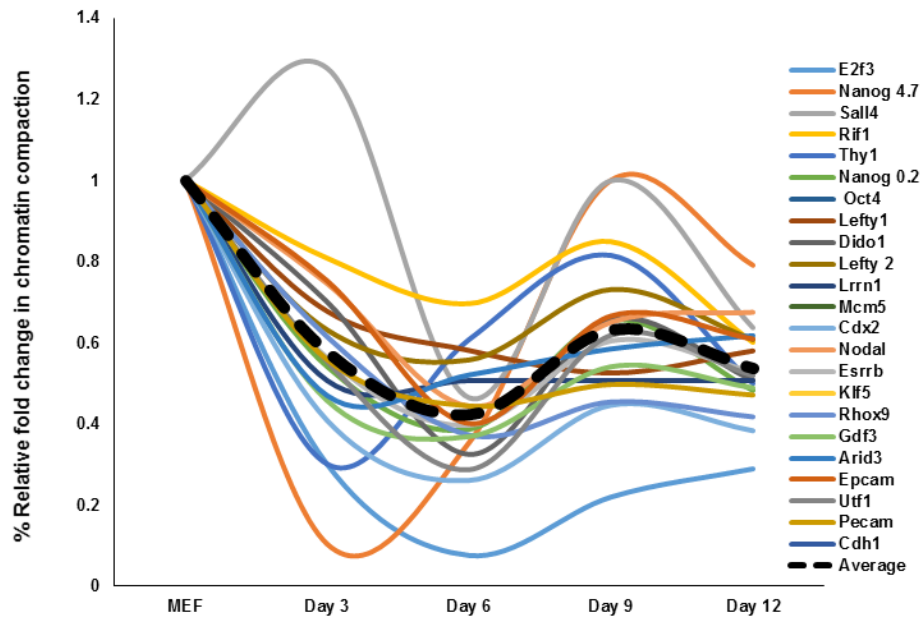


Figure 6. The average changes of chromatin compaction in Thy1+ cells during reprogramming. The 23 DNase I hypersensitive sites were examined in the Thy1+ population. The dotted line showed the average changes in these 23 target genes in terms of chromatin compaction in this population. Chromatin is accessible for DNase I digestion on day 3, with chromatin become closed from then onwards.

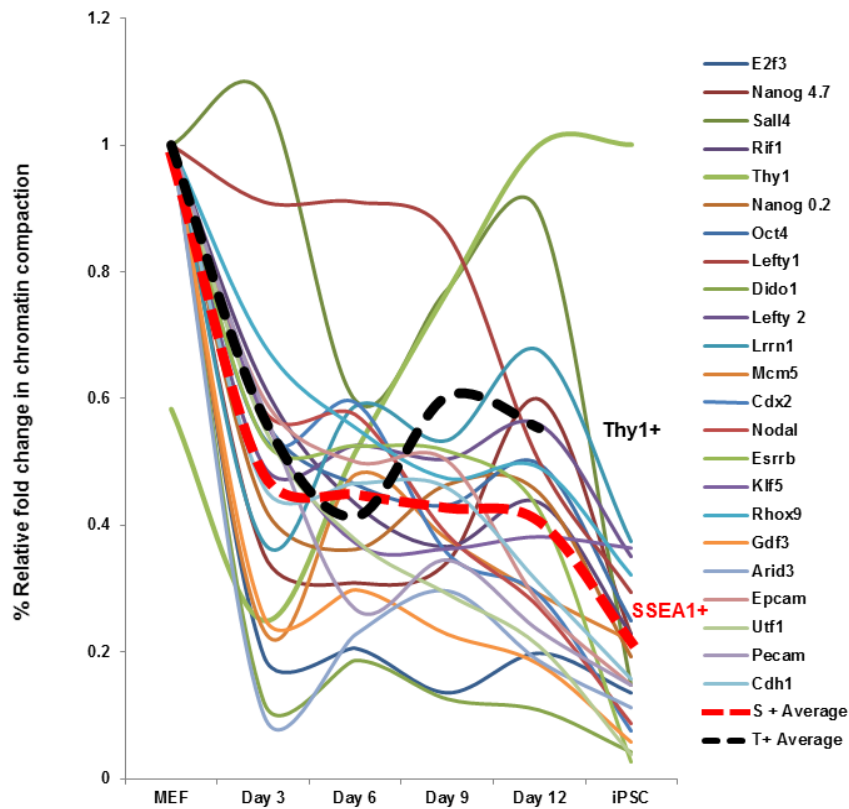


Figure 7. An overview of chromatin compaction in both populations. The results showed a similar trend in the Ssea1+ and Thy1+ populations at the early stages of reprogramming, until day 6. However the chromatin becomes more accessible in Ssea1+ cells (red dotted line), compared to the Thy1+ population where the chromatin become closed.

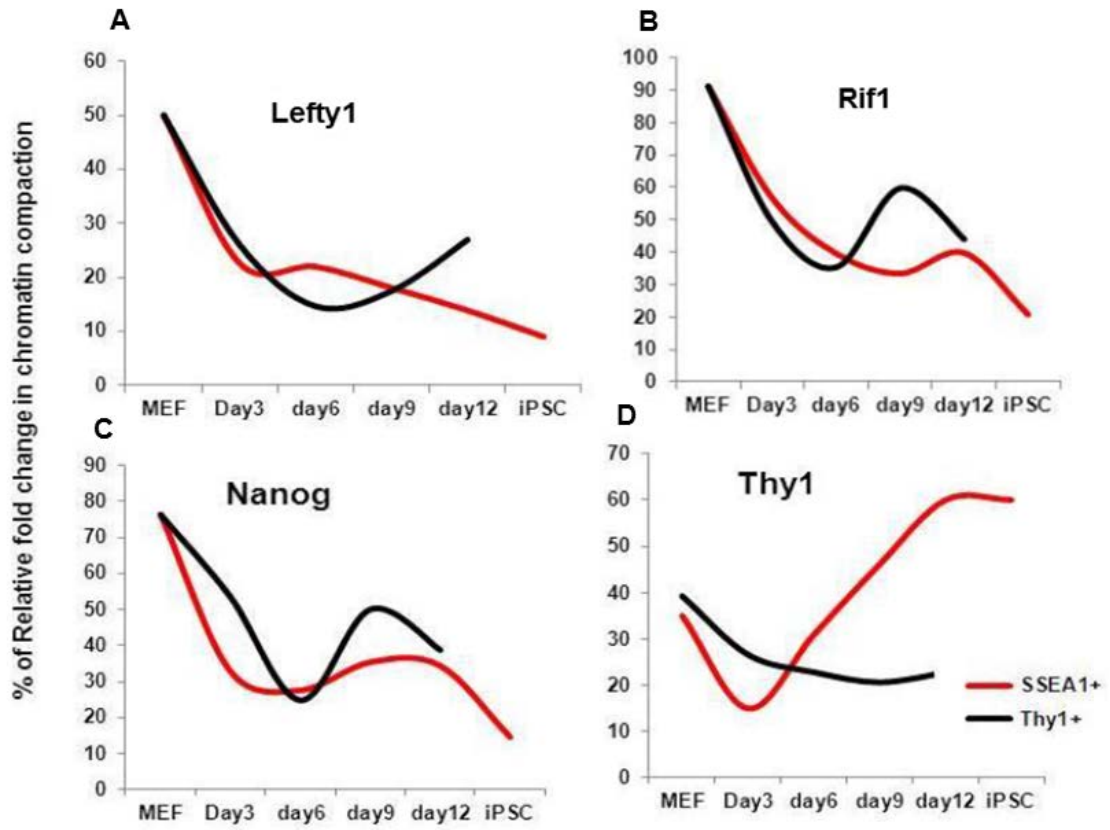


Figure 8. Chromatin compaction in positive and negative reprogramming markers. The Ssea1+ population showed low compaction in three positive target genes such as Lefty1, Rif1, Nanog (A, B, C) and high compaction observed in Thy1 gene as fibroblast gene (D). However, this trend is reversed in the Thy1+ population.

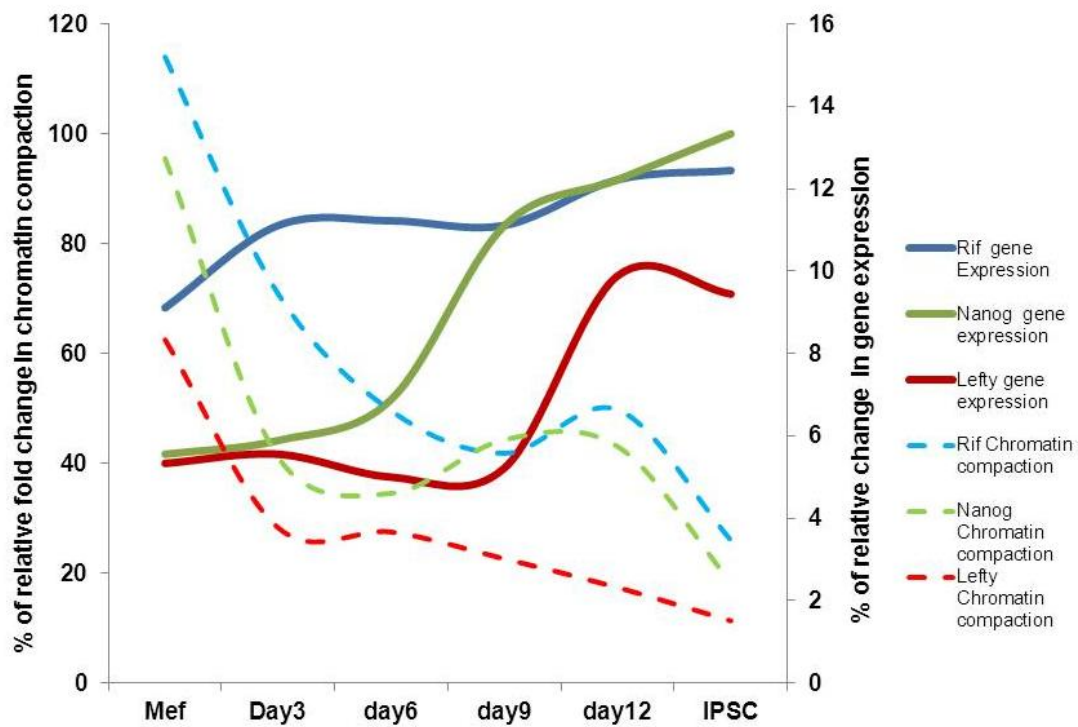


Figure 9. The relationship between gene expression and chromatin compaction. The comparison between gene expression and chromatin compaction in three Oct4 and Sox2 target genes such as Rif, Nanog and Lefty1. The more accessible the chromatin, the more the gene is expressed. The dotted line shows chromatin compaction and the thick line shows gene expression.

5.2.3 Identification of open region chromatin in refractory and reprogrammed cells during reprogramming

5.2.3.1. Changes in chromatin accessibility in refractory and reprogrammed cells on a genome wide scale in OSM targets during reprogramming

ATAC-seq was performed to identify regions of open chromatin on a genome-wide scale during reprogramming in refractory and reprogrammed cells. ATAC-seq is able to measure transcription factor occupancy, chromatin accessibility and nucleosome positioning (Buenrostro, Giresi et al. 2013). To study the differences between Thy1+ and Ssea1+ in chromatin accessibility at Oct4 and Sox2 target genes, 50,000 Ssea1+ and Thy1+ cells at different time points were sorted and treated with the transposase reaction, which is the integration step. In this step transposon integrates to open regions of chromatin. Then amplifiable DNA fragments in open regions were generated and followed by high-throughput sequencing. Thirty to 50 million sequencing reads give us the sufficient reads to identify spatial and temporal changes in regions of open chromatin. Moreover, ATAC-seq with higher numbers of sequencing reads (198 million paired-end reads) was performed in order to examine the nucleosome shifting during reprogramming. This detected nucleosomes within regulatory regions, as the most of reads are concentrated within open chromatin. The signal across all active TSS showed that the nucleosome-free fragments are enriched at a canonical nucleosome-free promoter region overlapping the TSS, while our nucleosome signal is enriched both upstream and downstream of the active TSS, and displays characteristic phasing of upstream and downstream nucleosomes. Therefore, a high frequency of mononucleosomes were observed close to TSS, suggesting that in these regions the chromatin is more relaxed and less compacted and consequently transcription activated.

Each peak may indicate regulatory elements where chromatin is accessible for the transposase, and therefore transcription factors or any other regulatory protein are able to bind. The results showed that the state of open and closed chromatin was different in refractory and reprogramming cells at various time points during reprogramming.

Analysis of the number of c-Myc target genes, such as E2f3, Mcm5, Zfp277, Rp139, and Kdelc1, were also investigated in Thy1+ and Ssea1+ cells. The data showed that the chromatin at the TSS is accessible in both Thy1+ and Ssea1+ cells for c-Myc targets (Figure 10). Then the chromatin accessibility in whole target genes of c-Myc was examined in Ssea1+ population, which showed that the TSSs in c-Myc targets are completely accessible. It appears that c-Myc acts as facilitator for both refractory (Thy1+ cells) and reprogrammed cells (Ssea1+ cells) (Figure 11A).

This result was consistent with gene expression profile, where a sharp downregulation of c-Myc targets at day 6 in Thy1+ cells was previously observed (Polo, Anderssen et al. 2012) (Figure 11B). This was consistent with the chromatin accessibility data showing that the chromatin becomes closed on day 6, suggesting that regulatory proteins are unable to access this region, resulting in no gene activity.

The analysis of chromatin accessibility in 784 Oct4 and 540 Sox2 targets ordered based on iPSC genes in Ssea1+ population showed no significant differences during reprogramming in TSS (Figure 12). However, investigation of some of the known targets of Oct4 and Sox2 (Zic5, Nodal, Sox2, Nanog and Pou5f1) in more detail in Thy1+ and Ssea1+ populations showed the chromatin was not accessible in Thy1+ cells, however the opposite was observed in Ssea1+ cells during reprogramming. In Ssea1+ cells, the chromatin started to open on day 3 and it becomes more and more accessible on day 12 and in iPS cells. Gene

expression analysis also showed the gradual upregulation of all Oct4 and Sox2 target genes in Ssea1+ cells (Figure 12 A, B).

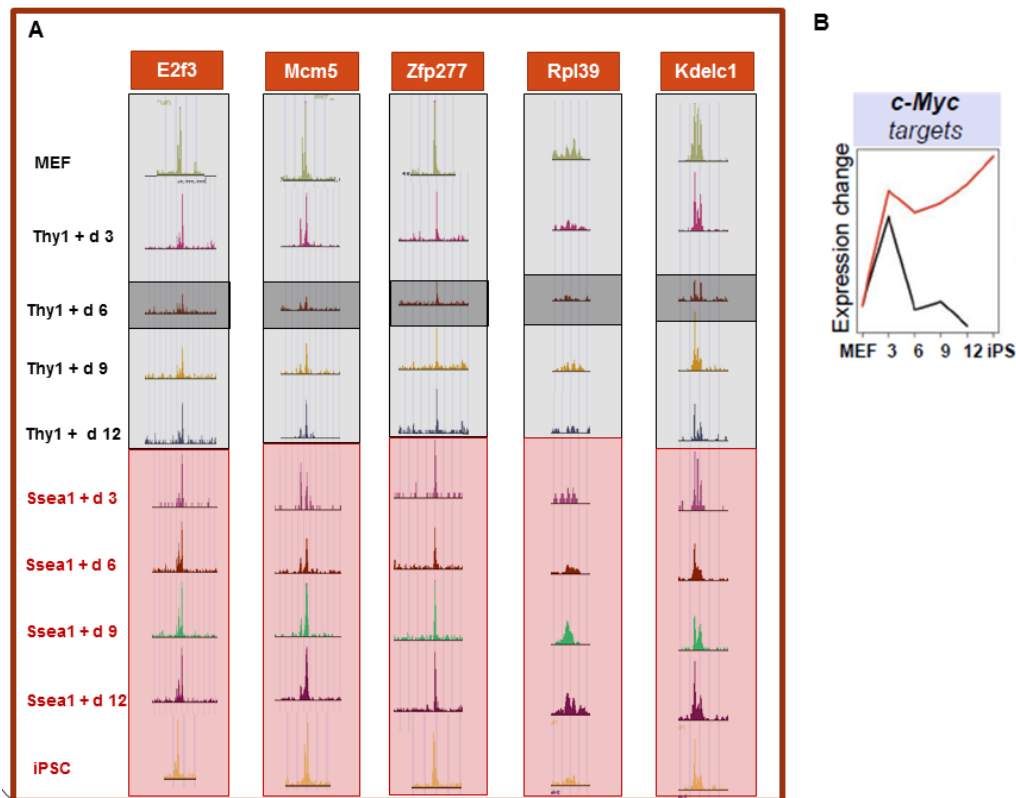


Figure 11. A. Chromatin accessibility in some of c-Myc targets in Ssea1+ and Thy1+ cells. The chromatin is accessible for c-Myc targets in Thy1+ and Ssea1+. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. **B.** Gene expression in all c-Myc targets (Polo, Anderssen et al. 2012).

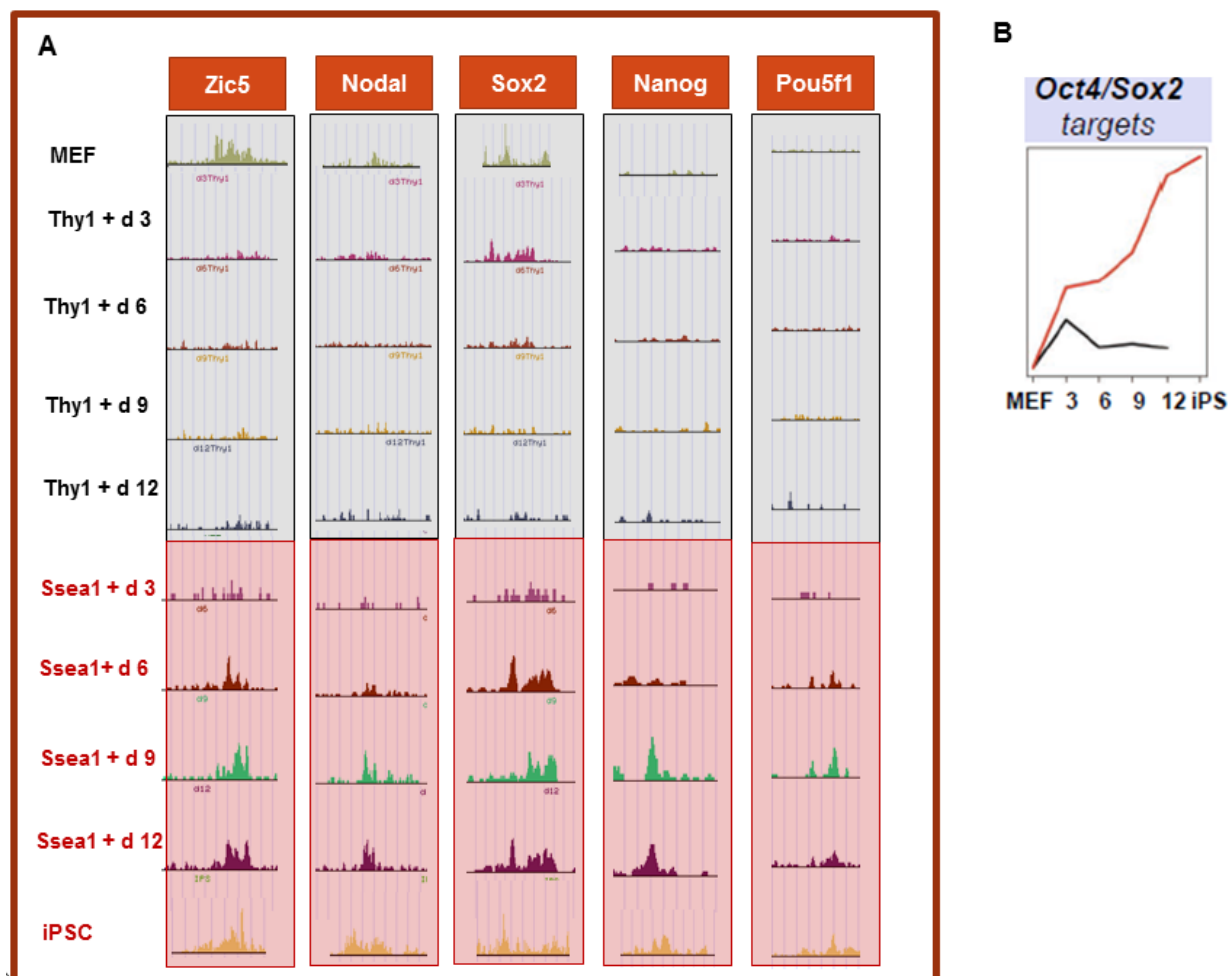
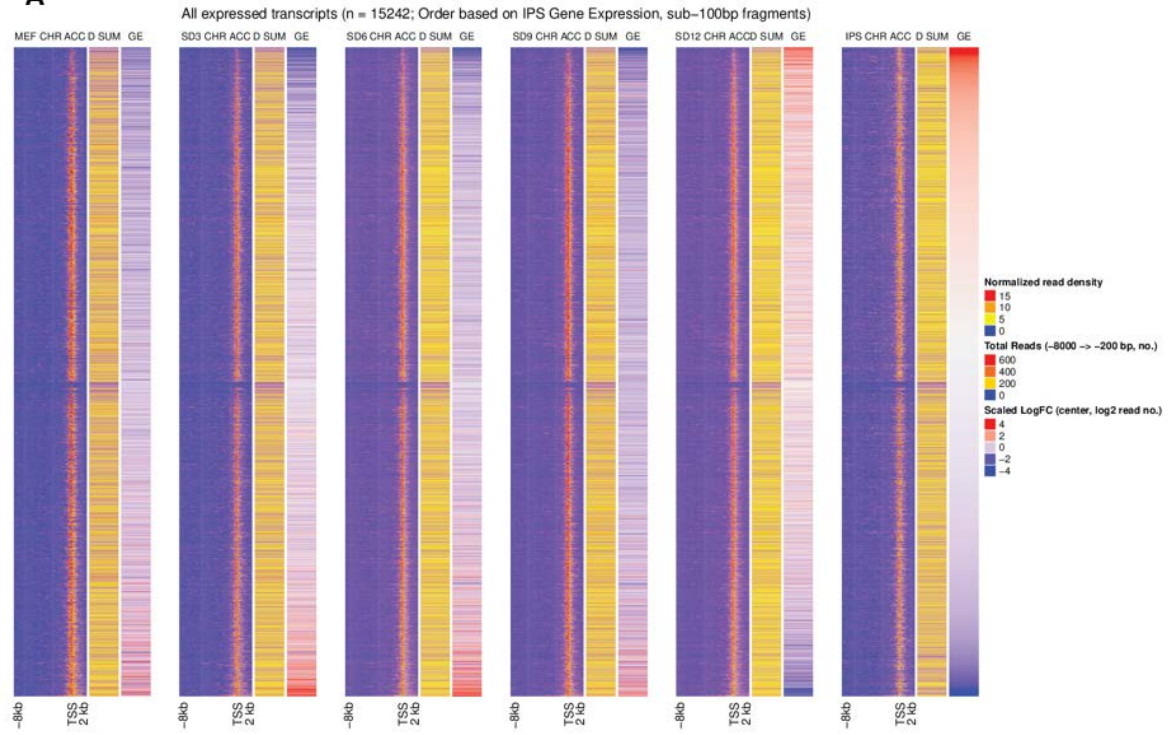


Figure 12. Chromatin accessibility in some of OSK targets in Ssea1+ and Thy1+ cells. A. The chromatin is not accessible for Oct4 and Sox2 targets in Thy1+ cells, however is open in Ssea1+ cells. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. **B.** Gene expression in all Oct4/Sox2 targets (Polo, Anderssen et al. 2012).

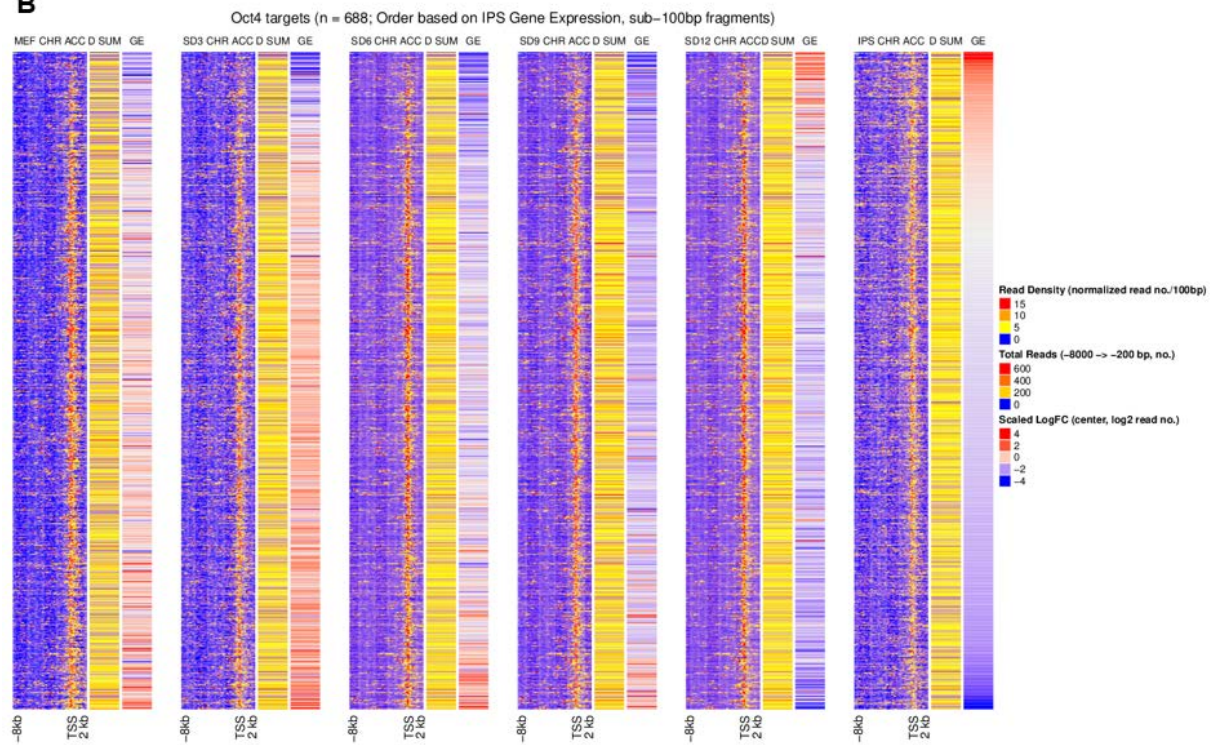
5.2.3.2. Is gene activity related to chromatin accessibility?

To examine the relation between the gene expression and chromatin accessibility, the open region chromatin based on gene activity of pluripotent genes in whole genome was analysed. In this study we analysed the area located 8kb upstream of TSS and 2kb downstream of TSS to determine all regulatory elements around TSS of OKSM target genes. The result showed that despite observed transcriptional changes in pluripotent and somatic genes, the changes in chromatin accessibility in TSS region were not strong. For Oct4 (n=688), Sox2 (n=560), Klf4 (n=1593) and c-Myc (n=3256) targets, we observed the downregulation of somatic cell-associated genes and upregulation of pluripotency-associated genes, however, chromatin accessibility in TSS did not show a direct correlation with transcriptional changes genome-wide. Interestingly, a closer look at the regulatory region between 8000bp and 200bp upstream of the TSS showed noteworthy changes in chromatin accessibility. The chromatin accessibility showed clear changes in Sox2 targets at regulatory regions (Figure 13B). We observed two waves of changes in chromatin accessibility in Sox2 targets: one wave, observed from MEF to day 9, and a second wave, observed in day 12 and iPSC. Although a lot of differences were observed in chromatin accessibility of Oct4, Klf4 and c-Myc targets between somatic cells and iPSC, there were not significant changes during different time points during reprogramming (Figure 13C, D, E).

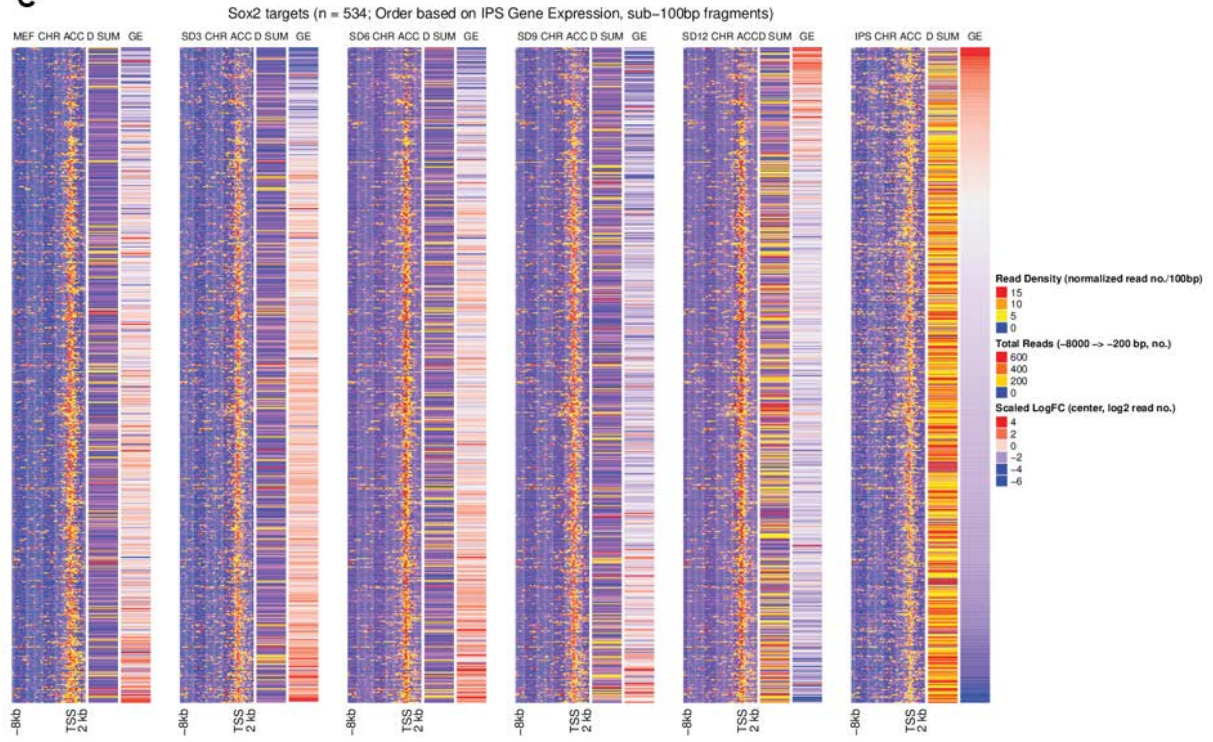
A



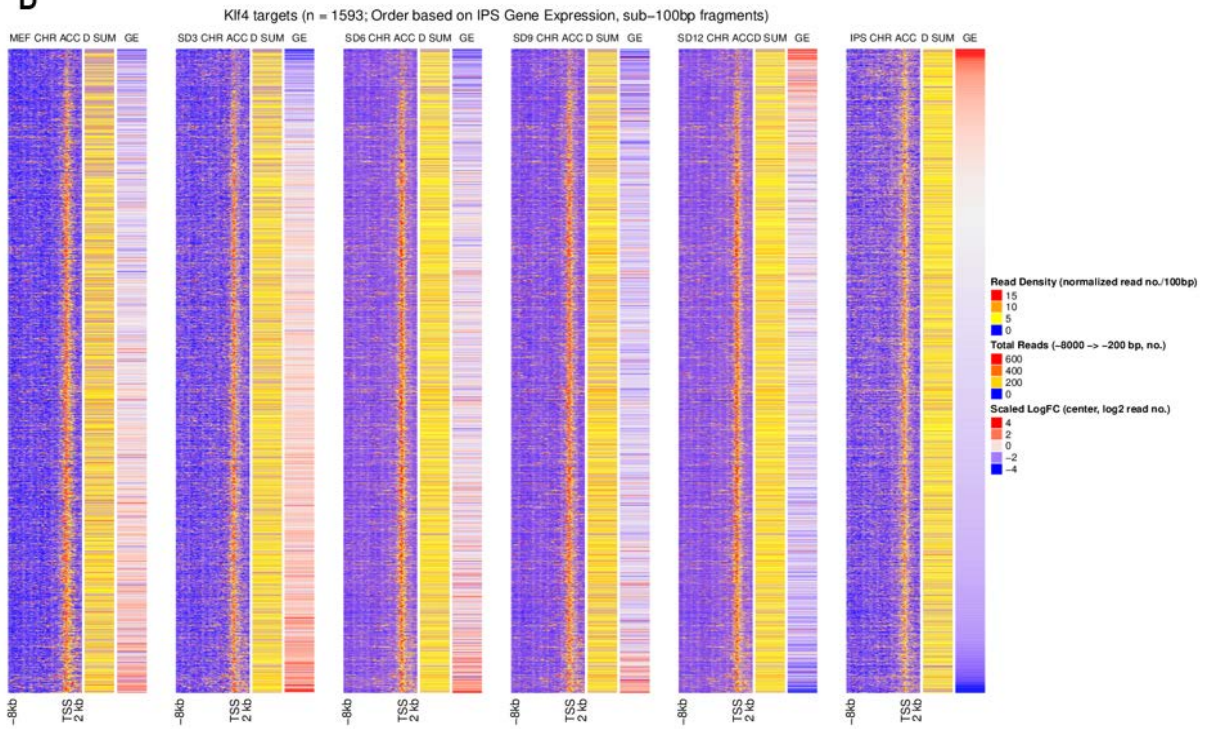
B



C



D



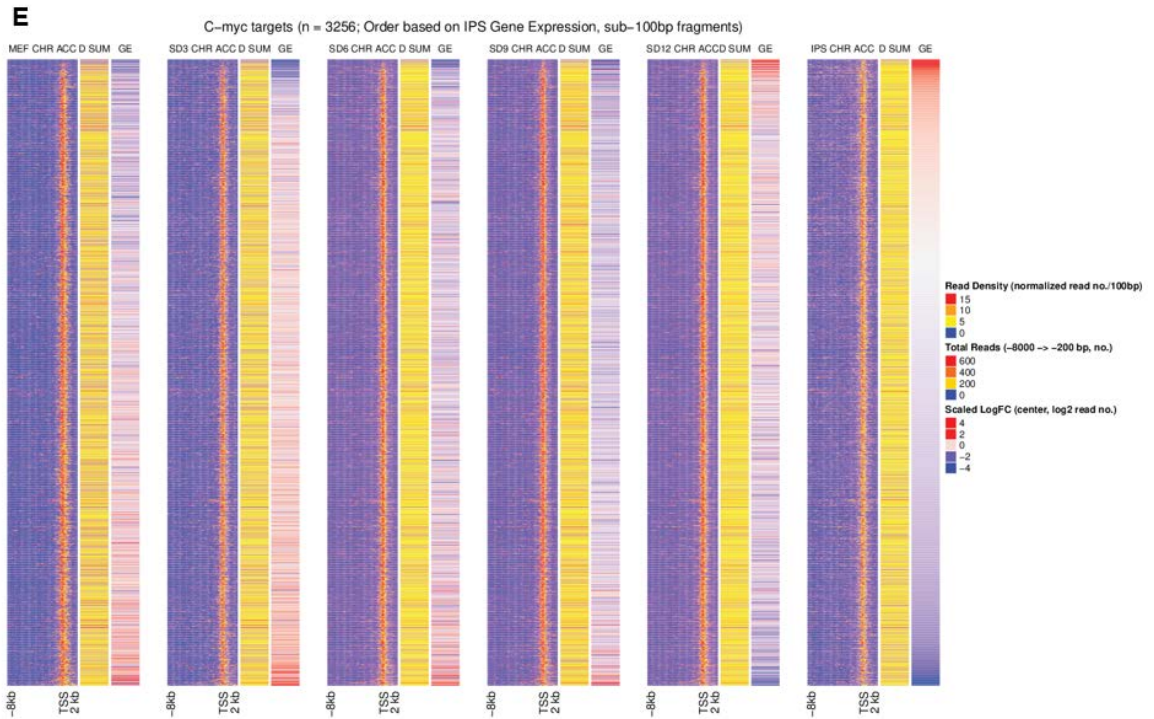


Figure 13: The correlation between gene activity and chromatin accessibility in whole genome wide in *Ssea1*⁺ cells. First column shows the heatmap with normalised read density. High density of open chromatin was shown in red and low density of open chromatin region is shown in blue. Second column shows the summation of chromatin status between -8kb and -200bp to focus on only the regulatory element in the promoter for each gene and the third column shows gene expression, all ordered according to iPSCs genes. G, E shows gene expression changes and CHR ACC showed chromatin accessibility. The changes from MEF, Day 3 (SD3), Day 6 (SD6), Day 9 (SD9), Day 12 (SD12). **A.** Chromatin accessibility and gene expression in all genes. **B.** Chromatin accessibility, gene expression in Oct4 target genes **C.** Chromatin accessibility and gene expression in Sox2 target genes. **D.** Chromatin accessibility and gene expression in Klf4 target genes. **E.** Chromatin accessibility and gene expression in c-Myc target genes.

Although we did not observe clear patterns at the genome-wide level, investigation of several target genes in more detail showed different expression patterns during reprogramming. The analysis of these target genes showed multiple peaks at promoters, and also upstream and downstream of the genes. For example, *Utf1* becomes upregulated at the end of reprogramming and two peaks were identified, one near the TSS and another one located downstream of TSS. This suggests that transcriptional activity of this gene may be regulated by (at least) two regulatory elements. *Vcam1* is downregulated during reprogramming, and four peaks were found close to TSS that might impact on its expression, three of them located downstream of the gene and another one located in TSS. Two peaks located downstream of the *Vcam1* disappeared from day 3 onwards, suggesting that chromatin at these loci is closed during early days of reprogramming, leading to downregulation of *Vcam1*. Interestingly, *Cxcr4* is one of the target genes that is upregulated in the early wave of reprogramming before becoming downregulated in the late wave of reprogramming. Two open chromatin peaks near the gene were found, one at promoter, and another one at downstream of TSS. Gene activity was linked to the peak located downstream of the gene, suggesting that this locus, not the one located close to TSS, may regulate transcriptional activity.

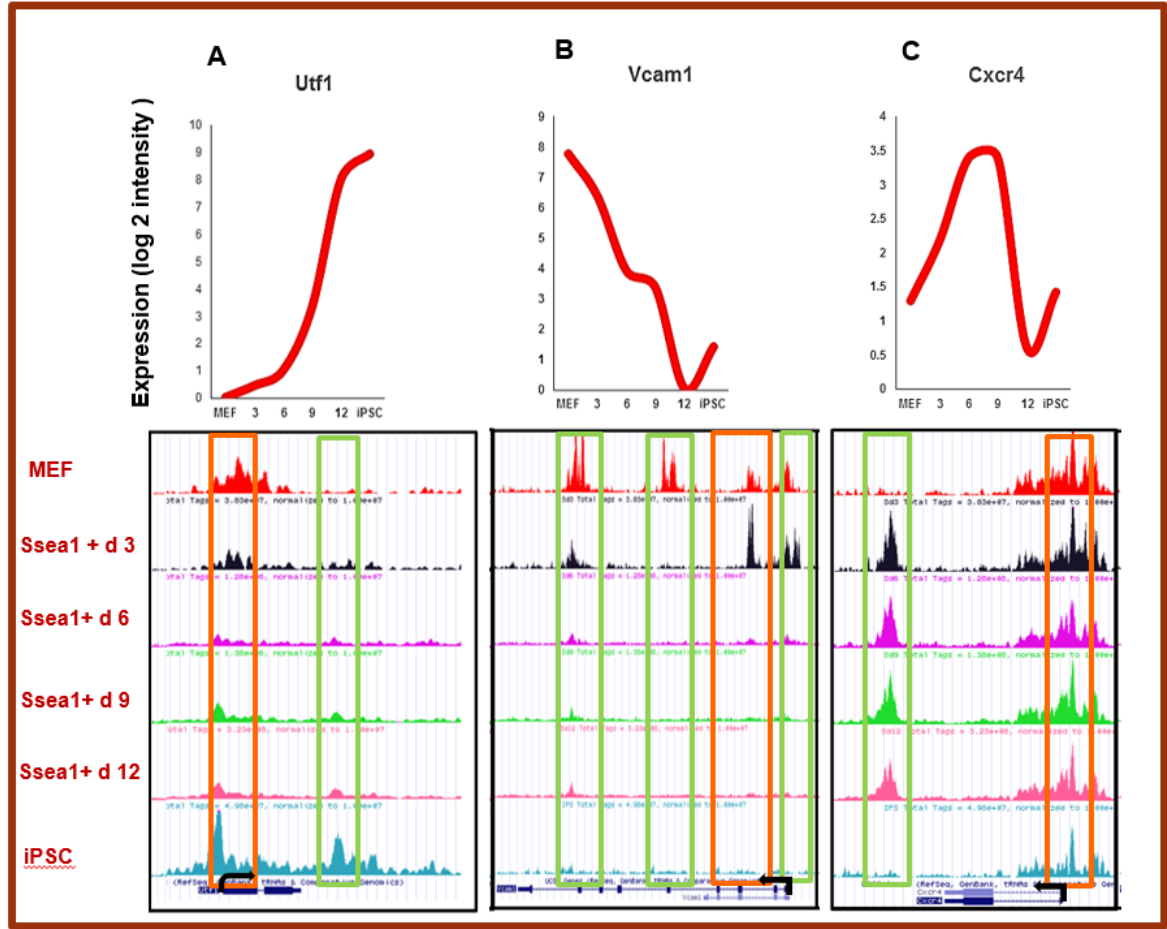


Figure 14. The association between gene activity and chromatin accessibility. Upper graphs depict transcriptional changes and the lower graphs the corresponding ATAC-seq tracks. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. The TSS region (red box) and other regulatory peaks (green box) are indicated. **A.** *Utf1* showed upregulation at the late stage of reprogramming and gradual accessibility in the promoter region. **B.** *Vcam1* with downregulation and low accessibility in regulatory region during reprogramming. **C.** *Cxcr4* with upregulation at the early wave of reprogramming with downregulation at the end of reprogramming. The data supports the downstream regulatory region regulating gene expression.

5.2.3.3. The binding sites for Oct4 and Sox2 are located in regions of open chromatin in iPSC

In Chapter 4, the binding of Oct4 and Sox2 to their targets during reprogramming showed differences in binding at different time points (see Chapter 4: Figure 13, 14). However, it was not clear whether these differences in binding were associated with changes in accessibility to their target genes. In order to address this, the chromatin accessibility at several Oct4 and Sox2 target genes with different chromatin and binding patterns were studied. The binding sites of Oct4 and Sox2 were located in regions of open chromatin in iPSCs, however at some time points the chromatin was not accessible. In the next section we have analysed the association between open regions of chromatin in Oct4 and Sox2 binding sites for some of their targets.

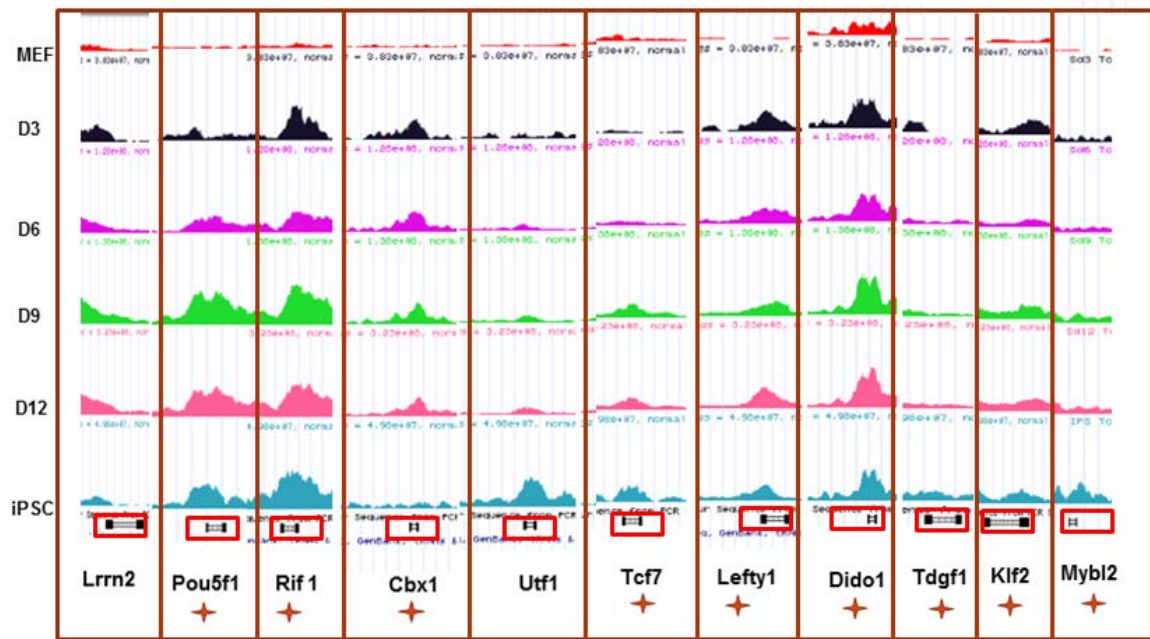


Figure 15. The comparison between location of Oct4 and Sox2 binding at some of their targets and chromatin accessibility. ChIP was performed on the genes marked with the star symbol (See Chapter 4: Figure 13, 14). In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment.

5.2.3.4. The correlation between Oct4 and Sox2 binding sites and transcriptional activity based on chromatin accessibility

Several genes were studied to better understand the relationship between binding site accessibility and gene expression. Mybl2 is an example of consistency between gene expression, TF binding and chromatin accessibility. It is regulated by gradual accessibility of Oct4 and Sox2 to their target loci (Figure 17B, C), which is consistent with gradual upregulation of Mybl2 during reprogramming (Figure 17D). The existence of an Oct4 and Sox2 binding site located 2.2Kb upstream of the Mybl2 TSS was previously shown (Kim, Chu et al. 2008). The analysis of chromatin accessibility and Oct4 and Sox2 binding showed that the Mybl2 regulatory region was bound gradually by both factors during reprogramming (Figure 17A). Although another regulatory region was found close to the TSS that might have an effect on Mybl2 regulation, it seems that control of this gene is most likely associated with the region where Oct4 and Sox2 bind. Moreover, the high frequency of mononucleosomes was consistent with gene activity. Therefore, Mybl2 is an example of correlation between chromatin accessibility and gene activity.

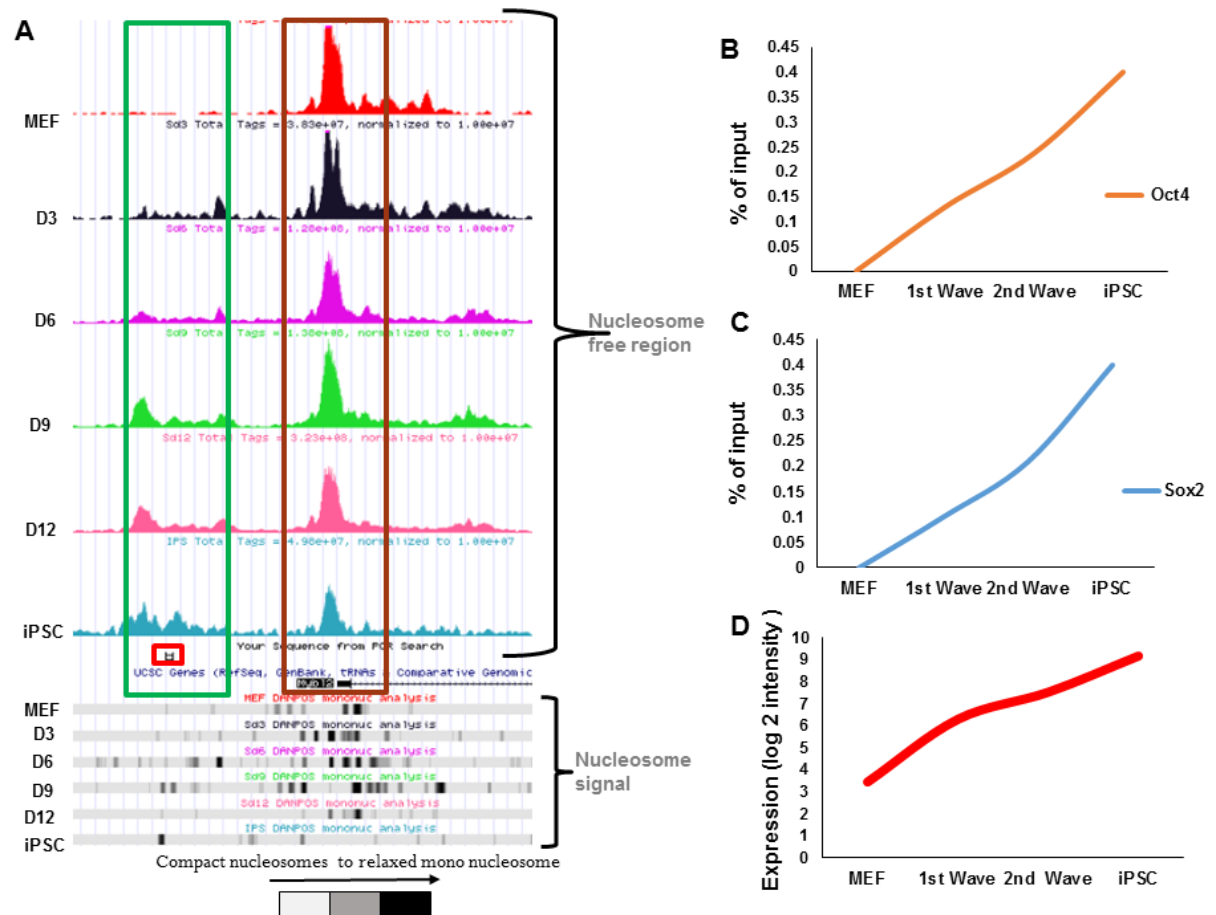


Figure 16. Gradual accessibility of Oct4 and Sox2 targets. The consistency among chromatin accessibility, binding site and gene expression was observed in several targets such as *Mybl2*. **A.** Chromatin accessibility, red box showed the binding sites for Oct4 and Sox2 at *Mybl2* locus. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. **B.** Oct4 signal in *Mybl2* regulatory region. **C.** Sox2 signal in *Mybl2* regulatory region. **D.** Gene expression.

The direct correlation between Nodal activity and chromatin accessibility is not as straightforward. Nodal is enriched for Oct4 and Sox2 during the first wave of reprogramming and chromatin was accessible for both factors from day 3 onwards. However, during the second wave, low enrichment for both factors was observed although the chromatin was open in this region. Gene activity showed gradual upregulation until the end of reprogramming, suggesting other factors might be involved in Nodal expression. Therefore, Nodal expression at early days of reprogramming may be because of strong binding of only two reprogramming factor (Oct4 and Sox2), however expression in the late stage of reprogramming might be because of other reprogramming factors binding such as Klf4, Nanog etc. The high frequency of mononucleosomes observed in binding sites suggests the chromatin becomes relaxed and this region is more accessible to regulatory proteins (Figure 16).

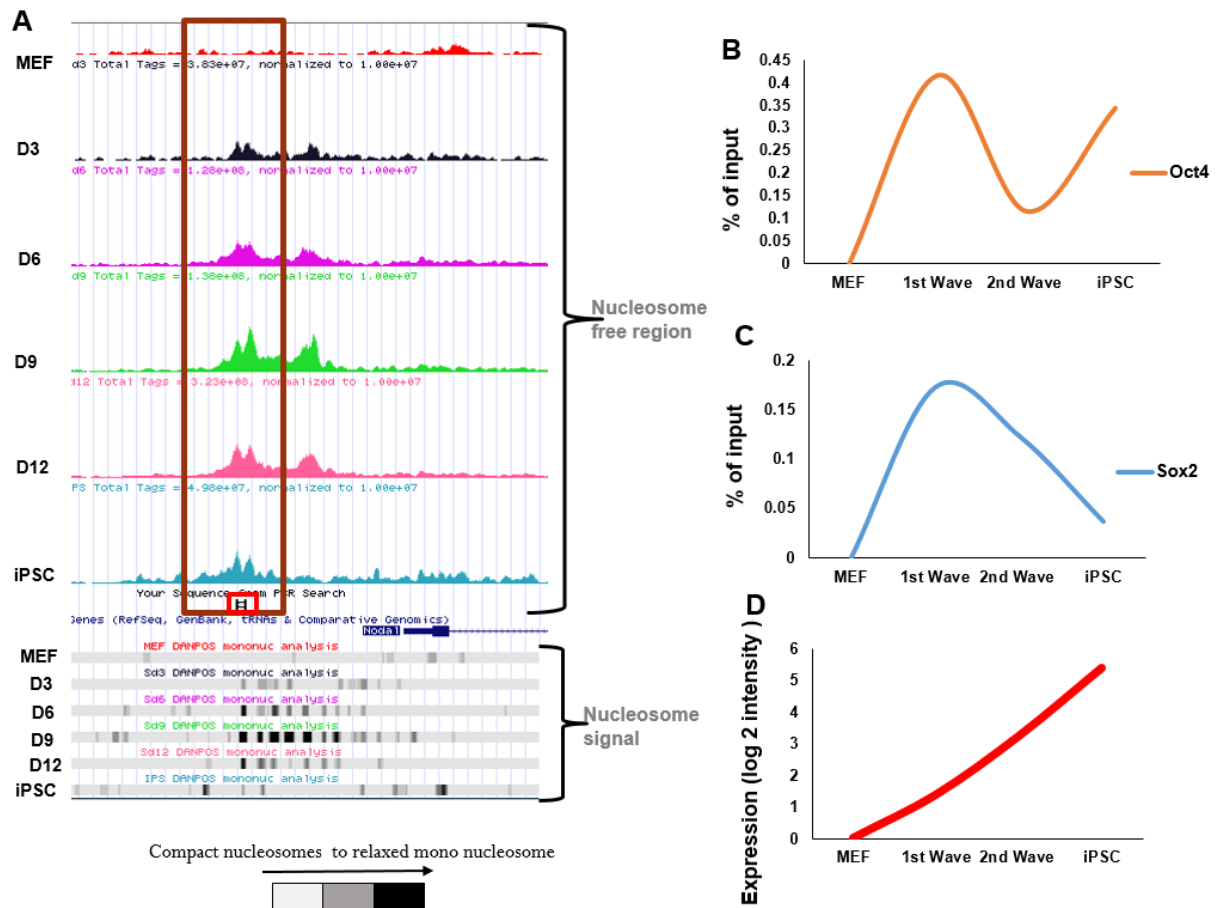


Figure 17. The association of binding site, chromatin accessibility and gene expression.

Nodal is one example that shows that the open region of chromatin might be because of complex of protein A. Chromatin accessibility in a nucleosome-free region during reprogramming. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. **B.** ChIP signal in Oct4 binding site. **C.** ChIP signal in Oct4 binding site. **D.** Gene expression

Another correlation between chromatin accessibility and Oct4 and Sox2 binding exemplified by Lefty1 that showed gradual binding of Oct4 and Sox2 with little change during the second wave of reprogramming and chromatin become accessible from day 3 onward. However Lefty1 is highly expressed at the end of reprogramming on day 12 and in iPSC. Although the chromatin is open and accessible from day 3 onwards and Oct4 and Sox2 bind to their target gradually, no significant expression of Lefty1 was observed, suggesting that Oct4 or Sox2 might play as gene repressors during the first wave of reprogramming and other regulatory proteins activate Lefty1 expression. We observed another regulatory region upstream of the TSS which is rapidly closed after initiation of reprogramming, suggesting this region might be related to somatic cell-associated elements. Furthermore, high frequency of mononucleosomes were observed at late days of reprogramming which is consistent with gene activity (Figure 18).

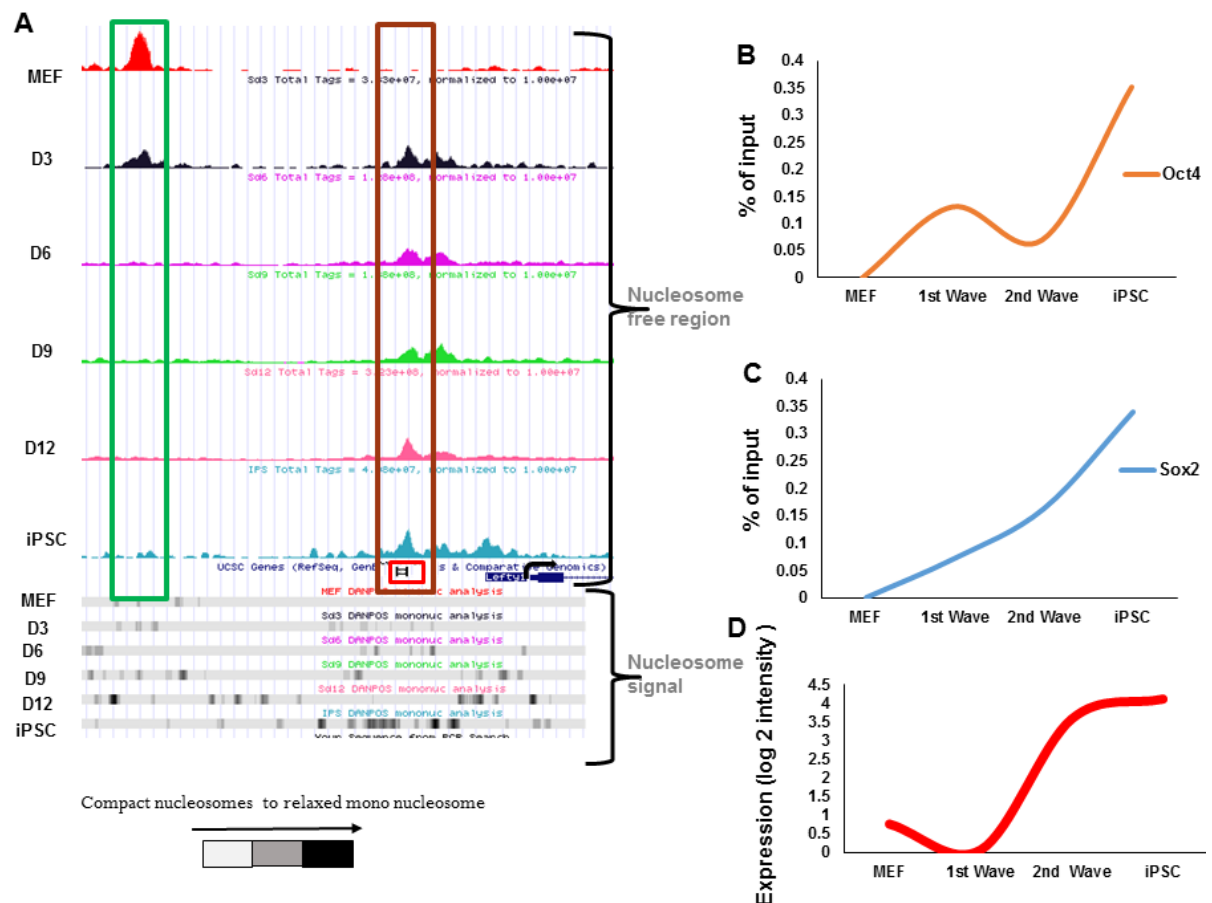


Figure 18. The association of binding site, chromatin accessibility and gene expression.

Chromatin in Lefty1 is accessible however low binding of Oct4 and Sox2 was observed **A**. the chromatin accessibility in nucleosome free region during reprogramming. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. **B**. ChIP signal in Oct4 binding site **C**. ChIP signal in Oct4 binding site **D**. Gene expression

One interesting example is observed at the Pou5f1 promoter, where two regulatory peaks can be detected. One located close to the TSS and the other one at a distal region. Although Pou5f1 activity is regulated by the distal region in iPSC, our results showed that its expression might be controlled by another regulatory element located close to the TSS in intermediates during reprogramming. Therefore, gene expression might be regulated by a switching between these two regulatory elements during reprogramming. The ChIP result showed gradual binding of Oct4 and Sox2 which is consistent with chromatin accessibility from day 3 onwards. These results suggested that the regulatory region located at the binding site of Oct4 and Sox2 might regulate Pou5f1 expression in intermediates during reprogramming (Figure 19).

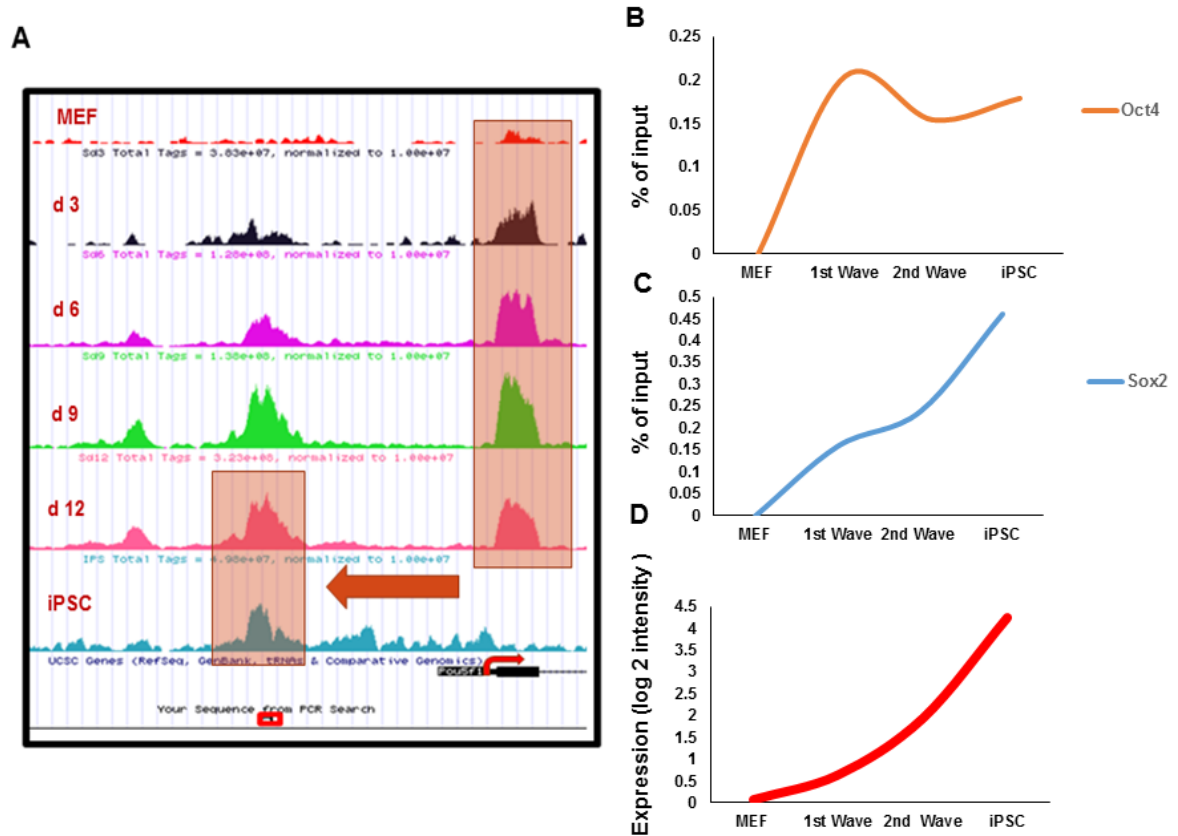


Figure 19. The regulation of Pou5f1 by two regulatory regions. A. Chromatin accessibility at Oct4 and Sox2 binding site and in TSS. The red box showed the Oct4 and Sox2 binding site. In the Y axis ATAC-seq signal is depicted in different intermediates during reprogramming. The signal is scaled from the minimal to the maximum signal across the entire experiment. **C.** Binding pattern of Oct4 and Sox2 during reprogramming. **D.** Transcriptional activity

In summary, for some target genes there was a correlation among chromatin accessibility, transcription factors binding and gene activity. This was not observed for all target genes, however, suggesting that Oct4 and Sox2 binding and chromatin accessibility are not sufficient factors for gene regulation.

5.3. Conclusion:

The examination of chromatin accessibility during reprogramming showed a number of interesting features. First, chromatin was accessible at most of c-Myc target genes in Thy1+ and Ssea1+ populations, suggesting that c-Myc acts as a chromatin facilitator. Second, although chromatin showed accessibility in the Ssea1+ population at Oct4 and Sox2 targets, chromatin at the same loci was compacted in Thy1+ population, which may explain the lack of success of Thy1+ population in iPSC formation. Third, the correlation between gene activity and chromatin accessibility is not always straightforward, suggesting a role for other factors in gene regulation. Therefore using different methods such as DNase-MLPA with high number of cells (1-2 million cells) and ATAC-seq with a low number of cells (50,000 cells) showed the chromatin becomes accessible in Ssea1+ population compared to Thy1+ population during reprogramming. Moreover, the mononucleosomes concentrated in TSS regions suggested that these chromatin regions are more relaxed and accessible for transcription factors. The examination in genome-wide datasets showed changes in chromatin accessibility at regulatory regions (8kb upstream and 2kb downstream) in OKSM target genes between MEF and iPSCs. The changes during reprogramming in chromatin accessibility showed that the Sox2 target genes become accessible at regulatory region downstream of TSS.

CHAPTER 6

General discussion and future directions

6.1. Different reprogrammable mouse models

When this study was initiated, the only known OKSM reprogrammable mouse was the M2rtTA OKSM mouse model. Based on previous studies from us and others, several facts about reprogramming of M2rtTA OKSM reprogrammable mouse were known: 1. The low efficiency of het-het OKSM M2rtTA with 1-5% efficiency (Stadtfeld, Maherali et al. 2008, Stadtfeld, Maherali et al. 2010, Polo, Anderssen et al. 2012), 2. Three different cell populations could be defined during reprogramming of mouse embryonic fibroblasts based on the surface markers Thy1 and Ssea1: Thy1⁺ cells, Thy1⁻ cells and Ssea1⁺ cells (Stadtfeld, Maherali et al. 2008, Polo, Anderssen et al. 2012), 3. It was shown that only Ssea1⁺ cells give rise to iPSC whereas the Thy1⁺ and Thy1⁻ populations are unable to generate iPSC and were consequently referred to as refractory cells (Polo, Anderssen et al. 2012), 4. The level of the OKSM protein in Thy1⁺ cells is lower than in Ssea1⁺ cells suggesting this might be the reason for these cells being refractory to reprogramming (Polo, Anderssen et al. 2012). This consequently suggests that the levels of the OKSM factors play an important role in reprogramming. This was further supported by a study which showed that low amounts of Klf4 were able to repress somatic cell associated genes but it was not sufficient for completing reprogramming (Wei, Gao et al. 2013) and consequently it might impact on the generation of partially reprogrammed cells and increase the number of refractory cells. The level of Oct4 also impacted on reprogramming by cooperation with several complexes such as the polycomb complex to maintain pluripotency (van den Berg, Snoek et al. 2010). Moreover, the amount of Oct4 affects the chromatin structure to

maintain stem cell identity by regulating the expression of genes which encode H3K9 demethylases such as Jmjd1a and Jmjd2c (Ding, Xu et al. 2012). Therefore the presence of just one of the reprogramming factors can result in a cascade of downstream events. Due to the impact of the level of Oct4 and other transcription factor expression in reprogramming, we hypothesised the overexpression of the four OKSM factors might rescue the refractory cells. The results revealed that refractory cells transduced with high levels of OKSM were able to form iPSCs (Polo, Anderssen et al. 2012). Our result showed that with high level of OKSM the refractory cells could be rescued. Interestingly, the level of Oct4 mRNA was similar in Ssea1+ and Thy1+, however, the levels of the OKSM proteins were lower in Thy1+ cells (Polo, Anderssen et al. 2012) which might be because of low stability of the proteins or their degradation.

Although we successfully rescued the refractory cells in the M2rtTA mouse model (Polo, Anderssen et al. 2012), the efficiency of reprogramming was low with a low percentage of reprogramming cells (Ssea1+ cells) in the het-het M2rtTA mouse model. Despite our previous study which revealed the homo-homo M2rtTA mouse model generates a higher number of Ssea1+ cells compared to het-het M2rtTA mouse model (Polo, Anderssen et al. 2012), the high levels of Oct4 (Gidekel, Pizov et al. 2003, Campbell, Perez-Iratxeta et al. 2007, Radzisheuskaya and Silva 2014) and also c-Myc (Thompson, Challoner et al. 1985) within homo-homo mice leads to tumourigenesis and consequently it was hard to maintain this mouse. Given our finding that refractory cells can be rescued by over expression of the four OKSM factors and a previous study had shown that M3rtTA results in high level of gene expression (Dow, Nasr et al. 2014) we hypothesised that an M3rtTA and OKSM mouse would lead to a higher reprogramming efficiency than the M2rtTA mouse. We therefore generated this mouse model by a cross between M3rtTA and OKSM mice.

The levels of Oct4 and Sox2 was determined in the M3rtTA mouse model and in agreement with our hypothesis, it was found that the M3rtTA mouse had higher levels of Oct4 and Sox2 than the M2rtTA mouse. As predicted, the reprogramming efficiency increased via over expression of four factors. Therefore beside the various methods that are able to increase reprogramming efficiency using small-molecule compounds such as DNA methyltransferase inhibitor, 5-aza-cytidine (AZA), and histone deacetylase (HDAC) inhibitors, Valporic acid (VPA), Suberoylanilide hydroxamic acid (SAHA) (Huangfu, Maehr et al. 2008), and also depletion of Mbd3 (Brumbaugh and Hochedlinger 2013, Rais, Zviran et al. 2013, Xu, Liu et al. 2013), here we found the substitution of M3rtTA instead of M2rtTA in the reprogrammable mouse model is also able to increase the reprogramming efficiency. In addition to the over-expression of the OKSM factors in the M3rtTA mouse, the higher reprogramming efficiency could have several reasons such as changes in the epigenetic status in the M3rtTA mouse compared to the M2rtTA. The previous studies showed that histone modifications such as H3/H4 acetylation and H3K4me3 methylation are associated with active chromatin and H3K9, H3K27 me3 are related to gene repression (Turner 2002, Hahn, Wu et al. 2011, Rao, Dhele et al. 2015). The reprogramming efficiency is also associated with different chromatin modifiers such as UTX which is a H3K27 demethylase which cooperates with Oct4, Sox2 and Klf4 and eliminates the repression of H3K27 from early pluripotency gene and SETDB1 (H3K9 demethylases) that reset the epigenome of somatic cells (Biran and Meshorer 2012). Moreover, SWI/SNF which acts as a chromatin remodelling complex and the Wdr5 complex, which is the core of the Trithorax complex and interacts with Oct4 and facilitates Oct4 activation. Oct4 interacts with several chromatin modifying complexes and epigenetic regulatory protein complexes which are essential for iPSC generation (Ding, Xu et al. 2012). According to our research, the increase in the protein level of Oct4 and Sox2 in the M3rtTA mouse might have an

effect on the protein complexes that interact with Oct4 and histone methylation and demethylation, which lead to higher efficiency.

It is of interest to know why low levels of Oct4 are unable to mediate reprogramming. Is it because of a lack of transcription factor accessibility to the targets, or the overexpression of the four factors that causes nucleosome shifting and changes histone modification?

Therefore, our finding in generation of high efficiency reprogramming mouse will help us to dissect molecular events toward iPSCs, which is critical to understand the black box of reprogramming.

6.2. The effect of chromatin dynamics in transcription during reprogramming

The OKSM cause conversion of somatic cells to pluripotent cells and each of these factors might play different roles during reprogramming. Several studies have focused on changes in gene expression and histone modifications during reprogramming (Koche, Smith et al. 2011, Polo, Anderssen et al. 2012) and two waves of reprogramming were defined. It was revealed that the genes involved in the first wave of reprogramming were different from the genes involved in second wave of reprogramming (Polo, Anderssen et al. 2012): genes associated with proliferation and metabolism vary during the first wave; whereas the second wave is related to pluripotent associated genes. Also, different gene clusters were identified with upregulation and downregulation at different time points. Some of the clusters first showed upregulation and then downregulation, others showed a gradual upregulation, whilst some were downregulated initially and then became upregulated until the end of reprogramming (Polo, Anderssen et al. 2012). However the temporal and spatial binding of the OKSM transcription factors to their target genes during reprogramming was unclear (Polo, Anderssen et al. 2012). The question is whether there is any correlation

between these different gene clusters and the binding of transcription factors and chromatin accessibility.

In 2012, Soufi *et al.* showed the initial binding of OKSM in human reprogramming (Soufi, Donahue et al. 2012). The reprogramming transcription factors revealed significant differences in binding sites between initial days of reprogramming and iPSC. Oct4, Sox2 and Klf4 initially bind to distal conserved elements or enhancers prior to binding to the promoters (Soufi, Donahue et al. 2012). They also found that unlike Oct4, Sox2 and Klf4, c-Myc is unable to bind to closed chromatin on its own and requires other factors to assist its binding to target sites (Soufi, Garcia et al. 2015). Although it is well known that c-Myc is not essential for reprogramming (Wernig, Meissner et al. 2008, Maekawa, Yamaguchi et al. 2011), it facilitates the initial binding of the other three factors to several chromatin sites (Knoepfler 2008, Hanna, Saha et al. 2009). Oct4, Sox2, Klf4 bind with c-Myc at distal regions of closed chromatin, whereas in pre-iPS and ES cells, c-Myc binds to a distinct region of active genes. c-Myc targets are involved in proliferation and metabolism (Plath and Lowry 2011) so it seems these targets are involved in the first wave of reprogramming (Polo, Anderssen et al. 2012). However, Oct4, Sox2 and Klf4 targets are involved in the pluripotency network of the second wave of reprogramming (Plath and Lowry 2011, Polo, Anderssen et al. 2012). Recently, Chew *et al* monitored fluorescently tagged single Oct4 and Sox2 in living cells and showed that Sox2 guided Oct4 binding to its targets (Chew, Loh et al. 2005). In our study, I showed that the binding of these two transcription factors (Oct4 and Sox2) to their targets is different during the first and second waves of reprogramming. For example, four different patterns of binding of Oct4 and Sox2 were found. In some target genes, the binding of Oct4 and Sox2 was not consistent with gene expression, for example Cbx7 showed high enrichment in ChIP experiments by both factors in the first wave of reprogramming which is consistent with high transcriptional activity in

first wave, however, in the second wave of reprogramming, despite the up and down in Oct4 or Sox2 binding, Cbx7 is upregulated (See chapter 4). Although in many targets the Oct4 and Sox2 binding motifs are close to each other, we found Oct4 and Sox2 binding to their targets can vary from one to another target gene and it is controversial why there is a difference between binding and gene expression. As I mentioned previously in the introduction chapter, some of the targets located in open region chromatin in somatic cells bind to OKSM immediately to downregulate somatic cell associated genes and upregulate MET associated genes (Polo, Anderssen et al. 2012). Here we showed that some of our investigated genes such as Rif1, which is associated with the first wave of reprogramming, become activated by binding of transcription factors in the early days of reprogramming. However some other targets need more chromatin modification such as H3K4me1 to activate. For these genes, TFs bind to distal elements of the target genes and crosstalk with the promoter to result in gene activation. In this study we showed that, despite low binding of Oct4 and Sox2 in promoter region for some targets such as Sox13 in first wave of reprogramming, the gene becomes activated. It is suggested that the target gene becomes activated by chromatin modification and other epigenetic changes.

Therefore there are several possibilities to answer the question such as epigenetic changes include levels of expression of the exogenous transcription factors. It is well known that the gene expression varies from one cell type to another cell type due to the fact that although the cells have identical genomes, they have different epigenetic patterns (Berdasco and Esteller 2010). Therefore, analysis of the chromatin state and DNA modifications would help dissect the mechanism of changes during development or reprogramming. Changes in histone modifications at promoters and enhancers leads to changes in transcription factor binding and gene activation (Polo, Anderssen et al. 2012) and histone methylation impacts on DNA through methylation of CpG islands and plays an

important role in epigenetic regulation (Mostoslavsky, Alt et al. 2003). Hence, to address how the pluripotent genes can be activated or repressed during reprogramming, DNA methylation, histone methylation, variation in nucleosome movement and the chromatin complexes at enhancers or promoter regions must be examined.

Moreover, other epigenetic changes are the presence of bivalent promoters which have both active (H3K4me3) and inactive histone marks (H3K27) (Karlič, Chung et al. 2010), which can lead to the activation or repression of genes during differentiation or reprogramming. It suggested there is a crosstalk between all epigenetic features. For example, in inactive genes, repressive histone marks such as H3K9me3 are accompanied with DNA methylation of the promoter (Hahn, Wu et al. 2011) and consequently it impacts transcription factor binding. We previously observed the bivalent domains are established gradually after the first wave of reprogramming whereas DNA methylation occurs after the second wave of reprogramming (Polo, Anderssen et al. 2012) suggesting the different binding patterns in the first and second wave of reprogramming might be associated with different epigenetic changes in each reprogramming stage.

Another question is: how can the chromatin be disassembled to bind the transcription factor during reprogramming? Is there any correlation between chromatin accessibility and transcription factor binding and gene activity? As reprogramming is a step-wise process and follows two waves (Stadtfield, Maherali et al. 2008, Polo, Anderssen et al. 2012), the dynamics of chromatin varies in each day of reprogramming. In this study we found that the transcriptional activity in some genes correlated with chromatin accessibility and for some others is not. For example Mybl2 becomes gradually upregulated and there was also a correlation between gradual binding of Oct4 and Sox2 and chromatin accessibility. However, Nodal and Lefty1 were different. In Nodal, despite the chromatin being accessible, some binding of Oct4 and Sox2 was observed suggesting that other factors such

as DNA methylation or histone methylation (H3K9me3 and H3K27me3) might impact on Oct4 and Sox2 binding and consequently gene activity. In Lefty1, although the chromatin was accessible and Oct4 and Sox2 was bound to the promoter region, the transcriptional activity was low suggesting the protein complexes required for transcriptional activity in some target genes were not completely assembled. Multiple protein complexes play an important roles in ES cells and iPSC by modification in histones such as methylation, acetylation, and ubiquitylation (Niwa 2007). Protein complexes including SWI/SNF, Tip60-p400, NuRD, Polycomb group protein (PcG), HMTs, HDMs, CHD, DNMTs are required to maintain pluripotency in embryonic stem cells (Orkin and Hochedlinger 2011). Therefore, different target genes were activated at different time points that might need different protein complexes for their activation. Depletion of PRC1 and PRC2 subunits of Polycomb group protein (PcG) reduced induced pluripotency underscoring the necessity of H3K27me3-mediated gene expression for pluripotency. On the other hand, ATP dependent enzymes that act as chromatin remodellers such as SWI/SNF impact on chromatin structure and chromatin regulation suggesting chromatin accessibility are associated with many factors such as histone modification and ATP dependent remodellers.

Furthermore, nucleosome occupancy of promoters showed that there is a difference at regulatory loci in different cell types. A nucleosome occupancy profile around enhancers in somatic cells and pluripotent cells revealed that the average occupancy in ESCs was lower than somatic cells (West et al. 2014). Here we showed that nucleosome occupancy from MEF to iPSC at different time points (MEF, day 3, day 6, day 9, day 12, iPSC) and in pluripotent associated genes such as Lefty1 and Nanog the nucleosome occupancy from MEF to iPSC varies from high to low. Low and high level of nucleosome occupancy at the promoters relied on H3K4me3 and H3K27me3 respectively and it was consistent with the observation that nucleosome occupancy inhibits transcription. Our data also was consistent

with previous reports and showed that mononucleosome signals around TSS was higher which means this region is more accessible and transcriptionally is active.

Moreover, the open region chromatin in somatic cells is required for transcription factors binding or histone methylation (H3K4 me2, H3K4 me3) in the early days of reprogramming. Additionally, downregulation of somatic cell associated genes and MET occurs in the early days of reprogramming. Therefore, successful reprogramming cells need to pass MET. E-Cadherin (E-CAD) acts as a hallmark marker of MET (Chantzoura, Skylaki et al. 2015) which is essential for iPSC generation suggesting E-CAD might play a role as roadblock to reprogramming in the Thy1⁺ population. Over-expression of E-CAD in refractory cells might remove this roadblock and make the chromatin become accessible in Thy1⁺ cells like in Ssea1⁺ cells, this study showed several regulatory elements for activation of the target genes. However, it is not clear which one plays strong effect on gene activation. Moreover, in our study in open chromatin regions close to TSS we found the E-Box motif (CACGTG), which can be bound by c-Myc suggesting that c-Myc might bind to open region located close to TSS and facilitates the opening of the chromatin in another regulatory elements.

Expression of Oct4 is regulated by two regulatory elements located upstream of the Oct4 gene named proximal enhancer (PE) and distal enhancer (DE). There are multiple potential binding sites within these two enhancers that activate or repress Oct4 expression (Pan, Chang et al. 2002). Moreover, DNA methylation in these regions led to inhibition of Oct4 expression. Another study on Oct4 regulation in mouse ES cells showed that the distal enhancer plays an important role in Oct 4 expression (Nazarov, Krasnoborova et al. 2014). The distal enhancer includes two elements, distal enhancer a and b. Oct4 interacts with Sox2 and binds to distal enhancer b to regulate Oct4 expression (Nazarov, Krasnoborova et al. 2014). There is no evidence for the role of PE in Oct4 expression in ES cells. In our

study, one distal enhancer for Oct4 regulation in iPSC was found which is consistent with Oct4 expression in mouse ES cells, however ATAC-seq in Ssea1+ cells at different time points from MEF to iPSC revealed the use of two regulatory elements. Our results suggest that from day 3 to day 12 of the reprogramming process two regulatory elements are being used to regulate Oct4 whereas in iPSC, only the distal regulatory element is responsible for Oct4 regulation. Therefore the PE seems to be important only during the first stages of the reprogramming process.

Moreover it was previously shown that MEFs treated by OKSM and some small molecules are able to generate pancreatic progenitor-like cells (PPLC) and these PPLCs generated all three pancreatic lineages including functional insulin secreting β -like cells (Li, Zhu et al. 2014). Also with 4 days exposure of reprogramming factors to MEFs, MEFs can be reprogrammed to differentiated cardiomyocytes in fast and efficient manner (Efe, Hilcove et al. 2011). In this study we showed that the chromatin becomes more accessible by day 3 for both Thy1+ and Ssea1+ cells which might potentially be a reason why these cells are able to transdifferentiate to any cell types easier and faster. On the other hand, a recent study showed that Klf4 and Klf5 knockdown increases the differentiation toward endoderm and mesoderm respectively (Aksoy, Giudice et al. 2014). Our research on Klf4 target genes during reprogramming showed that the chromatin is less accessible in Thy1+ cell on day 6, 9 and therefore, the refractory cells might be potential to differentiate to endoderm and mesoderm at these time points.

Many transcriptional coactivators are known chromatin modifiers and they actively impact on nucleosome positioning and modify the histones via phosphorylation, acetylation and methylation (Farthing, et al. 2008, Ruan, et al. 2015). Therefore, like any developmental process, there is a cascade of events that affect transcriptional activity of pluripotent-associated genes during reprogramming, however, the order of these events is unknown

and it is not clear which event is the commencement point of this cascade. Does the chromatin become accessible and transcription factors bind, or do the transcription factors force the chromatin to become accessible in these regions?

6.3. Future direction

1. This study showed the rescue of refractory cells by overexpression of the four OKSM factors, however, the molecular mechanism in transcription factor binding pattern and chromatin changes in the refractory cells is still unknown.
2. This research showed the binding pattern of Oct4 and Sox2 in M2rtTA reprogramming MEFs, however, the role of Klf4, c-Myc and other reprogramming factors in transcriptional activity of pluripotent associated genes needs to be examined.
3. Our finding led to generation of an efficient reprogramming mouse model with high number of reprogramming cells to study protein complexes and transcription factor mapping using ChIP seq.
4. There are many other transcription factors that impact on transcriptional activity of target genes during reprogramming, such as protein complexes and chromatin remodelers, that require us to understand how the pluripotent associated gene are regulated.

Appendix: List of primers, probes and target genes

List of ChIP primers used in Chapter 4.

Gene	Forward Primer	Reverse Primer
Klf2	ACTTCAGCTCACTCCCCCTAC T	AATAGATCAGTAGCTCAGAGCC AGA
Sall1	CCCAAGCCCTAAGTCTAGAA AAG	TTAACCACCTGGGGTTTATTTGT
Slc20a1	TGTGCTTTACACACAATTCCA AC	CTTGAACAAACCCTAGGTCTCCT
TdGF1	AGTGTGGACAAGTCCTGAGA GAG	TGGTAAATAACTGAGCCCTGAA A
Id3	TGCCTCCAAGGACTGAAGAT	TGGGAGTCCTGACAAGCAAT
Rest	GTCACACTGAGGAAATGAAG GAC	CCCAGTCTCTGAAACAGAAAAG A
Ssbp3	AGACAGTGCCGACAATGTTTT AT	TCCAAATCAAAAACCTGGTGAGA T
Cbx1	GCACCCTGAATTCTTTCTTT AT	AATCGTAACTCCAGGACATCTGA
Phc1	AGGTTGGGGAAATAAGGACA TAA	GGGCTTCTATAGGAAAACCTCAA CA
Sox2	GCAATGCTGAGAAATTCCAGT T	GTTCCCCTCCTCTCCTAATCTC

Usp44	ATGGAATGTGGAGGGGTAAA	TATGCCGTCCTTGCTTTTGT
Nkx 2-2	AAGGGTGACGAACAGAACTC TC	GACTAGGAGGGCAAGAGAACT C
Rp 13b	TGGGTGTATTTTTCAGGGTTT TA	GTTCTCATCCTGCAAGGGTAAAT
Rps17	GCTTTCCTCCCTGGTTATTTCT A	GCCTATCCCACCTTCTCATACTT
Tcf 7	AGGGAGAGCCAAGTCTTCTG	AGCCTGGAGGGTTCCTTTTA
Cdx2	GCCTCAGCTTTCATCCCCTT	TTACTTCCAAAGCAGGCGC
Arhgap2	TTCCCTCCACACTAGCACAA	TGAAATGTCTAACAGGTCCTTGG
Hox13	CTTAACCTTCAACTTGGCCTT GA	GGAGGAGTCTGTAGCCTTAGAG C
Utf1	GGTGCAGGTAGAGGAAAGGA	GATGGGCCCAAGAATTTGTAA
Kdelc1	CAGGATTTCCGCTAGTCCTAT G	GTCTTCTGCTAGTGGATGAGTGG
Tcl	AGGACAGACTGAAGGTGACA GAG	TCAGCTACTTTGTGTGACTGCTC
Klf9	TTAAAATGCCCTTGTTGTTTC TG	CCCATTTTAAAAACACCTCTTC
Pou5f1	CCCAGGGAGGTTGAGAGTTC	AGGAAGGGCTAGGACGAGAG

Lefty1	AGCTGCTCTTCTGCATAACAA A	AGACAATCTTAGTCGGGGGATA G
Mybl2	AAGGAATAGCCTTGATTGGA AG	CTCAGAACGGATAACAGAATTG C
Rif1	CGTTGTTTAAGGGTTTACTGT CG	TGCCTAGAATTGGTTTACTTCCA
Cbx7	TTATTTAGAAGGGGCTTCCTT TG	GGGTGTTTGTAAGGAGGGATA G
Nodal	TGGGGACACATCCTACTAGGT AA	TCAGAAGTGGAATTTGGAGAGA G
Sox13	TTTTGTCTGAAGTTCCTGAGA GG	GCTGAGGGCAACAATTACATAA C
Klf4	CTCCTCTACAGCCGAGAATCT G	AGGAGCTCAGCCACGAAG
Dido1	GGAAGTTGGAGGCTCTTTAAC AC	AATGACTGAATTTGCATTACCT
Zfp148	CTCTCAGGAAGTTCCCAATAA CC	AGGGAGGGCTTACAGTTAATGA C

List of MLPA probes used in Chapter 5.

MLPA probes based on positive target genes of different transcription factors (OKSM) designed in UCSC genome browser based on DNase I hypersensitive sites.

Probe	L	R	Length
Nanog	GGGTTCCCTAAGGGTTGGAGG ATAAGGAATGTAGCAAGCCTG CCTTTT	CCAGCCACCTCTTCGCTC GGATCTTTCACCTAGATT GGATCTTGCTGGCAC	105
Sall4	GGGTTCCCTAAGGGTTGGAGA GACAGGATTTCTGTGTAGCCCT GGCTAC	GTAAATCACCCGGTATAC CAGGCTGGCCTTCTAGAT TGGATCTTGCTGGCAC	108
Klf9	GGGTTCCCTAAGGGTTGGAGA ACGACCAGTGCCGCTCAAGGA CAGGGC	CATGGCAGTGAGCAGAGG GTGGTGAAGAGGTCTAGA TTGGATCTTGCTGGCAC	111
Rif	GGGTTCCCTAAGGGTTGGAGCC ACTAGGTCCTGCTTCCCTGACC TTCT	GAAACAATCTTGCGGGTT GTGGAGAGTGCTCTAGAT TGGATCTTGCTGGCAC	112
Sox13	GGGTTCCCTAAGGGTTGGACTT GATGTGGCTGCTGTTCCGAGAC TCGGA	GGAGGATGGGTAGGTAA GAGTCAGGCAGGTCTAGA TTGGATCTTGCTGGCAC	117
E2F3	GGGTTCCCTAAGGGTTGGAGCT GCTGCTGACAATGAATGAAGG GATACG	TCCCGCGAAACTCAGATT TCCCGCCGGCTTTCTAGA TTGGATCTTGCTGGCAC	120

Nanog	GGGTTCCCTAAGGGTTGGACTC	GACCTGAAACTTCCCACT	124
0.2	TTTCATGTCTGTAGAAAGAATG	AGAGATCGCCAGGTCTAG	
	GAA	ATTGGATCTTGCTGGCAC	
Oct4	GGGTTCCCTAAGGGTTGGACGC	GGAGAAGTGGGTGGAGG	128
	TTCGAGGCCTTGCAGCTCAGCC	AAGCCGACAACAATCTAG	
	TTAAG	ATTGGATCTTGCTGGCAC	
Lefty	GGGTTCCCTAAGGGTTGGACGT	CCAATCTCAAGCCTGACC	132
	GTCAGAAGCTGCAGACTTCATT	TGGGTCAGACCTTCTAGA	
	CCAGG	TTGGATCTTGCTGGCAC	
Rex1	GGGTTCCCTAAGGGTTGGAGGT	GAGACTCAGTCACATCTC	136
	GGGGCAAAGACTCTTGCTCAGT	AGTACCTCTGTTTCTAGAT	
	TATGC	TGGATCTTGCTGGCAC	
Dido	GGGTTCCCTAAGGGTTGGACGG	CACCTCCGGTAGACAATG	141
	ACGGACCAACATTCCTTCCAGA	GACACATAAATATCTAGA	
	AGCAT	TTGGATCTTGCTGGCAC	
Lrrn2	GGGTTCCCTAAGGGTTGGACTC	CCACCTTCCCTTTCCTTC	145
	CCTCTCTTCCACTTCCATGAG	TTCGCAGCCTCTCTAGATT	
	GCTCT	GGATCTTGCTGGCAC	
Zic5	GGGTTCCCTAAGGGTTGGACTG	GGGCTTTTTATTTTGTAGA	148
	AGAATAGGCCTAGCTACTTGGT	CGCCCTTTTTGTCTAGATT	
	CATCA	GGATCTTGCTGGCAC	
Mcm5	GGGTTCCCTAAGGGTTGGACGA	GAGGGCTGAGGTTCGTGT	151
	ACCAATAGGAGCGCAGAGACC	GGCGCGCCGCTGTCTAGA	
	GCGG	TTGGATCTTGCTGGCAC	

Rest	GGGTTCCCTAAGGGTTGGAGAC CTCTGGCCTGCATGTCATGATA ATTTG	GAATGCAGCCGCCATGAG AGCCATTGGCCTTCTAGA TTGGATCTTGCTGGCAC	154
Rhof	GGGTTCCCTAAGGGTTGGAGG ACTGGTAAAATGGCTCAGCTGT TAAAAA	GTACATGCAGTGCCAGCA GAGGCCAGAAGATCTAGA TTGGATCTTGCTGGCAC	160
Sox2	GGGTTCCCTAAGGGTTGGACTG GTGGTCGTCAAACCTCTGCTAAT TAGCA	CCTTCATTCCATAAGGA GATTAGGAGAGGTCTAGA TTGGATCTTGCTGGCAC	164
Nodal	GGGTTCCCTAAGGGTTGGAGA GCACCTGCGAGTCTCAGTTTGC CCATCT	CCCCCAGAAACGCTCTGC CCAGAACCGGGTCTAGAT TGGATCTTGCTGGCAC	168
Cbx1	GGGTTCCCTAAGGGTTGGACTC AGAAATCTGCCTGCCTCTGCCT CCCAA	CAACCAATACTGGGGGTT GGAGAGATGGTCTAGATT GGATCTTGCTGGCAC	171

List of ATAC primers used in Chapter 5.

Primer Name	Sequence
Ad1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCA GATGTG
Ad2.1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCT CGGAGATGT
Ad2.2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCT CGGAGATGT
Ad2.3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTC GGAGATGT
Ad2.4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCT CGGAGATGT
Ad2.5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCT CGGAGATGT
Ad2.6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCT CGGAGATGT
Ad2.7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCT CGGAGATGT
Ad2.8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCT CGGAGATGT
Ad2.9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCT CGGAGATGT

Ad2.10	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCT CGGAGATGT
Ad2.11	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTC GGAGATGT
Ad2.12	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCT CGGAGATGT

References:

Aksoy, I., et al. (2014). "Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells." Nature communications **5**.

Aksoy, I., et al. (2013). "Sox transcription factors require selective interactions with Oct4 and specific transactivation functions to mediate reprogramming." Stem cells **31**(12): 2632-2646.

Alaei-Shehni, S., et al. (2014). "Epigenetic Changes During Reprogramming." Australian Biochemist [P] **45**(1): 17-20.

Amabile, G. and A. Meissner (2009). "Induced pluripotent stem cells: current progress and potential for regenerative medicine." Trends in molecular medicine **15**(2): 59.

Barrero, M. J., et al. (2010). "Epigenetic mechanisms that regulate cell identity." Cell Stem Cell **7**(5): 565-570.

Bauer, U. M., et al. (2002). "Methylation at arginine 17 of histone H3 is linked to gene activation." EMBO reports **3**(1): 39-44.

Beato, M. and K. Eisefeld (1997). "Transcription factor access to chromatin." Nucleic acids research **25**(18): 3559-3563.

Berdasco, M. and M. Esteller (2010). "Aberrant epigenetic landscape in cancer: how cellular identity goes awry." Developmental cell **19**(5): 698-711.

Biran, A. and E. Meshorer (2012). "Concise review: chromatin and genome organization in reprogramming." Stem Cells **30**(9): 1793-1799.

Boyer, L. A., et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-956.

Brumbaugh, J. and K. Hochedlinger (2013). "Removing reprogramming roadblocks: Mbd3 depletion allows deterministic iPSC generation." Cell stem cell **13**(4): 379-381.

Buenrostro, J. D., et al. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nature methods **10**(12): 1213-1218.

Campbell, P. A., et al. (2007). "Oct4 targets regulatory nodes to modulate stem cell function." PLoS One **2**(6): e553.

Campolo, F., et al. (2013). "Essential role of Sox2 for the establishment and maintenance of the germ cell line." Stem Cells **31**(7): 1408-1421.

Carey, B. W., et al. (2009). "Reprogramming of murine and human somatic cells using a single polycistronic vector." Proceedings of the National Academy of Sciences **106**(1): 157-162.

Carey, B. W., et al. (2011). "Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells." Cell stem cell **9**(6): 588-598.

Chantzoura, E., et al. (2015). "Reprogramming Roadblocks Are System Dependent." Stem cell reports **5**(3): 350-364.

Chew, J.-L., et al. (2005). "Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells." Molecular and cellular biology **25**(14): 6031-6046.

Chong, J. J., et al. (2014). "Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts." Nature **510**(7504): 273-277.

Cirillo, L. A., et al. (2002). "Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4." Molecular cell **9**(2): 279-289.

Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annual review of biochemistry **78**: 273-304.

Colman, A. and O. Dreesen (2009). "Pluripotent stem cells and disease modeling." Cell stem cell **5**(3): 244-247.

Cornacchia, D., et al. (2012). "Mouse Rif1 is a key regulator of the replication - timing programme in mammalian cells." The EMBO journal **31**(18): 3678-3690.

Cowan, C. A., et al. (2004). "Derivation of embryonic stem-cell lines from human blastocysts." New England Journal of Medicine **350**(13): 1353-1356.

Creyghton, M. P., et al. (2008). "H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment." Cell **135**(4): 649-661.

Dimos, J. T., et al. (2008). "Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons." science **321**(5893): 1218-1221.

Ding, J., et al. (2012). "Oct4 links multiple epigenetic pathways to the pluripotency network." Cell research **22**(1): 155-167.

Dominko, T., et al. (1999). "Bovine oocyte cytoplasm supports development of embryos produced by nuclear transfer of somatic cell nuclei from various mammalian species." Biology of reproduction **60**(6): 1496-1502.

Dow, L. E., et al. (2014). "Conditional reverse tet-transactivator mouse strains for the efficient induction of TRE-regulated transgenes in mice." PloS one **9**(4): e95236.

Draper, J. S. and P. W. Andrews (2006). "Surface Antigen Markers." Essentials of stem cell biology: 337.

Drawnel, F. M., et al. (2014). "Disease modeling and phenotypic drug screening for diabetic cardiomyopathy using human induced pluripotent stem cells." Cell reports **9**(3): 810-820.

Ebert, A. D. and C. N. Svendsen (2010). "Human stem cells and drug screening: opportunities and challenges." Nature Reviews Drug Discovery **9**(5): 367-372.

Ebert, A. D., et al. (2009). "Induced pluripotent stem cells from a spinal muscular atrophy patient." Nature **457**(7227): 277-280.

Efe, J. A., et al. (2011). "Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy." Nature cell biology **13**(3): 215-222.

Eggan, K., et al. (2004). "Mice cloned from olfactory sensory neurons." Nature **428**(6978): 44-49.

Eminli, S., et al. (2009). "Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells." Nature genetics **41**(9): 968-976.

Eminli, S., et al. (2008). "Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression." Stem Cells **26**(10): 2467-2474.

Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.

Fazio, T. G., et al. (2008). "An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity." Cell **134**(1): 162-174.

Filipescu, D., et al. (2013). "Developmental roles of histone H3 variants and their chaperones." Trends in Genetics **29**(11): 630-640.

Flemming, W. (1882). Zellsubstanz, kern und zelltheilung, Vogel.

Gaspar-Maia, A., et al. (2010). "Open chromatin in pluripotency and reprogramming." Nature reviews Molecular cell biology **12**(1): 36-47.

Gidekel, S., et al. (2003). "Oct-3/4 is a dose-dependent oncogenic fate determinant." Cancer cell **4**(5): 361-370.

Gifford, C. A., et al. (2013). "Transcriptional and epigenetic dynamics during specification of human embryonic stem cells." Cell **153**(5): 1149-1163.

González, F., et al. (2011). "Methods for making induced pluripotent stem cells: reprogramming a la carte." Nature Reviews Genetics **12**(4): 231-242.

Gurdon, J. (1960). "The developmental capacity of nuclei taken from differentiating endoderm cells of *Xenopus laevis*." Journal of embryology and experimental morphology **8**(4): 505-526.

Gurdon, J. and D. Melton (2008). "Nuclear reprogramming in cells." Science **322**(5909): 1811-1815.

Gurdon, J. B. (1962). "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles." Journal of embryology and experimental morphology **10**(4): 622-640.

Hahn, M. A., et al. (2011). "Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks." PLoS One **6**(4): e18844.

Hanna, J., et al. (2009). "Direct cell reprogramming is a stochastic process amenable to acceleration." Nature **462**(7273): 595-601.

Heitz, E. (1928). Das heterochromatin der moose, Bornträger.

Hochedlinger, K. and R. Jaenisch (2002). "Monoclonal mice generated by nuclear transfer from mature B and T donor cells." Nature **415**(6875): 1035-1038.

Hochedlinger, K. and K. Plath (2009). "Epigenetic reprogramming and induced pluripotency." Development **136**(4): 509-523.

Hockemeyer, D., et al. (2008). "A drug-inducible system for direct reprogramming of human somatic cells to pluripotency." Cell stem cell **3**(3): 346-353.

Hu, G. and P. A. Wade (2012). "NuRD and Pluripotency: A Complex Balancing Act." Cell stem cell **10**(5): 497-503.

Huangfu, D., et al. (2008). "Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds." Nature biotechnology **26**(7): 795-797.

Jerabek, S., et al. (2014). "OCT4: dynamic DNA binding pioneers stem cell pluripotency." Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1839**(3): 138-154.

Kadonaga, J. T. (2004). "Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors." Cell **116**(2): 247-257.

Kaji, K., et al. (2006). "The NuRD component Mbd3 is required for pluripotency of embryonic stem cells." Nature cell biology **8**(3): 285-292.

Karlič, R., et al. (2010). "Histone modification levels are predictive for gene expression." Proceedings of the National Academy of Sciences **107**(7): 2926-2931.

Kim, J., et al. (2008). "An extended transcriptional network for pluripotency of embryonic stem cells." Cell **132**(6): 1049-1061.

Kimbrel, E. A. and R. Lanza (2015). "Current status of pluripotent stem cells: moving the first therapies to the clinic." Nature Reviews Drug Discovery.

Knoepfler, P. S. (2008). "Why myc? An unexpected ingredient in the stem cell cocktail." Cell Stem Cell **2**(1): 18-21.

Koche, R. P., et al. (2011). "Reprogramming factor expression initiates widespread targeted chromatin remodeling." Cell stem cell **8**(1): 96-105.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Lachner, M., et al. (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature **410**(6824): 116-120.

Li, B., et al. (2007). "The role of chromatin during transcription." Cell **128**(4): 707-719.

Li, D.-Q., et al. (2012). "Metastasis-associated protein 1/nucleosome remodeling and histone deacetylase complex in cancer." Cancer Research **72**(2): 387-394.

Li, K., et al. (2014). "Small molecules facilitate the reprogramming of mouse fibroblasts into pancreatic lineages." Cell stem cell **14**(2): 228-236.

Li, R., et al. (2010). "A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts." Cell stem cell **7**(1): 51-63.

Lin, C. H., et al. (2009). "Myc - regulated microRNAs attenuate embryonic stem cell differentiation." The EMBO journal **28**(20): 3157-3170.

Liu, Y., et al. (2014). "The death-inducer obliterator 1 (Dido1) gene regulates embryonic stem cell self-renewal." Journal of Biological Chemistry **289**(8): 4778-4786.

Loh, Y. H., et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nature genetics **38**(4): 431-440.

Lorvellec, M., et al. (2010). "B - Myb is Critical for Proper DNA Duplication During an Unperturbed S Phase in Mouse Embryonic Stem Cells." Stem Cells **28**(10): 1751-1759.

Luo, M., et al. (2013). "NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells." Stem Cells **31**(7): 1278-1286.

Maekawa, M., et al. (2011). "Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1." Nature **474**(7350): 225-229.

Maherali, N., et al. (2008). "A high-efficiency system for the generation and study of human induced pluripotent stem cells." Cell stem cell **3**(3): 340-345.

Maherali, N. and K. Hochedlinger (2008). "Guidelines and techniques for the generation of induced pluripotent stem cells." Cell stem cell **3**(6): 595-605.

Maherali, N., et al. (2007). "Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution." Cell stem cell **1**(1): 55-70.

Margueron, R. and D. Reinberg (2011). "The Polycomb complex PRC2 and its mark in life." Nature **469**(7330): 343-349.

Matsui, Y., et al. (1992). "Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture." Cell **70**(5): 841-847.

Meissner, A., et al. (2008). "Genome-scale DNA methylation maps of pluripotent and differentiated cells." Nature **454**(7205): 766-770.

Mekhoubad, S., et al. (2012). "Erosion of dosage compensation impacts human iPSC disease modeling." Cell stem cell **10**(5): 595-609.

Mikkelsen, T. S., et al. (2008). "Dissecting direct reprogramming through integrative genomic analysis." Nature **454**(7200): 49-55.

Miller, J. A. and J. Widom (2003). "Collaborative competition mechanism for gene activation in vivo." Molecular and cellular biology **23**(5): 1623-1632.

Miller, R. A. and F. H. Ruddle (1976). "Pluripotent teratocarcinoma-thymus somatic cell hybrids." Cell **9**(1): 45-55.

Mizuguchi, G., et al. (2004). "ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex." Science **303**(5656): 343-348.

Mo, A., et al. (2015). "Epigenomic signatures of neuronal diversity in the mammalian brain." Neuron **86**(6): 1369-1384.

Mostoslavsky, R., et al. (2003). "Chromatin dynamics and locus accessibility in the immune system." Nature immunology **4**(7): 603-606.

Nair, S. S., et al. (2013). "A core chromatin remodeling factor instructs global chromatin signaling through multivalent reading of nucleosome codes." Molecular cell **49**(4): 704-718.

Nazarov, I., et al. (2014). "Transcription regulation of Oct4 (Pou5F1) gene by its distal enhancer." Cell and Tissue Biology **8**(1): 27-32.

Nefzger, C. M., et al. (2014). "Cell Surface Marker Mediated Purification of iPS Cell Intermediates from a Reprogrammable Mouse Model." JoVE (Journal of Visualized Experiments)(91): e51728-e51728.

Ng, R. K. and J. Gurdon (2008). "Epigenetic memory of an active gene state depends on histone H3. 3 incorporation into chromatin in the absence of transcription." Nature cell biology **10**(1): 102-109.

Nichols, J., et al. (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-391.

Nishimura, K., et al. (2014). "Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming." Stem cell reports **3**(5): 915-929.

Niwa, H. (2007). "Open conformation chromatin and pluripotency." Genes & development **21**(21): 2671-2676.

Ohnesorg, T., et al. (2009). "Rapid high-throughput analysis of DNaseI hypersensitive sites using a modified Multiplex Ligation-dependent Probe Amplification approach." BMC genomics **10**(1): 412.

Ohnesorg, T., et al. (2012). "Detecting DNaseI-hypersensitivity sites with MLPA." Methods in molecular biology (Clifton, NJ) **786**: 201.

Olins, D. E. and A. L. Olins (2003). "Chromatin history: our view from the bridge." Nature reviews Molecular cell biology **4**(10): 809-814.

Orkin, S. H. and K. Hochedlinger (2011). "Chromatin connections to pluripotency and cellular reprogramming." Cell **145**(6): 835-850.

Pan, G. J., et al. (2002). "Stem cell pluripotency and transcription factor Oct4." Cell research **12**(5): 321-329.

Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." Nature Reviews Genetics **10**(10): 669-680.

Plath, K. and W. E. Lowry (2011). "Progress in understanding reprogramming to the induced pluripotent state." Nature Reviews Genetics **12**(4): 253-265.

Polach, K. and J. Widom (1996). "A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites." Journal of molecular biology **258**(5): 800-812.

Polo, J. M., et al. (2012). "A molecular roadmap of reprogramming somatic cells into iPS cells." Cell **151**(7): 1617-1632.

Quina, A., et al. (2006). "Chromatin structure and epigenetics." Biochemical pharmacology **72**(11): 1563-1569.

Radziskeuskaya, A. and J. C. Silva (2014). "Do all roads lead to Oct4? The emerging concepts of induced pluripotency." Trends in cell biology **24**(5): 275-284.

Rais, Y., et al. (2013). "Deterministic direct reprogramming of somatic cells to pluripotency." Nature **502**(7469): 65-70.

Rao, R. A., et al. (2015). "Ezh2 mediated H3K27me3 activity facilitates somatic transition during human pluripotent reprogramming." Scientific reports **5**.

Reardon, S. and D. Cyranoski (2014). "Japan stem-cell trial stirs envy." Nature **513**(7518): 287-288.

Samavarchi-Tehrani, P., et al. (2010). "Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming." Cell stem cell **7**(1): 64-77.

Schneuwly, S., et al. (1987). "Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia."

Segal, E. and J. Widom (2009). "What controls nucleosome positions?" Trends in Genetics **25**(8): 335-343.

Shi, Y., et al. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." cell **119**(7): 941-953.

Singh, A. M. and S. Dalton (2009). "The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming." Cell stem cell **5**(2): 141-149.

Singhal, N., et al. (2010). "Chromatin-remodeling components of the BAF complex facilitate reprogramming." Cell **141**(6): 943-955.

Sommer, C. A., et al. (2009). "Induced pluripotent stem cell generation using a single lentiviral stem cell cassette." Stem cells **27**(3): 543-549.

Soufi, A., et al. (2012). "Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome." Cell.

Soufi, A., et al. (2012). "Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome." Cell **151**(5): 994-1004.

Soufi, A., et al. (2015). "Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming." Cell **161**(3): 555-568.

Sridharan, R., et al. (2009). "Role of the murine reprogramming factors in the induction of pluripotency." Cell **136**(2): 364-377.

Stadtfield, M. and K. Hochedlinger (2010). "Induced pluripotency: history, mechanisms, and applications." Genes & development **24**(20): 2239-2263.

Stadtfield, M., et al. (2009). "A reprogrammable mouse strain from gene-targeted embryonic stem cells." Nature methods **7**(1): 53-55.

Stadtfield, M., et al. (2010). "A reprogrammable mouse strain from gene-targeted embryonic stem cells." Nature methods **7**(1): 53-55.

Stadtfeld, M., et al. (2008). "Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse." Cell stem cell **2**(3): 230-240.

Strizzi, L., et al. (2009). "Nodal as a biomarker for melanoma progression and a new therapeutic target for clinical intervention."

Taberlay, P. C., et al. (2011). "Polycomb-repressed genes have permissive enhancers that initiate reprogramming." Cell **147**(6): 1283-1294.

Tada, M., et al. (2001). "Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells." Current Biology **11**(19): 1553-1558.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." cell **126**(4): 663-676.

Tanaka, Y., et al. (1999). "B-myb is required for inner cell mass formation at an early stage of development." Journal of Biological Chemistry **274**(40): 28067-28070.

Thompson, C. B., et al. (1985). "Levels of c-myc oncogene mRNA are invariant throughout the cell cycle." Nature **314**(6009): 363-366.

Tsubouchi, T., et al. (2013). "DNA Synthesis Is Required for Reprogramming Mediated by Stem Cell Fusion." Cell **152**(4): 873-883.

Turner, B. M. (2002). "Cellular memory and the histone code." Cell **111**(3): 285-291.

Utikal, J., et al. (2009). "Immortalization eliminates a roadblock during cellular reprogramming into iPS cells." Nature **460**(7259): 1145-1148.

van den Berg, D. L., et al. (2010). "An Oct4-centered protein interaction network in embryonic stem cells." Cell stem cell **6**(4): 369-381.

Vangapandu, H. and W. Ai (2009). "Kruppel-like factor 4 (KLF4): a transcription factor with diverse context-dependent functions." Gene Ther Mol Biol **13**: 194-203.

Wang, Y., et al. (2006). "SOX13 exhibits a distinct spatial and temporal expression pattern during chondrogenesis, neurogenesis, and limb development." Journal of Histochemistry & Cytochemistry **54**(12): 1327-1333.

Wei, Z., et al. (2013). "Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency." Cell stem cell **13**(1): 36-47.

Wei, Z., et al. (2009). "Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming." Stem cells **27**(12): 2969-2978.

Werner, M. H., et al. (1995). "Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex." Cell **81**(5): 705-714.

Wernig, M., et al. (2008). "A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types." Nature biotechnology **26**(8): 916-924.

Wernig, M., et al. (2008). "c-Myc is dispensable for direct reprogramming of mouse fibroblasts." Cell stem cell **2**(1): 10-12.

Wernig, M., et al. (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature **448**(7151): 318-324.

Wilmut, I., et al. (1999). "Viable offspring derived from fetal and adult mammalian cells." Clones and Clones: Facts and Fantasies About Human Cloning: 21.

Wright, A. J. and P. W. Andrews (2009). "Surface marker antigens in the characterization of human embryonic stem cells." Stem cell research **3**(1): 3-11.

Xu, Y., et al. (2013). "Cell reprogramming: into the groove." Nature materials **12**(12): 1082-1084.

Yamanaka, S. and H. M. Blau (2010). "Nuclear reprogramming to a pluripotent state by three approaches." Nature **465**(7299): 704-712.

Yao, S., et al. (2004). "Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications." Molecular Therapy **10**(1): 27-36.

You, J. S., et al. (2011). "OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes." Proceedings of the National Academy of Sciences **108**(35): 14497-14502.

Young, Richard A. (2011). "Control of the Embryonic Stem Cell State." Cell **144**(6): 940-954.

Zhang, J., et al. (2011). "A human iPSC model of Hutchinson Gilford Progeria reveals vascular smooth muscle and mesenchymal stem cell defects." Cell stem cell **8**(1): 31-45.

Zhu, J., et al. (2013). "Genome-wide chromatin state transitions associated with developmental and environmental cues." Cell **152**(3): 642-654.

