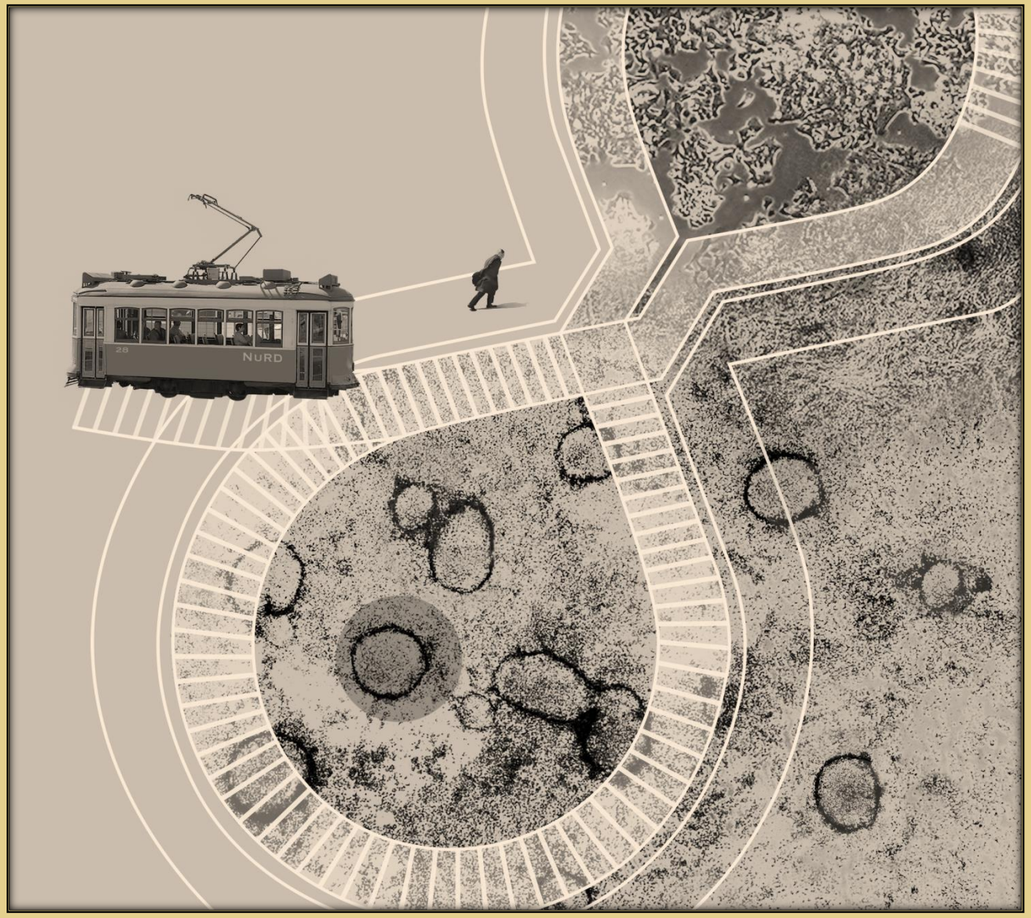
Mechanisms of induced pluripotency:
The role of the Nucleosome Remodelling and Deacetylase complex

Rodrigo Luiz dos Santos

UNIVERSIDADE DE COIMBRA



Rodrigo Luiz dos Santos

MECHANISMS OF INDUCED PLURIPOTENCY:
THE ROLE OF THE NUCLEOSOME REMODELLING AND DEACETYLASE COMPLEX

Tese de Doutoramento em Biologia Experimental e Biomedicina no ramo de Biologia Molecular, Celular e do Desenvolvimento, orientada pelo Doutor José Silva do Wellcome Trust – MRC Stem Cell Initiative, Universidade de Cambridge e pelo Professor Doutor João Ramalho-Santos do Departamento de Ciências da Vida da Faculdade de Ciências e Tecnolologia da Universidade de Coimbra, e apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra.

Agosto de 2014

· U C ·

UNIVERSIDADE DE COIMBRA

• U  C •

Mechanisms of induced pluripotency: The role of the Nucleosome Remodelling and Deacetylase complex

by

Rodrigo Luiz dos Santos

Tese de Doutoramento em Biologia Experimental e Biomedicina no ramo de Biologia Molecular, Celular e do Desenvolvimento, orientada pelo Doutor José Silva do Wellcome Trust – MRC Stem Cell Initiative, Universidade de Cambridge e pelo Professor Doutor João Ramalho-Santos do Departamento de Ciências da Vida da Faculdade de Ciências e Tecnolologia da Universidade de Coimbra, e apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra.

Doctoral thesis in Biomedicine and Experimental Biology in the field of Molecular, Cellular and Developmental Biology, supervised by Dr José Silva from the Wellcome Trust – MRC Stem Cell Initiative, University of Cambridge, and by Dr João Ramalho-Santos from the Department of Life Sciences of the Faculty of Sciences and Technology of the University of Coimbra, and presented to the Institute of Interdisciplinary Research of the University of Coimbra.

The study presented in this dissertation was supported by a PhD Studentship from the Portuguese Foundation for Sciences and Technology, FCT (SFRH/BD/51198/2010), attributed by the PhD Programme in Experimental Biology and Biomedicine (PDBEB)

This dissertation is the result of my original work and corresponds to the final version. Any work done in collaboration is specifically stated in the text.

Acknowledgements

We are guided by the wisdom of our ancestors. - Random post card found in Vancouver.

I was very fortunate to work under the supervision of José Silva at the Wellcome Trust – MRC Stem Cell Initiative, University of Cambridge. José gave all the freedom when I need it; he let me find my own path; and then, when things really mattered, he was the one who believed me, and the one who challenged me to reveal my true potential. Thank you for accepting me into your lab, for your help, support and guidance.

I would like to thank João Ramalho-Santos who allowed me to follow my dreams. Thank you for trusting me. I hope I made you proud.

I am very grateful to the PhD programme in Biomedicine and Experimental Biology and to the Portuguese Foundation for Sciences and Technology which made my PhD possible.

I would like to thank my collaborators Brian Hendrich and Keisuke Kaji. With their expertise and feedback, my work reached a whole new level.

Many thanks to the present and past members of the Silva lab: Chibeza Agley, Lawrence Bates, Yael Costa, Moyra Lawrence, Aliaksandra Radzisheuskaya, Hannah Stuart, Thorold Theunissen and Anouk van Oosten. Thank you all for your help, friendship and support; you have been pivotal for my success and happiness. Thank you Aliaksandra Radzisheuskaya for being such an amazing colleague throughout my stay in Cambridge. My PhD would have been very different without you there at the right moments. I will miss our chats.

I wouldn't have been able to conduct some of the experiments here without the help of Luca Tosti and Isabel M. Caballero. Thank you both. I would also like to thank the core facilities of the Wellcome Trust – MRC Stem Cell Initiative for their help.

Thanks to all my good friends in Cambridge and Coimbra. Thank you for all the good moments and for the support during the bad ones. Thank you Mário Pereira e Sofia Ribeiro for always been there for me, even if through skype. Thank you Pedro Albuquerque, Rui Pereira and Raquel Damas for being the best house mates ever. Our scientific discussions were epic! And thank you Raquel for designing the best thesis cover ever!

Special thanks to my mum for all the great support throughout my life, whatever decisions I have made. I am also immensely grateful to my brothers, all other members of my family and Joana's family, for their love and trust in me. And thank you dad. I know you are there, looking after me.

And finally, thank you Joana. Thank you for sharing your life with me. Thank you for making me a better person. And thank you for your unconditional love and support during all the ups and downs of my PhD.

Abstract

The reprogramming of somatic cells to naïve pluripotency can be robustly achieved by the forced expression of four reprogramming factors: Oct4, Sox2, Klf4 and c-Myc. Among these reprogramming transcription factors, Oct4 plays a central role, as it is sufficient and essential for the induction of pluripotent cells. Oct4 interactome studies in embryonic stem cells (ESCs) revealed members of the Nucleosome Remodelling and Deacetylase (NuRD) complex as its highest confidence interactors. Mbd3 is an essential subunit of the NuRD complex, in the absence of which the complex is not assembled. Embryos lacking Mbd3 die shortly after implantation and *Mbd3*-null ESCs are viable but show severely impaired lineage commitment. Since the NuRD complex is a high confidence interactor of Oct4 and a key regulator of developmental cell state transitions, I investigated if NuRD is also involved in the biological process of the induction of pluripotency.

I have addressed this question by means of genetic deletion or siRNA to manipulate Mbd3 levels in somatic cells. I demonstrated that the reprogramming of neural stem cells (NSCs) to pre-induced pluripotent stem cells (preiPSCs) is impacted by Mbd3 genetic deletion, and that *Mbd3*-null cells have delayed reprogramming kinetics. Likewise, using an inducible Mbd3 deletion strategy, I showed that the longer Mbd3 is intact the more NSCs reprogram to preiPSCs and finally to iPSCs. Using post-implantation epiblast-derived stem cells (EpiSCs), I showed that the reduction of Mbd3 by RNAi results in a complete impairment of Klf4-dependent reprogramming to iPSC and a 6 fold reduction in efficiency during Klf2-Nanog-dependent reprogramming.

Moreover, I also performed the inverse experiments to examine the impact of Mbd3 overexpression on reprogramming. I observed that Nanog-mediated reprogramming of MEF-derived preiPSCs is facilitated in rate and extent by increased Mbd3, which results in increased NuRD complex levels. This increase in efficiency seems to be the result of a synergistic function of Nanog and Mbd3 in inducing the transcription of key pluripotency genes prior to the induction of reprogramming. A similar outcome is observed for Nanog-dependent reprogramming of EpiSCs, where increased Mbd3 expression leads to increased efficiency.

In summary, my results identify a key role of the Mbd3/NuRD complex in the induction of pluripotency and show that a chromatin complex which is required for cell differentiation also facilitates the reversion of these cells back to a pluripotent cell state.

Sumário

A reprogramação de células somáticas a celulas estaminiais pluripotentes induzidas (iPSCs) pode ser conseguida através da sobreexpressão de quatro factores: Oct4, Sox2, Klf4 e c-Myc. Oct4 desempenha um papel fundamental, sendo suficiente e essencial para a geração de células pluripotentes induzidas. Análise do interactoma do Oct4 revelou que as subunidades do complexo NuRD (do inglês *Nuclessome remodelling and Deacetylase complex*) são os seus interactores principais. A subunidade Mbd3 é essencial para a geração do complexo, já que a sua ausência leva a que o complexo não se forme. Delecção do Mbd3 não é compatível com desenvolvimento embrionário e, apesar de células estaminais embrionárias poderem existir sem Mbd3, estas apresentam problemas de diferenciação celular. Sendo o complexo NuRD um dos principais interactores do Oct4 e um complexo chave na regulação de transições celulares durante o desenvolvimento embionário, decidi investigar qual o papel desempenhado pelo NuRD no processo de pluripotência induzida.

Para responder a esta questão fiz uso de diferentes delecções genéticas e ARN de interfência para manipular os níveis de Mbd3 em células somáticas. Nesta tese demonstrei que a remoção de Mbd3 leva a uma reduzida eficiência de reprogramação de células neuronais estaminais (NSC) a preiPSCs e, que as células que eventualmente reprogramam demoram mais tempo a fazê-lo. Através da remoção do Mbd3 em diferentes períodos, mostrei que quanto mais tarde o Mbd3 for removido, maior é o número de NSC que reprogramam em preiPSCs e eventualmente em iPSCs. Igualmente, usando células estaminais pluripotentes obtidas do epiblasto após implantação do blastocisto (EpiSCs), demonstrei que a redução da expressão de Mbd3 por ARN de interferência impede a reprogramação mediada pela sobreexpressão do factor de reprogramação Klf4, e que a delecção genética do Mbd3 reduz até seis vezes a eficiência de reprogramação mediada pelos factores Klf2 e Nanog.

Nesta tese investiguei também o impacto da sobreexpressão do Mbd3 na reprogramação celular. Demonstrei que a reprogramação mediada pelo Nanog de preiPSCs, obtidas de fibroblastos embrionários, a iPSCs, é facilitada pela sobreexpressão de Mbd3, que conduz a maiores níveis de complexo NuRD. Este aumento de eficiência parece ser fruto da sinergia entre o Nanog e o Mbd3 na indução de transcrição de genes necessários para pluripotência, antes da reprogramação das células em iPSCs. Resultados semelhantes foram obtidos na reprogramação de EpiSCs a iPSCs, onde a co-sobreexpressão de Nanog e Mbd3 leva a um aumento da eficiência da reprogramação celular.

Nesta tese demonstrei que o complexo NuRD desempenha um papel vital na formação de células estaminais pluripotentes induzidas e que um complexo previamente associado a diferenciação celular também desempenha um papel na reversão a um estado pluripotente.

Table of contents

List of figures

List of tables

List of common abbreviations

2i medium - defined medium containing inhibitors of both Erk and inhibitor of GSK3β

2i/LIF medium - defined medium containing inhibitors of both Erk and Inhibitor of GSK3β supplemented with leukaemia inhibitory factor

4-OHT - 4-hydroxytamoxifen

5-hmC- 5-hydroxymethylcytosine

5-mC - 5-methylcytosine

ac - acetylation

Act.A – activin A

Aid - activation-induced cytidine deaminase

AP - alkaline phosphatase

BAH - bromo-adjacent homology

BMP - bone morphogenetic protein

CC - coiled-coil motif

CD - chromodomains

cDNA - complementary DNA

ChIP-seq - chromatin immunoprecipitation followed by massively parallel sequencing

Chiron or CHIR99021 - inhibitor of glycogen synthase kinase-3β

Cit - citrullination

Co-F - co-factors

CR - conserved regions

d.p.t. - days post-transduction

D/B - driver or blocker of induced pluripotency

DMSO - dimethyl sulfoxide

DOX - doxycycline

DUF - domains of unknown function

DVA - doxycycline, vitamin C and Alki

E - embryonic day; or Glutamic acid (depending on context)

EB - embryoid body

ECC - Embryonic Stem Cell

EGF - epidermal growth factor

Egf+Fgf2 - defined medium supplemented with epidermal growth factor and fibroblast growth factor 2

EGG - Embryonic Germ Cell

ELM - egl-27 and MTA1 homology domain

EMT - epithelial-to-mesenchymal transition

EOS-GiP - GFPiresPuromycin under the control of early transposon promoter and Oct-4 and Sox2 enhancers

EpiSC - Epiblast-derived Stem Cell

esBAF - Embryonic Stem Cell-specific BAF complex

ESC - Embryonic Stem Cell

EST - expressed sequence tags

EtOH - ethanol

EV - empty vector

ex1 - exon 1

F - phenylalanine

FACs - fluorescence-activated cell sorting

FCS - foetal calf serum

FGF - fibroblast growth factor

Fgf2/Act.A - defined medium supplemented with fibroblast growth factor 2 and activin A

fl - flox

Gatad - GATA zinc finger domain

GATA-ZnF - GATA zinc finger

GFP - green fluorescent protein

GOF-18 - genomic Oct4 fragment 18 kb

GOF-18ΔPE - genomic Oct4 fragment 18 kb lacking Oct4 proximal enhancer

gp130 - glycoprotein 130

GR - glycine-arginine

GSK3β - glycogen synthase kinase-3β

H - histone

HDAC - histone deacetylase

ICM - inner cell mass

Id - inhibitors of differentiation (Id)

IGF - insulin-like growth factor

iMKOS - inducible plasmid encoding the reprogramming factors c-Myc, Klf4, Oct4 and Sox2

iOKSM - inducible plasmid encoding the reprogramming factors Oct4, Klf4, Sox2 and c-Myc

iPSC - induced Pluripotent Stem Cell

JAK/STAT3 - Janus kinase- signal transducer and activator of transcription 3

K - lysine

K2N - plasmid encoding Klf2 and Nanog

KD - knockdown

KO –knockout

KSR - knockout serum replacement

LIF - leukaemia inhibitory factor

LIR-R - leukaemia inhibitory factor receptor

MAPK/Erk - mitogen-activated protein kinase/ extracellular-signal-regulated kinase

Mbd - methyl-CpG-binding domain 3

Mbd3 - methyl-CpG-binding domain 3

me - methylation

MEF - mouse embryonic fibroblasts

MET - mesenchymal-to-epithelial transition

miR - microRNA

mRNA - messenger RNA

MTA - metastasis associated protein

N2B27 - basal media complemented with N2 and B27

Nanog-GFP - reporter where GFP is expressed under the control of Nanog

NEAA - non-essential amino acid

NSC - Neural Stem Cell

NuRD - Nucleosome remodelling and Deacetylase complex

Oct4-GFP - reporter where GFP is expressed under the control of Oct4

OE - overexpression

OSK - reprogramming factors Oct4, Sox2 and Klf4

OSKM - reprogramming factors Oct4, Sox2, Klf4 and c-Myc

PBS - phosphate buffered saline

PCR - polymerase chain reaction

PD03 or PD0325901- inhibitor of extracellular-signal-regulated kinase pathway

PE - proximal enhancer

Pen/Strep -penicillin/ streptomycin

PHD - plant homeodomain

PI3K/Akt - phosphoinositide-3-kinase- protein kinase B/Akt

PKC - protein kinase C

pMX-GFP - retroviral green fluorescent protein

pMXs - retroviruses

POU - Pit-Oct-Unc

pPB - piggyBac plasmids

PRC - polycomb repressive complex

preiPSC - pre- induced Pluripotent Stem Cell

qRT-PCR - quantitative reverse transcriptase - polymerase chain reaction

R - arginine

RB - retinoblastoma

Rbbp - retinoblastoma binding protein

rGFP - retroviral green fluorescent protein

rMKO - retroviruses encoding the reprogramming factors Oct4, Sox2, Klf4 and c-Myc

RNAi - RNA interference

S+LIF - serum-containing medium supplemented with leukaemia inhibitory factor

SCNT - somatic cell nuclear transfer

SCNT-ESCs - Embryonic Stem Cells generated from somatic cell nuclear transfer

shRNA - short hairpin RNA

siRNA - small interference RNA

SKM - reprogramming factors Sox2, Klf4 and c-Myc

S-LIF - serum-containing medium

SRY - sex-determining region Y

SSC - saline-sodium citrate

STDEV - standard deviation

TDG - thymine DNA glycosylase

TDR - transcriptional repression domain

TET - ten-eleven translocation methylcytosine dioxygenase

TGF-β - transforming growth factor-β

TSS - transcription start site

Vit.C - vitamin C

WCE - whole cell extracts input

WD - tryptophan-aspartic acid

WT - wild type

Y - tyrosine

Yamanaka cocktail or the Yamanaka factors - reprogramming factors Oct4, Sox2, Klf4 and c-Myc

RNA-seq - RNA sequencing

Publications arising from this work

dos Santos, R.L., Tosti, L., Radzisheuskaya, A., Caballero, I.M., Kaji, K., Hendrich, B., and Silva, J.C. (2014). MBD3/NuRD Facilitates Induction of Pluripotency in a Context-Dependent Manner. Cell Stem Cell 15, 102-110.

Radzisheuskaya, A., Chia Gle, B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol 15, 579-590.

Note: Full texts of the manuscripts are included in the Appendix section.

CHAPTER 1 – General Introduction

1.1 – Pluripotency in the embryo and in culture

Mammalian development is a unidirectional process paralleled with a progressive loss of developmental potential. It begins with the formation of a unicellular zygote, and as cells divide and their numbers multiply, it ends with more than 200 specialized cell types being generated. Since all cells carry the same genome but have different functions, a key question is how these different stable cell states are established and maintained, and whether these states can be perturbed resulting in changes in cell fate. In this Introduction I will describe how pluripotency arises during mammalian development; how this cell state can be captured *in vitro* in the form of Embryonic Stem Cells (ESCs); the key gene regulatory network that controls the pluripotent state; the nuclear reprogramming of adult cells to a pluripotent state; and the role of the Nucleosome remodelling and Deacetylase complex (NuRD) during development and ESC maintenance.

1.1.1 – The establishment of naïve pluripotency in the mouse embryo

Mouse development progresses from a state of totipotency in the zygote, which is formed after the fertilization of the oocyte by sperm. The zygote undergoes 3-4 rounds of cleavage divisions, leading to the generation of a structure containing 8-16 totipotent blastomeres, called morula. At this stage, blastomeres start to compact and the first lineage choice takes place, resulting in the generation of two cell lineages with restricted developmental potential: trophectoderm (outside) and inner cell mass (ICM, inside) (Schrode et al., 2013) (Figure 1.1.1A). The trophectoderm lineage, which will form part of the placenta, is specified by expression of trophectoderm specification factors Cdx2 and Gata3, which are activated by the transcription factor Tead4 (Nishioka et al., 2009; Ralston et al., 2010). Both cell polarity and differential activation of Hippo signalling in the cells of the compacted morula are implicated in the process of trophectoderm segregation. Outside cells that will become the trophectoderm are polarized, having inactive Hippo pathway, allowing Yap/Taz to be translocated to the nucleus, where they activate Tead4 (Nishioka et al., 2009). The Hippo signalling is activated in the unpolarised cells of ICM, which results in the phosphorylation and degradation of Yap/Taz, which keeps Tead4 repressed (Nishioka et al., 2009). On the other hand, ICM cells express the homeodomain protein Nanog and the Pit-Oct-Unc (POU) domain protein Oct4 (also known as Pou5f1), which are repressed in trophectoderm cells (Nichols and Smith, 2012; Niwa et al., 2005). Since Oct4 and Cdx2 reciprocally inhibit each other, the down-regulation of Oct4 in trophectoderm cells due to Cdx2 expression and Cdx2 down-regulation in ICM cells due to Oct4 expression is likely to be a mechanism underlying the stabilization

of this first lineage specification (Nichols and Smith, 2012; Niwa et al., 2005) (Figure 1.1.1B).

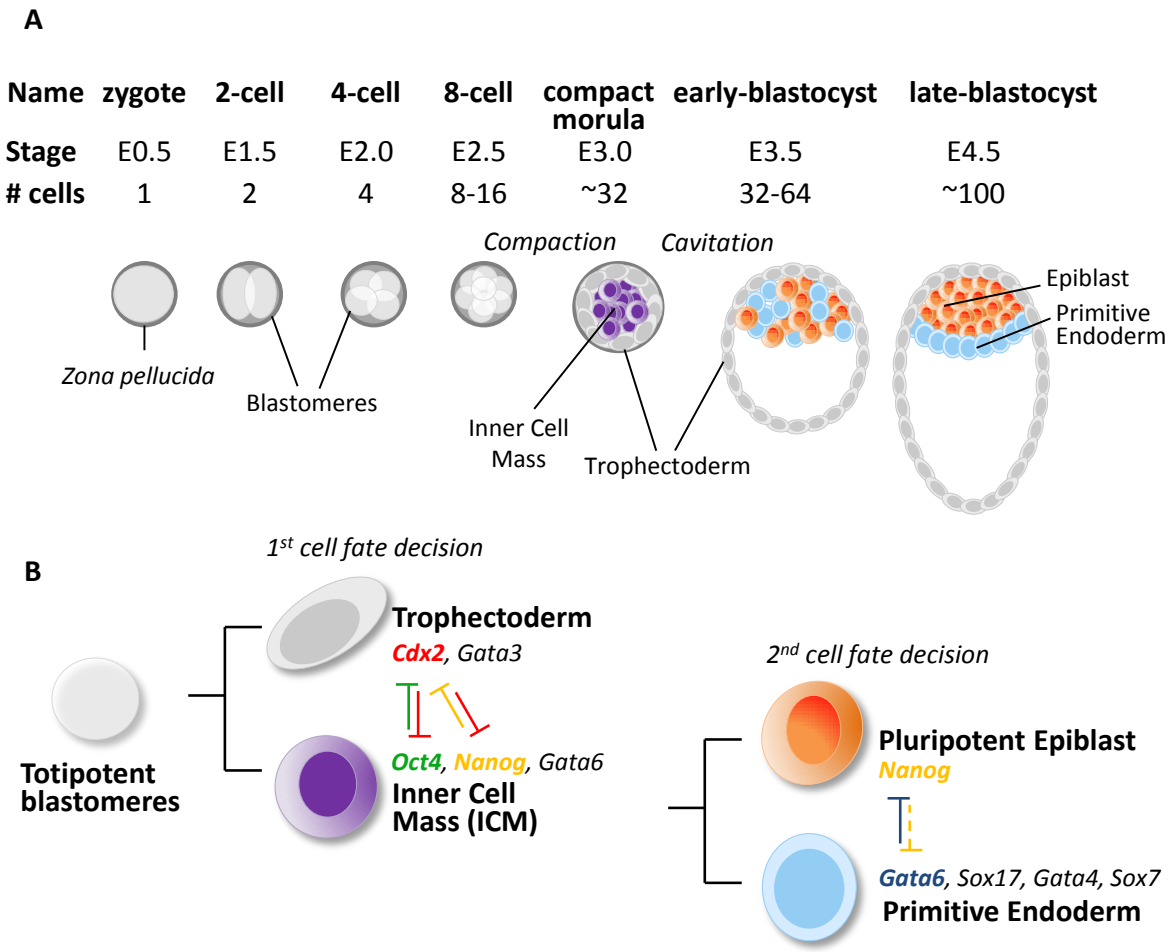


Figure 1.1.1 – Schematic representation of mouse pre-implantation development.

(A) After fertilization, the totipotent zygote undergoes several rounds of division leading to the generation of a compacted morula that consists of two cell lineages: trophectoderm (outside) and inner cell mass (ICM, inside). At the blastocyst stage, the cells from inner cell mass undergo a second fate decision, resulting in the segregation of the naïve pluripotent epiblast and primitive endoderm. **(B)** Schematization of the two lineage choices that take place during pre-implantation mouse development. The first gives rise to trophectoderm and ICM and the second leads to the allocation of epiblast and primitive endoderm. Typical markers of each cell lineage are indicated. Reciprocal inhibition between Oct4 (and possibly Nanog) and Cdx2 during the first lineage specification, and between Nanog and Gata6 during the second, are indicated. E: embryonic day. Adapted from (Schrode et al., 2013).

Approximately 12 hours after the first lineage segregation (embryonic day E3.5), there is a formation of a cavity inside compacted morula, resulting in the generation of an early blastocyst where the ICM is pushed to one side. The cells within the ICM start expressing lineage-specific transcription factors, Nanog and Gata6, which are further restricted until a distribution of cells exhibiting mutually restricted transcription factor emerges (Chazaud et

al., 2006; Nichols and Smith, 2011). A subset of ICM cells down-regulate Gata6, while still expressing Nanog, priming them towards an epiblast state. The remaining ICM cells down-regulate Nanog but maintain Gata6 expression and subsequently up-regulate Sox17, which biases them for a primitive endoderm state (Artus et al., 2011). At embryonic day 4.5 (E4.5), this "salt-and-pepper" distribution of epiblast- and primitive-endoderm primed cells is established, with primitive-endoderm cells sorted to form an epithelized layer positioned adjacent to the epiblast at the interface with the blastocyst cavity that will form the yolk sac (Nichols and Smith, 2011). Although different models have been hypothesized to explain this second lineage specification, the mechanisms are not fully understood. It is likely that a mechanism involving Gata6-induced repression of Nanog expression in the cells that will become primitive endoderm is in place (Niwa, 2007). However, there is no direct evidence to support this suggestion. The pre-implantation epiblast, also termed naïve epiblast, represents the pluripotent compartment of the embryo and will form all somatic and germline cells of the developing embryo (Nichols and Smith, 2011). Oct4 also plays a pivotal role in this lineage decision, since Oct4 deletion at the morula to blastocyst transition blocks ICM segregation into epiblast and primitive endoderm, implicating Oct4 in both epiblast and primitive endoderm differentiation (Le Bin et al., 2014).

1.1.2 – Capturing naïve pluripotency *in vitro*

The pluripotent cells of the pre-implantation epiblast can be captured in culture in the form of Embryonic Stem Cells (ESCs) (Evans and Kaufman, 1981; Martin, 1981). The initial derivations of ESCs were made possible due to the previous discovery and study of embryonic carcinoma cells (ECCs), another type of pluripotent cells generated from explants of teratocarcinomas (Hogan, 1976). Based on the same protocol as for ECCs isolation, ESCs were initially derived by the culture of blastocysts or ICM on a feeder layer of mitotically inactivated fibroblasts using media containing foetal calf serum (FCS) (Evans and Kaufman, 1981; Martin, 1981). Mouse ESCs share many characteristics with the ICM cells, exhibiting expression of key pluripotency genes such as Oct4, Sox2 and Nanog, and two active X chromosomes in female cells. Mouse ESCs have two key features: self-renewal capacity, as they are able to be maintained indefinitely in culture; and developmental potential, as they can efficiently be differentiated into all three embryonic lineages in culture when subjected to differentiation stimuli (Doetschman et al., 1985; Nichols and Smith, 2012), or when introduced into the mouse kidney capsule, resulting in teratomas. ESCs can also be differentiated into germ cells *in vitro* (Geijsen et al., 2004; Toyooka et al., 2003), and can recapitulate full developmental potential when injected into blastocysts, contributing to the

three germ layers and to the germline (Bradley et al., 1984). Furthermore, it is possible to derive liveborn mice that are entirely composed of ESC derivatives (Nagy et al., 1993).

1.1.3 – Key signalling pathways of ESCs

Maintaining the pluripotent state of ESCs depends on key signalling pathways (Figure 1.1.2). ESCs were first derived and cultured on a feeder layer in presence of serum (Evans and Kaufman, 1981; Martin, 1981). Subsequent studies identified the soluble signalling molecule leukaemia inhibitory factor (LIF) (Smith et al., 1988; Williams et al., 1988), which allows both the derivation and maintenance of ESCs in the absence of a feeder layer (Nichols et al., 1990; Pease et al., 1990). LIF signalling is transduced through the LIF receptor (LIF-R) and glycoprotein 130 (gp130) and activates distinct signalling pathways involved in maintenance and differentiation of ESCs (Burdon et al., 1999; Matsuda et al., 1999; Niwa et al., 1998). The most potent signalling pathway activated is the Janus kinase- signal transducer and activator of transcription 3 (JAK/STAT3) pathway (Matsuda et al., 1999; Niwa et al., 1998), which promotes self-renewal by maintaining the expression of key transcription factors, such as Klf4 (Niwa et al., 2009), Gbx2 (Tai and Ying, 2013), and Tfcp2l1 (Martello et al., 2013; Ye et al., 2013). LIF also induces the phosphoinositide-3-kinase- protein kinase B/Akt (PI3K/Akt) pathway, mediating the transcription of Tbx3, which enhances Nanog expression, stimulating self-renewal (Niwa et al., 2009). The PI3K/Akt pathway can also be activated by insulin and insulin-like growth factors (IGFs) (Hemmings and Restuccia, 2012). However, LIF also stimulates the mitogen-activated protein kinase/ extracellular-signal-regulated kinase (MAPK/Erk) pathway (Burdon et al., 1999), which acts to promote differentiation. This differentiation effect of MAPK/Erk pathway seems to be compensated by the promotion of self-renewal caused by the JAK/STAT3 and PI3K induction. Other activators of the MEPK/Erk pathway include fibroblast growth factors (FGFs), mainly Fgf2 and Fgf4, whose addition to ESC culture triggers them to differentiate (Kunath et al., 2007).

Further studies to identity the key components of the foetal serum, used during ESC culture, revealed that one of the most important active growth factors is bone morphogenetic protein 4 (BMP4) (Ying et al., 2003). Cultivation of ESCs in defined medium (serum-free) supplemented with BMP4 allows maintenance of ESCs in the absence of a feeder layer. Mechanistically, BMP4 seems to synergize with the LIF signalling, supporting self-renewal of ESCs by the induction of inhibitors of differentiation (Id) genes (Ying et al., 2003).

Derivation of ESCs using serum and feeders is very inefficient and restricted to particular mouse strains, particularly to the 129 strain (Evans and Kaufman, 1981; Martin, 1981). The efficiency of ESC derivation is increased when LIF is added to medium (Smith et al., 1988;

Williams et al., 1988), and combination of LIF, BMP4 and the pharmacological inhibition of Erk was shown to allow ESC derivation from other mouse strains which had been refractory to serum-containing medium derivation, such as C57BL/6 (Batlle-Morera et al., 2008). Though the inhibition of Erk signalling plays a pivotal role in ESC derivation and culture, Erk inhibition alone is not sufficient to maintain ESCs self-renewal (Ying et al., 2008). However, the addition of a selective glycogen synthase kinase-3β (GSK3β) inhibitor to the defined medium containing the Erk inhibitor allows the culture of self-renewing ESCs without any sign of cell differentiation (Ying et al., 2008). GSK3 inhibition leads to the de-repression of the canonical Wnt signalling pathway, which results in the release of Tcf3-mediated repression of pluripotency genes (Wray et al., 2011), such as Esrrβ (Martello et al., 2012). Interestingly, the exogenous supply of Wnt ligand, Wnt3a, is not sufficient to maintain self-renewal in ESCs in which Erk signalling is blocked (Ying et al., 2008). Thus, the effect of GSK3β inhibition is thought not only to enhance the Wnt signalling but also to improve the general biosynthetic capacity of the cell (Sato et al., 2004). This defined medium containing inhibitors of both Erk (PD0325901, abbreviated PD03) and GSK3β (CHIR99021, abbreviated Chiron) is named 2i, and, even though LIF is not required for ESC culture using this media, the addition of LIF enhances self-renewal (Marks et al., 2012; Ying et al., 2008). Importantly, the usage of 2i/LIF medium allowed highly efficient derivation of ESC from embryos from various mouse strains (Nichols et al., 2009; Ying et al., 2008), mouse embryonic germ cells (EGCs) (Leitch et al., 2010), and also of pluripotent cells from other rodent species, such as germline competent rat ESCs (Buehr et al., 2008; Li et al., 2008) and rat EGCs (Leitch et al., 2010). Lastly, it was recently reported that one more signalling pathway plays an important role in ESC self-renewal. It was demonstrated that inhibition of protein kinase C signalling (PKC) is sufficient for derivation and culture of ESCs in defined medium, an effect that seems dependent on the inhibition NF-kB activity, and is independent of the signalling pathways described above (Dutta et al., 2011).

In summary, the culture of self-renewing ESCs is possible when self-renewal-associated pathways such as JAK/STAT3, PI3K/Akt or canonical Wnt signalling are stimulated; or when differentiation associated pathways such as MAPK/Erk pathway, GSK3β and PKC are inhibited. All these signalling pathways are schematized in Figure 1.1.2.

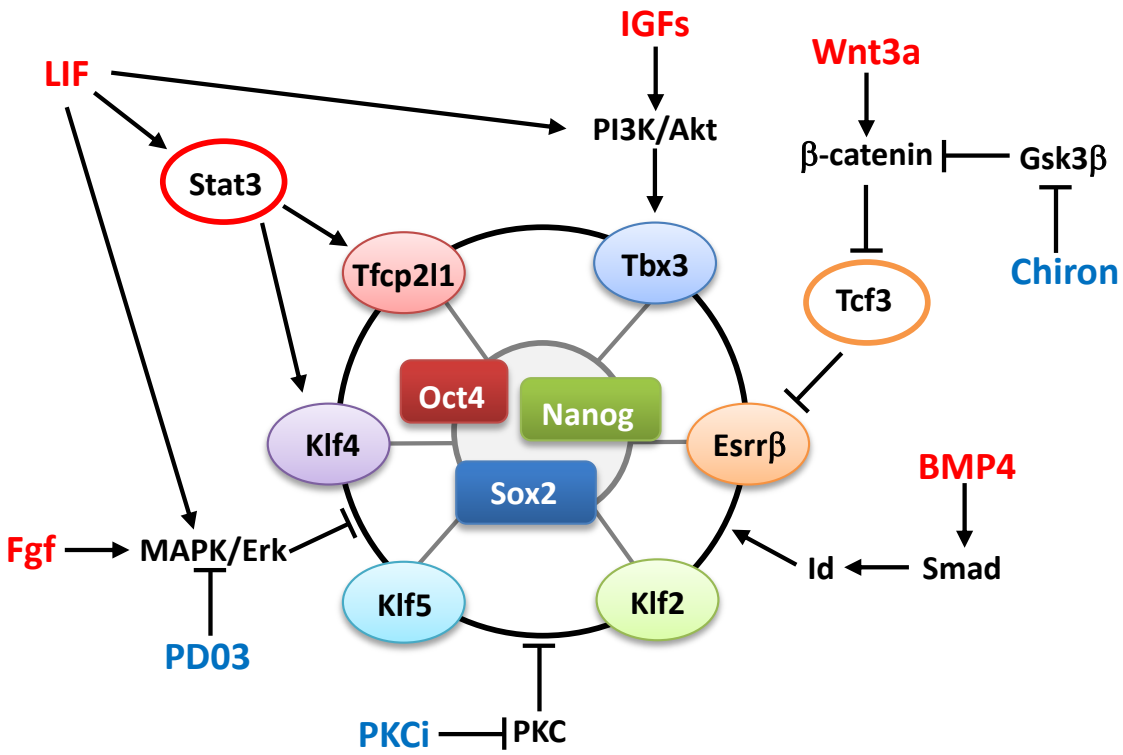


Figure 1.1.2 – Signalling pathways and key transcription factors in mouse ESCs.

The main signalling pathways and transcription factors involved in ESC self-renewal are represented. The exogenous soluble proteins that activate the corresponding signalling cascades are in indicated in red, and the selective pharmacological inhibitors of corresponding enzymes are in indicated in blue. In the inner circle are the core pluripotency transcription factors (Oct4, Nanog and Sox2) and in the outer circle are other transcription factors that play a very important role in the maintenance of pluripotency. Of note, some transcription factors such as Stat3, Tcf3 and Tbx3 are regarded as the main bridges between JAK/STAT3, Wnt and PI3K/Akt signalling pathways inputs, and the transcriptional factor core of naïve pluripotency. For further details about the signalling pathways, please refer to the text.

1.1.4 – Core transcriptional factor network of naïve pluripotency

Naïve pluripotent ESCs are characterized by a unique network of transcription factors, with Oct4, Sox2 and Nanog being considered the main keepers of pluripotency.

Oct4 (also known as *Pou5f1*) is a Pit-Oct-Unc (POU) family transcription factor. It was first identified as a transcription factor specifically expressed in ECCs, being down-regulated upon differentiation (Lenardo et al., 1989; Scholer et al., 1989; Scholer et al., 1990). Oct4 expression was further identified as being restricted to pluripotent and germ-line cells (Scholer et al., 1990), and required for the establishment of the naïve pluripotent compartment in the embryo (Nichols et al., 1998). *Oct4*-null ESCs cannot be recovered from outgrowths of *Oct4*$^{-/-}$ blastocysts (Nichols et al., 1998). *Oct4* depletion in ESCs triggers their conversion into

trophectoderm, and its overexpression induces ESC differentiation into a mixture of lineages (Niwa et al., 2000). Interestingly, recently published work by our and the Chambers laboratory revealed that a defined level of Oct4 is essential not only for pluripotency maintenance, but also for the exit from pluripotency into differentiation, both *in vitro* and *in vivo* (Karwacki-Neisius et al., 2013; Radzisheuskaya et al., 2013). Therefore, the Oct4 levels have to be tightly controlled to sustain, and allow the exit from, pluripotency.

The best-characterized Oct4 partner is the sex-determining region Y (SRY)-related box protein Sox2 (Rizzino, 2009). The cooperation between Oct4 and Sox2 was first described in ECCs, where their interaction was found to be required for cell-type-restricted activity of the *Fgf4* enhancer (Dailey et al., 1994). Subsequent studies revealed that a similar mechanism exists in ESCs (Yuan et al., 1995), and that an Oct4/Sox2 binding element is present upstream of many pluripotent-associated genes (Chambers and Tomlinson, 2009). *Sox2*$^{-/-}$ embryos fail to establish a naïve pluripotent epiblast at E4.5 and die shortly after implantation. ESCs cannot be derived from these embryos (Avilion et al., 2003). The conditional deletion of Sox2 in ESCs results in trophectoderm differentiation (similar to Oct4 deletion), a phenotype that can be rescued by constitutive expression of Oct4 at wild type levels (Masui et al., 2007).

Nanog is a homeodomain-containing protein, which was initially discovered by two groups that used two different approaches: Smith and colleagues isolated Nanog in a functional complementary DNA (cDNA) screen to identify factors that confer LIF-independent self-renewal on ESC (Chambers et al., 2003); and Yamanaka and colleagues identified Nanog by comparing expressed sequence tags (EST) between ESCs and somatic cells (Mitsui et al., 2003). Similar to Oct4 and Sox2 deletion, *Nanog*$^{-/-}$ embryos failed to establish a naïve pluripotent compartment and died before implantation, and no ESCs can be derived from these blastocysts (Mitsui et al., 2003). Yet, pluripotent *Nanog*-null ESCs can be generated and maintained *in vitro* using targeting by homologous recombination, despite being highly prone to differentiation. They contribute to all tissues in chimeric mice upon blastocyst injection (Chambers et al., 2007). However, *Nanog*$^{-/-}$ ESCs fail to mature in the genital ridge, failing to contribute to the germline (Chambers et al., 2007).

Oct4, Sox2 and Nanog constitute the core of transcription factors which are associated with the pluripotency state. They are essential for the establishment of pluripotent cells *in vivo* and for ESC maintenance and function. Other transcription factors have been identified and functionally validated as playing pivotal roles in ESC self-renewal and differentiation (Table 1.1.1; more information about their roles in ESCs can be found in (Nichols and Smith, 2012)). These are regarded as regulators of pluripotency, since they are individually dispensable for

ESC maintenance, but work together to regulate the expression of the core transcription factors, and thus, the pluripotent state (Figure 1.1.2).

Table 1.1.1 – Phenotypes associated with key mouse ESC transcription factors.

Transcription factor	Embryonic phenotype	ESC loss of function phenotype	ESC gain of function phenotype	References
Esrrβ	Placental failure; mid-gestation lethality	None (in 2i/LIF)	LIF-independent self-renewal	(Festuccia et al., 2012; Luo et al., 1997; Martello et al., 2012)
Klf2	Viable	None (by RNAi)	LIF-independent self-renewal	(Hall et al., 2009; Jiang et al., 2008; Kuo et al., 1997)
Klf4	Viable	None (by RNAi)	LIF-independent self-renewal	(Hall et al., 2009; Jiang et al., 2008; Segre et al., 1999)
Klf5	Early Implantation defect; embryonic lethality	None	LIF-independent self-renewal	(Ema et al., 2008; Hall et al., 2009; Jiang et al., 2008)
Nanog	Loss of epiblast and primitive endoderm	Prone to differentiation, but continue self-renewing	LIF-independent self-renewal	(Chambers et al., 2003; Chambers et al., 2007; Mitsui et al., 2003; Silva et al., 2009)
Nr5a1/2	Early post-implantation lethality; failed gastrulation	None	No evident role	(Gu et al., 2005; Guo and Smith, 2010)
Nr0b1	Not reported	Differentiation, mainly to trophectoderm	No evident role	(Khalfallah et al., 2009; Niakan et al., 2006; van den Berg et al., 2010)
Oct4	ICM becomes trophectoderm	Differentiation to trophectoderm	Differentiation	(Nichols et al., 1998; Niwa et al., 2000)
Rex1	Viable	None	No evident role	(Climent et al., 2013; Masui et al., 2008; Scotland et al., 2009)
Sall4	Early post-implantation lethality	Tendency for spontaneous differentiation	No evident role	(Sakaki-Yumoto et al., 2006; Tsubooka et al., 2009; Yuri et al., 2009; Zhang et al., 2006)
Sox2	Early post-implantation lethality	Differentiation to trophectoderm	Not reported	(Avilion et al., 2003; Masui et al., 2007)

Stat3	Post-gastrulation lethality	None (in 2i/LIF)	Exogenous activation of LIF signalling substitutes for LIF	(Do et al., 2013; Takeda et al., 1997; Ying et al., 2008)
Tbx3	Mid-gestation lethality	Partially impaired self-renewal (RNAi)	LIF-independent self-renewal	(Davenport et al., 2003; Ivanova et al., 2006; Niwa et al., 2009)
Tcf3	Disrupted axial patterning; lethality at gastrulation	Reduced LIF dependence	Differentiation	(Guo et al., 2011; Pereira et al., 2006; Wray et al., 2011)
Tfcp2l1	Viable	None (in 2i/LIF)	LIF-independent self-renewal	(Martello et al., 2013; Yamaguchi et al., 2005; Ye et al., 2013)
Tfe3	Viable	None (in 2i/LIF)	Nuclear expression can cause cells to withstand differentiation	(Betschinger et al., 2013; Steingrimsson et al., 2002)

RNAi – RNA interference; Esrrβ: Estrogen-related receptor β; Klf2/4/5: Krüppel-like Factor 2/4/5; Nr5a1/2: Nuclear receptor subfamily 5 group A member ½; Nr0b1: Nuclear receptor subfamily 0 group B member 1; Oct4: Octamer-binding transcription factor 4; Rex1: Reduced expression 1; Sox2: Sex-determining region Y (SRY)-box2; Stat3: Signal transducer and activator of transcription 3; Tbx3: T-box transcription factor 3; Tcf3: Transcription factor 3; Tfcp2l1: Transcription factor CP2-like 1; Tfe3: Transcription factor E3.

Pluripotent ESCs are thought to be maintained due to the regulation of pluripotency transcription factors by positive-feedback loops, which enables their ongoing expression until this core network is dismantled (Boyer et al., 2005; Jaenisch and Young, 2008; Young, 2011). In addition to positive auto-regulation, these pluripotency factors act cooperatively with members of repressive chromatin-modifying machinery (like Polycomb repressive complexes, PRCs) to repress the expression of lineage-specification genes (Young, 2011). Moreover, several transcription factors that regulate the pluripotent state are known mediators of key ESC signalling pathways (Figure 1.1.2), demonstrating that the pluripotent state is maintained due to an intricate connection between the environment (culture conditions) and transcription factors.

In summary, pluripotent cells, both in the embryo and in culture, are sustained by a flexible transcription factor network which is responsive to different signals, which can either

sustain pluripotency or disassemble the naïve pluripotency associated transcriptional network, leading to cell differentiation.

1.1.5 – Post-implantation epiblast and primed pluripotent stem cells

As described above, at E4.5 the blastocyst is composed of three structures: the naïve pluripotent epiblast, the primitive-endoderm epithelia and the trophectoderm. This mature blastocyst then implants into the uterus, a process that is paralleled with the morphological transformation of the naïve epiblast into a cup-shaped structure, known as the egg cylinder (Figure 1.1.3). Since the embryo is now physically constrained, the continuous growth of the epiblast is possible due to its movement into the blastocyst cavity. At this stage, female epiblast cells undergo random X chromosome inactivation, naïve pluripotency markers such as *Rex1*, *Klf4* and *Essrβ* are down-regulated and lineage-commitment genes such as *Fgf5* and *T-brachyury* start to be expressed (Nichols and Smith, 2009).

Since Oct4 and Nanog are expressed in the post-implantation epiblast, this structure is regarded as pluripotent. Yet, attempts to derive ESCs from the post-implantation epiblast using standard ESC conditions have failed (Nichols and Smith, 2009). However, post-implantation epiblast-derived stem cells (EpiSCs) can derived if Fgf2 and Activin A-containing serum-free medium is used (Brons et al., 2007; Tesar et al., 2007). EpiSCs are dependent on Fgf signalling, being morphologically and molecularly distinguishable from ESCs (Figure 1.1.3). Like the cells of the post-implantation epiblast, they express *Oct4* but not *Rex1* and *Klf4*, and up-regulate expression of *Fgf5* and *T-brachyury* (Nichols and Smith, 2009). Genetic studies indicated that Oct4 expression is required for post-implantation pluripotency, whereas Nanog is not (Osorno et al., 2012). EpiSCs can be differentiated into all embryonic lineages (Brons et al., 2007; Tesar et al., 2007) including primordial germ cells (Hayashi and Surani, 2009), but they do not contribute efficiently to chimeras when injected into blastocysts. However, if EpiSCs are injected in the post-implantation epiblast explanted in culture, they can contribute to all tissues and germline (Huang et al., 2012). EpiSCs can also be established by the culture of ESCs (Guo et al., 2009) or the pre-implantation epiblast (Najm et al., 2011) in Fgf2/Activin A, which triggers the differentiation of these naïve cells into the primed pluripotent state.

Interestingly, although human ESCs are isolated from pre-implantation blastocysts like mouse ESCs (Thomson et al., 1998), they share more characteristics with mouse EpiSCs than mouse ESCs (Hanna et al., 2010b; Nichols and Smith, 2009). The morphological and

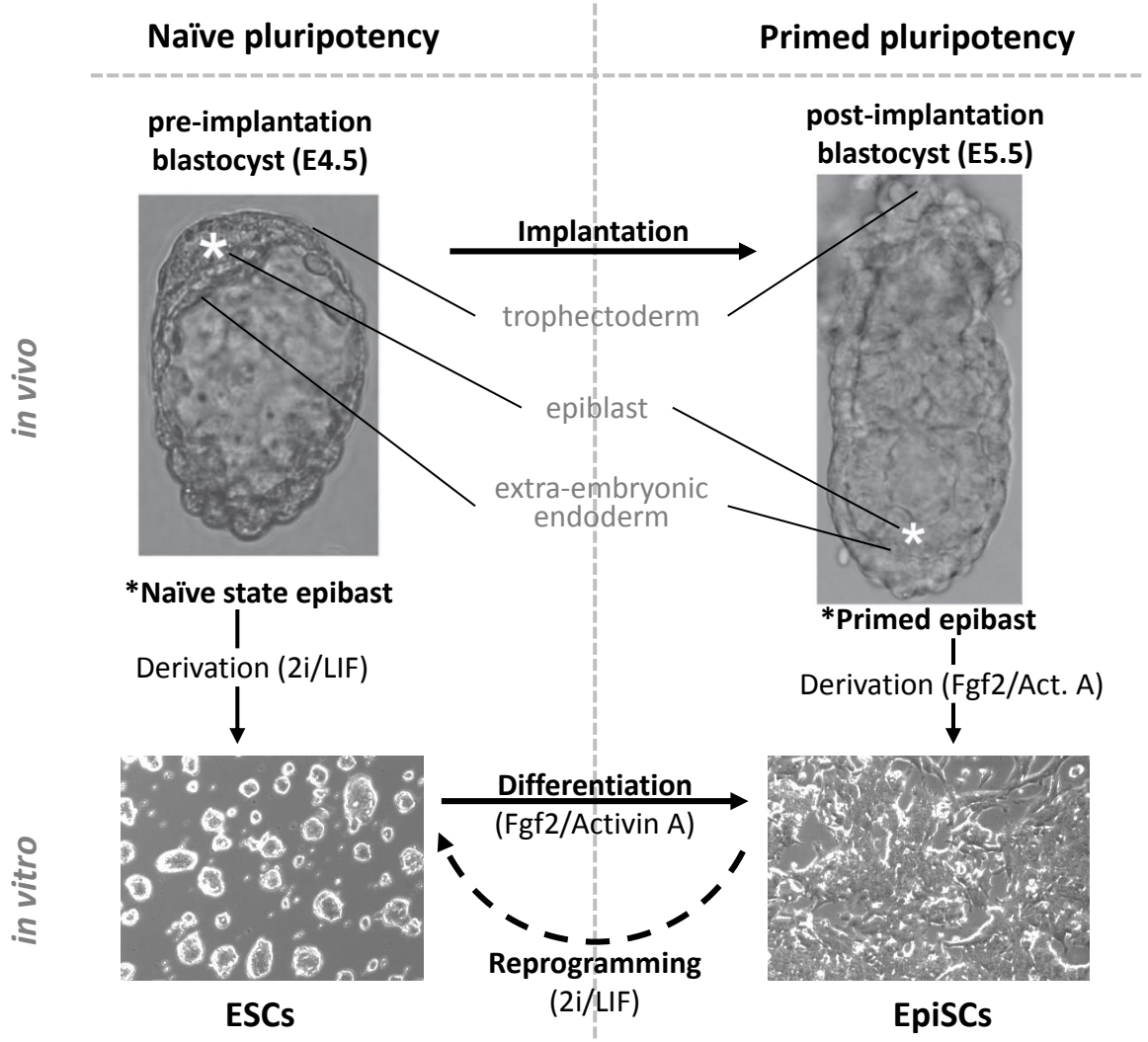
biological similarities between human ESCs and mouse EpiSCs suggest that human ESCs correspond to a primed state of pluripotency. Whether the naïve human pluripotent state exist remains open. Indeed, recent efforts to establish naïve human ESCs showed that different human ESCs can be established, which are more similar to mouse ESCs (Chan et al., 2013; Gafni et al., 2013; Hanna et al., 2010a). However, further characterization of these cells is necessary, addressing if these cells are true naïve pluripotent cells or correspond to a different primed state.

In summary, different pluripotent cells can be derived from the mouse embryo: ESCs, which are derived from the pre-implantation epiblast, and EpiSCs, derived from the post-implantation epiblast. Since ESCs fail to re-enter blastocyst development, they represent a more committed pluripotent state, called the primed pluripotent state, which is molecularly distinguishable from the naïve pluripotent state.

(figure on next page)

Figure 1.1.3 – Comparison of Naïve and Primed Pluripotent states.

Pluripotency comes in two flavours during embryo development: ground state naïve pluripotency, which is established in the epiblast of the mature blastocyst (E4.5) and can be captured *in vitro* in the form of ESCs (culture in 2i/LIF or serum+LIF); and primed pluripotency, which is established in the post-implantation epiblast (E5.5), a cup-shaped epithelium that becomes primed for lineage specification and commitment in response to external stimulus from the extra-embryonic tissues, and can be captured *in vitro* in the form of EpiSCs (culture in Fgf2 and Activin A). EpiSCs can also be generated by the differentiation of ESCs after their exposure to Fgf2 and Actvin A. EpiSCs can be reverted back after the forced expression of transcription factors, such as Klf4, Klf2 or Nanog, followed by culture in 2i/LIF (Guo et al., 2009; Silva et al., 2009). Tables indicate the differences and similarities between ESCs and EpiSCs. Embryo images from (Nichols and Smith, 2009)



Differences	ICM-like state	Epithelial epiblast-like state
	LIF-responsive	Fgf2/Activin A-responsive
	Self-renewal response to 2i	Differentiation/ death response to 2i
	High clonogenicity	Low clonogenicity
	Two active X chromosomes (XaXa)	One inactive X chromosome (XaXi)
	Chimera and germline competent	Little or no contribution to chimeras
	Express Rex1, Klf4, Klf2, Esrrβ, Nr0b1	Express Fgf5, Lefty, T (brachyury)
	Use of Oct4 distal enhancer	Use of Oct4 proximal enhancer
Similarities	Teratomas with derivatives of all three germ layers	
	Express Oct4, Nanog and Sox2	

.

1.2 – Nuclear reprogramming

During embryo development, pluripotent cells in the epiblast start to commit to become precursors of different germ layers (endoderm, ectoderm and mesoderm), which will then give rise to the cells and tissues of an adult. The process of cellular differentiation was thought to be irreversible and unidirectional, and was represented by Waddington as a marble rolling down the mountain of developmental potential. This classical view of development hypothesized that different cell types and states were defined by boundaries that could not be crossed. However, several studies challenged this view, demonstrating the even terminally differentiated cells are "plastic" and amenable to switch fate, being able to be converted back to a pluripotent state or to a different cell type, altering their gene expression, and therefore their fate. This process is termed nuclear reprogramming, and can be achieved *in vitro* by three different experimental approaches: somatic cell nuclear transfer, cell fusion, and transcription factor-mediated direct reprogramming.

Somatic cell nuclear transfer (SCNT) consists of the transfer of the nucleus from a somatic cell into an enucleated oocyte (an oocyte whose nucleus has been removed) (Figure 1.2.1A). The first SCNT experiments were performed in amphibians using either nuclei from early blastocysts (Briggs and King, 1952) or intestinal cells (Gurdon, 1962a, b; Gurdon et al., 1958). Results from Gurdon and colleagues showed that the exposure of somatic nuclei to the cytoplasm of oocyte is enough to reprogram this nucleus to a pluripotent state, which is able to re-enter development and create an entire normal adult frog (Gurdon et al., 1958). These seminal experiments proved that all the genes that are required to create an entire organism are present in the nucleus of all specialized cells, indicating that during development no permanent genetic changes occur, and that all cell boundaries are epigenetically defined. Forty years later, SCNT was successfully reported in mammals, with Dolly sheep being the first cloned mammal (Campbell et al., 1996; Wilmut et al., 1997). For that, the blastocyst formed after cell division of a manipulated oocyte (now with a "somatic" nucleus) was transferred into the uterus of a foster female sheep, where it could implant and continue embryonic development (Campbell et al., 1996). The same protocol was then used to generate the first cloned mouse (Wakayama et al., 1998) and to generate mouse ESCs (SCNT-ESCs) from adult mice (Munsie et al., 2000). SCNT-ESCs can re-enter blastocyst development and contribute to cloned mice (Hochedlinger and Jaenisch, 2002). Many efforts were made to derive patient-specific human SCNT-ESCs using this method. However, after many attempts, only recently did scientists succeed in establishing human SCNT-ESCs (Chung et al., 2014; Kang et al., 2014; Tachibana et al., 2013; Yamada et al., 2014).

Cell fusion consists of the fusion of two cells to form a single identity (Figure 1.2.1B). Cell fusion occurs in two steps. Firstly, short-lived products which do not proliferate, termed heterokaryons, are generated. These represent two nuclei in a common cytoplasm. Then, the nuclei fuse generating tetraploid hybrids through a very inefficient process that requires cell proliferation. Cell fusion was first developed to study the impact of one genome over the other, which led to the discovery of trans-acting repressors by the fusion of two somatic cells (Davidson et al., 1966; Harris et al., 1969). By fusing cells from different species, which leads to the formation of mixed-species heterokaryons, cell fusion also provided the first definitive evidence that silent genes could be activated (Blau et al., 1983; Blau et al., 1985). Further studies shown that nuclei of adult mouse thymocytes could be reprogrammed back to pluripotency by fusion with either EGCs (Tada et al., 1997) or ESCs (Tada et al., 2001). The tetraploid hybrids resulting from the fusion between thymocytes and ESCs exhibited the properties of pluripotent cells (gene expression, differentiation into all germ layers), and could contribute to chimeric embryos, although very inefficiently, due to their tetraploid genome (Tada et al., 2001). Human somatic cells can also be reprogrammed to pluripotent cells after fusion with human ESCs (Cowan et al., 2005). The efficiency of somatic cell-ESC cell fusion can be boosted if either the somatic cells or ESCs used overexpress Nanog (Silva et al., 2006). This study led to the discovery of the first reprogramming factor, which together with the ESC machinery can enhance reprogramming efficiency. However, the overexpression of Nanog alone in somatic cells is not sufficient to drive their reprogramming to a pluripotent state (Silva et al., 2006).

Transcription factor-mediated reprogramming consists of the direct alteration of cell fate by the forced expression of tissue-specific transcription factors. The fact that single transcription factors can trigger a cascade of events, leading to altered gene expression profiles that ultimately result in phenotypic alteration was first identified in the fly. Forced expression of key transcription factors in *D. melanogaster* larvae led to the development of an additional set of legs instead of antennae (Schneuwly et al., 1987), or to the formation of functional eyes on the legs, antennae and wings (Gehring, 1996). Alteration of cell fate by the ectopic expression of transcription factors was then reported in the mouse, where the forced expression of MyoD induces the conversion of fibroblasts or other cell types to myoblasts (Davis et al., 1987; Weintraub et al., 1989). This process of direct programming of a cell type into another, bypassing an intermediate pluripotent state, by the forced expression of transcription factors that are characteristic of the desired cell type is termed transdifferentiation (Graf and Enver, 2009). In 2006, Takahashi and Yamanaka demonstrated that a similar approach can be employed to reprogram mouse embryonic or adult fibroblast

back to pluripotency (Takahashi and Yamanaka, 2006). In this seminal study, 24 genes that are normally expressed in ESCs were cloned and delivered retrovirally into fibroblasts. These factors were then screened for their ability to collectively induce pluripotency, which was assessed by the reactivation of endogenous *F box only protein 15* (*Fbx15)* in transduced cells, a gene previously identified as being specific to ESCs (Tokuzawa et al., 2003). Two-weeks after the transduction of the 24 factors, Fbx15-positive colonies emerged, at a 0.01–0.1% efficiency, showing similar morphology, growth and gene expression to ESCs, forming teratomas when injected into mice (Takahashi and Yamanaka, 2006). Since these cells were induced back to pluripotency, they were named induced pluripotent stem cells (iPSCs). By the progressive elimination of factors, Takahashi and Yamanaka found a cocktail of four factors that is sufficient induce pluripotency from fibroblasts: Oct4, Sox2, Klf4 and c-Myc, which will be referred throughout the as reprogramming factors, OSKM, the "Yamanaka cocktail" or the "Yamanaka factors" (Takahashi and Yamanaka, 2006) (Figure 1.2.1C).

The initial iPSCs generated by Takahashi and Yamanaka are thought to be reprogramming intermediates and not truly pluripotent cells, since when injected into a mouse blastocyst, they do not contribute to adult chimeric animals (Takahashi and Yamanaka, 2006). However, when the emerging iPSCs are selected for *Nanog* or *Oct4* expression instead of *Fbx15* expression, truly naïve germ-line competent pluripotent cells can be isolated, and are able to re-enter blastocyst development and generate germ-line competent adult mouse chimeras (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007), or to give rise to "all-iPSC"-derived mice through tetraploid complementation (Boland et al., 2009; Kang et al., 2009; Zhao et al., 2009). Another hallmark of induced pluripotency is the reactivation of the inactivated in chromosome of somatic female cells (Stadtfeld et al., 2008b). Within a year of this discovery, human iPSCs were generated from human fibroblasts by the ectopic expression of the OSKM transcription factors (Park et al., 2008; Takahashi et al., 2007), or using a similar cocktail (Yu et al., 2007). As with mouse iPSCs, human iPSCs can be differentiated into all germ lineages.

Five decades have passed since the discovery that somatic nuclei can be reprogrammed back to a pluripotent state, and that reprogrammed cells are able to re-enter embryo development and generate an entirely new animal (Gurdon, 1962a, b; Gurdon et al., 1958). Now it is possible to convert somatic cells directly into pluripotent cells by the forced expression of ESC transcription factors, without the use of eggs or ESCs, or the destruction of embryos (Takahashi and Yamanaka, 2006). In the next section I will further discuss the process of induced pluripotency, mainly focusing on recent technical advances and the epigenetic changes that take place during this process.

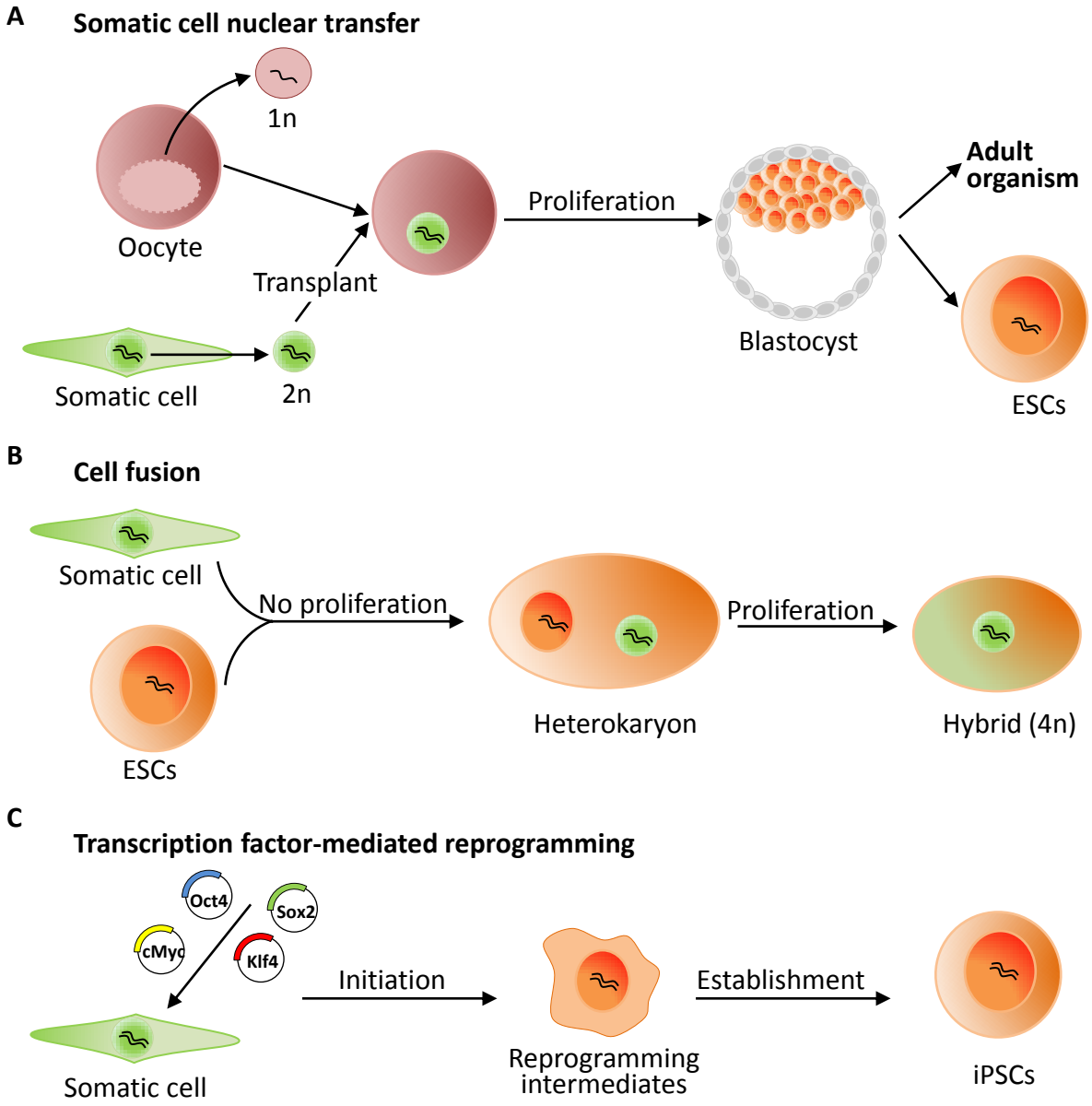


Figure 1.2.1 – Three approaches to nuclear reprogramming to pluripotency.

(A) *Somatic cell nuclear transfer*: a diploid somatic nucleus is transferred into an enucleated oocyte, and is reprogramed to a pluripotent state. Through cell division, a blastocyst is formed, which can be used for ESC derivation or can be re-introduced into a foster mother to generate a cloned animal. **(B)** *Cell fusion*: somatic cells are fused to pluripotent cells (like ESCs), which generates short-lived heterokaryons with two spatially-separated nuclei. When the nuclei fuse and cells start to proliferate, tetraploid hybrids are formed. **(C)** *Transcription factor-mediated reprogramming*: somatic cells are transduced with a set of defined transcription factors (OSKM), which induce reprogramming to an induced pluripotent state, leading to the generation of iPSCs. 1n: Haploid; 2n: diploid; 4n: tetraploid. Refer to text for further details.

1.3 – Induced pluripotent stem cells

The reprogramming of the epigenome (chromatin modifications that may affect gene expression) of somatic cells to an induced pluripotent state can be achieved by the ectopic expression of Oct4, Sox2, Klf4 and c-Myc (OSKM) (Takahashi and Yamanaka, 2006). The advent of iPSCs has opened unprecedented new perspectives in both regenerative medicine and drug discovery. However, before this technology can be used in a clinical setting, it is important to understand the molecular mechanisms underlying their generation. In this section I will review the major technical advances, the role of OSKM and the epigenetic changes during iPSC generation.

1.3.1 – Technical advances in iPSC generation

Numerous alterations to the original protocol have been made in order to increase the efficiency and safety of iPSC generation. Since they were first established from mouse and human somatic cells, iPSCs have been generated from a number of different species, for example rats (Li et al., 2009b) or rhesus monkeys (Liu et al., 2008), among others. They have also been established from different cell types (Stadtfeld and Hochedlinger, 2010), such as lymphocytes (Hanna et al., 2008), stomach and liver (Aoi et al., 2008), neural stem cells (Kim et al., 2009b; Kim et al., 2008), pancreatic β cells (Stadtfeld et al., 2008a), or haematopoietic cells (Eminli et al., 2009; Guo et al., 2014). These experiments showed that the efficiency of inducing pluripotency strongly depends on the origin of somatic cell, ranging from 0.01% when fibroblasts are used (Takahashi and Yamanaka, 2006), to almost 100% if ultra-proliferative granulocyte-monocyte progenitors are used (Guo et al., 2014). This data indicates that reprogramming efficiency is dependent on the proliferation rate of the original somatic cells.

Initially, factors were delivered into target cells by their transduction with constitutively active retroviral vectors (Takahashi and Yamanaka, 2006). This is a very clever delivery method, since retroviral transgenes become silenced when cells enter pluripotency, blocking their further expression upon iPSC generation, in a process that involves both DNA (Silva et al., 2008; Stadtfeld et al., 2008b) and histone (Lei et al., 1996) methyltransferases. In order to improve the safety of iPSC generation, different strategies were soon developed, such as the use of doxycycline-inducible polycystronic vectors encoding all reprogramming factors (Carey et al., 2009; Sommer et al., 2009), the use of piggyBac vectors containing loxP sites that can be subsequently excised (Kaji et al., 2009; Woltjen et al., 2009), and the delivery of reprogramming factors as recombinant proteins (Zhou et al., 2009) or synthetic modified messenger RNA (mRNA) (Warren et al., 2010), for example.

Other modifications to the initial protocol to increase the efficiency of transcription factor induced pluripotency consisted of different culture conditions and the use of small molecules. These small molecules can act as modulators of the epigenetic machinery or signalling pathways, which in some cases results in the replacement of some reprogramming factors. The inhibition of DNA methyltransferases with 5-aza-cytidine (Mikkelsen et al., 2008) or the inhibition of histone deacetylase (HDACs) with valproic acid (Huangfu et al., 2008) was shown to increase OSKM-mediated reprogramming. Similarly, the use of 2i/LIF conditions increases iPSC generation and allows the progression of cells blocked during reprogramming named preiPSCs, to naïve pluripotency (Silva et al., 2008). This enhancement of reprogramming efficiency is probably due to the combined action of 2i in both the establishment and maintenance of the transcription factor network of naïve pluripotency (Wray et al., 2011; Ying et al., 2008), and also the induction of PR domain zinc finger protein 14 (Prdm14) expression, which directly represses *de novo* methyltransferases Dnmt3a and *Dnmt3b* (Blaschke et al., 2013; Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013; Yamaji et al., 2013). The inhibition of transforming growth factor-β (TGF-β) signalling can also increase reprogramming efficiency, and replace the forced expression of Sox2 and c-Myc (Ichida et al., 2009; Maherali and Hochedlinger, 2009), probably due to the endogenous activation of Nanog expression. The addition of vitamin C to the reprogramming medium was also shown to enhance the induction of pluripotency from fibroblasts (Esteban et al., 2010). Mechanistically, vitamin C leads to increased expression of histone demethylases Jhdm1a/1b, with concomitant reduced histone 3 lysine 36 di- and tri-methylation (H3K36me2/3), alleviating cell senescence by removing H4K36me2/3 from the *lnk4/Arf* (p16/p19) locus, repressing its expression (Liang et al., 2012; Wang et al., 2011). Indeed, the expression of the *Ink4/Arf* locus has been shown by many groups to be a major barrier for induced pluripotency (Hong et al., 2009; Kawamura et al., 2009; Li et al., 2009a; Marion et al., 2009a; Utikal et al., 2009). Jhdm1b also cooperates with Oct4 to activate the expression of mesenchymal-to-epithelial transition (MET) genes, pluripotency-associated microRNAs (miRs) cluster 302/367, and pluripotency genes (Liang et al., 2012; Wang et al., 2011). Vitamin C was also shown to induce global DNA demethylation in mouse ESCs by enhancing ten-eleven translocation methylcytosine dioxygenase (TET) activity (Blaschke et al., 2013). Lastly, vitamin C was reported to play a positive role in induced pluripotency by regulating H3K9 methylation status at the core pluripotency genes (Chen et al., 2013a), and also by maintaining the correct imprinting status of the *Dlk-Dio3* locus by regulating H3K4me3 which attenuates its expression levels (Stadtfeld et al., 2012).

After a long quest to develop transgene-free reprogramming methods (Li et al., 2013), Deng and colleagues showed that the combined inhibition of the signalling pathways and some components of the epigenetic machinery is sufficient for the induction of pluripotency in mouse fibroblasts (Hou et al., 2013). They demonstrated that medium containing valproic acid (HDAC inhibitor), Chiron (Gsk3β inhibitor), 616452 (Tgfβ inhibitor), TTNPB (synthetic retinoic acid receptor ligand), forskolin (cyclic adenosine monophosphate agonist) and DZNep (S-adenosylhomocysteine hydrolase inhibitor), followed by 2i culture (medium named "VC6TFZ plus 2i"), reprograms mouse fibroblasts to germline-competent iPSCs (Hou et al., 2013). It remains to be tested if similar conditions can also reprogram human somatic cells.

Less than a decade has passed since the discovery of iPSCs and significant advances have already been made. The manipulation of reprogramming methodologies has not only enabled the more efficient generation of better iPSCs, but has also provided some insights into the mechanisms governing induced pluripotency. These technologies can now be used to understand the biology of induced pluripotency.

1.3.4 – The role of OSKM during induced pluripotency

The initial Yamanaka cocktail was constituted of the reprogramming factors Oct4, Sox2, Klf4 and c-Myc (OSKM). Oct4 and Sox2 are part of the transcription factor core of the naïve pluripotency (Young, 2011) and are essential for embryo development (Avilion et al., 2003; Nichols et al., 1998). The rationale behind Klf4 and c-Myc is more obscure, since both factors are not required for embryo development and ESC maintenance (Baudino et al., 2002; Ivanova et al., 2006; Segre et al., 1999). Ablation of ectopic c-Myc was the first modification made to the initial reprogramming cocktail (Nakagawa et al., 2008; Wernig et al., 2008), which showed that iPSCs could be generated in the absence of the oncogene c-Myc, but that comes at the expense of efficiency. All the Yamanaka factors can be replaced by other transcription factors, for example: SKM can be replaced by their family homologs (Nakagawa et al., 2008); Klf4 and c-Myc can be replaced by Nanog and Lin28 (Yu et al., 2007), or by Esrrβ (Feng et al., 2009); c-Myc can replaced by Glis1 (Maekawa et al., 2011); and Oct4 can be replaced by Nr5a1 or Nr5a2 (Heng et al., 2010) or by Tet1 (Gao et al., 2013). Moreover, the recently acquired understanding of the downstream effects of each OSKM factor led to the generation of new cocktails that can replace OSKM entirely, such as Esrrβ, Lin28 and Sall4, plus Nanog or Dppa2 (Buganim et al., 2012). More surprisingly, it was recently demonstrated that lineage-specific transcription factors could replace both Oct4 (replaced by Gata4/6, Sox7, C/EBPα, HNF4a or Grb2) and Sox2 (replaced by Sox1/3 or Rcor2) during the reprogramming

of mouse (Shu et al., 2013) or human fibroblasts (Montserrat et al., 2013). This shows that there is a high degree of flexibility among the factors that induce pluripotency.

As part of an effort to develop reprogramming methods that allow reprogramming factor expression without the need for genomic integration of foreign DNA, two independent groups have shown that both mouse and human somatic cells can be reprogrammed to pluripotent stem cells by forced expression of miRs, completely eliminating the need for ectopic OSKM expression (Anokye-Danso et al., 2011; Miyoshi et al., 2011). Two cocktails of miRs that are normally highly expressed in ESCs were used: five miRs from the 302/367 cluster (Anokye-Danso et al., 2011), or seven miRs belonging to the miR302, 200, and 369 families (Miyoshi et al., 2011). However, although many other groups reported that reprogramming efficiency can be enhanced or that individual "Yamanaka factors" can be replaced by forced expression of miRs (Judson et al., 2013; Parchem et al., 2014; Sridharan and Plath, 2011; Subramanyam et al., 2011), no further reporters were published about miR-only mediated reprogramming. In fact, these two reports failed to be reproduced by different laboratories all over the world.

Little is known about how the ectopic expression of OSKM drives the conversion of somatic cells to a pluripotent state. The oncogene c-Myc has been shown to act predominantly during the earlier phases of reprogramming (Polo et al., 2012; Sridharan et al., 2009), a phase marked by increased proliferation and the initiation of MET (Li et al., 2010; Mikkelsen et al., 2008; Samavarchi-Tehrani et al., 2010). Mechanistically, c-Myc induces the down-regulation of *Tgf-β1/2* receptors, resulting in reduced Tgf-β signalling, a repressor of MET (Li et al., 2010; Liu et al., 2013). OSK factors have also have been implicated in the MET regulation. Oct4/Sox2 suppress the epithelial-to-mesenchymal transition (EMT) by repressing both *SnaiI* (Li et al., 2010) and *Zeb2* expression, through the activation of the miRNA-200 family that represses *Zeb2* (Wang et al., 2013a). Klf4 was found to enhance MET by inducing epithelial genes such as E-cadherin (Li et al., 2010). However, although OSK are involved in a facilitation of MET, these factors are more associated with transcriptional changes that occur in the later phases of reprogramming (Polo et al., 2012), characterized by the activation of core pluripotency circuitry, transgene silencing and complete epigenetic resetting (Apostolou and Hochedlinger, 2013; Buganim et al., 2013; Papp and Plath, 2013; Plath and Lowry, 2011).

How do OSKM work to assist genome-wide reprogramming? The four OSKM factors have been shown to co-occupy open chromatin immediately after transduction, binding the promoters of both active and repressed genes (Koche et al., 2011; Sridharan et al., 2009). Adding to this fact, the three factors, OSK, bind to distal elements with high nucleosome occupancy that lack pre-existing histone modifications (Soufi et al., 2012). These genomic

sites are different from the ones bound by these pluripotency factors in pluripotent cells (Koche et al., 2011), which led to the notion that OSK are "pioneer factors" that guide reprogramming through promiscuous binding to distal enhancers (Soufi et al., 2012) (Figure 1.3.1). The "pioneer" OSK factors might be responsible for chromatin remodelling and the activation of genes that are essential for the establishment and maintenance of pluripotency (Soufi et al., 2012). This process is likely to be facilitated by the recruitment of other factors like c-Myc, which has been recently hypothesized to act as general amplifier of gene expression by increasing the transcription of active promoters (Lin et al., 2012; Nie et al., 2012). In addition to the possibility that Oct4 and Klf4 might act as "pioneer factors" to induce pluripotency, they also have an established function in promoting the establishment and maintenance of the core pluripotency circuitry (Hall et al., 2009).

Although the molecular mechanisms underlying induced pluripotency are still very much a black box, some evidence suggest that OSK drive genome-wide remodelling through their promiscuous interaction with closed chromatin, leading to the activation of pluripotency genes (Soufi et al., 2012). This process seems to be facilitated by c-Myc, which not only leads to increased proliferation, which is extremely important for reprogramming, but also enhances transcription (Lin et al., 2012; Nie et al., 2012). Since recent evidence has shown that the OSKM factors can be replaced by lineage specifiers (Montserrat et al., 2013; Shu et al., 2013), or by a completely different set of transcription factors (Buganim et al., 2012), or even by chemical compounds (Hou et al., 2013), it is of interest to address if cells being reprogrammed by alternative approaches pass through the same sequence of events described above.

1.3.4 – Epigenetic changes during induced pluripotency

Induced pluripotency is fundamentally an epigenetic phenomenon. Cell specialization from pluripotent cells during embryonic development or during *in vitro* differentiation is characterized by extensive chromatin remodelling that results mainly in the deposition of repressive marks followed by chromatin compaction (Gifford et al., 2013; Xie et al., 2013; Zhu et al., 2013). For induced pluripotency to occur, this somatic epigenetic signature has to be erased and a new pluripotency epigenome has to be established. Changes include, but are not restricted to, global chromatin organization, the resetting of post-translational histone modifications, DNA demethylation of key pluripotency genes and the re-activation of inactive X chromosome (Mikkelsen et al., 2008; Polo et al., 2012; Stadtfeld et al., 2008b; Wernig et al., 2007). In this section I will review how chromatin evolves during induced pluripotency,

elucidating the roles of main chromatin marks and chromatin regulators during reprogramming.

For somatic cells to be stably reprogrammed to a pluripotent state, somatic chromatin has to be dismantled, meaning that the reprogramming factors have to engage with the epigenetic machinery. Recent studies shown that OSKM binding correlates with immediate changes in histone modification early after reprogramming initiation (Koche et al., 2011; Polo et al., 2012; Soufi et al., 2012; Sridharan et al., 2009). This leads to the hypothetical assumption that the binding of OSKM throughout the genome somehow marks regions that will ultimately be remodelled. Three different classes of genes can be defined, based on OSKM accessibility, histone marks and kinetics of chromatin remodelling and transcription: somatic and EMT genes, early pluripotency and MET genes, and late pluripotency genes (Figure 1.3.1). Somatic and EMT genes, such as the fibroblast-specific and EMT gene *SnaiI*, are transcriptionally active and marked by an open chromatin state and H3K4me2/3 activation-associated marks in somatic cells (Soufi et al., 2012). These are immediately bound by OSKM without the need of further chromatin remodelling, starting to lose H3K4me2/3 and become heterochromatic, acquiring H3K27me3 marks, which results in transcriptional silencing (Koche et al., 2011; Soufi et al., 2012). Early pluripotency-associated and MET genes, such *Fbx15* and *Sal-like protein 4* (*Sall4*), whose early activation is essentially required for induced pluripotency, are marked by repressive marks such as H3K27me3 and are transcriptionally inactive in somatic cells (Soufi et al., 2012). Hours after OSKM induction, OSK "pioneer" factors quickly occupy the enhancers of these genes, allowing the remodelling of these high-density nucleosome regions, leading to the deposition of H3K4me2/3 at enhancers and promoters, which results in gene re-activation a few days later (Koche et al., 2011; Soufi et al., 2012). A third class of genes include the late pluripotency genes, such as *Nanog* and *Sox2*, whose enhancers and promoters contain H3K27me3 and H3K9me2/3 repressive marks in somatic cells, and are refractory to OSKM binding (Buganim et al., 2012; Soufi et al., 2012). These genes require more chromatin remodelling for OSKM binding and transcriptional activation, and are re-activated only at later stages of the reprogramming process (Mikkelsen et al., 2008; Polo et al., 2012; Soufi et al., 2012). The generated iPSCs fully reactivate the expression of pluripotency genes, that now contain high levels of H3K4me2/3, and silence genes from the original somatic cells and mark them with H3K27me3 (Mikkelsen et al., 2008; Polo et al., 2012; Soufi et al., 2012).

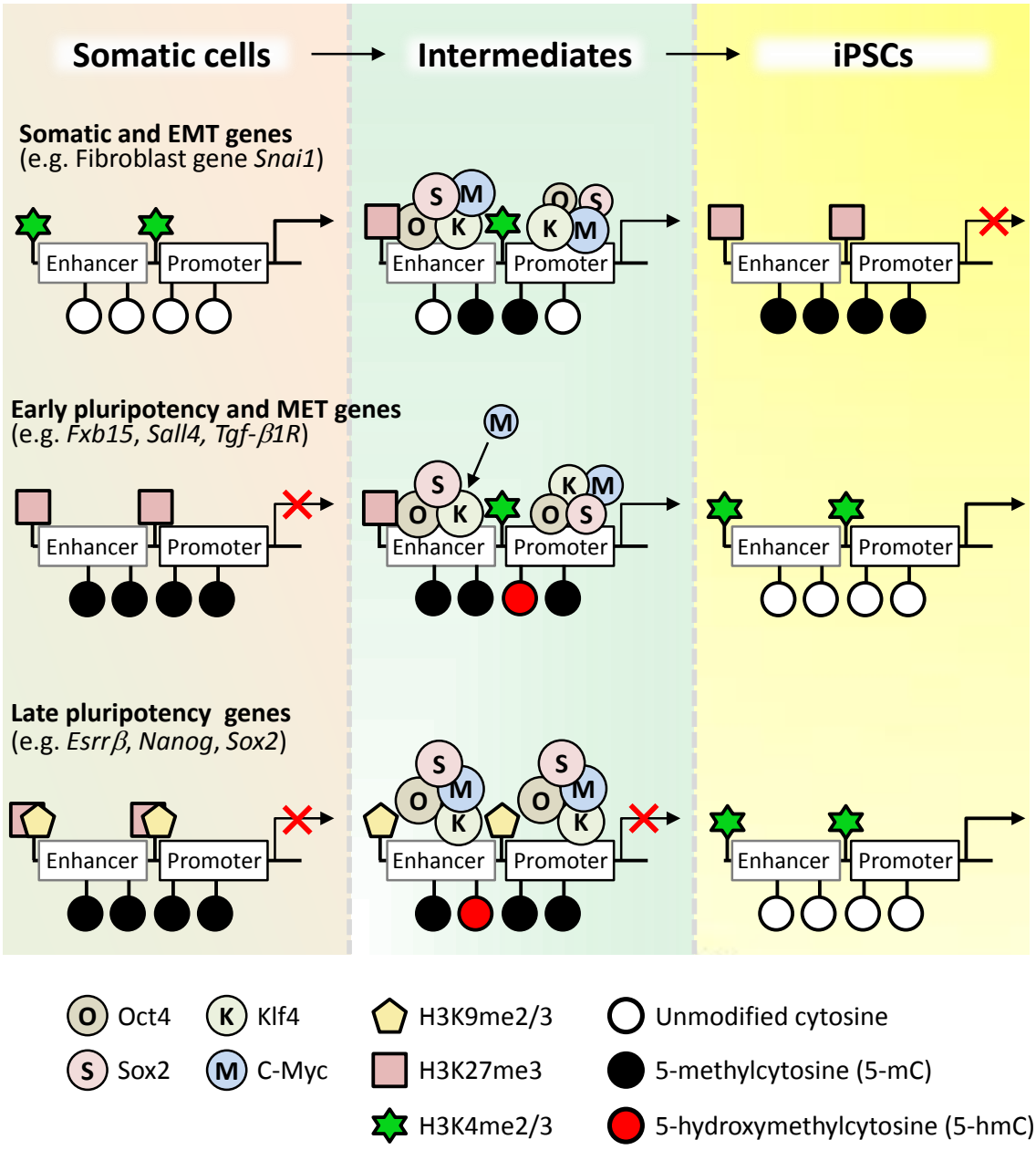


Figure 1.3.1 – OSKM binding and Chromatin remodelling during induced pluripotency.

Three classes of genes can be grouped based on their transcriptional activation/ repression kinetics: somatic and EMT genes, early pluripotency and MET genes, and late pluripotency genes. *Somatic and EMT genes* are marked by H3K4me2/3 in somatic cells, are immediately bound by OSKM after ectopic expression and start to accumulate H3K27me3 and methylated DNA, becoming transcriptionally inactive. *Early pluripotency and MET genes* which are hypermethylated and marked by H3K27me3, are transcriptionally inactive in somatic cells, are occupied by OSK "pioneer" factors immediately after OSKM induction, which allows their remodelling and transcriptional activation. *Late pluripotency genes* are enriched for H3K9me2/3 and H3K27me3 marks in somatic cells which repress OSKM binding, resulting in delayed transcriptional re-activation. Re-activation of pluripotency genes is marked by complete enhancer and promoter demethylation, which requires an intermediate step of 5- methylcytosine (5-mC) oxidation to 5-hydroxymethylcytosine (5-hmC) prior to pluripotency entry.

Stable reprogramming of somatic cells requires the full and persistent re-activation of pluripotency associated genes, together with the complete silencing of somatic genes. This is possible due to the complete demethylation of enhancers and promoters of ESC-related genes, and the hypermethylation of the somatic genes (Mikkelsen et al., 2008; Wernig et al., 2007) (Figure 1.3.1). Whereas changes in histone modification begin to occur early during reprogramming, driving the initiation of this process, DNA demethylation occurs only late in the process (Polo et al., 2012). This is in accordance with what is observed during cell differentiation, where changes in histone modification typically precede DNA methylation changes, indicating that DNA methylation is a more stable modification required for the stabilization of cellular phenotype (Gifford et al., 2013; Xie et al., 2013). Both passive (replication-dependent) and active demethylation mechanisms have been shown to play a pivotal role in induced pluripotency. Inhibition of DNA demethylation maintenance by DNA methyltransferase Dmnt1 inhibition enhances reprogramming efficiency (Mikkelsen et al., 2008). Active DNA demethylation, involving oxidation of methylcytosines (5-mC) to hydroxymethylcytosines (5-hmC) by Tet enzymes followed by thymine DNA glycosylase (TDG)-mediated base excision repair into unmodified cytosine (Kohli and Zhang, 2013), is required for efficient reprogramming. Tet2 mediates the priming of key pluripotency genes, such as *Esrrβ* and *Nanog*, early after OSKM induction by promoting active DNA demethylation (Di Stefano et al., 2014; Doege et al., 2012), and the ectopic expression of Tet1/2 with Nanog was shown to increase reprogramming efficiency (Costa et al., 2013). Due to Tet-mediated re-activation of the *Oct4* locus, Tet1 can substitute for ectopic Oct4 expression in reprogramming (Gao et al., 2013). Moreover, *Tet1/2/3* triple-knockout (KO) somatic cells show reduced reprogramming efficiency, maybe due to a failure in the activation of MET genes (Hu et al., 2014), although *Tet1*$^{-/-}$ (Dawlaty et al., 2011), *Tet1/2* double-KO (Dawlaty et al., 2013), or *Tet1/2/3* triple-KO mice (Wang et al., 2013b), which exhibit aberrant methylation levels, are viable and fertile. The poly ADP-ribose polymerase 1 (Parp1) was also shown to promote induced pluripotency, by the regulation 5-mC levels, and thus 5-hmC levels, which are required for *Oct4* activation (Doege et al., 2012). Lastly, one group reported that activation-induced cytidine deaminase-dependent (Aid) DNA demethylation is required for the generation of high-quality iPSCs (Kumar et al., 2013). However, further work from others groups revealed that Aid is dispensable for induced pluripotency (Habib et al., 2014; Shimamoto et al., 2014).

More than 100 different histone post-translational modifications have been identified (Bannister and Kouzarides, 2011; Bernstein et al., 2007; Kouzarides, 2007). Above I

described the kinetics of the deposition or removal of key chromatin markers after the induction of the reprogramming factors. In the Table 1.3.1 other selected chromatin marks that have been identified as playing an important role induced pluripotency are listed, as well as some examples of their dynamics during reprogramming.

Table 1.3.1 – List of selected chromatin marks and their role during induced pluripotency.

Chromatin mark	Assumed function	Dynamics during reprogramming (example)	References
H3K4me1	Marks active or poised enhancers	Increase at proliferation and metabolism genes	(Creyghton et al., 2010; Rada-Iglesias et al., 2011)
H3K4me2	Marks active or poised enhancers and promoters	Decrease at somatic genes; Increase at pluripotency genes	(Koche et al., 2011; Soufi et al., 2012; Sridharan et al., 2009)
H3K4me3	Marks TSS of active and poised genes	Decrease at somatic genes; Increase at proliferation, metabolism and pluripotency genes	(Koche et al., 2011; Maherali et al., 2007; Soufi et al., 2012; Sridharan et al., 2009)
H3K9me3	Marks heterochromatic regions	Decrease at late pluripotency genes	(Chen et al., 2013b; Soufi et al., 2012; Sridharan et al., 2013)
H3K9ac	Marks TSS of active genes	Increase at Oct4 targets	(Esch et al., 2013; Singhal et al., 2010)
H3K27me3	Marks repressed genes	Increase at somatic and EMT genes; Decrease at pluripotency genes	(Koche et al., 2011; Maherali et al., 2007; Mansour et al., 2012; Polo et al., 2012; Sridharan et al., 2009)
H3K27ac	Marks open chromatin and active enhancers	Global increase during reprogramming, mainly at pluripotency genes	(Creyghton et al., 2010; Mattout et al., 2011; Rada-Iglesias et al., 2011)
H3K36me2	Marks potential regulatory regions	Increase at early pluripotency genes	(Liang et al., 2012; Wang et al., 2011)
H3K36me3	Marks active genes	Increase at pluripotency genes	(Koche et al., 2011; Liang et al., 2012; Wang et al., 2011)
H3K79me2	Marks active genes	Decrease at MEF and EMT genes	(Onder et al., 2012)
H1citR54	Marks chromatin decondensation	Increase at pluripotency genes	(Christophorou et al., 2014)
5-mC	Marks repressed genes	Deposited at somatic genes; Erased from pluripotency genes	(Mikkelsen et al., 2008; Polo et al., 2012; Wernig et al., 2007)
5-hmC	Marks gene reactivation and priming	Increase at MET and pluripotency genes (*Oct4*, *Esrrβ*)	(Chen et al., 2013a; Costa et al., 2013; Doege et al., 2012; Gao et al., 2013; Hu et al., 2014)

H: Histone; K: Lysine; R: Arginine; me: methylation; ac: acetylation; cit: Citrullination; 5-mC: 5-methylcytosine; 5-hmC: 5-hydroxymethylcytosine; TSS: transcription start site; MET: mesenchymal-to-epithelial transition; EMT: epithelial-to-mesenchymal transition; MEFs: mouse embryonic fibroblasts.

As seen in Table 1.3.1, nuclear reprogramming is accompanied by a complete change in histone modification patterns, indicating that induced pluripotency is controlled by the kinetics of histone modification. This means that the efficiency and kinetics of nuclear reprogramming can be enhanced by manipulating the expression of chromatin modifying enzymes and proteins involved in the "writing", "reading" or "erasing" of chromatin marks. Based on the outcome after depletion and/ or overexpression of different chromatin modulators, they can be considered as "drivers" (D) or "blockers" (B) of induced pluripotency (Table 1.3.2). In the context of this thesis, "drivers" of reprogramming are chromatin modulators whose activity facilitate or is required for induced pluripotency. Consequently, their pharmacological inhibition or protein deletion usually results in reduced reprograming efficiency, while their pharmacological activation or protein overexpression results in enhanced reprogramming efficiency. Conversely, "blockers" of reprogramming are chromatin modulators that inhibit reprogramming, which means that their depletion or inhibition increases reprogramming efficiency, and/ or their overexpression or enhanced activation decreases it. By applying these criteria, several "drivers" and "blockers" of induced pluripotency were identified (Table 1.3.2). They were divided into different categories: histone demethylases, histone methyltransferases, histone variants, chromatin remodellers, proteins involved in the 3D chromatin organization, DNA methylation, and others. Table 1.3.2 also includes a description of their roles in reprogramming.

Table 1.3.2 – List chromatin modulators and their main role during induced pluripotency.

	Chromatin modulator	D/B	Function and role in reprogramming	References
Histone demethylases	**Jmjd1a/b (Kdm3a/b), Jmjd2c (Kdm4c)**	**D**	H3K9 demethylases; KD reduces reprogramming; OE enhances reprogramming; Regulates dynamic switch between open-close chromatin	(Chen et al., 2013b)
	Cbx3	**B**	H3K9me "reader"; KD reduces reprogramming by repressing *Nanog* expression together with H3K9 methyltransferases	(Sridharan et al., 2013)
	Jmjd3	**B**	H3K27 demethylase; KO enhances reprogramming; Suppresses *lnk4/Arf* locus (senescence) and targets PHF20, required for Oct4 expression, for ubiquitination	(Zhao et al., 2013)

	Utx (Kdm6a)	**D**	H3K27 demethylase; KO reduces reprogramming by inefficient H3K27me demethylation of pluripotency genes	(Mansour et al., 2012)
	Jhdm1a/b (Kdm2a/b)	**D**	H3K36 demethylases; KD reduces reprogramming; OE enhances reprogramming; Cooperate with Oct4, Required for early activation of MET and pluripotency genes, and repression of *lnk4/Arf* locus	(Liang et al., 2012; Wang et al., 2011)
Histone methyltransferases	**Ehmt1 (G9a), Ehmt2, Setdb1, (ESET), Suv39h1/2**	**B**	H3K9 methyltransferases; KD enhances reprogramming by allowing more efficient OSKM binding genes in somatic cells marked by H3K9me3	(Chen et al., 2013b; Onder et al., 2012; Sridharan et al., 2013)
	PRCs (Ezh2, Eed, Suz12, Bmi1, Ring1)	**D**	H3K27 methyltransferases; KD reduces reprogramming; Required for maintaining a transcriptional gene repression	(Fragola et al., 2013; Onder et al., 2012)
	Dot1L	**B**	H3K79 methyltransferase; KD or chemical inhibition enhances reprogramming by repressing expression of EMT genes	(Onder et al., 2012)
	Wdr5 complex	**D**	H3K4me "effector", member of mammalian Trithorax complex; KD reduces reprogramming; Interacts with Oct4 and facilitates activation of pluripotency genes	(Ang et al., 2011)
Histone variants	**macroH2A**	**B**	Histone variant; KD enhances reprogramming; OE reduces reprogramming; Maintain pluripotent genes in repressed state	(Gaspar-Maia et al., 2013; Pasque et al., 2012)
	TH2A/B	**D**	Histone variant; OE enhances reprogramming; Required to induce transcriptionally active open chromatin	(Shinagawa et al., 2014)
Chromatin remodellers	**Chd1**	**D**	Chromatin-remodeler; KD reduces reprogramming; Required for chromatin decondensation at pluripotent genes	(Gaspar-Maia et al., 2009)
	SWI/SNF (Baf155, Brg1, Brm)	**D**	Chromatin-remodelling complex (also known as BAF); OE enhances reprogramming; enhance Oct4 binding by increasing H3K4me3 and H3K9ac euchromatin marks and demethylation at pluripotency genes	(Esch et al., 2013; Singhal et al., 2010)
	INO80 complex	**D**	Chromatin-remodelling complex; KD reduces reprogramming; Required for Mediator and RNA polymerase II recruitment for pluripotency genes activation	(Wang et al., 2014)
3D Chromatin	**Cohesin-mediator complexes**	**D**	Architectural proteins; KD reduces reprogramming by blocking long-range chromatin interactions (*Oct4* intrachromosomal looping and *Nanog* locus genome-wide interactions) induced by OSKM that contribute for activation of pluripotency genes	(Apostolou et al., 2013; Wei et al., 2013; Zhang et al., 2013)

DNA methylation	**Dnmt1**	**B**	DNA methyltransferase; KD or chemical inhibition enhances reprogramming by enhancing demethylation of pluripotency genes	(Mikkelsen et al., 2008)
	Dnmt3a/b	**X**	*de novo* DNA methyltransferase; Double-KO cells reprogram with same efficiency as wild type cells; Dispensable for reprogramming	(Pawlak and Jaenisch, 2011)
	Tet1/2/3	**D**	Methylcytosine dioxygenases that oxidase 5-mC to 5-hmC; Triple-KO block reprogramming; Prime pluripotency (*Esrrβ*, *Oct4*) and MET genes activation; Tet1/2 OE enhance Nanog-mediated reprogramming	(Costa et al., 2013; Doege et al., 2012; Gao et al., 2013; Hu et al., 2014)
	Parp1	**D**	Poly ADP-ribose polymerase; KO blocks reprogramming; Regulates 5-mC and Oct4 recruitment at pluripotency genes	(Doege et al., 2012)
	Aid	**X**	DNA demethylation; KO cells reprogram with same efficiency as wild type cells; Dispensable for reprogramming	(Habib et al., 2014; Kumar et al., 2013; Shimamoto et al., 2014)
Others	**Hdacs**	**B**	Histone deacetylases; Chemical inhibition enhances reprogramming by activation of pluripotency genes	(Huangfu et al., 2008; Mali et al., 2010)
	Padi4	**D**	Citrullination of arginine residues; KD reduces reprogramming; Disrupts the binding of H1 from chromatin, resulting in decondensation	(Christophorou et al., 2014)
	Terc	**D**	Telomerase; KO reduces reprogramming; Required for rejuvenation of telomeres to ESC-like state; Decrease of H3K9m3 repressive mark at telemetric regions during reprogramming	(Marion et al., 2009b)
	TFIID	**D**	Regulates RNA polymerase II function; KD reduces reprogramming; Together with OSKM induce stable pluripotency genes expression	(Pijnappel et al., 2013)
	Brca1/2, Rad51	**D**	Homology-directed DNA repair; KO reduces reprogramming; OSKM increases DNA DSB marker γH2AX levels; p53 deletion rescues reprogramming defects of *Brca1/2*-null MEFs	(Gonzalez et al., 2013)
	SCC/XPC	**D**	Nucleotide excision repair complex; KD reduces reprogramming; co-activator of Oct4 and Sox2 for *Nanog* and *Oct4* activation	(Fong et al., 2011)

D/B: "Driver" (facilitates), or "Blocker" (inhibits) of induced pluripotency (see text for definition"). X – "Driver" or "Blocker" definition does not apply; OE: overexpression; KD: Knockdown; KO: Knockout; Jmjd1a/b: Jumonji domain-containing 1a/b; Cbx3: Chromobox homolog 3; Jhdm1a/b: Jumonji domain-containing histone demethylase 1a/b; Utx: Lysine-specific demethylase 6a; Ehmt1/2: Histone-lysine N-methyltransferas 1/2; Setdb1: SET domain, bifurcated 1; Suv39h1/2: Suppressor of variegation 3-9 homolog 1; PRCs: Polycomb repressive complexes; Ezh2: enhancer of zeste homolog 2; Eed: embryonic ectoderm development; Suz12: Suppressor of zeste 12; Bmi1: B lymphoma Mo-MLV insertion region 1 homolog; Dot1L: DOT-like 1; Wdr5: WD repeat domain 5;

TH2A/B: Testis-specific H2A/B; Chd1: Chromodomain helicase DNA binding protein 1; SWI/SNF: SWItch/Sucrose Non-Fermentable; Baf155: BRG1-associated factor 155; Brg1: Brahma-related gene 1; Brm: Brahma; Dnmt1: DNA methyltransferase 1; Dnmt3a/b: *de novo* DNA methyltransferase 3a/b; Tet1/2/3: Ten-eleven translocation methylcytosine dioxygenase 1/2/3; Parp1: Poly-(ADP-ribose) polymerase 1; Aid: Activation-induced cytidine deaminase; Hdac: Histone deacetylase; Padi4: Peptidylarginine deiminase 4; Terc: Telomerase RNA component; TFIID: Transcription factor IID; Brca1/2: Breast cancer type 1/2 susceptibility protein; SCC/XPC: stem cell co-activator complex/ Xeroderma pigmentosum complementation group c; PHD20: PHD finger protein 20; DSD: Double-strand break; γH2AX: phosphorylated histone H2AXH: Histone; K: Lysine; me: methylation; ac: acetylation; EMT: epithelial-to-mesenchymal transition; MET: mesenchymal-to-epithelial transition

In summary, all evidence suggests that the re-establishment of a pluripotent epigenetic state is the main feature during reprogramming. In addition to local chromatin, whose repressive environment presents a major barrier somatic cell reprogramming, 3D chromatin architecture has also been implicated in induced pluripotency. Indeed, by assessing *Oct4* locus intrachromosomal looping and *Nanog* locus genome-wide chromatin interactions, it was demonstrated which enhancer–promoter co-associations were unique to pluripotent cells and that these are required for transcriptional activation of pluripotency genes during reprogramming (Apostolou et al., 2013; Wei et al., 2013; Zhang et al., 2013). Further studies will aim to better understand the interaction between the induction of the reprogramming factors, and the resetting of the somatic and the establishment of the pluripotent epigenome. It will also be very important to understand the epigenetic events that take place during reprogramming using different reprogramming methodologies, such as chemical reprogramming, and compare them to transcription factor-mediated reprogramming. New clues about the molecular sequence of events might emerge from those comparisons.

1.3.5 – Role of Oct4 and Nanog on induced pluripotency

Once the pluripotent state is re-established after the forced expression of exogenous factors, it is maintained by the re-activation of the endogenous pluripotency-associated transcription circuitry. As seen in section 1.1.4, pluripotency is maintained by an intricate inter-connection between signalling pathways and different transcription factors which control cell function. Of these, Nanog, Oct4 and Sox2 are the main components of the pluripotency network, playing crucial roles in induced pluripotency.

Nanog was not a part of the original reprogramming cocktail, but selection of the emerging colonies for endogenous *Nanog* and *Oct4* expression enabled the isolation of germ-line competent iPSCs (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007). Nanog

overexpression enhances OSKM-mediated reprogramming (Hanna et al., 2009; Theunissen et al., 2011b), and when used in combination with other defined factors can completely substitute OSKM in reprogramming (Buganim et al., 2012). Tet1/2 and Nanog co-overexpression can boost reprogramming efficiency even further (Costa et al., 2013). Initial reports described that endogenous *Nanog* is not required for initial stages of reprogramming, being extremely required for final stages of reprogramming during 2i/LIF culture (Silva et al., 2009), which could be rescued by ectopic expression of mouse Nanog or its homologs and orthologues (Theunissen et al., 2011a). Further studies indicate that Nanog can be bypassed if its downstream targets are ectopically provided, such as *Esrrb* (Festuccia et al., 2012), or activated STAT3 in combination with *Esrrb* or *Klf4* (Stuart et al., 2014). Moreover, using different reprogramming protocols, it was recently shown that *Nanog*-null mouse embryonic fibroblasts (MEFs) can be reprogrammed to iPSCs, although at lower efficiencies than wild type MEFs (Carter et al., 2014), a process which can be enhanced if vitamin C is provided (Schwarz et al., 2014).

The role of Oct4 in induced pluripotency has been extensively studied since the discovery of iPSCs. Oct4 is the only "Yamanaka factor" that cannot be replaced by members of its family, Oct1 and Oct6 (Nakagawa et al., 2008), but can be replaced by its human and xenopus homologues, or medaka and axolotl ortologues (Esch et al., 2013; Tapia et al., 2012). Oct4 can also be replaced by Nr5a1/2 (Heng et al., 2010) or by Tet1 (Gao et al., 2013), but in both cases that seems to be possible due to the activation of endogenous *Oct4*. Importantly, Oct4 is the only "Yamanaka factor" that can induce reprogramming of somatic cells alone. Studies from the Scholer laboratory shown that Oct4 overexpression is sufficient to induce pluripotency from neural stem cell (Kim et al., 2009a; Kim et al., 2009b). Genetic *Oct4* deletion ablates reprogramming (unpublished data by our laboratory), which is in accordance with the fact that *Oct4*$^{-/-}$ ESCs cannot be maintained and differentiate into trophectoderm (Niwa et al., 2000). Interestingly, rescue experiments showed that independently of the Oct4 transgene levels in reprogramming intermediates, the resulting iPSCs exhibited a defined ESC-level of Oct4 (Radzisheuskaya et al., 2013). These results demonstrated that endogenous Oct4 is essential for induced pluripotency, which does not seem to be the case for SCNT-mediated reprogramming, since enucleated *Oct4*$^{-/-}$ oocytes can reprogram somatic cells back to pluripotency (Wu et al., 2013).

The role of Sox2 in induced pluripotency has been less characterized. Since Sox2 is indispensable for ESC maintenance (Masui et al., 2007), it is acceptable to say that the re-establishment of pluripotency is not possible in the absence of endogenous *Sox2*. However,

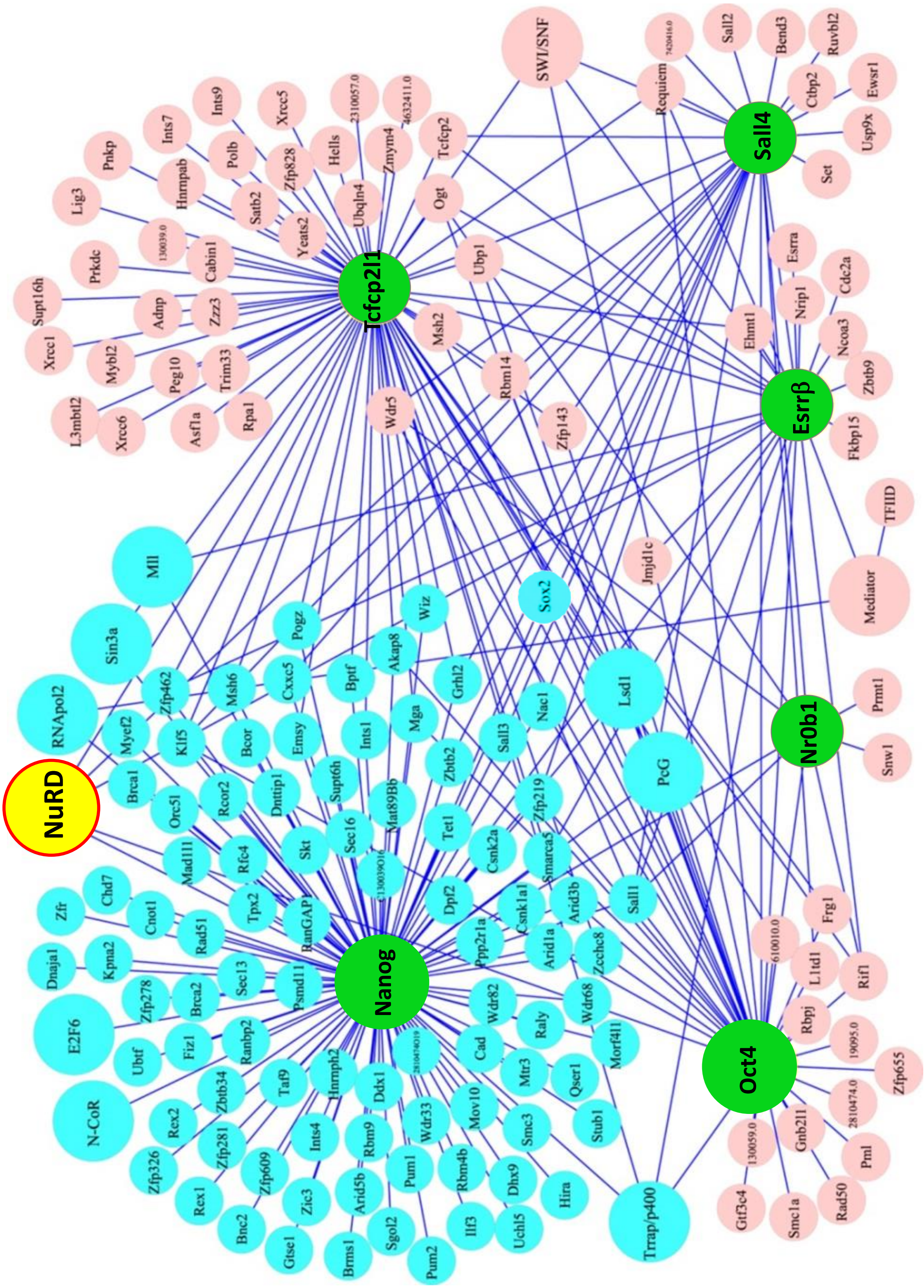
the Sox2 levels which are tolerable for induced pluripotency, and the time window where Sox2 is required, are questions which remain open.

As described above and in the previous sections, Nanog and Oct4 play essential roles in the establishment and maintenance of pluripotency, both *in vitro* and *in vivo*. This has stimulated many scientists to understand their mechanisms of action. One way to identify those mechanisms is to identify their protein interactome, which might reveal core interactions which control Nanog and Oct4 function. Nanog protein-protein interactions have been extensively studied. The Nanog protein interactome in ESCs was analysed by mass-spectrometry analysis and was first reported in 2006 (Wang et al., 2006). This was recently refined and expanded by two independent laboratories (Costa et al., 2013; Gagliardi et al., 2013). The Oct4 protein interactome in ESCs was unveiled in 2010 (Pardo et al., 2010; van den Berg et al., 2010), and then further expanded in 2012 (Ding et al., 2012). An interactome map of Nanog and Oct4 proteins, together with the identified interactomes of other transcription factors, such as Essrβ, Sall4, Nr0b1 and Tcfcp2l1, has been created by the Chambers laboratory and is shown in Figure 1.3.2.

(figure on next page)

Figure 1.3.2 – Nanog and Oct4 protein interactomes.

Map of identified protein-protein interaction centred on Nanog and Oct4, together with interaction centred on other ESC-specific transcription factors Essrβ, Sall4, Nr0b1 and Tcfcp2l1. Blue circles correspond to the Nanog interactions found by the Chambers laboratory (Gagliardi et al., 2013), whereas the pink circles are the protein interacts of the other protein-protein interaction studies centred on other transcription factors (green) discovered by the Poot and Chambers laboratories (van den Berg et al., 2010). The Nucleosome remodelling and Deacetylase complex (NuRD), a common interactor of many transcription factors, including Nanog and Oct4, is also indicated. Image adapted from (Gagliardi et al., 2013).


NuRD
Nanog
Tcfcp2l1
Sall4
Esrrβ
Nr0b1
Oct4
SWI/SNF
Mll
Sin3a
RNApol2
E2F6
N-CoR
Lsd1
PcG
Trrap/p400
Mediator
TFIID
Sox2
Jmjd1c
Requiem
Sall2
Bend3
Ruvbl2
Ctbp2
Ewsr1
Usp9x
Set
Tcfcp2
Ogt
Ubp1
Msh2
Rbm14
Zfp143
Wdr5
Esrra
Nrip1
Cdc2a
Ehmt1
Ncoa3
Zbtb9
Fkbp15
Prmt1
Snw1
Frg1
L1td1
Rif1
Rbpj
Zfp655
Gnb2l1
Pml
Rad50
Smc1a
Gtf3c4
Ints9
Xrcc5
Hells
Zmym4
Ints7
Polb
Zfp828
Ubqln4
Pnkp
Hnrnpab
Satb2
Yeats2
Lig3
Cabin1
Prkdc
Supt16h
Adnp
Zzz3
Xrcc1
Mybl2
Peg10
Trim33
L3mbtl2
Xrcc6
Asf1a
Rpa1
Wiz
Pogz
Akap8
Grhl2
Zfp462
Msh6
Cxxc5
Bptf
Mga
Nac1
Myef2
Klf5
Bcor
Emsy
Ints1
Sall3
Zbtb2
Brca1
Orc5l
Rcor2
Dnttip1
Supt6h
Mat89Bb
Tet1
Zfp219
Mad1l1
Rfc4
Skt
Sec16
Csnk2a
Smarca5
Chd7
Zfr
Cnot1
Tpx2
RanGAP1
Dpf2
Csnk1a1
Arid3b
Sall1
Kpna2
Rad51
Ppp2r1a
Arid1a
Zcchc8
Dnaja1
Zfp278
Sec13
Psmd11
Wdr82
Wdr68
Brca2
Ranbp2
Raly
Morf4l1
Ubtf
Fiz1
Cad
Mtr3
Qser1
Zbtb34
Taf9
Hnrnph2
Ddx1
Mov10
Smc3
Stub1
Rex2
Zfp281
Ints4
Rbm9
Wdr33
Zfp326
Zfp609
Pum1
Rbm4b
Dhx9
Hira
Rex1
Zic3
Arid5b
Bnc2
Gtse1
Sgol2
Ilf3
Uchl5
Brms1
Pum2

Analysis of all published Nanog and Oct4 protein interaction networks revealed that amongst their highest-confident interactors are not only other core transcription components, but also multiple proteins with an epigenetic function (Costa et al., 2013; Ding et al., 2012; Gagliardi et al., 2013; Pardo et al., 2010; van den Berg et al., 2010). This indicates that Nanog and Oct4 might control the pluripotent state by engaging with multiple components of the epigenetic machinery. Oct4 has been found to interact with multiple subunits of different chromatin-modifying complexes with documented roles in development and ESC function (Ding et al., 2012; Pardo et al., 2010; van den Berg et al., 2010). The Oct4 high-confidence interactors are the members of the Nucleosome Remodelling and Deacetylation (NuRD), such as ATP-dependent chromatin remodeler Mi-2β, the GATA zinc finger domain containing 2A and 2B (Gatad2a/b), the histone deacetylases 1 and 2 (Hdac1/2), the methyl-CpG-binding domain 3 (Mbd3), the metastasis associated proteins 1, 2 and 3 (Mta1/2/3) and the retinoblastoma binding protein 7 (Rbbp7) (Ding et al., 2012; Pardo et al., 2010; van den Berg et al., 2010). The same holds true for Nanog, where chromatin regulators rank very high in its protein network, which includes Tet1 and all components of the NuRD complex which interact with Oct4 except Rbbp7 and Hdac1 (Costa et al., 2013; Gagliardi et al., 2013). Of all the core members of the NuRD complex, only Rbbp4 was not found to interact with either Oct4 or Nanog. Although many other chromatin regulators have been studied in the context of induced pluripotency (previous section), the role the NuRD complex in reprogramming was not addressed until the start of this thesis.

Reprogramming is an epigenetic process where the epigenetic machinery plays a pivotal role in the establishment and maintenance of pluripotency. Taking into account that one the key components of the epigenetic machinery, the NuRD complex, is a high-confidence interactor of both Nanog and Oct4, I decided to investigate its function in induced pluripotency. In the next section I will review current knowledge about the role of the NuRD complex during early embryo development and in ESC biology.

1.4 – Nucleosome Remodelling and Deacetylase complex

The NuRD complex was first purified in 1998 using nuclear extracts of human cells (Tong et al., 1998; Xue et al., 1998; Zhang et al., 1998) and *Xenopus laevis* eggs (Wade et al., 1998). The NuRD complex was described after immunoprecipitation of the ATP-dependent chromatin remodeler Mi-2β, which was discovered three years earlier as an auto-antigen in dermatomyositis (Ge et al., 1995; Seelig et al., 1995). Closer inspection of the purified large multi-subunit complex led to the description of its initially identified subunits: Hdac1/2, Rbbp4 and 7 and Mta1. NuRD is the only complex with histone deacetylase and ATP-dependent chromatin remodelling activity, provided by Hdac1/2 and Mi-2a/β subunits, respectively (Lai and Wade, 2011; McDonel et al., 2009). Further characterization of the NuRD complex revealed the other core subunits: Mbd2 and 3 (Hendrich and Bird, 1998; Wade et al., 1999; Zhang et al., 1999), Mta2 (Zhang et al., 1999) and Mta3 (Fujita et al., 2004; Fujita et al., 2003), and the Gatad2a and 2b (Gatad2a/b) (Brackertz et al., 2006) (Figure 1.4.1A). Other proteins have been hypothesized to be a part of the NuRD complex, such as the cell-cycle inhibitor Cdk2ap1 (Le Guezennec et al., 2006) and the proteins of the Sall family (Lauberth and Rauchman, 2006; Lu et al., 2009; Yuri et al., 2009), which will not be discussed in this thesis. After its initial purification from xenopus and human cells, homologues of the complex were found in many different species, including *Arabidopsis thaliana* (Ahringer, 2000; Ogas et al., 1999), *Drosophila melanogaster* (Tweedie et al., 1999; Wade et al., 1999), and mouse (Hendrich and Bird, 1998; Wade et al., 1999; Zhang et al., 1999). Two NuRD complexes with different biochemical and functional properties can be isolated, Mbd2/NuRD or Mbd3/NuRD, and the main difference between them is the mutually exclusive use of either Mbd2 or Mbd3 (Le Guezennec et al., 2006). Generally, Mi-2β is the ATPase subunit assembled in the NuRD complex, but in some cases, this can be replaced by Mi-2α (Allen et al., 2013). In the context of this thesis, the NuRD complex refers to the Mbd3/Mi-2β/NuRD complex.

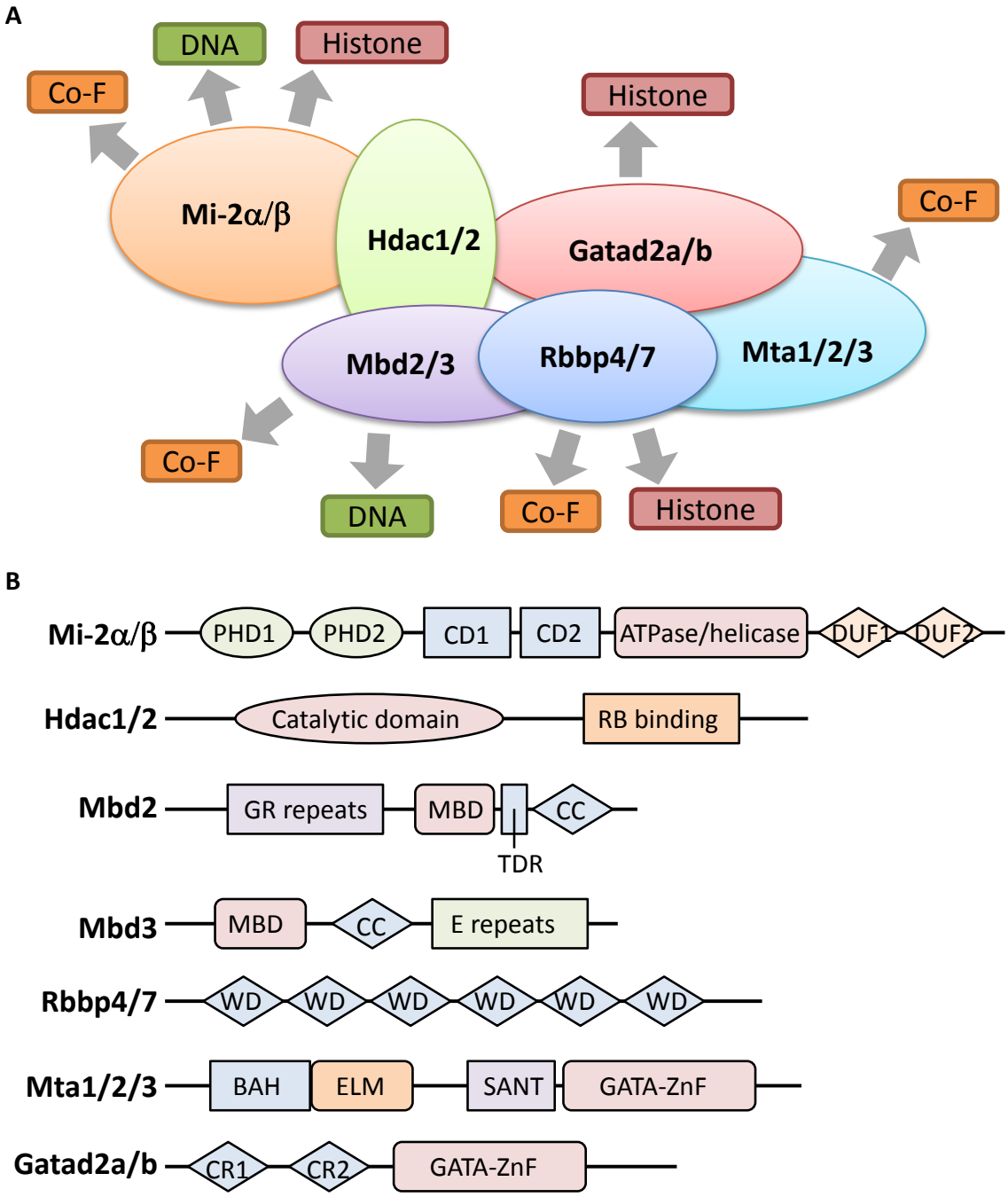


Figure 1.4.1 – The architecture of the NuRD complex.

(A) Schematic representation of the NuRD complex. The core components of the complex are represented, and the way they connect between each other reflects validated interactions. Several subunits have been shown to interact with DNA or histones, or with other co-factors (Co-F), which modulate NuRD function and specificity (Allen et al., 2013). **(B)** Schematic structure of the core subunits of the NuRD complex. PHD1/2: Plant homeodomain 1/2; CD1/2: Chromodomains 1/2; DUF1/2: Domains of unknown function 1/2; RB: Retinoblastoma; GR: Glycine-Arginine; MBD: Methyl-CpG-binding domain; TDR: Transcriptional repression domain; CC: Coiled-coil motif; E: Glutamic acid; WD: Tryptophan-aspartic acid; BAH: Bromo-adjacent homology; ELM: Egl-27 and MTA1 homology domain; SANT: domain found in the SWI-SNF, ADA, the co-repressor N-CoR and TFIIIB; GATA-ZnF: GATA zinc finger; CR1/2: Conserved regions 1/2.

Mi-2β (220kDa) is the largest component of the NuRD complex and is encoded by the *Chd4* gene. It was first identified as autoantigen in an autoimmune disease dermatomyositis (Ge et al., 1995; Seelig et al., 1995). It contains multiple functional domains: two plant homeodomains (PHD) which bind to H3K4 and H3K9me3 residues (Mansfield et al., 2011; Musselman et al., 2009); two chromodomains (CD) which bind to DNA (Bouazoune et al., 2002); an ATPase/ helicase domain which binds nucleosomes and uses energy provided from ATP hydrolysis to move them (Bouazoune et al., 2002); and two domains of unknown function (DUF) (Figure 1.4.1B). On top of being a defining member of the NuRD complex, much evidence suggest that Mi-2β also functions independently of NuRD, playing important roles in cell cycle regulation and the DNA-damage response (Amaya et al., 2013; O'Shaughnessy and Hendrich, 2013).

Hdac1 and 2 (50kDa) are histone deacetylases with one Hdac domain each and they catalyse the removal of acetyl groups from histone tails, resulting in chromatin compaction and transcriptional repression (Figure 1.4.1B). In addition to NuRD, they can also be a part of other complexes, such as Sin3A, CoREST, NCoR/SMRT and ESC-specific NODE (Allen et al., 2013).

Mbd3 (30kDa) is a member of the methyl-CpG-binding domain (MBD) family of proteins that, in contrast to the other four members (including Mbd2), does not bind methylated DNA (Hashimoto et al., 2012; Saito and Ishikawa, 2002). Yet, Mbd3 has been shown to bind to hydroxymethylated DNA (Hashimoto et al., 2012; Yildirim et al., 2011). On top of the MBD, Mbd3 contains a coiled-coil (CC) motif and a poly-glutamate (E) region, whose function is unknown (Allen et al., 2013) (Figure 1.4.1B). Mbd3 is essential for keeping all NuRD subunits together, acting as a scaffold protein, in the absence of which the complex is not assembled (Kaji et al., 2006).

Rbbp4 and 7 (50kDa) are composed of a region of tryptophan-aspartic acid (WD) repeats, which have been shown to bind to histone H4 (Murzina et al., 2008) and the co-factor friend of GATA 1 (FOG-1) (Lejon et al., 2011) (Figure 1.4.1B). They were first purified through their interaction with the retinoblastoma tumour suppressor (Qian et al., 1993), and bind to many chromatin-modifying complexes, such as NuRD, Sin3A and PRC2 (Allen et al., 2013).

Mta1, 2 and 3 (70-80kDa) were initially discovered as proteins involved in breast cancer metastasis (Toh et al., 1994), and are found in mutually-exclusive NuRD complexes, which is suggestive that the NuRD complex has different functions depending on its Mta subunit assembly (Bowen et al., 2004). Mta proteins contain a bromo-adjacent homology (BAH) domain which binds to histones (Armache et al., 2011); a Egl-27 and MTA1 homology domain (ELM) which might bind DNA or proteins; a SANT domain (found in SWI/SNF,

ADA, the co-repressor N-CoR and TFIIIB) which binds DNA and histones (Horton et al., 2007); and a Gata zinc finger (GATA-ZnF) domain that can interact with DNA and recruit proteins (Liew et al., 2005) (Figure 1.4.1B).

Gatad2a and 2b (70kDa) contain two conserved regions (CR) and a GATA-ZnF domain which are also responsible for histone binding, and bind to unmodified histones and to Mbd3 (Brackertz et al., 2006; Gong et al., 2006) (Figure 1.4.1B). Their function is not clear. Commonly, each Mbd3/Mi-2β/NuRD complex contains both Gatad2a and 2b proteins, both Hdac1 and 2, and both Rbbp4 and 7, but only one Mta protein (McDonel et al., 2009).

As mentioned above, the NuRD complex can be found in different species and in different tissues, showing that the biology of the NuRD complex goes beyond pluripotency. It has been shown that, depending on its subunit composition, the NuRD complex can have different functions related to DNA biology, ranging from chromatin organization and transcriptional control, to DNA repair and genomic stability (Denslow and Wade, 2007). Its different subunit composition leads to the interaction of the NuRD complex with different proteins, which is a major mechanism of controlling NuRD activity and localization and impacts its function (Bowen et al., 2004). Whether the transcriptional programme controls NuRD's architecture and its interactions, or whether it is the other way around, still remains to be understood. This is particularly evident in cancer, where NuRD activity has been shown to both promote and suppress tumorigenesis, depending on the cell type and microenvironment (Lai and Wade, 2011). This impedes the use of NuRD as therapeutic target. NuRD loss has been associated with aging-associated chromatin defects, such as heterochromatin loss and DNA damage (Pegoraro et al., 2009).

In summary, NuRD is a multifunctional complex that contains a dual enzymatic activity and whose aberration is associated with human disease. Although the NuRD complex has been shown to play different biological roles in different tissues and species, only the role of the NuRD complex in cell fate decisions will be discussed in this thesis. In the next section I will review the role of the NuRD complex during early embryo development and maintenance of the pluripotent state.

1.4.2 – The role of Mbd3 in early embryonic development and pluripotency

Loss-of-function studies revealed that the NuRD complex fails to assemble in the absence of the core component Mbd3 (Kaji et al., 2006), meaning that *Mbd3*-null cells can be considered NuRD-null cells. Mouse embryos lacking Mbd3 expression die early in embryonic development (Hendrich et al., 2001). *Mbd3*$^{-/-}$ embryos are indistinguishable from wild type until E4.5, exhibiting a morphologically normal segregation of the trophectoderm, epiblast

and primitive endoderm (Kaji et al., 2007). This indicates that the NuRD complex is not required for cell viability and for the two initial cell fate decisions of the blastocyst. This does not seem to be due to compensation by maternal Mbd3 contribution, since immunostaining reveals that Mbd3 disappears at the morula stage in *Mbd3*$^{-/-}$ embryos (Kaji et al., 2007). From the early post-implantation state (E5.5), *Mbd3*$^{-/-}$ embryos exhibit aberrant development, failing to differentiate and ultimately die (Kaji et al., 2007). E5.5 *Mbd3*-null embryos fail to form an organized extra-embryonic endoderm and visceral endoderm, and do not show maturation and expansion of the epiblast, exhibiting only a few Oct4-positive cells that remain proximally located as in E3.5 and E4.5. The failure in epiblast expansion does not seem to be the result of an arrest in proliferation or of the apoptosis of Oct4-positve cells, since these cells express markers of proliferating cells (Ki67), and no Oct4-Caspase-3 (marker of apoptosis) double-positive cells were found. Many of these Oct4-posive cells are also Gata4-positive, something never observed in E5.5 wild type embryos, but characteristic of E4.5 KO and wild type embryos (Kaji et al., 2007). It is therefore possible to conclude that Mbd3, and thus the NuRD complex, is necessary for the normal development of many embryonic and extra-embryonic tissues during early post-implantation development (Figure 1.4.2). Moreover, no ESC lines could be established from *Mbd3*$^{-/-}$ ICM outgrowths which contained few, if any, Oct4-positive cells, and could not be expanded (Kaji et al., 2007). Importantly, heterozygous Mbd3 embryos did not show any phenotype, being indistinguishable from wild type embryos at all stages analysed, and ESCs could be derived from their ICM outgrowths (Kaji et al., 2007) (Figure 1.4.2).

While Mbd3 is required for ESC derivation from the embryo, it is not required for ESC self-renewal, since *Mbd3*-null ESC can be stably generated by gene targeting (Kaji et al., 2006). *Mbd3*$^{-/-}$ ESC are viable, exhibiting slower proliferation, which is consistent with the role of NuRD in cell cycle progression (Lai and Wade, 2011; Sims and Wade, 2011). These cells are unable to exit self-renewal and undergo spontaneous differentiation upon LIF withdrawal (Kaji et al., 2006; Reynolds et al., 2012a) (Figure 1.4.2). *Mbd3*-null ESCs also fail to commit to differentiated lineages in an embryoid body (EB) assay that mimics embryonic development (Doetschman et al., 1985), failing to down-regulate the expression of pluripotency genes (such as *Oct4, Nanog, Rex1*) and up-regulate lineage-specific markers (such as *T brachyury, Gata6*) (Kaji et al., 2006). They also do not contribute to chimeric embryos (Kaji et al., 2006). Surprisingly, when *Mbd3*$^{-/-}$ ESC are treated with retinoic acid in the absence of LIF, they lose Oct4 expression and exit self-renewal (Kaji et al., 2006). This indicates that these cells are still responsive to certain stimuli, and that the NuRD complex might not be an absolute requirement for the exit from naïve pluripotency.

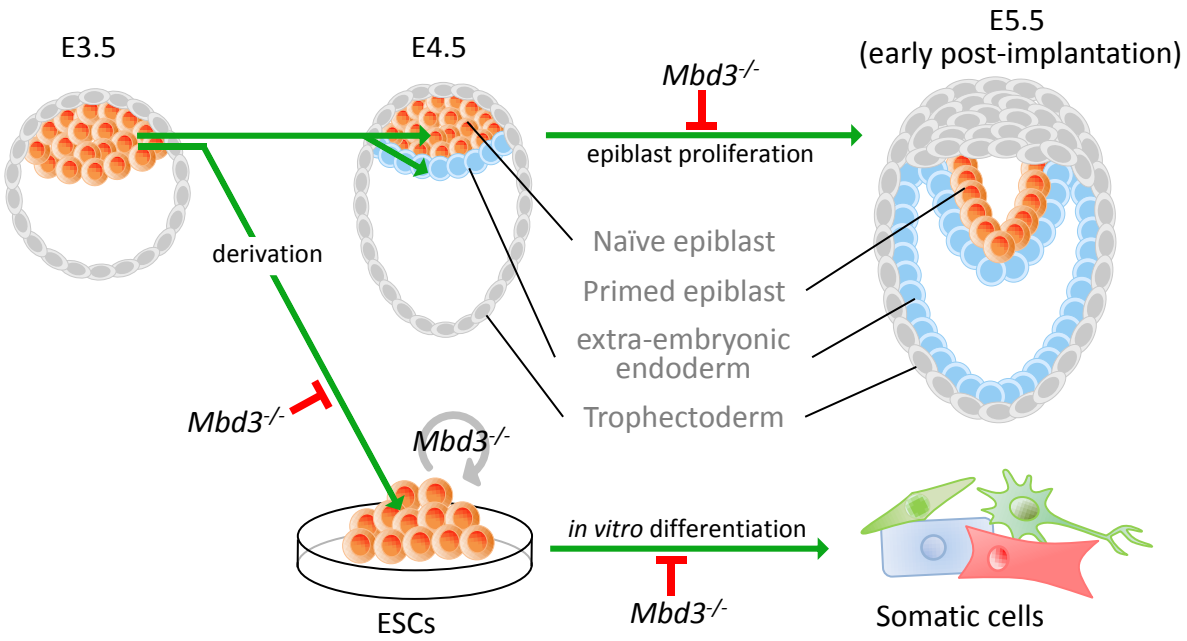


Figure 1.4.2 – In vivo and in vitro phenotypes associated with the lack of Mbd3.

From fertilization until E4.5, wild type and *Mbd3*-null embryos are morphologically indistinguishable. At E5.5 the embryo implants and starts undergoing morphological changes, including cavitation and the expansion of the epiblast, features that are absent in *Mbd3*$^{-/-}$ embryos, which eventually die. ESCs cannot be derived from *Mbd3*$^{-/-}$ embryos, but can be established *in vitro*, exhibiting impaired differentiation. Green arrows indicate normal development or differentiation; Red blunt arrows represent a requirement for Mbd3.

How the NuRD complex regulates pluripotency is still unknown. Two recent reports from the Hendrich laboratory have provided new insights into how the intricate balance between pluripotency and differentiation is maintained, uncovering an important role of the NuRD complex (Reynolds et al., 2012a; Reynolds et al., 2012b). *Mbd3*$^{-/-}$ ESCs cultured in normal culture conditions (serum plus LIF) express homogenous high levels of the pluripotency genes Klf4, Klf5, Tbx3 and Rex1 (Reynolds et al., 2012a), genes whose expression is usually heterogeneous in wild type ESCs (Toyooka et al., 2008). This is suggestive that the NuRD complex might act to repress the expression of these pluripotency genes, an action that might be a required step for the exit of self-renewal and differentiation. Indeed, KD of *Klf4* and *Klf5* resulted in a marked rescue of *Mbd3*$^{-/-}$ ESC differentiation (Reynolds et al., 2012a). In another study by the same laboratory, it was found that NuRD-mediated deacetylase activity (mediated by the Hdac1/2) contributes to reduced H3K27ac at NuRD target genes, which then become available for H3K27me3 action by PRC2 (Reynolds et al., 2012b). Mbd3 deletion, or Hdac inhibition, leads to high levels of H3K27ac at NuRD targets, resulting in the impaired binding of PRC2 components and decreased H3K27me3 levels. This indicates a relationship

between the acetylation levels controlled by NuRD and the methylation levels controlled by PRC2. Moreover, the loss of either NuRD or PRC2 activity create overlapping gene expression changes, including the up-regulation of genes involved in embryonic development (Reynolds et al., 2012b).

The results described above seem to indicate that ESC's ability to self-renew, while maintaining their differentiation potential, is dictated by a balance between opposing chromatin modifications which might be, at least in part, controlled by NuRD and PRC2. It is feasible to assume that this plastic state is induced by chromatin modifications and controlled by both the core transcription factor circuitry and external clues, transduced in the form of signalling pathways, such as LIF signalling. A model can then be created, where pluripotency is maintained by the positive action of pluripotency-associated transcription factors and LIF signalling and these compete with the repressive action of NuRD and PRC2 (Hu and Wade, 2012; Reynolds et al., 2012a; Reynolds et al., 2012b). Upon the loss of LIF signalling, the repressive action of NuRD and PRC2 dominate, resulting in differentiation. In agreement with this model, the deletion of PRC2 complex members Ezh2 (Shen et al., 2008), Suz12 (Pasini et al., 2007) and Eed (Boyer et al., 2006; Chamberlain et al., 2008; Leeb et al., 2010), or Mbd3 deletion (Kaji et al., 2006) blocks differentiation upon LIF withdrawal. This also seems to be in accordance with the observation that *Mbd3*$^{-/-}$ ESCs cultured in serum conditions seem to have a transcriptional program similar to wild type ESCs cultured in 2i medium (Reynolds et al., 2012a). Lastly, LIF signalling seems to be directly connected to chromatin status, since Stat3's binding to target genes can be potentiated by the chromatin remodelling ESC-specific BAF complex (esBAF) (Ho et al., 2009a; Ho et al., 2011; Ho et al., 2009b). The ATPase subunit of esBAF, Brg, was found to be required for Stat3 binding to target loci, mainly by repressing PRC2 binding (Ho et al., 2011). Disruption of the esBAF complex leads to rapid polycomb binding and H3K27me3-mediated silencing of many Brg-activated genes genome-wide (Ho et al., 2011). As with all models, this model does not fully describe all ESC behaviours, but it provides an integrative view of how transcription factors, signalling pathways and chromatin regulators might regulate the complicated ESC biology.

On top of its function controlling pluripotency-associated genes expression, the NuRD complex was also shown to maintain the boundary between embryonic and trophectoderm fates in ESCs (Latos et al., 2012; Zhu et al., 2009). *Mbd3*$^{-/-}$ ESCs have been shown to commit towards the normally inaccessible trophectoderm lineage in a EB assay (Kaji et al., 2006), and the efficiency of this differentiation can be enhanced by the culture of *Mbd3*$^{-/-}$ ESCs in trophoblast stem cell medium (Latos et al., 2012).

In summary, the Mbd3/NuRD complex plays a fundamental role in cell state transitions, and its activity is necessary for proper embryonic development, mainly for the expansion and differentiation of the epiblast, and for ESC commitment *in vitro*. Thus, the Mbd3/NuRD complex is an important component of the cellular machinery, but the reasons why it is so important are yet to be fully defined.

1.5 – Scope of the study

It has been less than a decade since the breakthrough discovery that somatic cells can be reprogrammed back to a pluripotent state by the forced expression of defined factors, and already important discoveries have been made to understand the mechanisms underlying this transition. As described in section 1.3.4, chromatin remodelling plays an important role in reprogramming to naïve pluripotency. Surprisingly, the role of the chromatin remodelling complex NuRD, a key regulator of developmental cell state transitions (section 1.4.2) and a high confidence interactor of both key pluripotency transcription factors Oct4 and Nanog (section 1.3.5), in induced pluripotency has never been addressed until the start of this thesis. Thus, the main aim of this thesis is to study the role of the NuRD complex in induced pluripotency. I will start by addressing the impact of the disruption of the NuRD complex on somatic cell reprogramming. I will move to assess how induced pluripotency progresses when NuRD activity is enhanced. I will finish by analysing how the NuRD complex might mechanistically function during reprogramming, studying its molecular interactions and their role in NuRD activity.

CHAPTER 2 – Materials and Methods

2.1 – Cell culture

2.1.1 – Culture media and routine cell line manipulations

PLAT-E, pre-induced Pluripotent Stem Cells (preiPSCs) and mouse embryonic fibroblasts (MEFs) were cultured in GMEM (Cat. No.: G5154, Sigma-Aldrich) supplemented with 10% FCS (Cat. No.: F7524, Sigma-Aldrich), 1x non-essential amino acid (NEAA; Cat. No.: M11-003, PAA), 1x Penicillin/ Streptomycin (Pen/Strep: Cat. No.: P11-010, PAA), 1 mM sodium pyruvate (Cat. No.: S11-003, PAA), 0.1 mM β-mercaptoethanol (Cat. No.: 31350-010, Life Technologies), 2 mM L-glutamine (Cat. No.: 25030-024, Life Technologies) and 20 ng/ml of leukemia inhibitory factor (LIF; Department of Biochemistry, University of Cambridge) - indicated as Serum plus LIF, *S+LIF*, medium throughout.

Embryonic Stem Cells (ESCs) and induced Pluripotent Stem Cells (iPSCs) were maintained in N2B27-based medium composed of DMEM/F12 (Cat. No.: 11320-033, Life Technologies) and Neurobasal (Cat. No.: 21103-049, Life Technologies) in 1:1 ratio with 1x Pen/Strep, 0.1 mM β-mercaptoethanol, 2 mM L-glutamine, 1:200 N2 supplement (Cat. No.: F005-004, PAA), and 1:100 B27 supplement (Cat. No.: 17504-044, Life Technologies) supplemented 20 ng/ml of LIF and 2i inhibitors: CHIR99021 (3 μM) and PD0325901 (1 μM) - indicated as *2i/LIF* throughout (Ying et al., 2008).

Neural Stem Cells (NSCs) were cultured in DMEM/F-12 with 1x NEAA, 0.1 mM β-mercaptoethanol, 1x Pen/Strep, 1:100 B27 supplement, 1:200 N2 supplement, 4.5 μM HEPES (Cat. No.: S11-001, PAA), 0.03 M glucose (Cat. No.: G528, Sigma-Aldrich), 120 μg/ml BSA (Cat. No.: 15260-037, Life Technologies), supplemented with 10 ng/mL of Epidermal growth factor (Egf; Cat. No.:315-09, Peprotech) and 20 ng/ml of fibroblast growth factor 2 (Fgf2; Department of Biochemistry, University of Cambridge) - indicated as E*gf+Fgf2* medium throughout.

Epiblast Stem Cells (EpiSCs) were maintained in N2B27-based medium containing 12 ng/ml of Fgf2 and 20 ng/ml of Activin A (Department of Biochemistry, University of Cambridge) - indicated as F*gf2/Act.A* medium throughout.

For selection of stable transgenic clones, cells were cultured in the presence of antibiotics. Hygromycin B (Cat. No.: P02-015, PAA) was used at 200 ug/ml, blasticidin (Cat. No.: A11139, Life Technologies) at 20 ug/ml, puromycin (Cat. No.: P9620, Sigma-Aldrich) at 1 ug/ml and Zeocyn (Cat. No.: R25001, Life Technologies) at 100 μg/ml.

For Cre-mediated transgene excision, the cells were treated with 500 nM of 4-hydroxytamoxifen (4-OHT; Cat. No.: H7904, Sigma-Aldrich).

EpiSCs were cultured on plastic which had been coated for 30 min with 10 μg/mL fibronectin (Cat. No.: FC-010, Millipore), and NSCs were cultured on plastic coated for 3

hours with 10 μg/mL laminin (Cat. No.: L2020, Sigma-Aldrich) in phosphate buffered saline (PBS). PLAT-E cells were cultured on plastic without substrate. All other cell types were grown on 0.10% gelatine (Cat. No.: G1890, Sigma-Aldrich).

For cell passaging, cells cultured in serum-containing media were dissociated with trypsin (Cat. No.: 15090046, Life Technologies) after washing with PBS, and cells cultured in serum-free media were dissociated with accutase (Cat. No.: L11-007, PAA). After their treatment with dissociation enzyme, the cells were centrifuged and replated at the desired density in fresh media.

All cell types were manipulated in a sterile BioMAT Class II Microbiological Safety Cabinet (Thermo Scientific) and maintained in a humidified Sanyo incubator (MCO-18M) at 37° and 7% CO_2.

All cell types were frozen in S+LIF supplemented with 10% dimethyl sulfoxide (DMSO; Cat. No.: A3372,0100, AppliChem) in cryovials. Cells were first put in a -80^{0}C freezer and transferred to a liquid nitrogen tank 2-14 days later.

2.1.2 – Cell derivation

NSCs: brains from *Mbd3*$^{fl/fl}$ and *Mbd3*$^{ex1fl/ex1fl}$ E13.5 embryos were dissected, dissociated in Egf+Fgf2 medium and plated onto the laminin-coated cell culture flasks. *Mbd3*$^{fl/-}$ NSCs were derived from ESCs as described (Pollard et al., 2006). Briefly, ESCs were seeded onto gelatinized 10 cm dishes in N2B27 medium for 7 days. After this period, cells were trypsinized and plated on non-gelatinized dishes for 3 days in Egf+Fgf2 medium. The emergent neurospheres were then seeded on gelatinized plates and maintained in monolayer in Egf+Fgf2 medium. For Cre-mediated excision of the Mbd3 floxed allele, *Mbd3*$^{fl/-}$ NSCs were nucleofected with a pCAG-Cre-ires-Puro plasmid and clonal lines of *Mbd3*$^{-/-}$ NSCs were expanded.

MEFs: organ-free carcasses from E12.5 or 13.5 embryos were dissociated into small pieces, trypsinised and plated in S+LIF medium.

EpiSCs: *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ EpiSCs were derived from ESCs as previously described (Guo et al., 2009). Briefly, ESCs transfected with pPB-EOS-GFP-ires-Puro (EOS-GiP - GFPiresPuro under the control of early transposon promoter and Oct-4 and Sox2 enhancers) and were cultured in Fgf2/Act.A medium for at least 10 passages before analysis. To obtain a pure EpiSC culture, GFP$^+$ cells were removed by fluorescence-activated cell sorting (FACS). Genotypes were analysed by polymerase chain reaction (PCR) using the *Mbd3* genotyping primer pair (Table 2.1).

2.1.3 – Generation of transgenic cell lines

NSCs and preiPSCs were transfected using Amaxa Nucleofection Technology (Lonza AG). 2 x 10^6 cells were used per transfection and program T-020 was used. ESCs, iPSCs and EpiSCs were transfected in suspension using Lipofectamine 2000 (Cat. No.: 11668-019, Life Technologies). Both protocols were performed according to the manufacturer's instructions. piggyBac transposon (pPB) plasmids were co-transfected with piggybac transposase expression vector pBase (mixture of 1:1) to generate stable cell lines. Selection for transgenes was applied for at least 3 passages before experiments were set up.

2.1.4 – siRNA knockdown

Knockdown was carried out using the Flexitube siRNAs (Qiagen) listed in Table 2.1. For Mbd3 knockdown, four different small interference RNAs (siRNAs) (Mm_Mbd3_1, 2, 3 and 5) were individually tested and three of them were chosen to be used as a pool in subsequent experiments: siRNA #1, #3 and #5. AllStars Negative Control siRNA (Cat. No.: 1027280, QIAGEN) was used as a control. The final concentration of siRNAs for transfection was 0.2 μM/cm^2. For reprogramming experiments, EpiSCs were transfected with the siRNAs using the Lipofectamine RNAiMAX transfection reagent (Cat. No.: 13778-150, Life Technologies) according to the manufacturer's protocol. The media was switched to 2i/LIF 24h after transfection.

2.1.4 – NSC, MEF and preiPSC reprogramming experiments

To generate retroviruses containing the reprogramming factors, 2 x 10^6 PLAT-E cells (per transfection) were seeded in 10 cm dishes and transfected the next day with 9 μg of pMXs plasmid (MKO or MKOS in case of NSCs and MEFs, respectively; where indicated, pMXs-GFP was also used) using the FuGENE 6 transfection reagent (Cat. No.: E2691, Promega) according to the manufacturer's instructions. The medium was switched to S+LIF the next day. The retrovirus-containing supernatants from PLAT-E cultures were collected 48 hours post-transfection and filtered using 0.45 μm filters. Polybrene (Cat.No.:H9268**,** Sigma-Aldrich) was added to the filtered supernatants to a final concentration of 4 μg/ml. The mixture was then applied to the plated NSCs or MEFs. In case of NSCs, 24h after incubation, the virus-containing medium was replaced with Egf+Fgf2 medium for 2-3 days, after which the cells were switched to S+LIF medium in the case of NSCs, or to S+LIF in the case of MEFs, to enable preiPSC (reprogramming intermediate) formation. MEFs were maintained in S+LIF throughout. The emergent preiPSCs were then switched to 2i/LIF medium to induce

complete reprogramming. Where indicated in the text, preiPSCs were passaged and stably transfected at the preiPSC stage, plated in S+LIF and switched to in 2i/LIF conditions 1-2 days later. *Nanog*-GFP MEF derived preiPSCs were chosen to address the synergy between Nanog and Mbd3 during reprogramming since they very inefficiently convert to naïve pluripotency in 2i/LIF conditions, unless transfected with additional factors (cells used in (Costa et al., 2013)). Where cells contained a reprogramming reporter (*Nanog*-GFPiresPuro or *Oct4*-GFPiresPuro), 1 ug/ml puromycin was added to 2i/LIF cultures six days after the medium switch. Reprogramming experiments were ended 12 days after the medium switch to 2i/LIF. The number of NSCs, MEFs or preiPSCs plated differ from experiment to experiment and are indicated in the figure legends.

2.1.5 – EpiSC reprogramming

Transgenic EpiSCs were plated in Fgf2/Act.A medium and switched to 2i/LIF conditions the next day. Once the medium is switched to 2i/LIF the EpiSCs no longer proliferate, unless they undergo reprogramming, making the resulting iPSC colonies representative of the initial plated EpiSC numbers. Where cells contained a reprogramming reporter (*Oct4*-GFPiresPuro of EOS-GFPiresPuro), 1 ug/ml puromycin was added to 2i/LIF cultures six days after the medium switch. Reprogramming experiments were ended 12 days after the medium switch to 2i/LIF. The number of EpiSCs plated differs from experiment to experiment and in indicated in the figure legends.

2.1.6 – piggyBac transposon reprogramming

The PB-TAP IRI attP2LMKOSimO or PB-TAP IRI tetO-STEMCCAimO (500 ng), pPB-CAG-rtTA (500 ng) and pCyL43 piggyBac transposase expression vector (500 ng) were introduced into MEFs with the *Nanog-GFP* reporter (Chambers et al., 2007), which were seeded at 1.0 x 10^5 cells per well in a 6-well plate on the day before transfection and transfected using 6 μl of FugeneHD (Cat. No.: E2311, Promega). Twenty-four hours later, the culture medium was changed to S+LIF medium supplemented with 1.0 μg/ml doxycycline (DOX) (Cat. No.: 198955, MP Biomedicals), 10 μg/ml vitamin C (Cat. No.: A4403, Sigma-Aldrich) and 500 nM Alk inhibitor A 83-01 (Cat. No.: 2939, TOCRIS Bioscience) (+DVA). This medium was changed every two days until day 13 of reprogramming. Lentiviral infection with pLKO.1 encoding short hairpin RNA (shRNA) expression against Mbd3 (shMbd3.2 and shMbd3.5), Hygromycin and Zeocin resistant genes (shHyg and shZeo) was carried out 24 hours after DOX administration. Cell lysates from one of the triplicate wells at day 7 of reprogramming were used for western blotting analysis to confirm Mbd3

knockdown. For Mbd3 exon1 deletion, the retroviral Cre-ERt2 expression vector was infected 24 hours after PB-TAP MKOSimO piggybac transfection with Fugene. At the same time the culture medium was changed to S+LIF +DVA medium. 4-OHT was added at this point or 48 hours later, and kept in the culture medium for 48 hours. Cell lysates from one of the triplicate wells at day 7 of reprogramming were used for Western blotting analysis to confirm Mbd3 depletion.

2.1.7 – Cell differentiation

For embryoid body differentiation, 1.5 x 10^6 cells were plated in non-adherent 10 cm bacterial dishes in serum - LIF medium. Samples were collected at day 3, 5 and 7 of differentiation and analysed by quantitative reverse transcriptase PCR (qRT-PCR). After the last time-point, embryoid bodies were plated on 0.10% gelatine and the generation of beating cells was analysed over the next 2-3 days. For assessing ESCs/iPSCs differentiation in monolayer, 0.2 x 10^6 cells were plated on 0.10% gelatine and cultured in serum - LIF medium. Samples were collected at day 2, 4, 6, 8 and 10 of differentiation and analysed by flow cytometry. Clonal LIF-independent self-renewal was assessed by plating 2000 ESCs/iPSCs per well in a 6-well plate in serum minus LIF medium. AP-staining was performed at day 3, 5 and 7.

2.1.8 – Cell proliferation analysis

Cell proliferation was assessed by counting cells every 24h using the Vi-Cell XR Cell Viability analyser (Cat. No.: 731050, Beckman Coulter). For iPSCs proliferation analysis, 2000 cells were plated per 96-well plate and assayed with CellTiter 96 AQueous One Solution Cell Proliferation Assay (G3580, Promega) every 24h.

2.2 – Plasmids

pMX-Klf4 (Cat. No.: 13370), pMX-Oct3/4 (Cat. No.: 13366), pMX-c-Myc (Cat. No.: 13375), pMX-Sox2 (Cat. No.: 13367) and pLKO.1 (Cat. No.: 8453) were obtained from Addgene repository; pCyL43 (PBase) and pPB-CAG-rtTA were obtained from Sanger Institute's plasmid repository; pDONR211 (Cat. No.:12536-017) was obtained from Life Technologies; pPB-CAG-Nanog-pA-pgk-hph, pPB-CAG-Klf4-pA-pgk-hph, pPB-CAG-DEST-pA-pgk-hph and pMX-GFP were kindly provided by Dr Thorold Theunissen; pPB-CAG-Klf2.2A.Nanog-Cherry-ires-zeo, pPB-CAG-DEST-ires-zeo, pPB-CAG-Tet1-pA-pgk-hph and pPB-CAG-Tet2-ires-bsd were kindly provided by Dr Yael Costa; pPB-CAG-Nr5a2-pA-pgk-hph and pPB-CAG-DEST-ires-bsd were kindly provided by Moyra Lawrence;

pCAG-Cre-ires-Puro and pCAG-Mbd3bireshygro were kindly provided by Dr Brian Hendrich; pCAG-CreERt2^{NLS}-IRES-BSD was kindly provided by Dr Joerg Betschinger; pPB-EOS-GFP-ires-Puro (pPB-EOS-GiP) was kindly provided by Dr Ge Guo. PB-TAP IRI attP2LMKOSimO, PB-TAP IRI tetO-STEMCCAimO (iOKSM; Oct4-F2A-Klf4-ires-Sox2-E2A-c-Myc) and PB-TAP MKOSimO (iMKOS) were kindly provided by Dr Keisuke Kaji. They were generated by transferring STAMCCA reprogramming cassette into PB-TAP IRI piggyback backbone with ires-mOrange (O'Malley et al., 2013; Sommer et al., 2009), and MKOSimO cassette into PB-tetO backbone (Woltjen et al., 2009), respectively.

For cloning of the pPB-CAG-Mbd3b-ires-bsd and pPB-CAG-Mbd3c-ires-bsd plasmids, Mbd3b was amplified from pCAG-Mbd3bireshygro and Mbd3c from mouse ESC cDNA, using primers introducing attB arms for Gateway cloning (Table 2.1). The amplified transgenes were purified from 1% agarose gels (Cat. No.: A9539, Sigma-Aldrich) and 300 ng was incubated with 150 ng of pDONR211 vector and 3 μl of BP Clonase II enzyme mix (Cat. No.: 11789-020, Life Technologies) in TE buffer to final volume of 15 μl (BP reaction - incubation at room temperature, overnight). After incubation, 300 ng of pPB-CAG-DEST-ires-bsd destination plasmid and 3 μl of LR Clonase II enzyme mix (Cat. No.: 11791-020, Life Technologies) were added to BP reaction (LR reaction - incubation at room temperature, 5 hours). After 5 hours, Proteinase K was added to the LR (incubation at 37^0C, 10 min). 2 μl of the final reaction were transformed into chemically competent E. cloni bacterial cells (Cat. No.: 60106, Lucigen), according to the manufacturer's protocol. All constructs were sequenced by Sanger-sequencing (Department of Biochemistry, University of Cambridge) using internal primers (Table 2.1).

2.3 – RNA isolation, cDNA synthesis and qRT-PCR

Total RNA was isolated from cells using the RNeasy mini kit (Cat. No.: 74104, QIAGEN) in accordance to the manufacturer's protocol. After purification, 1 μg of total RNA was reverse transcribed into cDNA using the SuperScript III First-Strand Synthesis SuperMix kit (Cat. No.: 18080-400, Life Technologies). 10 ng of cDNA was used for qRT-PCR reactions that were set up in triplicates using either TaqMan Universal PCR Master Mix (Cat. No.: 4352042, Life Technologies) or Fast SYBR Green Master Mix (Cat. No.: 4385612, Life Technologies). TaqMan gene expression assays (Life Technologies) or specific primers (see Table 2.1) were used for each gene analysed. qRT-PCR experiments were performed using StepOnePlus Real Time PCR System (Cat. No.: 4376600, Life Technologies). Delta Ct values to *Gapdh* were calculated and brought to power -2. Where indicated, the values were

normalized to the highest value. Error bars represent standard deviation (STDEV) of technical triplicates.

Table 2.1 – Primers and si/shRNA oligos used in the study.

Mbd3 genotyping primers	
Mbd3 gen. FP	ACTGCTCCAGCTTGGTACAG
Mbd3 gen. RP	AATCAGATCACTTCAGCTCC
Cloning primers	
attB-Mbd3b FP	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CACCATGGAGCGGAAGAGCCCCA
attB-Mbd3b RP	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTACACTCGCTCTGGCTCC
attB-Mbd3b inter. FP	CTGGCACGTTACCTGGGCGGAT
attB-Mbd3 inter. RP	TGCTGCGGAACTTCTTCCCGCT
attB-Mbd3c FP	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CACCATGGCGCGCATTTGGTTTGG
attB-Mbd3c RP	GGGGACCACTTTGTACAAGAAAGCTGGGTC CTACACTCGCTCTGGCTCC
attB-Mbd3c inter. FP	ACGGGGAGGCACCACTGGACAA
attB-Mbd3 inter. RP	CGGGCAAGCTCCTCCACATGAG
Primers used with SYBR green	
Mbd3 FP	AGAAGAACCCTGGTGTGTGG
Mbd3 RV	TGTACCAGCTCCTCCTGCTT
Pl-1 FP	ATTTTGACTACCCTGCTTGGTCT
Pl-1 RP	TCTACATAACTGAGGAGGGGAAAG
Gapdh FP	CCCACTAACATCAAATGGGG
Gapdh RP	CCTTCCACAATGCCAAAGTT
Olig2 FP	CTGCTGGCGCGAAACTACAT
Olig2 RP	CGCTCACCAGTCGCTTCAT
Blbp FP	AGACCCGAGTTCCTCCAGTT
Blbp RP	ATCACCACTTTGCCACCTTC
Sox2 FP	TCCAAAAACTAATCACAACAATCG
Sox2 RP	GAAGTGCAATTGGGATGAAAA
Nr5a2 FP	CCAGAAAACATGCAAGTGTCTCAA
Nr5a2 RP	CGTGAGGAGACCGTAATGGTA
Applied Biosystems Taqman probes	
Nanog	Mm02384862_g1
Rex1	Mm03053975_g1
Klf4	Mm00516104_m1
Klf2	Mm01244979_g1
Fgf5	Mm00438919_m1

Lefty1		Mm00438615_m1
T-Brachyury		Mm01318252_m1
Gapdh		4352339E
Esrrβ		Mm00442411_m1
Gata4		Mm00484689_m1
Nr0b1		Mm00431729_m1
Tet1		Mm01169087_m1
Tet2		Mm00524395_m1
Applied Biosystems custom Taqman probes		
Retroviral Klf4 FP		TGGTACGGGAAATCACAAGTTTGTA
Retroviral Klf4 RP		GAGCAGAGCGTCGCTGA
Retroviral Klf4 probe		FAM-CCCCTTCACCATGGCTG-MGB
Retroviral Oct4 FP		TGGTACGGGAAATCACAAGTTTGTA
Retroviral Oct4 RP		GGTGAGAAGGCGAAGTCTGAAG
Retroviral Oct4 probe		FAM-CACCTTCCCCATGGCTG-MGB
Retroviral cMyc FP		TGGTACGGGAAATCACAAGTTTGTA
Retroviral cMyc RP		GGTCATAGTTCCTGTTGGTGAAGTT
Retroviral cMyc probe		FAM-CCCTTCACCATGCCCC-MGB
Retroviral Sox2 FP		TGGTACGGGAAATCACAAGTTTGTA
Retroviral Sox2 RP		GCCCGGCGGCTTCA
Retroviral Sox2 probe		FAM-CTCCGTCTCCATCATGTTAT-MGB
Endogenous Oct4 FP		TTCCACCAGGCCCCC
Endogenous Oct4 RP		GGTGAGAAGGCGAAGTCTGAAG
Endogenous Oct4 probe		FAM-CCCACCTTCCCCATGGCT-MGB
siRNAs from Qiagen		
Mm_Mbd3_1 - SI00206836		CGGAAAGATGTTGATGAACAA
Mm_Mbd3_2 - SI00206843		ACCGGTGACCAAGATCACCAA
Mm_Mbd3_3 - SI00206850		CAGGACCATGGACTTGCCCAA
Mm_Mbd3_5 - SI02740045		AAGTCACTTTCCTTCAATAAA
AllStars Negative Control siRNA		Cat. No.: 1027280
Upper strand oligos used for lentiviral knockdown vectors		
shMbd3.2	Upper strand	CCGGGCGCTATGATTCTTCCAACCACTCGA GTGGTTGGAAGAATCATAGCGCTTTTT
	Bottom strand	AATTAAAAAGCGCTATGATTCTTCCAACCA CTCGAGTGGTTGGAAGAATCATAGCGC
shMbd3.5	Upper strand	CCGGAAGTCACTTTCCTTCAATAAACTCGA GTTTATTGAAGGAAAGTGACTTTTTTT
	Bottom strand	AATTAAAAAAAGTCACTTTCCTTCAATAAA CTCGAGTTTATTGAAGGAAAGTGACTT
shHyg	Upper strand	CCGGGCGAAGAATCTCGTGCTTTCACTCGA GTGAAAGCACGAGATTCTTCGCTTTTT

	Bottom strand	AATTAAAAAGCGAAGAATCTCGTGCTTTCA CTCGAGTGAAAGCACGAGATTCTTCGC
shZeo	Upper strand	CCGGGCCAAGTTGACCAGTGCCGTTCTCGA GAACGGCACTGGTCAACTTGGCTTTTT
	Bottom strand	AATTAAAAAGCCAAGTTGACCAGTGCCGTT CTCGAGAACGGCACTGGTCAACTTGGC

2.4 – Blastocyst injection, chimera generation and germline transmission assessment

For blastocyst injection, standard microinjection methodology using host blastocysts of C57BL/6 strain was employed. Floxed pPB transgenes were excised using TAT-Cre treatment before injection. Injected blastocysts were transferred to recipient mice to assess the contribution to chimeras. Generated chimeras were back-crossed with C57BL/6 mice to assess germline transmission.

2.5 – Alkaline Phosphatase staining

Cells were fixed with a citrate-acetone-formaldehyde solution and stained for 30 min using the Alkaline Phosphatase (AP) kit (Cat. No.: 86R-1KT, Sigma-Aldrich) according to manufacturer's instructions.

2.6 – Flow cytometry and imaging

Flow cytometry analysis was performed using a fully calibrated BD LSRFortessa analyser (488nm, 635nm, 406nm and 561nm). GFP was excited by a 488nm laser and detected using a 530/30 filter. DAPI (406nm 450/50) was used to exclude dead cells. Acquisition gates were set using FACS DIVA software and post-acquisition analysis was performed using FlowJo software (Treestar). Briefly, live, single, and intact cells were identified based on viability dye exclusion, SSC-W, and FSC-A v SSC-A parameters, respectively. Lastly, GFP positive cells were identified based on the untransfected control. A minimum of 10,000 single intact live cells was acquired. Cell sorting was performed using a Beckman Coulter MoFlo Legacy Cell Sorter (488nm, 355nm and 648nm), following a similar gating strategy as described above. Live cells were imaged with inverted Olympus IX51 microscope supplied with the Leica DFC310 FX digital colour camera, and processed with Leica software.

2.7 – Western blotting

Protein extracts were obtained by incubation of cells with RIPA buffer – PBS (Cat. No.: D8537, Sigma-Aldrich), 1% IGEPAL (Cat. No.: I3021, Sigma-Aldrich), 0.1% SDS (Cat. No.: BP1311-1, Fisher Scientific), 0.1 mM EDTA (Cat. No.: 15575-038, Life Technologies) supplemented with proteinase inhibitors (Cat. No.: 05 892 970 001, Roche). Total protein was quantified using BCA protein assay kit (Cat. No.: 23227, Thermo Scientific). Protein samples were denatured using NuPAGE LDS sample buffer (Cat. No.: NP0007, Life Technologies) supplemented with 10% DTT (Cat. No.: 43816, Sigma-Aldrich). The same amount of protein was loaded into Novex NuPAGE 10% Bis-Tris gels (NP0301BOX, Life Technologies) for all the samples. Protein transfer was carried out with the iBlot 2 system (Cat. No. IB21001, Life Technologies) using nitrocellulose iBlot 2 transfer stacks (Cat. No.: IB23001, Life Technologies). The following primary antibodies used: rabbit polyclonal Mbd3 antibody (Cat.No.: A302-528A, Bethyl, 1:2000), rabbit polyclonal Nanog antibody (Cat. No.: A300-397A, Bethyl, 1:5000) goat polyclonal Mta2 antibody (C-20) (Cat. No.: sc-9447, Santa Cruz, 1:1000) and mouse monoclonal antibody α-tubulin (Cat. No.: ab7291, Abcam, 1:5000). Secondary HRP-conjugated antibodies used: anti-rabbit (Cat. No.: NA934VS, GE Healthcare, 1:10000), anti-mouse (Cat. No.: NA931VS, GE Healthcare, 1:10000) and anti-goat (Cat. No.: sc-2020, Santa Cruz, 1:2000). Blocking was carried out using 5% milk (Cat. No.: 70166, Sigma-Aldrich)/0.1% tween-20 (Cat. No.: P1379, Sigma-Aldrich) in PBS (blocking solution) for 1 hour at room temperature. Primary antibodies were diluted in blocking solution and membranes were incubated with antibody overnight at 4^0C. Three washing steps of 10 min were carried out with blocking solution after primary antibody incubation. Secondary antibodies were diluted in blocking solution and incubated with membranes for 1 hour at room temperature. Then three washing steps of 20 min were performed with 0.1% tween-20 in PBS. Membranes were developed using the ECL Prime detection Kit (Cat. No.: RPN2232, GE Healthcare) according to the manufacturer's instructions. Membrane re-probing with another primary antibody was carried out after stripping the membrane with 100 mM β-Mercaptoethanol, 2% SDS and 62.5 mM Tris-HCl solution in water for 30 min at 50^0C.

2.8 – DNA dot blot

To prepare DNA samples, 20X Saline-Sodium Citrate (SSC – 3M Sodium chloride (Cat. No.: 31434, Sigma-Aldrich) and 0.3M triSodium citrate dihydrate (Cat. No.: S1804, Sigma-Aldrich) in water, pH 7.0), and water were added to the DNA to give a final concentration of 6X SSC in the minimum possible volume. DNA was denatured by incubation the samples for 10 min at 100^0C. Samples were then placed on ice and diluted by adding an equal volume of

20X SSC. To prepare the membrane for application of the samples, a grid of 1 cm x 1 cm squares was marked in uncharged hybond-C extra nitrocellulose membranes (Cat. No.: 45-000-930, GE Healthcare) with a blunt pencil and the membrane was incubated in 6X SSC for 10 min at room temperature. After incubation, the membranes were allowed to air-dry for 10 min and samples were applied. Up to 2 μl were applied each time, with each spot being allowed to dry before the next aliquot was applied on top, if more sample was required. After all samples were applied, the membranes were placed between two sheets of Watman paper and cross-linked in a Hoefer UVC 500 ultraviolet cross-linker (Cat. No.: 80-6222-31, GE Healthcare) using 1200 $\mu J/cm^2$ energy. The membranes were then rehydrated by incubation in 2X SSC for 5 min at room temperature. After air-drying for 10 min, the membranes were incubated overnight at 4^0C with blocking solution (10% milk/ 1% bovine serum albumin / 0.1% tween-20 in PBS). For 5-hydroxymethylcytosine (5-hmC) detection, the membranes were incubated with anti-5-hmC rabbit polyclonal antibody (Cat. No.: 39769, Active Motif, 1:1000) for 1.5 hours at room temperature and washed 4 x 10 min with 0.1% tween-20 in PBS. After washing off the primary antibody, the membranes were incubated with HRP-conjugated anti-rabbit (Cat. No.: NA934VS, GE Healthcare, 1:10000) for 1 hour at room temperature. Primary and secondary antibodies were diluted in blocking solution. The membranes were washed 3 x 20 min in 0.1% tween-20 in PBS and developed using the ECL Prime detection Kit (Cat. No.: RPN2232, GE Healthcare) according to the manufacturer's instructions.

CHAPTER 3 – Investigation of the role of Mbd3/ NuRD complex in the initiation of reprogramming

3.1 – Introduction

3.1.1 – NuRD complex during the initiation of reprogramming

Our laboratory has recently discovered that Oct4 is of extreme importance for somatic cell reprogramming, and *Oct4*$^{-/-}$ somatic cells fail to be reprogrammed to naïve pluripotency if no exogenous source of Oct4 expression is provided (Radzisheuskaya et al., 2013). Furthermore, Oct4 interactome studies in ESCs revealed members of NuRD complex, an essential chromatin complex for embryo development and pluripotent cell differentiation, as its highest confidence interactors (Ding et al., 2012; Liang et al., 2008; Pardo et al., 2010; van den Berg et al., 2010). Taking these observations into account, we decided to assess the role of the NuRD complex during transcription-factor induced pluripotency.

To study the role of the NuRD complex in nuclear reprogramming, we have made use of the biochemical phenomena described by the Hendrich laboratory where the NuRD complex fails to assemble, if the 32kD subunit Mbd3 is deleted, as Mbd3 acts as an essential scaffold protein of the NuRD complex (Kaji et al., 2006). The process of somatic cell reprogramming mediated by the retroviral-delivery of transcription factors can be divided into two distinct and sequential phases: initiation and establishment of reprogramming. To better dissect the role that the NuRD complex plays during the process of reprogramming, I have looked closely at the impact of Mbd3 removal during both phases of reprogramming. In this chapter, I focused my attention to the role of the NuRD in the initial phase of reprogramming. The role of the NuRD complex in the establishment of reprogramming will be addressed in the chapter 4.

The initiation phase of reprogramming consists of the transduction of somatic cells with retroviruses (pMXs) encoding the reprogramming factors Oct4, Klf4 c-Myc and Sox2 (Figure 3.1.1). If Neural Stem cells are used, Sox2 is not necessary since this cell type already expresses it (Silva et al., 2008). If transduced cells are maintained in serum-containing conditions (S+LIF), highly-proliferative reprogramming intermediates (pre-induced Pluripotent Stem cells, preiPSCs) start to emerge and take over the culture. The preiPSCs show downregulation of lineage specific genes from the original somatic cell type and express alkaline phosphatase, amongst other pluripotency markers, but are dependent on the continuous expression of the transgenes (Silva et al., 2008). PreiPSCs do not express key pluripotency genes, such as Nanog and Esrrβ and do not exhibit full differentiation potential (Okita et al., 2007; Silva et al., 2008; Wernig et al., 2007). This system allows the detailed study of the impact of the NuRD complex during the initial stages of reprogramming, since its readout (the amount of preiPSC colonies) is a direct measure of this phase and not of the overall process.

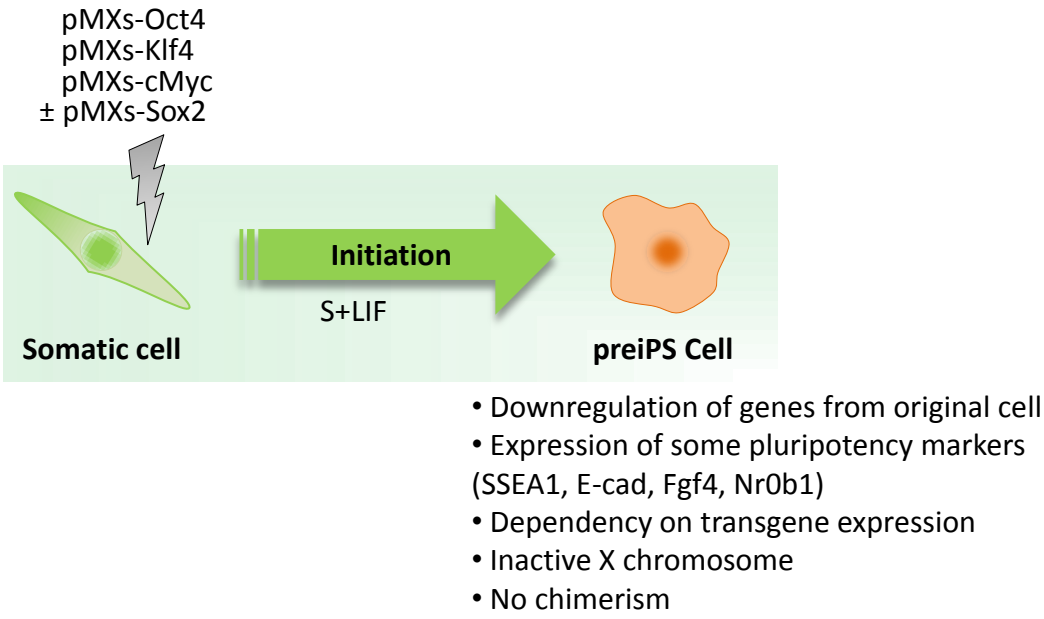


Figure 3.1.1 – Initiation phase of reprogramming.

Schematic representation of the experimental design used to assess the requirement of the NuRD complex in the initiation of reprogramming. During this phase, downregulation of genes from the original somatic cell type is observed, but resulting preiPSCs fail to activate endogenous naïve pluripotency transcriptional programming, being dependent on transgene expression (Silva et al., 2008).

3.1.2 – Aim of the chapter

Since chromatin remodelling plays an important role in reprogramming to naïve pluripotency (Apostolou and Hochedlinger, 2013; Papp and Plath, 2013) and, as the NuRD complex is a high confidence interactor of Oct4 and a key regulator of developmental cell state transitions, I investigated if NuRD is also involved in the reverse biological process of the induction of pluripotency. I aimed to address not only whether NuRD is required for somatic cell reprogramming to pluripotency, but also if the initiation of reprogramming kinetics is altered after the disruption of the complex.

3.2 – Results

3.2.1 – Mbd3/NuRD complex is required for efficient initiation of reprogramming

To study the impact of Mbd3/NuRD deletion during somatic cell reprogramming, *Mbd3*-null somatic cells had to be generated. Since the NuRD complex is required for embryo development and *in vitro* differentiation of ESCs, I employed a strategy where Mbd3 could be conditionally deleted after differentiation. For that, *Mbd3* alleles were targeted in ESCs in order to flank them with loxP sites, which allowed future excision induced by Cre recombination (Ramirez-Solis et al., 1995; Zheng et al., 2001). After targeting, selection and clonal expansion, $Mbd3^{fl/-}$ (one allele floxed; one allele removed) ESCs were generated (Kaji et al., 2006). Upon Cre-mediated excision of loxP-flanked *Mbd3* allele, $Mbd3^{-/-}$ ESCs were established containing no Mbd3 protein (none of the isoforms were detected, Figure 3.2.1A). Phenotypically, $Mbd3^{fl/-}$ ESCs are indistinguishable from their wild type parental ESC line, showing the same level of Mbd3 protein as assessed using western blot analysis (Figure 3.2.1B). The same cell line was used in different published reports as a control for Mbd3-deleted line (Kaji et al., 2006; Reynolds et al., 2012a; Reynolds et al., 2012b). A different genetic targeting was also designed, where only Mbd3 exon1 (ex1) was *loxP*-flanked ($Mbd3^{ex1fl/-}$ ESCs; Figure 3.2.1C), resulting in a loss of Mbd3a and Mbd3b, but retaining residual hypomorphic Mbd3c expression (Figure 3.2.1D) (Kaji et al., 2006).

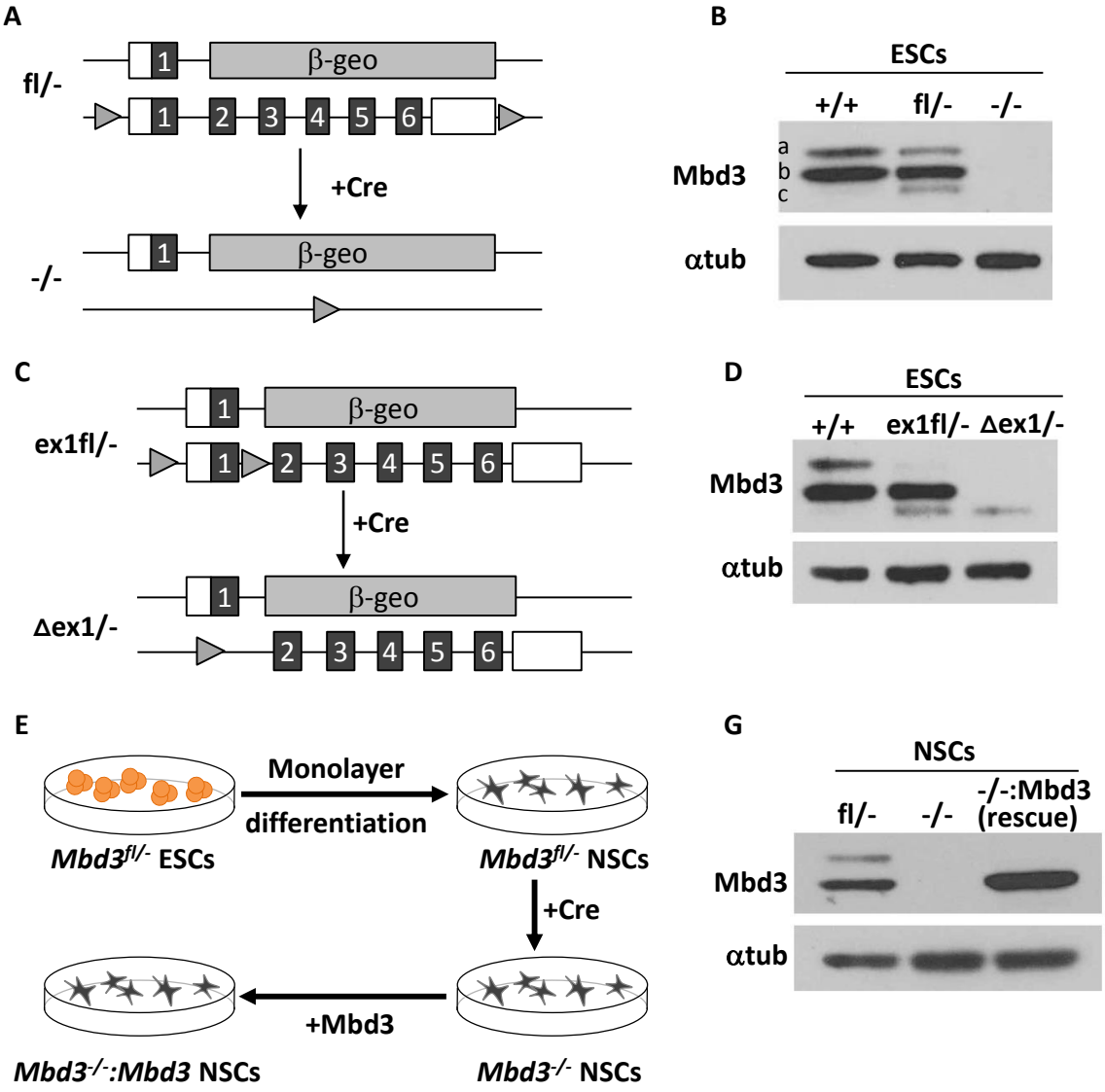
Since NSCs constitute a somatic cell system that can grow clonally and be maintained for long periods in culture, enabling efficient and quick gene targeting and manipulation, they were chosen as a model to study the role of Mbd3/ NuRD in reprogramming (Silva et al., 2008). These features of NSCs make them an optimal cell type for my studies, allowing the generation of expandable *Mbd3*-null clonal lines. For the establishment of $Mbd3^{-/-}$ NSCs, $Mbd3^{fl/-}$ ESCs were subjected to an adherent monolayer differentiation protocol (Pollard et al., 2006), and resulting NSCs were treated with Cre-recombinase to induce the deletion of the floxed-*Mbd3* allele (Figure 3.2.1E). The gene targeting strategies and cell lines described above were generated and kindly provided by the Hendrich laboratory.

In order to establish an *Mbd3* rescue NSC line, $Mbd3^{-/-}$ NSCs were stably transfected with an *Mbd3* transgene. As stated above, there are three Mbd3 isoforms, which differ in their N-terminus: Mbd3a is 32 amino acids bigger than Mbd3b, and the latter is 8 amino acids bigger than Mbd3c (Figure 3.2.1F). Mbd3b is the most abundant *Mbd3* isoform in ESCs (Figure 3.2.1B), and this was the isoform used to generate the rescue $Mbd3^{-/-}$*:Mbd3* NSC line (Figure 3.2.1G). For ease of reading, Mbd3b will be referred as *Mbd3* throughout the rest of the thesis, unless stated otherwise.

(figure on next page)

Figure 3.2.1 – Generation of *Mbd3*$^{-/-}$ Neural Stem Cells.

(A) Schematic representation of *Mbd3* loci in ESCs *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$. Exons are indicated as dark grey boxes, non-coding sequences are indicated as unfilled boxes, and light grey triangles represent loxP sites. **(B)** Western blot analysis of Mbd3 and α-Tubulin (αtub) protein levels in ESC lines with different *Mbd3* genotypes cultured in 2i/LIF conditions: *Mbd3*$^{+/+}$, *Mbd3*$^{fl/-}$, and *Mbd3*$^{-/-}$ (derived from *Mbd3*$^{fl/-}$ after Cre-mediated deletion). Mbd3 isoforms a, b and c are indicated. **(C)** Schematic representation of *Mbd3* loci in *Mbd3*$^{ex1fl/-}$ and *Mbd3*$^{\Delta ex1/-}$ ESCs. Exons are indicated as dark grey boxes, non-coding sequences as unfilled boxes and light grey triangles represent loxP sites. Exon 1 deletion creates a hypomorphic Mbd3 protein, with lower molecular weight. **(D)** Western blot analysis of Mbd3 and α-Tubulin (αtub) protein levels in ESC lines with different *Mbd3* genotypes cultured in 2i/LIF conditions: *Mbd3*$^{+/+}$, *Mbd3*$^{ex1fl/-}$ and *Mbd3*$^{\Delta ex1/-}$ (derived from *Mbd3*$^{ex1fl/-}$ after Cre-mediated deletion). **(E)** Schematic representation of the strategy used for generation of *Mbd3*$^{-/-}$ NSCs. *Mbd3*$^{fl/-}$ ESCs were differentiated in vitro in *Mbd3*$^{fl/-}$ NSCs. *Mbd3*$^{-/-}$ NSCs were clonally generated from the latter after Cre treatment. An *Mbd3*$^{-/-}$ rescue NSC line (*Mbd3*$^{-/-}$*:Mbd3*) was generated by stable transfection of an *Mbd3b* transgene. **(F)** ClustalW2 sequence alignment between Mbd3a, Mbd3b (cDNA used this study unless otherwise stated) and Mbd3c isoforms (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Phenylalanine 34 (F34) is indicated in red. **(G)** Western blot analysis of Mbd3 and αtub protein levels in the *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* (rescue) NSCs used for reprogramming experiments.



F

```
Mbd3a  MERKRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFDFRTGKML 70
Mbd3b  MERK--------------------------------SPSGKKFRSKPQLARYLGGSMDLSTFDFRTGKML 38
Mbd3c  MARIWFGG-------------------------WEISVPDRPPATIQALAKHLPGPSNPP---------- 35
       * *                                * ..:   :   **::* *. : .

Mbd3a  MNKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQPRQLFWE 130
Mbd3b  MNKMNKSRQRVRYDSSNQVKGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQPRQLFWE 98
Mbd3c  WTPVGAARCRVFSP-----QGKPDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQPRQLFWE 100
        . :. :* **        :**************************************************

Mbd3a  KKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIASALHTSTLPITGQLSAAVEKNPGVWLNT 210
Mbd3b  KKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIASALHTSTLPITGQLSAAVEKNPGVWLNT 178
Mbd3c  KKLSGLSAFDIAEELVRTMDLPKGLQGVGPGCTDETLLSAIASALHTSTLPITGQLSAAVEKNPGVWLNT 180
       ************************************************************************

Mbd3a  AQPLCKAFMVTDDDIRKQEELVQQVRKRLEEALMADMLAHVEELARDGEAPLDKACAEEEEEEEEEEPEPERV 285
Mbd3b  AQPLCKAFMVTDDDIRKQEELVQQVRKRLEEALMADMLAHVEELARDGEAPLDKACAEEEEEEEEEEPEPERV 253
Mbd3c  AQPLCKAFMVTDDDIRKQEELVQQVRKRLEEALMADMLAHVEELARDGEAPLDKACAEEEEEEEEEEPEPERV 245
       ************************************************************************
```

Morphologically, *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* rescue NSCs were indistinguishable (Figure 3.2.2A), expressing normal levels of three canonical NSC markers, *Olig2*, *Blbp*, and *Sox2* (Figure 3.2.2B). I observed that *Mbd3*$^{-/-}$ NSCs, which are robustly self-renewing, proliferate 1.5 fold slower than *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$*:Mbd3* NSCs, consistent with previous reports in *Mbd3*$^{-/-}$ ES cells (Kaji et al., 2006) (Figure 3.2.2C). In fact, it has previously been shown that the Mbd3/NuRD complex function is required for normal S phase progression in Human Burkitt's lymphoma cell lines (Sims and Wade, 2011).

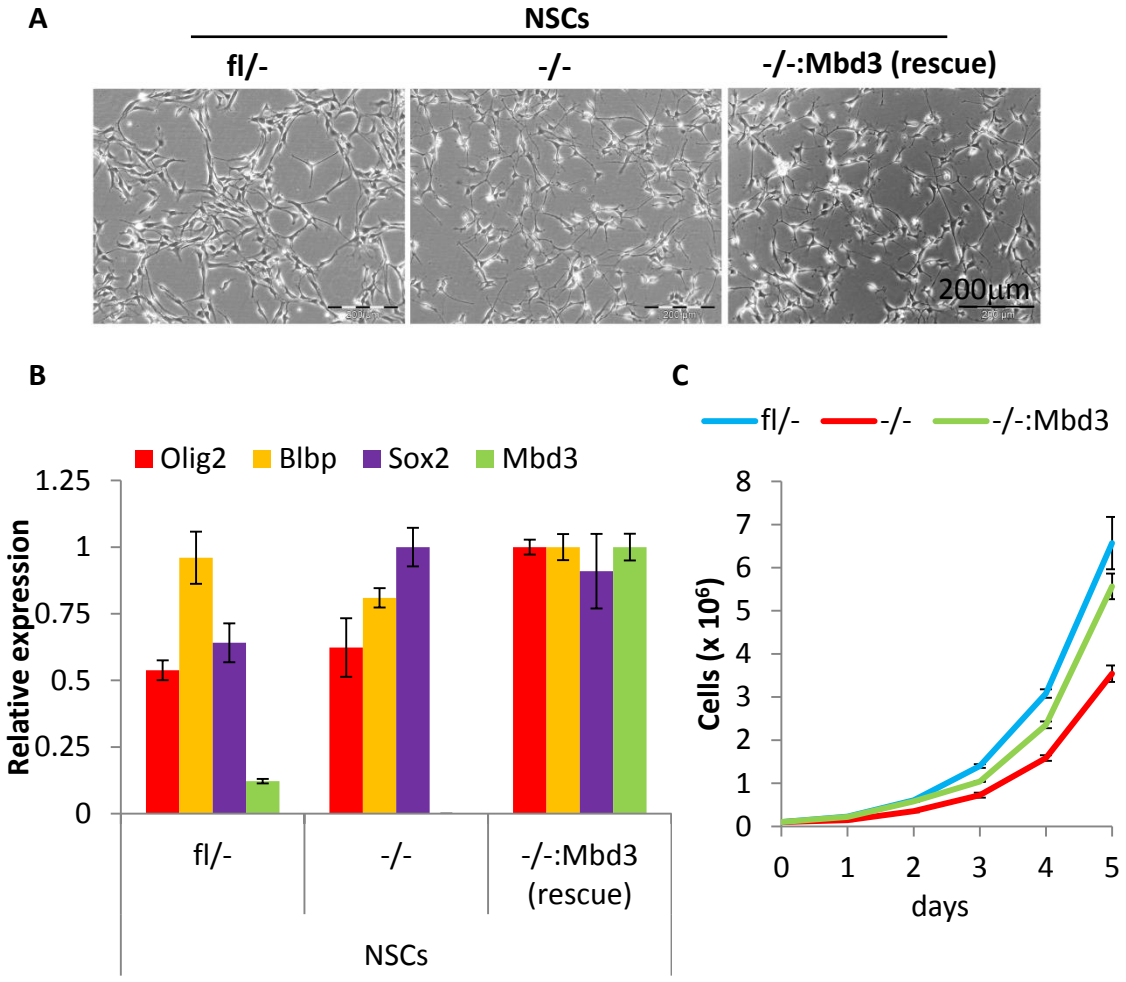
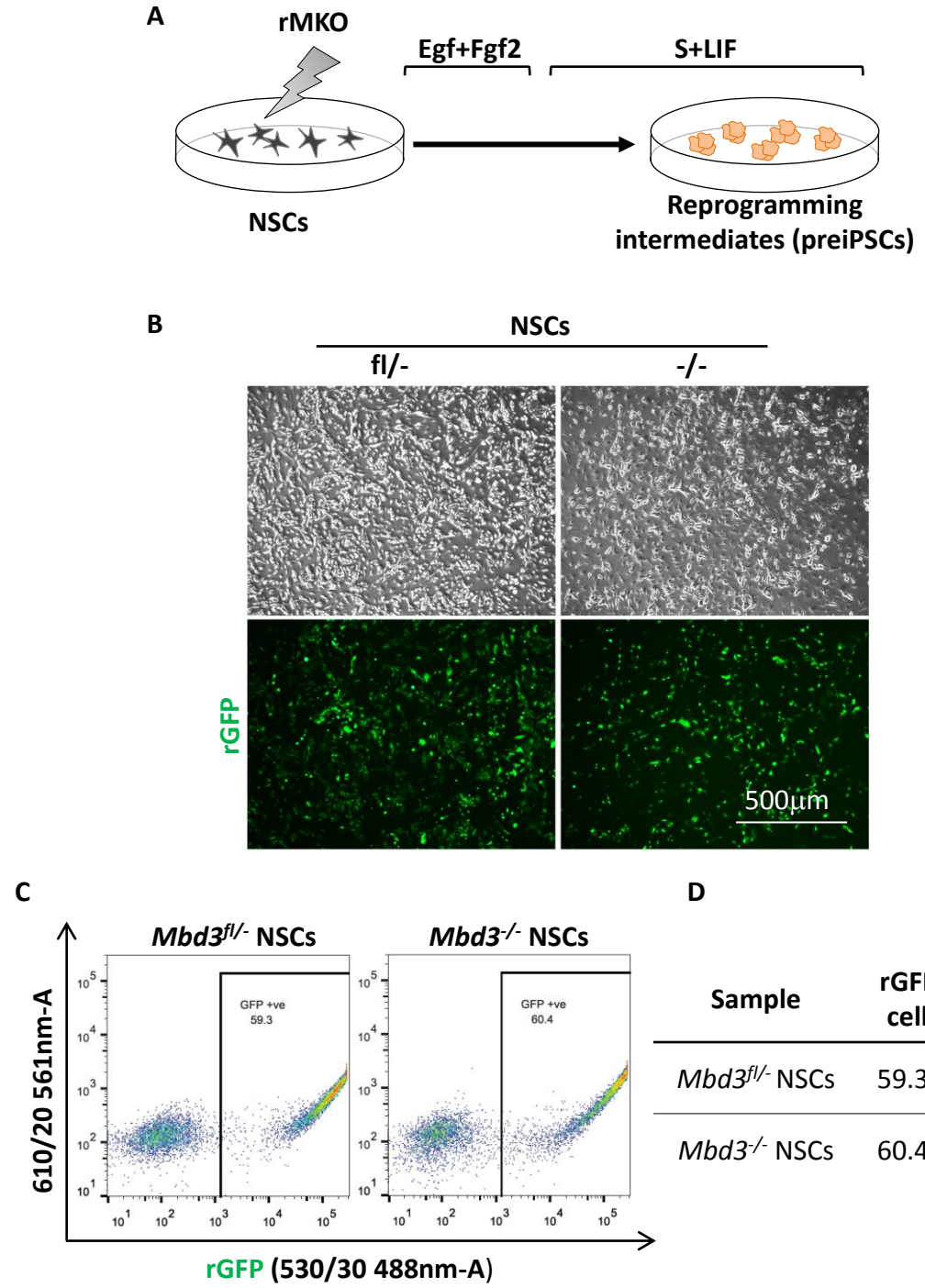


Figure 3.2.2 – Characterization of generated *Mbd3*$^{-/-}$ Neural Stem Cells.

(A) Phase images of the *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* (rescue) NSCs used for reprogramming experiments. **(B)** qRT-PCR analysis of NSC markers (*Sox2*, *Olig2*, *Blbp*) and *Mbd3* expression levels in NSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. **(C)** Cell proliferation analysis of *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and rescue NSC lines. The error bars indicate STDEV.

To initiate reprogramming, the generated NSC lines were transduced with retroviruses encoding cMyc, Klf4 and Oct4 (rMKO), kept in NSCs conditions (Egf+Fgf2) for 3 days and then switched to serum plus LIF (S+LIF) conditions (Figure 3.2.3A), which typically results in the formation of highly proliferative reprogramming intermediates, or preiPSCs (Silva et al., 2008). To assess the transduction efficiency after *Mbd3* deletion, $Mbd3^{fl/-}$ and $Mbd3^{-/-}$ NSCs were transduced with retroviruses encoding green fluorescent protein (rGFP; Figure 3.2.3B). Equal percentages (~60%) of GFP^{+} cells were observed 72h after transduction in both conditions (Figure 3.2.3C). $Mbd3^{-/-}$ NSCs showed higher mean expression of GFP (86944, compared to 78640 in the case of $Mbd3^{fl/-}$ NSCs) (Figure 3.2.3D), indicating that Mbd3 deletion does not affect transduction efficiency.



Sample	rGFP⁺ cells	rGFP mean
Mbd3$^{fl/-}$ NSCs	59.3%	78640
Mbd3$^{-/-}$ NSCs	60.4%	86944

Figure 3.2.3 – Transduction efficiency of *Mbd3*$^{-/-}$ NSCs.

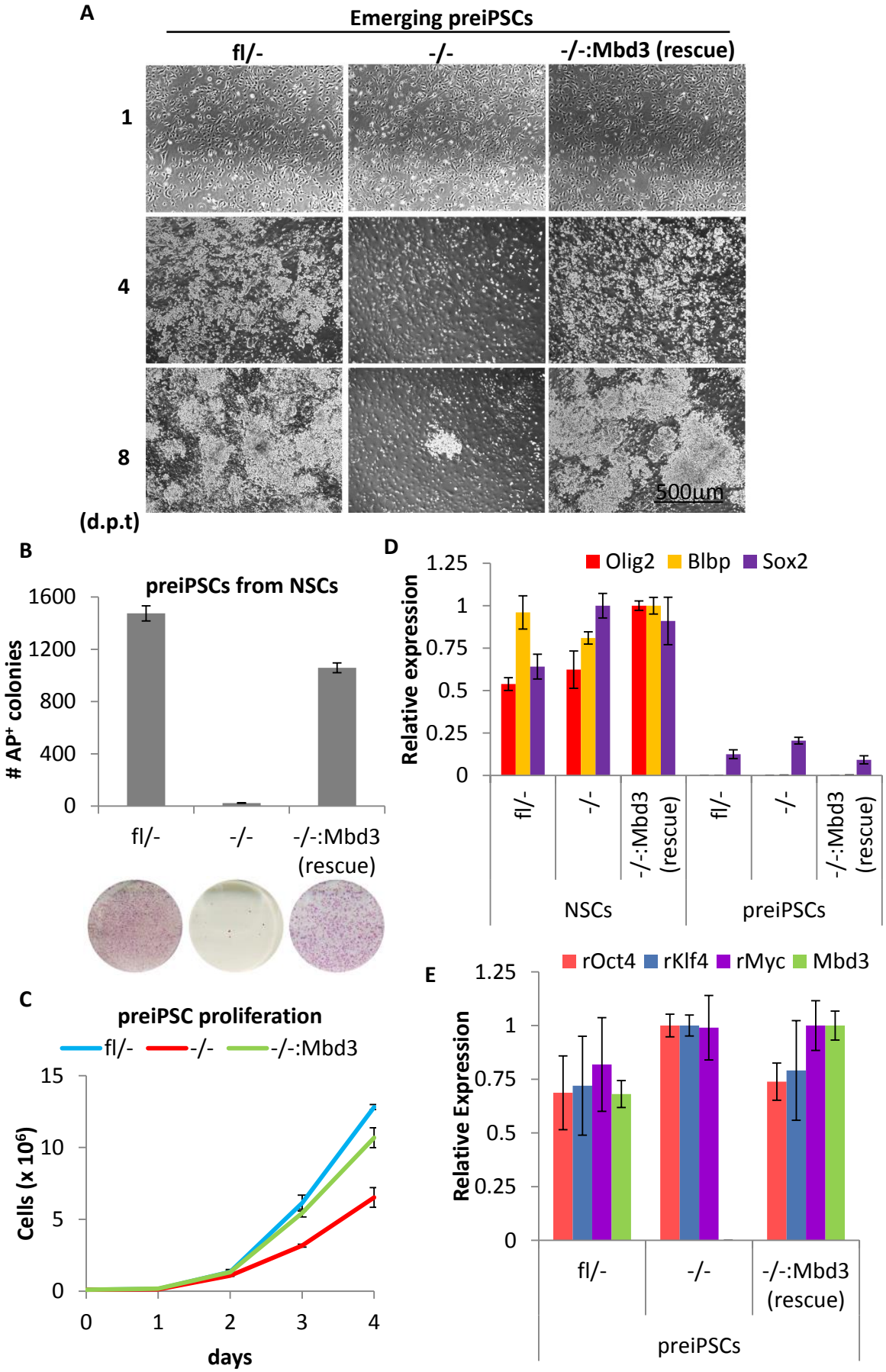
(A) Experimental design used to address the kinetics and efficiency of initiation of reprogramming in NSCs with different Mbd3 genotypes. NSCs were transduced with retroviruses (pMXs) encoding cMyc, Klf4, Oct4 (rMKO), maintained in Egf+Fgf2 medium for three days and then switched to S+LIF medium. **(B)** Phase and GFP images of *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ NSCs 72h after transduction with retroviruses encoding GFP (pMX-GFP). **(C)** Retroviral GFP (rGFP) expression 72h after transduction of *Mbd3*$^{fl/-}$ or *Mbd3*$^{-/-}$ NSCs with pMX-GFP, assessed by flow cytometry. GFP^{+} gates are shown. **(D)** Table indicates the percentages of rGFP^{+} cells and rGFP mean 72h after transduction of *Mbd3*$^{fl/-}$ or *Mbd3*$^{-/-}$ NSCs with pMX-GFP.

Strikingly, the kinetics of the preiPSC emergence after rMKO transduction was markedly delayed in the case of *Mbd3*$^{-/-}$ cells. While *Mbd3* expressing preiPSCs dominated the culture by day 4 post-transduction (d.p.t), *Mbd3*$^{-/-}$ preiPSCs emerged only by 7-8 d.p.t (Figure 3.2.4A). In addition, the number of emerging alkaline phosphatase positive (AP^{+}) *Mbd3*$^{-/-}$ preiPSC colonies was significantly reduced (65 fold less colonies) compared to parental and rescue cell lines (Figure 3.2.4B). Nevertheless, it was possible to establish and expand *Mbd3*$^{-/-}$ preiPSCs, although less efficiently and with delayed kinetics. *Mbd3*-null preiPSCs exhibit the same slower proliferation phenotype as the NSCs from which they were generated (they proliferate 1.5 fold slower than *Mbd3*$^{fl/-}$ and rescue preiPSCs) (Figure 3.2.4C). All generated preiPSC lines exhibited expected downregulation of the original NSC marker genes, such as *Olig2*, *Blbp* and *Sox2* (Figure 3.2.4D). *Mbd3*$^{-/-}$ preiPSCs were found to express slightly higher levels of retroviral transgenes (up to 25% higher), assessed 12 days after transduction, compared to control cells (Figure 3.2.4E), suggesting that the dosage of reprogramming factors is not the reason for the observed reduced efficiency of reprogramming initiation.

(figure on next page)

Figure 3.2.4 – The absence of Mbd3/NuRD decreases the efficiency and kinetics of initiation of reprogramming.

(A) Phase images of the reprogramming intermediates (preiPSCs) emerging from *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* (rescue) NSCs at different days post-transduction (d.p.t). **(B)** Efficiency of preiPSC colony formation per 2.5 x 10^5 NSCs assessed by alkaline phosphatase (AP) staining at day 9 post-transduction. Representative alkaline phosphatase (AP) stainings are shown. **(C)** Cell proliferation analysis of *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* preiPSCs lines. **(D)** qRT-PCR analysis of NSC markers (*Sox2*, *Olig2*, *Blbp*) and *Mbd3* expression levels in NSCs and corresponding preiPSCs. **(E)** qRT-PCR analysis of retroviral transgenes (rOct4, rKlf4 and rMyc) and *Mbd3* expression in the obtained preiPSCs maintained in S+LIF. Three independent NSCs transductions were carried out and gene expression was assessed 12 days after transduction. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.


A
Emerging preiPSCs
fl/-
-/-
-/-:Mbd3 (rescue)
1
4
8
(d.p.t)
500μm
B
preiPSCs from NSCs
AP+ colonies
1600
1200
800
400
0
fl/-
-/-
-/-:Mbd3 (rescue)
C
preiPSC proliferation
fl/-
-/-
-/-:Mbd3
Cells (x 10^6)
15
10
5
0
0
1
2
3
4
days
D
Olig2
Blbp
Sox2
Relative expression
1.25
1
0.75
0.5
0.25
0
fl/-
-/-
-/-:Mbd3 (rescue)
fl/-
-/-
-/-:Mbd3 (rescue)
NSCs
preiPSCs
E
rOct4
rKlf4
rMyc
Mbd3
Relative Expression
1.25
1
0.75
0.5
0.25
0
fl/-
-/-
-/-:Mbd3 (rescue)
preiPSCs

To further dissect the requirement of Mbd3 in the initiation of reprogramming, I analysed the effect of *Mbd3* deletion at different experimental time-points. To do this, I made use of a NSC line derived from E13.5 embryos, where both *Mbd3* alleles are flanked by loxP sites. These $Mbd3^{fl/fl}$ NSCs were stably transfected with Cre-ERt2, which enabled Cre-mediated excision of the floxed *Mbd3* alleles upon addition of tamoxifen (4-OHT; Figure 3.2.5A). Around 90% reduction in Mbd3 protein levels was observed 72h after 4-OHT treatment of $Mbd3^{fl/fl}$*:Cre-ERt2* NSCs (Figure 3.2.5B). As with $Mbd3^{-/-}$ clonal NSCs, slower proliferation (up to 2 fold) was observed in cells when Mbd3 was deleted (Figure 3.2.5C). 4-OHT treatment, with the resulting deletion of *Mbd3,* did not alter the gene expression signature typical of NSCs (Figure 3.2.5D). $Mbd3^{fl/fl}$ NSCs could be maintained in culture in the presence of 4-OHT (5 passages tested), excluding other effects (e.g. cytotoxicity) beyond Cre-mediated *Mbd3* excision. An $Mbd3^{ex1fl/ex1fl}$ NSC line was also derived from E13.5 embryos, which exhibited similar phenotypes upon Cre-mediated *Mbd3* exon 1 excision, resulting in the removal of all but a small amount of a truncated Mbd3 protein isoform (Aguilera et al., 2011; Kaji et al., 2006) (Mbd3c; Figure 3.2.5E).

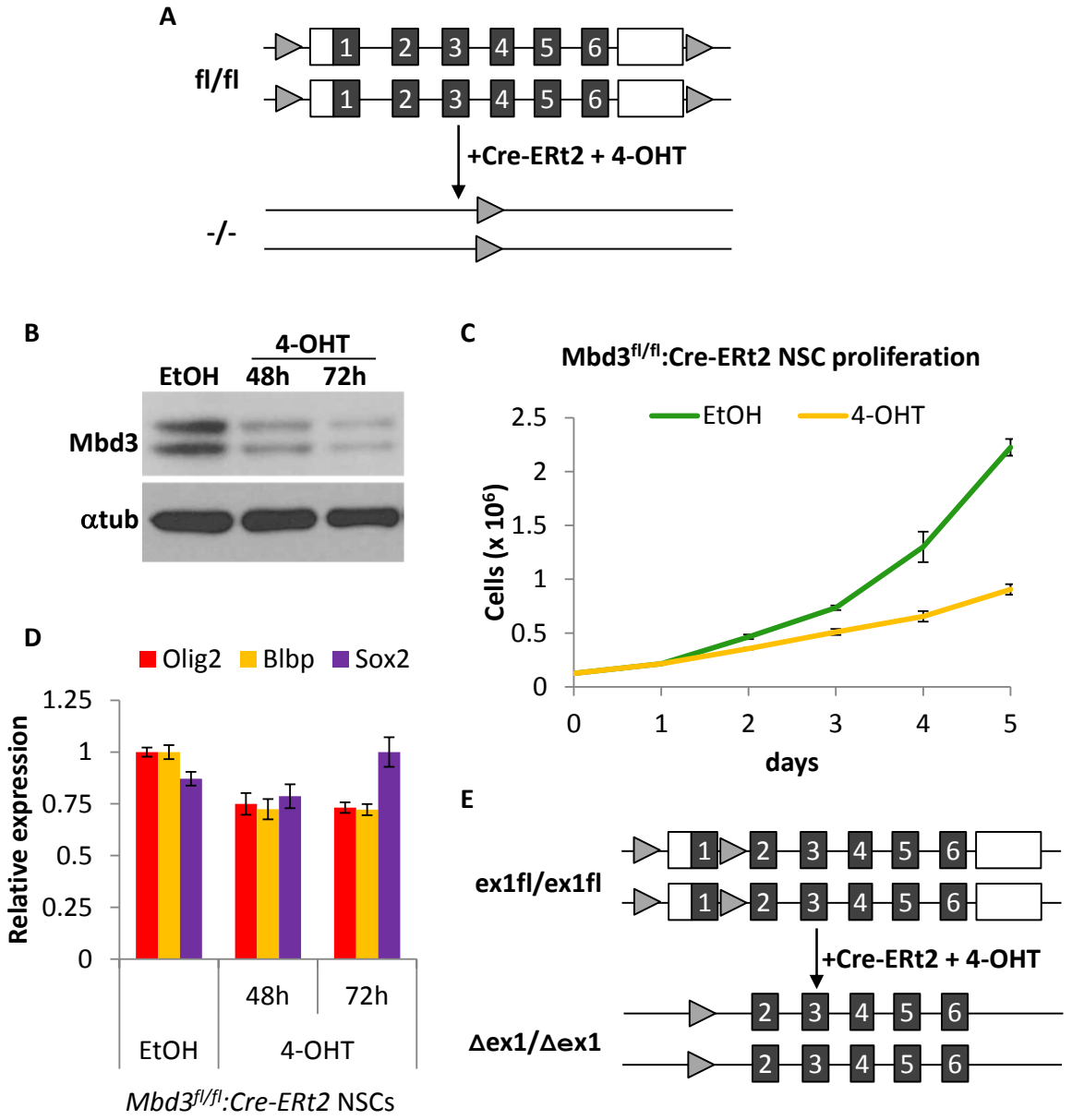


Figure 3.2.5 – Characterization of *Mbd3*$^{fl/fl}$ Neural Stem Cells

(A) Schematic representation of *Mbd3* loci in *Mbd3*$^{fl/fl}$ NSCs before and after the transfection with the pCAG-CreERt2 transgene and treatment with 4-OHT. Exons are indicated as dark grey boxes, non-coding sequences are indicated as unfilled boxes, and light grey triangles represent loxP sites. **(B)** Western blot analysis of Mbd3 and α-Tubulin (αtub) protein levels in the *Mbd3*$^{fl/fl}$*:Cre-ERt2* NSCs treated with tamoxifen (4-OHT) or ethanol (EtOH). **(C)** Cell proliferation analysis of NSC *Mbd3*$^{fl/fl}$*:Cre-ERt2*, in the presence of 4-OHT or EtOH. **(D)** qRT-PCR analysis of NSC markers expression in *Mbd3*$^{fl/fl}$*:Cre-ERt2* NSCs treated with 4-OHT or EtOH. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV. **(E)** Schematic representation of *Mbd3* loci in *Mbd3*$^{ex1fl/ex1fl}$ NSCs before and after transfection with pCAG-Cre-ERt2 transgene and treated with 4-OHT. Exons are indicated as dark grey boxes, non-coding sequences are indicated as unfilled boxes, and light grey triangles represent loxP sites.

To study the impact of Cre-mediated *Mbd3* deletion at different time-points, *Mbd3*$^{fl/fl}$*:Cre-ERt2* or *Mbd3*$^{ex1fl/ex1fl}$*:Cre-ERt2* NSCs were transduced with rMKO and 4-OHT was added to the media for different periods. 4-OHT was applied two days before transduction and kept throughout in the condition #1, or just applied six days after transduction in the condition #6 (Figure 3.2.6). Since 4-OHT is ethanol-based (EtOH), EtOH-only was used as a control for the experiment. As a result, I observed that the earlier *Mbd3* was removed, the fewer preiPSC colonies were formed. An average of 40 colonies could be scored when 4-OHT was applied at the time of transduction, compared to an average of 1200, when 4-OHT was not applied. Similar results were obtained upon the conditional deletion of the *Mbd3* exon1. By timely removal of Mbd3 during the initiation of reprogramming process, I was able to recapitulate the reprogramming phenotype observed in the KO cells.

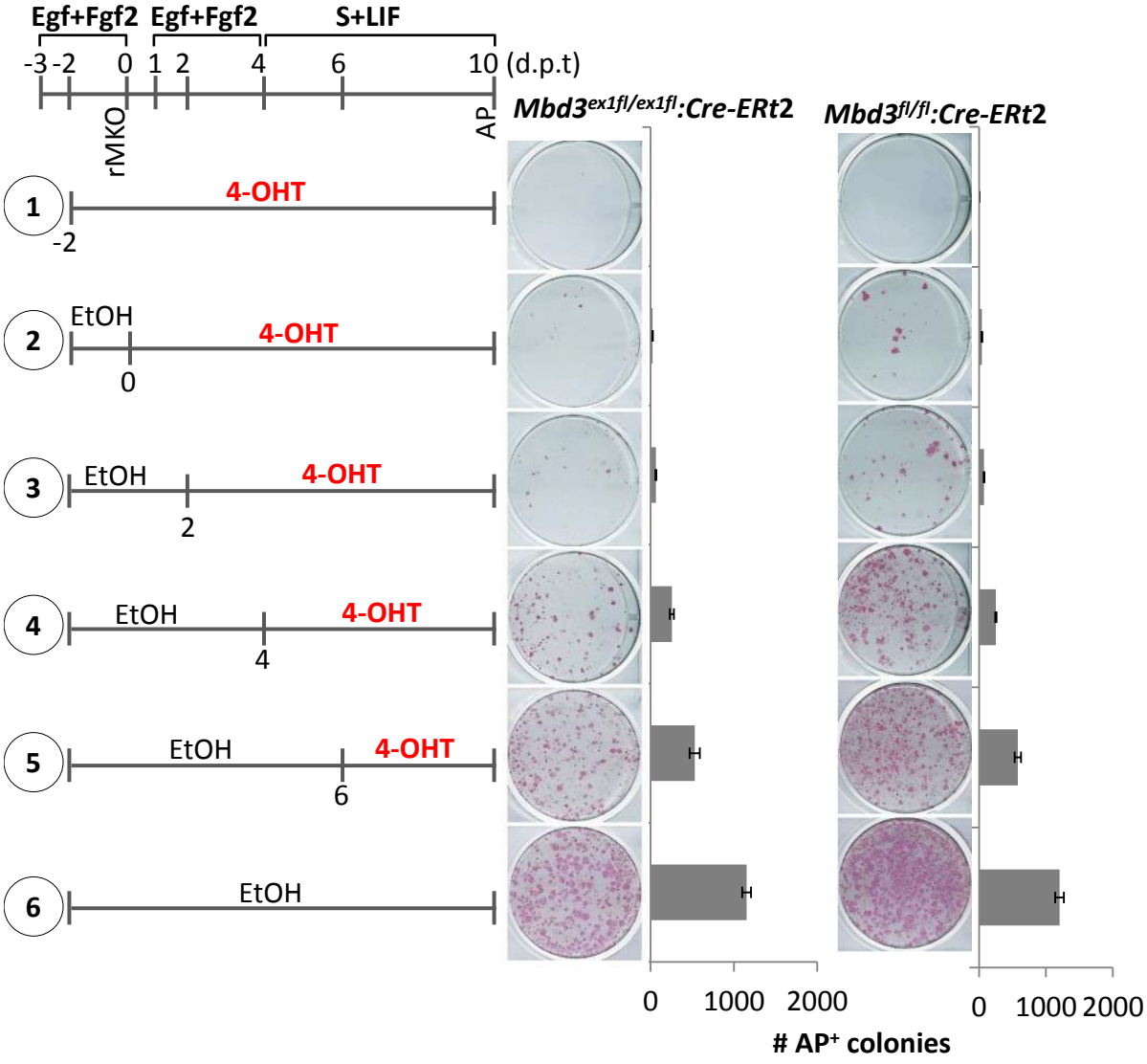


Figure 3.2.6 – Time-course of Mbd3 requirement during the initiation of reprogramming.

Mbd3$^{fl/fl}$ and *Mbd3*$^{ex1fl/ex1fl}$ NSCs were stably transfected with pCAG-CreERt2 transgene, transduced with retroviral transgenes MKO and treated with 4-OHT at indicated time points to induce Cre-mediated deletion of the floxed alleles during reprogramming. Ethanol (EtOH) was used as a control. The encircled numbers correspond to different conditions. PreiPSC colony formation was assessed by AP staining at day 10 post-transduction and is presented as the number of colonies per 7.5 x 10^4 NSCs. The error bars indicate STDEV.

Independent *Mbd3*$^{-/-}$ cells exhibit the same reprogramming defect when compared to either *Mbd3*$^{fl/-}$, *Mdb3*$^{fl/fl}$ or *Mbd3*$^{-/-}$ plus Mbd3 (rescue) NSCs. To gain insights into the mechanisms by which the NuRD complex acts during reprogramming, I investigated the described relationship between the NuRD complex and the Ten-eleven translocation (Tet) family of enzymes (Yildirim et al., 2011), which catalyse the oxidation 5-methylcytosine (5-mC) in 5-hydroxymethylcytosine (5-hmC). It was recently reported that Mbd3 KD results in

lower 5-hmC levels, leading to reduced expression of 5-hmC marked genes in ESCs (Yildirim et al., 2011). I observed that the *Mbd3*$^{-/-}$ NSCs have 2 fold higher expression of *Tet1/2* compared to *Mbd3*$^{fl/-}$ or the rescue line, a feature that is lost after the transition to a preiPSC state (Figure 3.2.7A). Those apparently higher levels of *Tet1/2* are not translated into the higher bulk 5-hmC levels. *Mbd3*$^{-/-}$ NSCs exhibit lower levels of 5-hmC than the parental or rescue cell line (Figure 3.2.7B), a feature that is maintained in their corresponding preiPSCs (Figure 3.2.7C). To test if increased levels of Tet1/2 could rescue the *Mbd3*-null reprogramming phenotype, *Mbd3*$^{-/-}$ NSCs were stably transfected with Tet1 or/and Tet2. By scoring the amount of AP^{+} preiPSC colonies formed 9 days after the transduction with rMKO, I observed that the initiation of reprogramming was not rescued by the overexpression of either *Tet1* or *Tet2*, alone or in combination (Figure 3.2.7D). This indicates that although the 5-hmC levels are lower in *Mbd3*$^{-/-}$ NSCs, overexpressing of Tet enzymes is not sufficient to alleviate their reprogramming block.

Taken together, these results demonstrate that lack of a functional NuRD complex strongly impairs the initiation of reprogramming from NSCs.

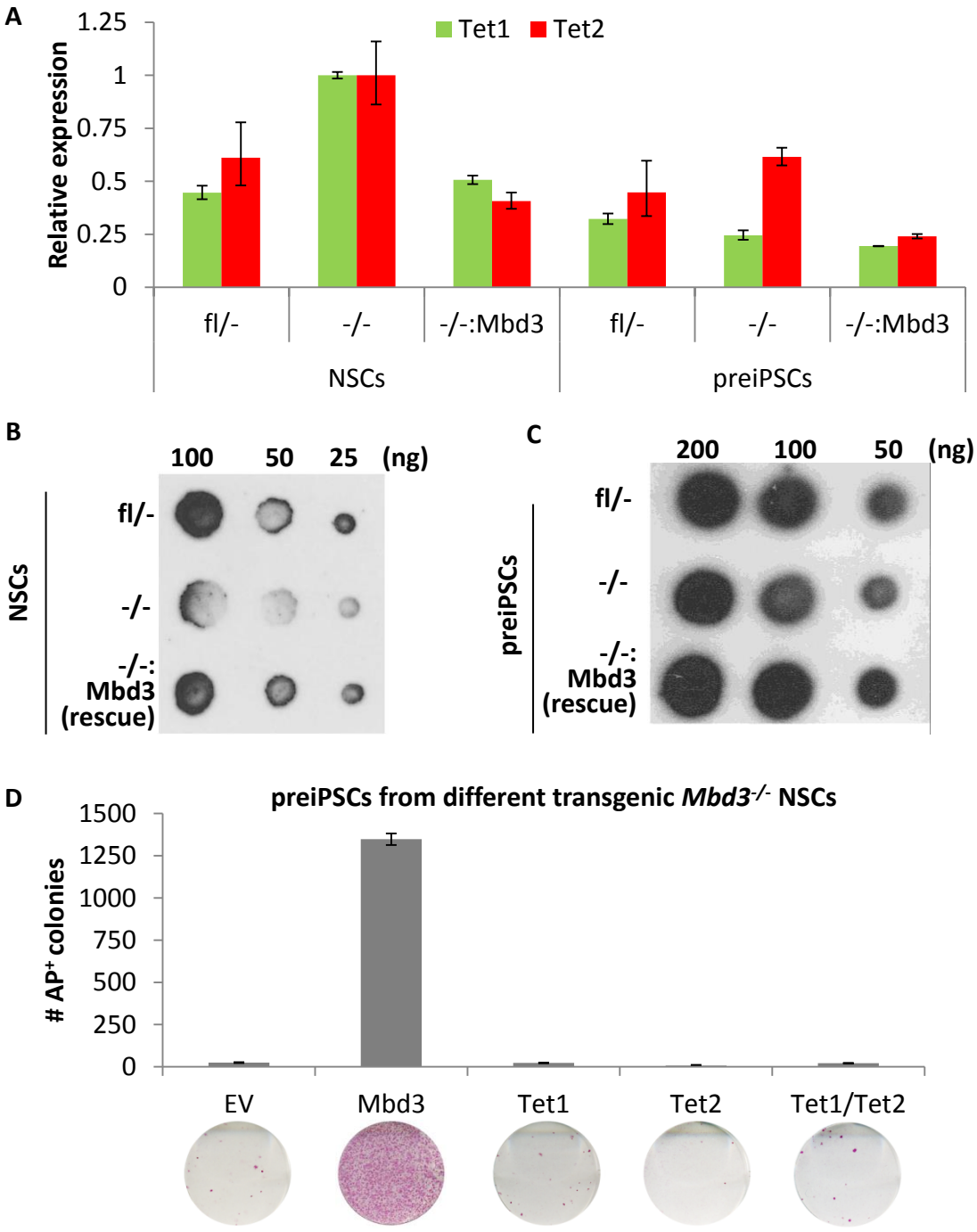


Figure 3.2.7 – Lack of NuRD complex leads to loss of 5-hydroxymethylation.

(A) qRT-PCR analysis of Tet1/2 expression levels in *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* NSCs and corresponding preiPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV. **(B-C)** DNA dot blot analysis of 5-hydroxymethylation (5-hmC) bulk levels in (B) NSCs and (C) corresponding preiPSCs. **(D)** Quantification of AP^{+} preiPSC colonies formed 9 days post-transduction of *Mbd3*$^{-/-}$ NSCs transfected with the indicated transgenes with MKO and culture in S+LIF conditions. Colony number is per 2.5 x 10^{5} NSCs. Representative alkaline phosphatase (AP) staining are shown. The error bars indicate STDEV.

3.3 – Discussion

3.3.1 – Mbd3 facilitates the initiation of reprogramming from neural stem cells

In this chapter I analysed the role of the NuRD complex during transcription-factor mediated reprogramming of NSCs to naïve pluripotency. I demonstrated that the lack of a functional NuRD complex not only strongly impairs the initiation of reprogramming resulting in up to 65 fold less AP^+ colonies, but also reduces the kinetics of this process, phenotypes that are not due to a different dosage of reprogramming factors between the cell lines used (Figure 3.2.4). I also found that depletion of Mbd3/NuRD leads to decreased global 5-hmC levels in *Mbd3*$^{-/-}$ NSCs and preiPSCs compared to *Mbd3*$^{fl/-}$ and rescue cell lines (Figures 3.2.7B-C). This is in agreement with a recent report where it was shown that *Mbd3* KD results in reduction of bulk levels of 5-hmC in ESCs, and that Tet1 KD impairs Mbd3 recruitment to target genes (Yildirim et al., 2011). This shows that Mbd3 is both dependent upon 5-hmC for DNA binding and is necessary for normal levels of 5-hmC within the genome. How this relationship between NuRD and Tet1 could occur is not understood, but it is possible that Mbd3 could bind to a 5-hmC rich region and recruit Tet enzymes to oxidate adjacent methylcytosines, or Mbd3 could bind to 5-hmC *loci* and protect them from further steps in a demethylation pathway. Both these mechanisms conjecture that Mbd3 can bind to hydroxymethylated DNA, although Mbd3 is not capable of binding methylated DNA (Hendrich and Bird, 1998; Zhang et al., 1999). Comparison between the methyl binding domain sequences of Mbd3 and remaining methyl binding domain family members revealed that one of the few differences is a substitution of tyrosine (Y) for phenylalanine (F) residue (Y34/F34 in Mbd/Mbd3) (Figure 3.2.1F) (Yildirim et al., 2011). Since the structural difference between phenylalanine and tyrosine is the absence of a hydroxyl group in the former, the additional hydroxyl group in 5-hmC relative to 5-mC may allow a direct binding of Mbd3 to hydroxymethylated DNA in a manner structurally analogous to the binding of remaining Mbd members to 5mC. Although I also observed a correlation between the lack of the NuRD complex and lower global 5-hmC levels, this observation does not seem to be due to a direct binding of Mbd3 to 5-hmC through the methyl binding domain. I observed that the levels of 5-hmC in *Mbd3*$^{-/-}$ NSCs and preiPSCs could be rescued by the expression of Mbd3b, an Mbd3 isoform that lacks part of the annotated methyl binding domain, including the phenylalanine 34 (Y34) (Figure 3.2.1F and Figures 3.2.7B-C). Irrespectively of the mechanism through which NuRD and Tet1 interact, Mbd3 depletion results in lower global hydroxymethylation. This epigenetic mark has been previously associated with primed expression of key pluripotency genes before reprogramming to naïve pluripotency (Costa et al., 2013). Taking all these data into account, I hypothesize that the NuRD complex might act

as a mediator of 5-hmC's effects on gene expression, enabling gene expression priming, acting as a facilitator of the induction of pluripotency. This would explain, at least in part, the observed Mbd3 deletion phenotype, since key genes might fail to become hydroxymethylated, repressing the specific gene activation signature that is required for efficient reprogramming.

3.3.2 – Conclusions

In this chapter, I demonstrated that disruption of the NuRD complex strongly impairs the initiation of reprogramming from NSCs, both in rate and extent. I also show that the earlier *Mbd3* is deleted, the lower the number of preiPSC colonies formed. Although further molecular mechanistic information about the role of the NuRD complex during reprogramming is necessary, our data indicates that the hydroxymethylation of chromatin may play a part in NuRD-mediated genome-wide reprogramming.

CHAPTER 4 - Dissection of the requirement of Mbd3/ NuRD complex for iPSC generation

4.1 – Introduction

4.1.1 – NuRD complex in the establishment of naïve pluripotency

In the previous chapter I analysed the role of the NuRD complex in the initiation of reprogramming using retroviral delivery of transcription factors. In this chapter I will assess the requirement for Mbd3 at the later stages of reprogramming, as well as its overall role in reprogramming using different cell systems.

PreiPSCs are generated after the transduction of somatic cells and their culture in serum plus LIF containing medium. PreiPSCs can undergo full conversion to naïve pluripotency by their culture in serum-free medium containing LIF and inhibitors of both mitogen-activated protein kinase (Mek/Erk signalling – PD0325901) and glycogen synthase kinase-3 signalling (GSK3β - CHIR99021) - 2i/LIF media (Silva et al., 2008; Ying et al., 2008). This phase is termed the establishment phase of reprogramming. During this phase retroviral transgenes become silenced and the endogenous naïve pluripotent transcriptional program gets fully reactivated (Silva et al., 2008) (Figure 4.1.1). The generated iPSCs are now capable of re-entering embryonic development, and contributing to all the tissues in the generated chimeras, including the germline (Okita et al., 2007; Silva et al., 2008; Silva et al., 2009).

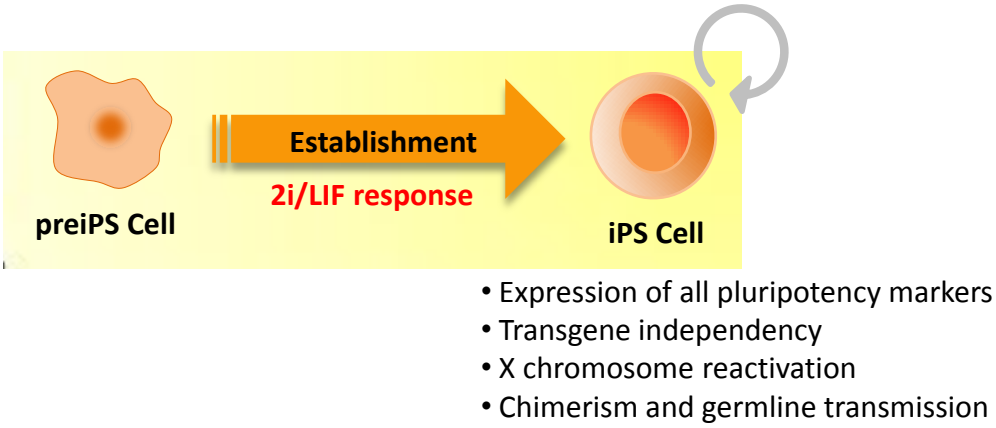


Figure 4.1.1 – Establishment of reprogramming.

Schematic representation of the experimental design used to assess the requirement of the NuRD complex during the establishment of reprogramming. Upon transduction of somatic cells with the reprogramming factors Oct4, Klf4 and c-Myc (and Sox2) and culture in serum-containing media (S+LIF), preiPSCs are generated, being characterized for their exogenous retroviral transgene dependency. Exposure to serum-free medium containing LIF and inhibitors of both Mek/Erk and GSK3β (2i/LIF) leads to the establishment of iPSCs, which are capable to re-enter embryo development and contribute that all germ layers (Silva et al., 2008).

Other cell systems can be used to study nuclear reprogramming, such as the reprogramming to naïve pluripotency of primed pluripotent stem cells, termed epiblast-derived stem cells (EpiSCs) (Brons et al., 2007; Tesar et al., 2007). As discussed in section 1.1.5, EpiSCs are isolated from the post-implantation epiblast (E5.5) and constitute a primed pluripotent state that fail to re-enter development when injected into a blastocyst (Nichols and Smith, 2009). EpiSCs can be obtained by the culture of the post-implantation Epiblast in medium containing Fgf2 and Actinin A (Brons et al., 2007; Tesar et al., 2007), and also obtained by the direct differentiation of ESCs, by culture of these in Fgf2 and Activin A (Guo et al., 2009). Conversely, they can be reprogrammed to naïve pluripotency by combining the overexpression of at least one transcription factor, such as Klf4, Klf2 or Nanog, with the use of serum-free 2i/LIF medium which not only promotes the reprogramming of EpiSCs but also blocks their self-renewal (Guo et al., 2009; Silva et al., 2009). Forced activation of the LIF signalling pathway in EpiSCs is also capable of inducing their reprogramming to naïve pluripotency (Stuart et al., 2014; van Oosten et al., 2012; Yang et al., 2010). The resulting iPSCs are able to be injected into the blastocyst and to re-enter embryo development (Guo et al., 2009; Silva et al., 2009). A complete comparison of naïve pluripotency (ESCs/ iPSCs) and primed pluripotency (EpiSCs) can be found in Figure 1.1.3.

A third system used in this chapter to address the role of the NuRD complex during induced pluripotency involves the use of piggyBac (PB) transposon/ transposase vectors to deliver the required transcription factors into mouse embryonic fibroblasts (MEFs) (Kaji et al., 2009; Woltjen et al., 2009). Taking into account that PB vectors remain active throughout the reprogramming process, including in the generated iPSCs, this system can allow the study of somatic cell reprogramming as a whole (Kaji et al., 2009; O'Malley et al., 2009; Woltjen et al., 2009).

4.1.2 – Aim of the chapter

In the previous chapter I showed that the lack of a functional NuRD complex strongly impairs the efficiency and kinetics of the initiation of reprogramming from NSCs. In this chapter I aimed to address whether NuRD is required for the later stages of reprogramming, and whether the deletion of *Mbd3* at later stages of the reprogramming process mimics its removal at earlier stages. I also aimed to assess the role of the NuRD complex in the reprogramming of other cells types, namely EpiSCs and MEFs.

4.2 – Results

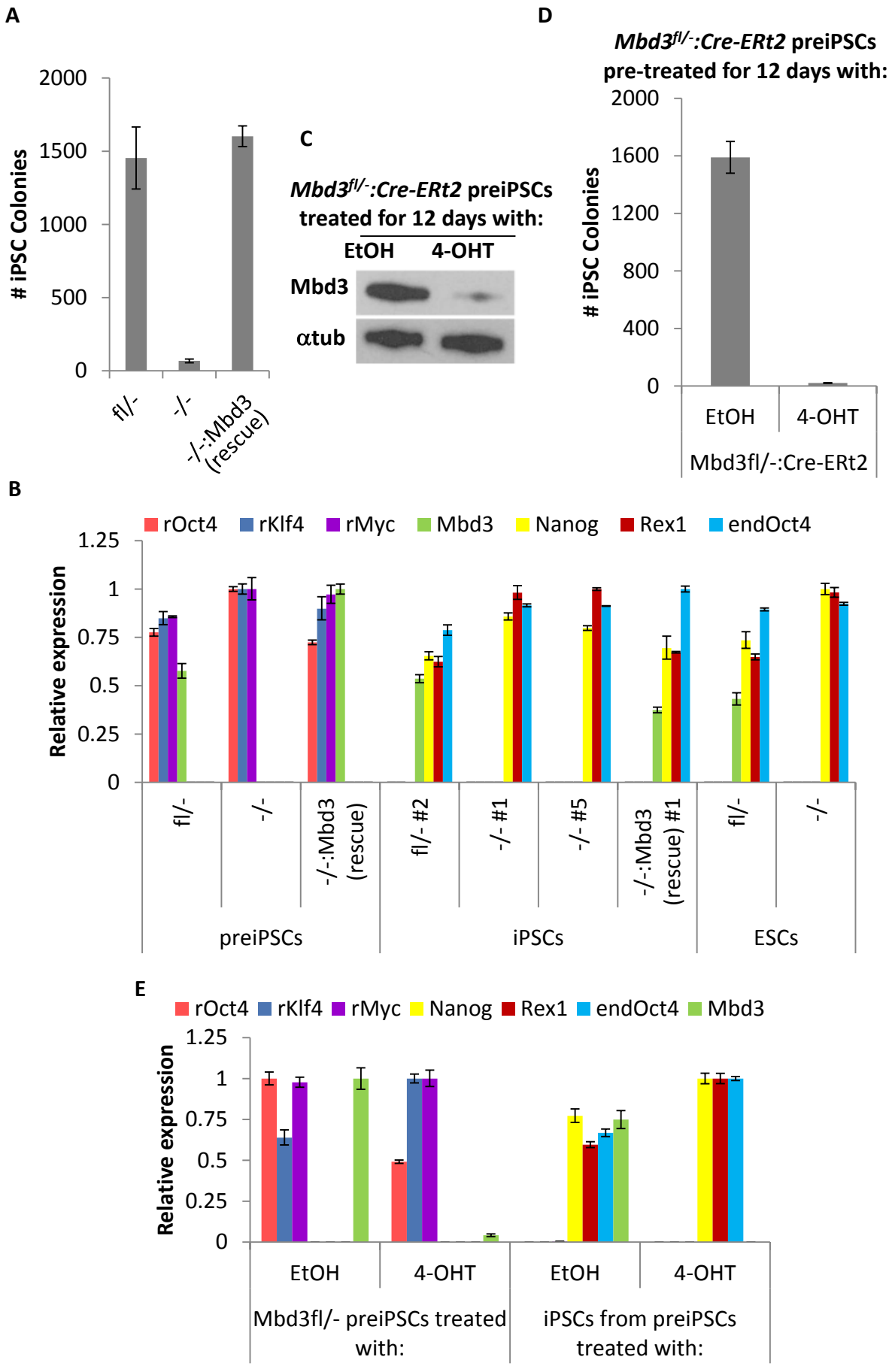
4.2.1 – Mbd3/NuRD complex requirement to the establishment of pluripotency

To induce the completion of the reprogramming process, *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and rescue *Mbd3*$^{-/-}$*:Mbd3* preiPSCs obtained after MKO transduction (Figure 3.2.4) were cultured in serum-free 2i/LIF medium, and the resulting iPSC colonies were scored 12 days afterwards (Figure 4.2.1A). I observed that the efficiency of the conversion to naïve pluripotency of *Mbd3*$^{-/-}$ preiPSCs is strongly reduced (20 fold) compared to *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$*:Mbd3* preiPSCs. The obtained *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* iPSCs could be clonally expanded in 2i/LIF conditions, exhibiting the expected reactivation of the pluripotency-associated transcriptional program and the silencing of retroviral reprogramming promoters (Figure 4.2.1B). From this experiment it is unclear whether the lower efficiency of iPSC generation from *Mbd3*$^{-/-}$ preiPSCs was due to the requirement for the NuRD complex in those reprogramming intermediates or if they were epigenetically disturbed due to the lack of the NuRD complex during their generation. To answer this question, *Mbd3*$^{fl/-}$ preiPSCs were stably transfected with Cre-ERt2 and treated for 12 days (3-4 passages) with 4-OHT in EtOH, which results in the depletion of Mbd3 protein levels, or EtOH only (Figure 4.2.1C). I plated the same number of *Mbd3*$^{fl/-}$ preiPSCs in 2i/LIF medium and scored the resulting iPSCs colonies 12 days later (Figure 4.2.1D). I observed that *Mbd3* deletion in this established preiPSC line, prior to 2i/LIF medium switch, decreases reprogramming efficiency by up to 80 fold, indicating that *Mbd3* affects both the initiation and the intermediate stages of reprogramming. iPSCs generated from both *Mbd3*$^{fl/-}$ preiPSC pre-treated with 4-OHT or EtOH showed a strong down-regulation of the retroviral transgenes and an up-regulation of pluripotency markers (Figure 4.2.1E).

(figure on next page)

Figure 4.2.1 – Conversion to naïve pluripotency of *Mbd3*$^{-/-}$ preiPSCs is strongly impaired.

(A) Quantification of iPSCs colonies generated from *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* (rescue) preiPSCs after 2i/LIF culture for 12 days. Colony number is per 1.0 x 10^5 preiPSCs. **(B)** qRT-PCR analysis of retroviral transgenes, *Mbd3*, and pluripotency-associated in preiPSCs and corresponding derived iPSCs. **(C)** Western blot analysis of Mbd3 and α-Tubulin (αtub) protein levels in the *Mbd3*$^{fl/-}$*:Cre-ERt2* preiPSC treated with 4-OHT or EtOH for 12 days. **(D)** Quantification of iPSCs colonies generated from preiPSCs *Mbd3*$^{fl/-}$ preiPSCs stably transformed with pCAG-Cre-ERt2 and treated with 4-OHT or EtOH for 12 days while cultured in S+LIF. Medium was switched from S+LIF to 2i/LIF 24h after plating. No 4-OHT or EtOH was added during 2i/LIF culture. Colony number is per 1.0 x 10^5 preiPSCs. **(E)** qRT-PCR analysis in *Mbd3*$^{fl/-}$*:Cre-ERt2* preiPSCs treated with 4-OHT (or EtOH control) and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.


A
iPSC Colonies
2000
1500
1000
500
0
fl/-
-/-
-/-:Mbd3 (rescue)
C
Mbd3fl/-:Cre-ERt2 preiPSCs treated for 12 days with:
EtOH
4-OHT
Mbd3
αtub
D
Mbd3fl/-:Cre-ERt2 preiPSCs pre-treated for 12 days with:
iPSC Colonies
2000
1600
1200
800
400
0
EtOH
4-OHT
Mbd3fl/-:Cre-ERt2
B
rOct4
rKlf4
rMyc
Mbd3
Nanog
Rex1
endOct4
Relative expression
1.25
1
0.75
0.5
0.25
0
fl/-
-/-
-/-:Mbd3 (rescue)
fl/- #2
-/- #1
-/- #5
-/-:Mbd3 (rescue) #1
fl/-
-/-
preiPSCs
iPSCs
ESCs
E
rOct4
rKlf4
rMyc
Nanog
Rex1
endOct4
Mbd3
Relative expression
1.25
1
0.75
0.5
0.25
0
EtOH
4-OHT
EtOH
4-OHT
Mbd3fl/- preiPSCs treated with:
iPSCs from preiPSCs treated with:

I previously observed that the earlier *Mbd3* was removed during the initiation of reprogramming, the fewer preiPSC colonies were formed (Figure 3.2.6). To determine if a similar relationship was also observed during the establishment phase of reprogramming, I performed a time-course experiment to define the time window for *Mbd3* requirement in NSC reprogramming. For that, *Mbd3*$^{fl/fl}$*:Cre-ERt2* NSCs were transduced with rMKO and also with retroviruses encoding GFP and treated with 4-OHT at different experimental time points (Figure 4.2.2). Media was changed to S+LIF 4 days after transduction and, 4 days later, to 2i/LIF. The number of iPSC colonies exhibiting silencing of retroviral GFP expression was assessed 12 days after 2i/LIF medium switch.

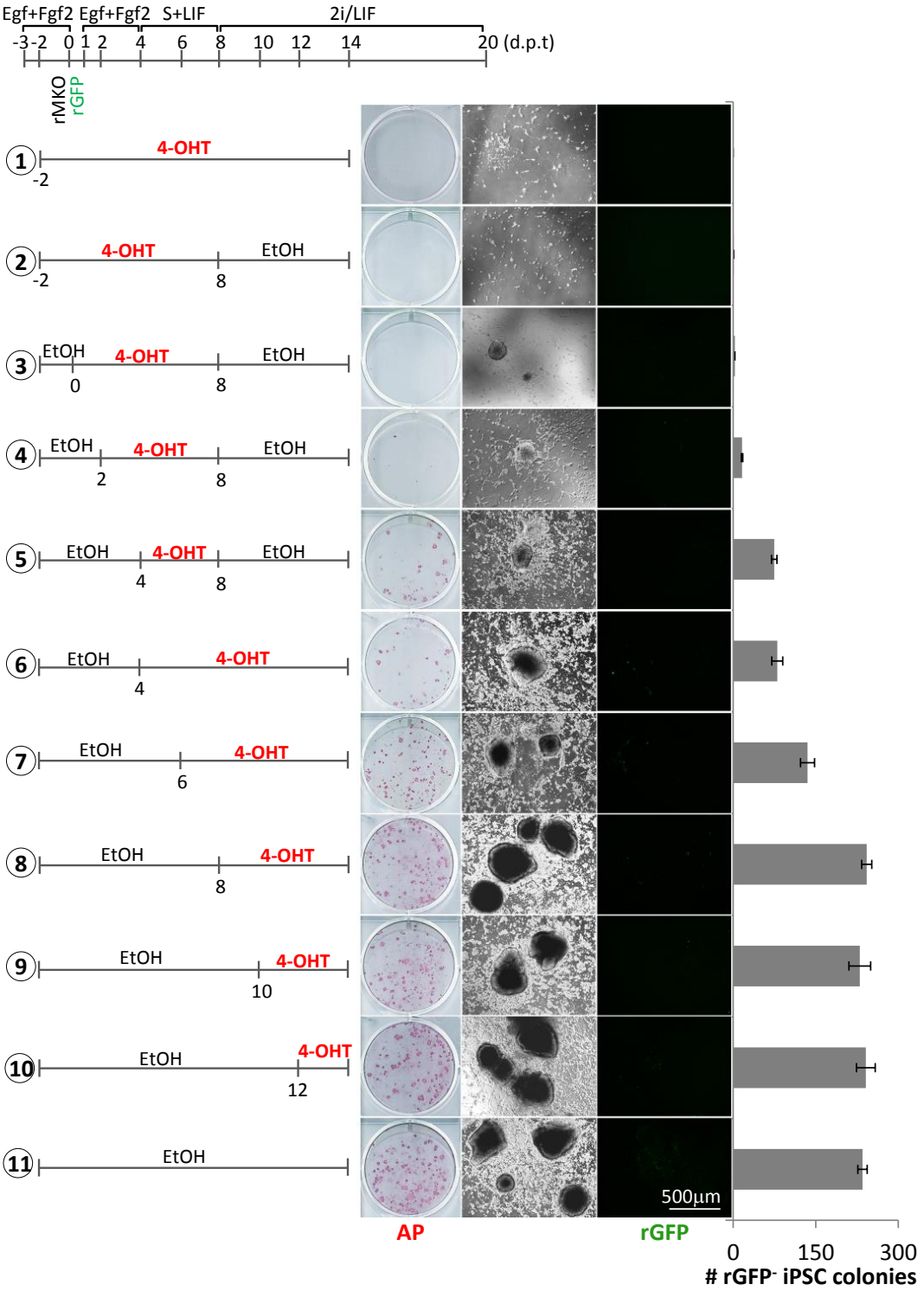


Figure 4.2.2 – Mbd3 is specifically required for the initiation of pluripotency.

Mbd3$^{fl/fl}$*:Cre-ERt2* NSCs were transduced with rMKO and rGFP, maintained in Egf+Fgf2 medium for three days, switched to S+LIF for more four days to allow preiPSC emergence and then switched to 2i/LIF conditions to induce iPSC formation. 4-OHT was added at different time-points (before or after preiPSC emergence) to induce Mbd3-floxed alleles excision. The encircled numbers correspond to different conditions. At day 20 after transfection, GFP$^-$ iPSC colonies were counted and subsequently stained for AP. The number of colonies is presented per 7.5 x 10^4 NSCs. The error bars indicate STDEV.

Although retroviral GFP silencing is not a mainstream pluripotency readout, its combination with stringent culture conditions, serum-free 2i/LIF medium, makes it a valuable marker to study reprogramming. In this experimental setup, only cells undergoing reprogramming and already reprogrammed ones survive. This explains why no GFP is visible in Figure 4.2.2, despite flow cytometry analysis showing that more than 80% of emerging preiPSCs 8 days after transduction are $rGFP^+$ (Figure 4.2.3A-C). I observed that the number of $rGFP^-$ iPSC colonies formed was proportional to the amount of time cells had expressed *Mbd3* during the initiation phase of reprogramming (prior to 2i/LIF culture). The control condition had an average of 230 colonies, a number that dropped down to 1 colony when 4-OHT was applied at the time of transduction. I observed neither a reduction nor a gain of reprogramming efficiency when *Mbd3* was deleted at the 2i/LIF stage. Regardless of the stage of *Mbd3* deletion, the generated iPSCs presented a pluripotency-associated transcriptional signature similar to wild type ESCs (Figure 4.2.3D). Analyses were performed after one passage on a pool of colonies from the entire well, preventing further culture enrichment for naïve pluripotent cells. Since transduction efficiency is maintained upon *Mbd3* deletion (Figure 3.2.3B-D), and similar retroviral expression of reprogramming transgenes is observed between cells expressing or not *Mbd3* (Figure 3.2.4E), I can conclude that the indicated results are a direct measure of the impact of *Mbd3* excision on reprogramming.

Together, these results indicate that *Mbd3* specifically affects the initiation and intermediate stages, preiPSC formation, of reprogramming but not the 2i/LIF-mediated establishment of pluripotency.

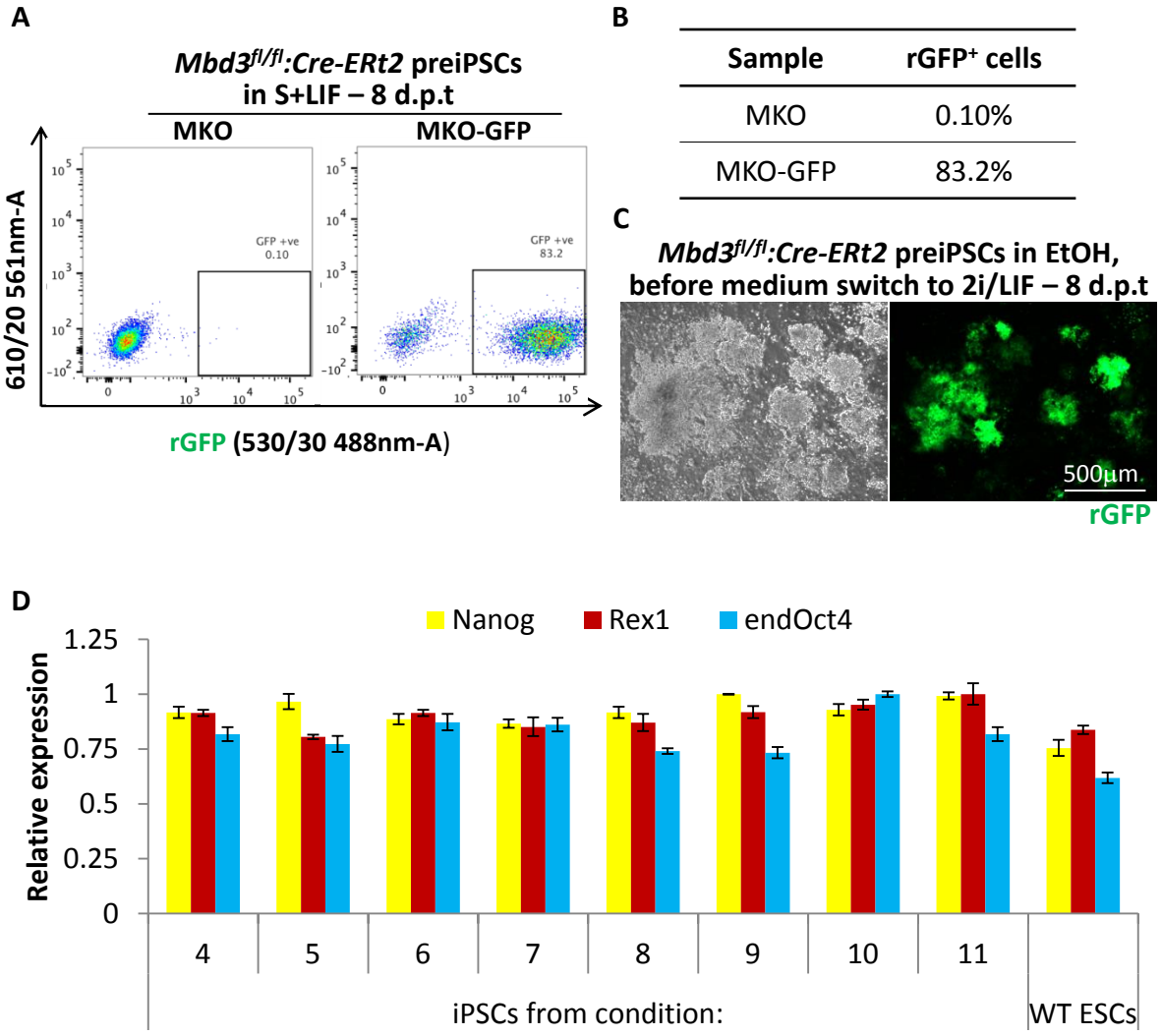


Figure 4.2.3 – Characterization of the iPSCs generated from *Mbd3$^{fl/fl}$:Cre-ERT2* Neural Stem Cells.

(A) Retroviral GFP (rGFP) expression 8 days post transduction of *Mbd3$^{fl/fl}$:Cre-ERt2* NSCs with rMKO or rMKO-rGFP. Percentage of rGFP^{+} cells was assessed by flow cytometry. GFP^{+} gates are shown. **(B)** Table indicates the percentage of rGFP^{+} cells 8 days post transduction of *Mbd3$^{fl/fl}$:Cre-ERt2* NSCs with MKO or MKO-GFP. **(C)** Phase and GFP images of *Mbd3$^{fl/fl}$:Cre-ERt2* preiPSC colonies 8 days after transduction with rOKM + rGFP in presence of EtOH. Images acquired before medium switch to 2i/LIF, day 8. **(D)** qRT-PCR analysis of pluripotency-associated genes in iPSCs generated from *Mbd3$^{fl/fl}$:Cre-ERt2* NSCs with Mbd3 deletion at different time points of reprogramming, which conditions refer to Figure 2C. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.

4.2.2 – Comparison of *Mbd3*$^{-/-}$ iPSCs to *Mbd3*$^{-/-}$ ESCs

To address if *Mbd3*$^{-/-}$ iPSCs, where *Mbd3* has been removed at a somatic cell state, showed similar behaviour to *Mbd3*$^{-/-}$ ESCs, where *Mbd3* was removed at a pluripotent state, different phenotypic assays were employed. As previously reported, *Mbd3*$^{-/-}$ ESCs can self-renew in culture, expressing relatively normal levels of pluripotency markers (Kaji et al., 2006; Reynolds et al., 2012a). However, *Mbd3*-null ESCs proliferate 2-times slower, a phenotype consistent with the known involvement of *Mbd3* in cell cycle regulation (Kaji et al., 2006; Sims and Wade, 2011). Similar to *Mbd3*-null ESCs, different clonal *Mbd3*$^{-/-}$ iPSC lines also proliferate 2-times slower than *Mbd3*$^{fl/-}$ or rescue lines (Figure 4.2.4A).

Another previously reported phenotype of *Mbd3*$^{-/-}$ ESCs is their lack of differentiation potential. When seeded for *in vitro* differentiation, *Mbd3*$^{-/-}$ ESCs fail to exit the self-renewing program, continuing to express high levels of pluripotency genes (Kaji et al., 2006). Moreover, when injected back into the embryo, they fail to contribute to the germ lineage (Kaji et al., 2006; Kaji et al., 2007). I have assessed if *Mbd3*$^{-/-}$ iPSCs also fail to undergo *in vitro* differentiation using different differentiation assays. First, two clonal *Mbd3*$^{-/-}$ iPSC lines, *Mbd3*$^{fl/-}$ and rescue iPSC lines, and ESC controls were clonally seeded in LIF-depleted serum media (S-LIF). AP^{+} colonies could still be observed after seven days in S-LIF culture in the case *Mbd3*$^{-/-}$ iPSCs/ESCs, but not in Mbd3-expressing lines (Figure 4.2.4B), indicating that although *Mbd3*-expressing cells fail to remain pluripotent when cultured in the absence of LIF, *Mbd3*-null cells can resist differentiation and clonally maintain a pluripotent status.

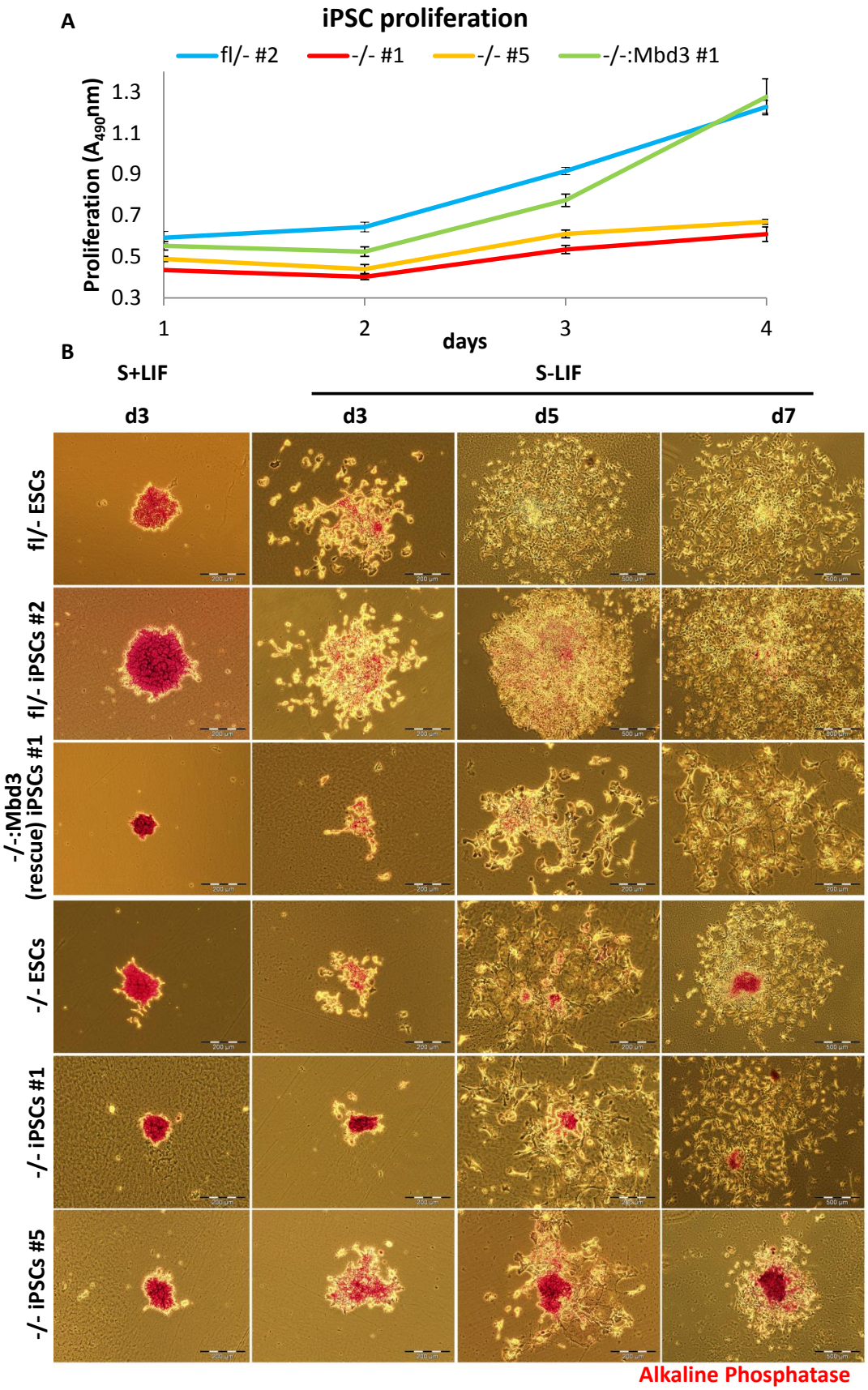


Figure 4.2.4 – iPSCs proliferation and clonal differentiation analysis.

(A) Cell proliferation analysis of *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{-/-}$*:Mbd3* iPSCs clonal lines. **(B)** Alkaline phosphatase staining after culture in LIF-deprived serum-containing media (S-LIF) of clonal seeded ESCs or iPSCs. 2000 ESCs/iPSCs were seeded in the indicated media.

All above indicated cell lines were then cultured in S+LIF or S-LIF media to assess if *Mbd3*-deletion allows not only clonal growth in the absence of LIF but also extensive culture. Whereas *Mbd3*$^{fl/-}$ iPSCs, rescue iPSCs, and *Mbd3*$^{fl/-}$ ESCs were lost after two passages in S-LIF, *Mbd3*$^{-/-}$ iPSCs could be expanded in both conditions for more than four passages (Figure 4.2.5A), exhibiting only a small reduction in pluripotency-associated gene expression (Figure 4.2.5B).

To further assess the lack of differentiation in *Mbd3*$^{-/-}$ iPSCs when cultured in a monolayer culture in the absence of LIF, I measured the differentiation potential of all of the above cell lines during a time-course of 10 days using a pluripotency reporter. For that, I made use of a construct where GFP and puromycin resistance are expressed under the control of the *Oct4* distal enhancer (*EOS*-GiP), meaning that only naïve pluripotent cells are GFP^{+}. After stable transfection of the *EOS*-GiP construct into iPSCs/ ESCs and their culture in 2i/LIF containing puromycin, I observed that 100% of the cells of all genotypes are GFP^{+} (Figure 4.2.5C-D). Since only pluripotent cells are GFP^{+}, I used these stably transfected cell lines to indirectly study the exit from pluripotency, since differentiated cells will become GFP^{-}.

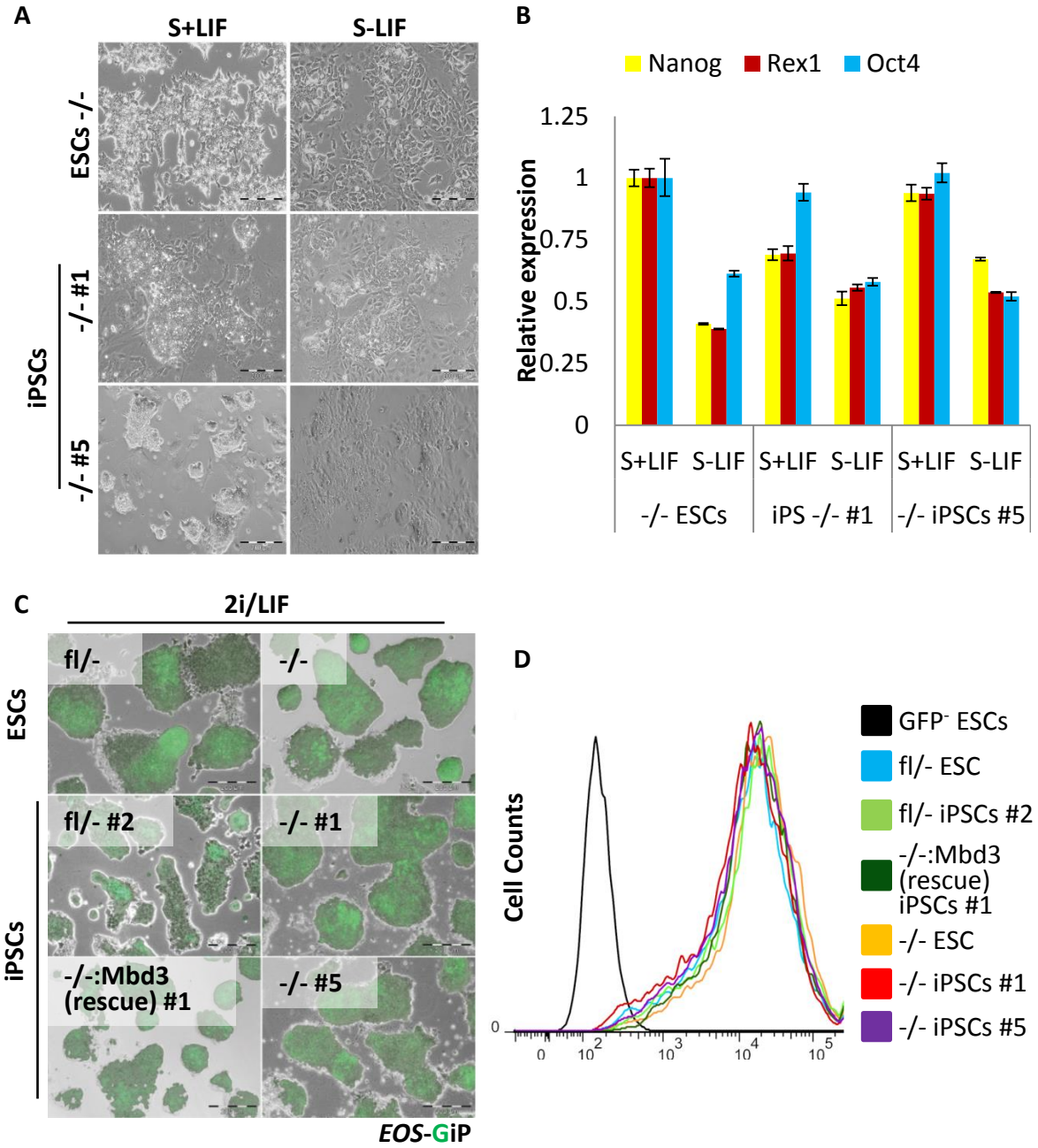


Figure 4.2.5 – *Mbd3*$^{-/-}$ ESCs/iPSCs can be cultured in the absence of pluripotency culture requisites.

(A) Phase images of *Mbd3*$^{-/-}$ ESCs/iPSCs cultured in the presence or absence of LIF for 4 passages (12 days). *Mbd3*$^{fl/-}$ ESCs/iPSCs and rescue cell lines were lost after 2 passages in S-LIF. **(B)** qRT-PCR analysis of pluripotency-associated in *Mbd3*$^{-/-}$ ESCs/iPSCs cultured in the presence or absence of LIF for 4 passages (12 days). qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV. **(C)** ESCs/iPSCs with different genotypes were transfected with pB-EOS-GiP (GFPiresPuro under the control of early transposon promoter and Oct-4 and Sox2 enhancers) and selected using puromycin. Shown are merged phase and GFP images. **(D)** *EOS*-GiP levels in the different ESCs/iPSCs after stable transfection with pB-EOS-GiP. *Mbd3*$^{fl/-}$ untransfected ESCs were used as GFP^{-} control.

The generated *EOS*-GiP cell lines were then cultured in S-LIF media and GFP expression was analysed by flow cytometry every other day for 10 days (Figure 4.2.6A). More than 50% of *Mbd3*$^{-/-}$ iPSCs/ ESCs remained *EOS*-GiP^{+} after 10 days of culture in the absence of LIF, whereas more than 80% of *Mbd3*$^{fl/-}$ iPSCs/ ESCs exited self-renewal and differentiated, becoming *EOS*-GiP^{-}. A time-course representation of all time points and duplicates collected can be found in Figure 4.2.6B. This indicates that although a minority of *Mbd3*$^{-/-}$ iPSCs/ ESCs will eventually become *EOS*-GiP^{-} (around 30%), this process is very inefficient and exhibits delayed kinetics.

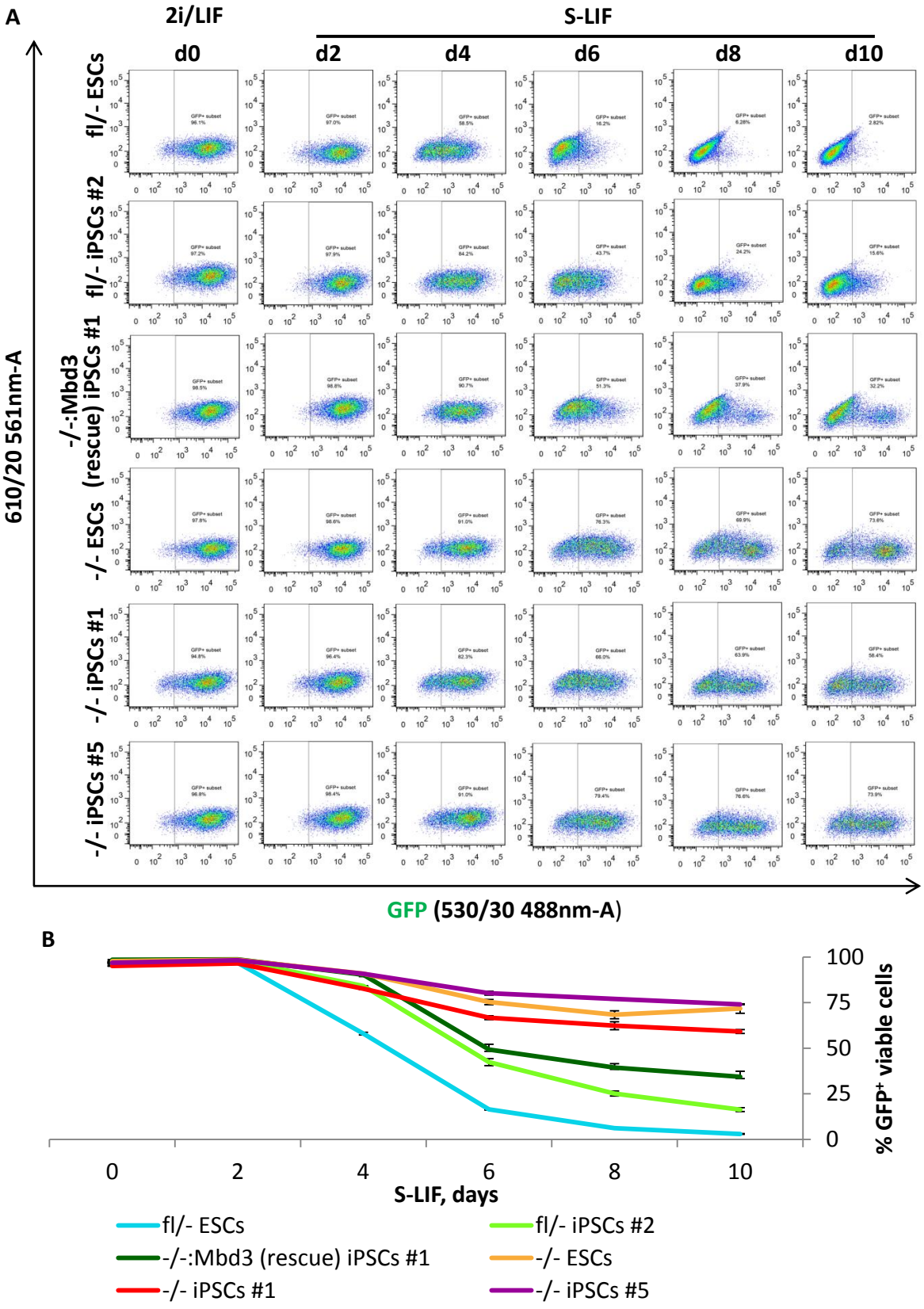


Figure 4.2.6 – *Mbd3*$^{-/-}$ iPSCs fail to down-regulate Oct4 reporter expression upon LIF-withdrawal.

(A) Time-course of *Oct4*-GFP (*EOS*-GiP) expression upon during culture in LIF-deprived serum-containing media. GFP expression was assessed by flow cytometry. GFP^{+} gates are shown. **(B)** Graphical representation of GFP^{+} gates from panels above. The error bars indicate STDEV.

The gold-standard of *in vitro* differentiation assays is the differentiation of pluripotent cells as embryoid bodies (EBs), the protocol that more closely resembles *in vivo* development (Doetschman et al., 1985). For the generation of EBs, ESCs are grown in low-attachment bacterial dishes in S-LIF, and after 3 to 8 days of culture the cells form complex EBs that are morphologically similar to morulas and early blastocysts (Doetschman et al., 1985), containing cells that are committed for differentiation into all germ lineages: endoderm, mesoderm and ectoderm. I tested how the iPSCs/ ESCs with different *Mbd3* genotypes performed during EB differentiation, collecting samples for gene expression analysis at day 3, 5 and 7 of culture in S-LIF. Key pluripotency-associated genes such a *Nanog*, *Rex1*, *Esrrβ* and *Klf2* were found to be significantly down-regulated after 3 days in S-LIF in *Mbd3*$^{fl/-}$ iPSCs/ ESCs and *Mbd3*$^{-/-}$*:Mbd3* iPSCs (Figure 4.2.7A). *Oct4* expression was kept high during the first 5 days of differentiation, and down-regulated only after 7 days, which is in accordance with the recently described role of *Oct4* in the exit from pluripotency (Karwacki-Neisius et al., 2013; Radzisheuskaya et al., 2013). All these pluripotency markers were found to be highly expressed throughout the EB differentiation protocol in *Mbd3*$^{-/-}$ cells. Moreover, whereas key lineage specification genes such as *Fgf5* (primitive ectoderm), *Gata4* (endoderm) and *T-brachyury* (mesoderm) could be highly detected after 3 to 5 days of differentiation in *Mbd3*-expressing cells, these genes failed to be properly up-regulated during S-LIF culture of *Mbd3*-null cells. From our panel of differentiation genes, the only gene that was found to be up-regulated after EB differentiation of *Mbd3*$^{-/-}$ cells was *Pl-I* (trophectoderm), a phenomena previously described during the differentiation of *Mbd3*$^{-/-}$ ESCs (Kaji et al., 2006). Those results indicate that a lack of functional NuRD seems to be priming pluripotent cells to differentiate into trophectoderm, an extra-embryonic lineage that wild type ESCs cannot generate. The fact that disruption of the NuRD complex prompts cells to differentiate into trophectoderm might explain why ~30% of *Mbd3*$^{-/-}$ iPSCs/ ESCs become *EOS*-GiP^{-} after 10 days in culture in S-LIF (Figure 4.2.6B). Indeed, cells resembling trophoblast giant cells were visible after 6-8 days of culture of *Mbd3*$^{-/-}$ iPSCs/ ESCs in S-LIF (data not shown). Additionally, beating cells were visible 3 to 4 days after plating the EBs generated from *Mbd3*$^{fl/-}$ iPSCs/ ESCs and *Mbd3*$^{-/-}$*:Mbd3* iPSCs on gelatine, but not *Mbd3*-null cells. In fact, AP^{+} colonies could still be seen upon 4 days of S-LIF culture on gelatine of the *Mbd3*$^{-/-}$ 7 day-old EBs on gelatine (Figure 4.2.7B), indicating how recalcitrant to differentiation these cells are. All together, the data presented above indicates that *Mbd3*$^{-/-}$ iPSCs are phenotypically indistinguishable from previously reported *Mbd3*-null ESCs (Kaji et al., 2006), exhibiting slower proliferation and a lack of differentiation capacity, marked by impaired embryoid body differentiation.

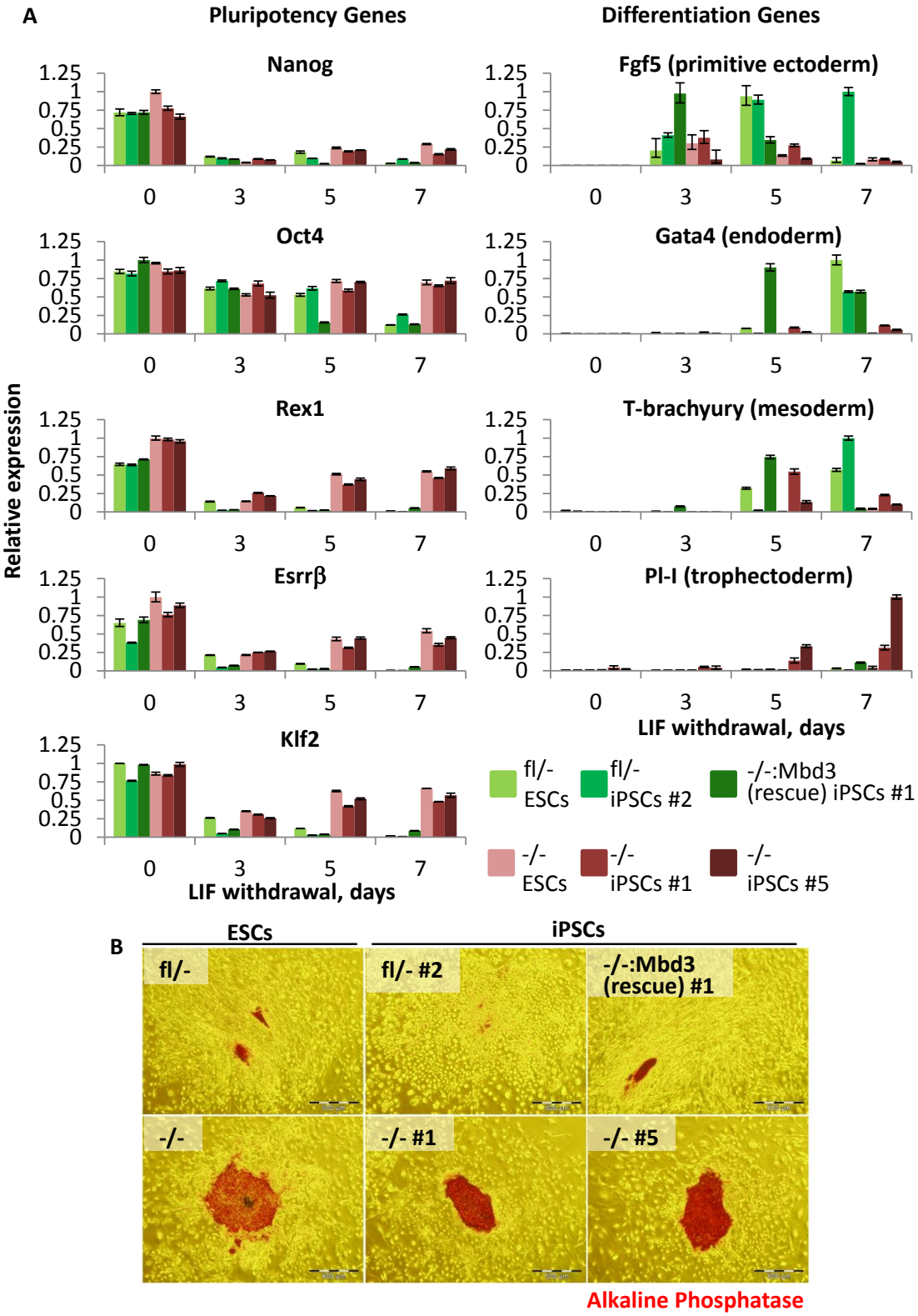


Figure 4.2.7 – *Mbd3*$^{-/-}$ iPSCs exhibit impaired embryoid body differentiation.

(A) qRT-PCR analysis of pluripotency- and differentiation-associated genes during differentiation as embryoid bodies of *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$, *Mbd3*$^{-/-}$*:Mbd3* iPSCs or ESCs controls. The error bars indicate STDEV. **(B)** Alkaline phosphatase staining 4 days after plating of 7-day matured embryoid bodies in gelatine-coated dishes. Beating cells were observed after differentiation of *Mbd3*$^{fl/-}$ ESCs/iPSCs and rescue line, but not *Mbd3*$^{-/-}$ ESCs/iPSCs.

In the chapter 3, I showed that *Mbd3*$^{-/-}$ NSCs and preiPSCs exhibit lower global levels of 5-hmC compared to *Mbd3*-expressing cells (Figure 3.2.7). Since this relationship between *Mbd3* and 5-hmC levels was first reported in ESCs (Yildirim et al., 2011), I tested if a similar phenotype is observed in iPSCs. As seen before in the case of NSCs and preiPSCs, similar *Tet1/2* expression levels were detected in both *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ iPSCs (Figure 4.2.8A). Yet, *Mbd3*$^{-/-}$ iPSCs display lower levels of 5-hmC, to a similar extent than observed in *Mbd3*-null ESCs (Figure 4.2.8B), revealing that, although *Mbd3*$^{-/-}$ iPSCs could be established, these maintain a lower level of 5-hmC, which might explain, at least in part, the requirement for NuRD in somatic cell reprogramming (discussed in section 3.3.1).

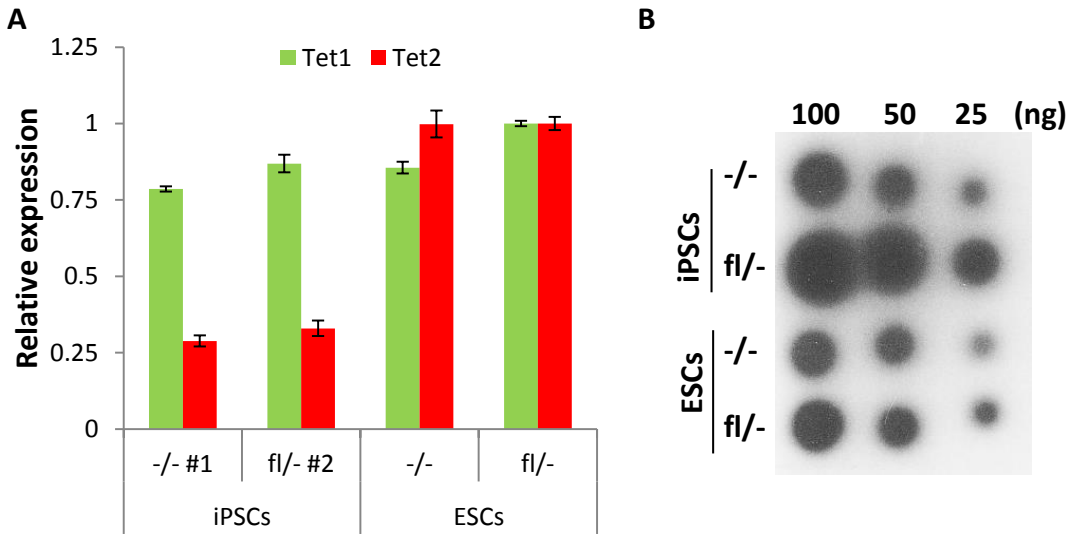


Figure 4.2.8 – *Mbd3*$^{-/-}$ ESCs/iPSCs show impaired levels of 5-hydroxymethylation.

(A) qRT-PCR analysis of *Tet1/2* expression levels in *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs/iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV. **(B)** DNA dot blot analysis of 5-hydroxymethylation (5-hmC) global levels in *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs/iPSCs.

4.2.3 – The NuRD complex is required for Epiblast Stem Cell reprogramming

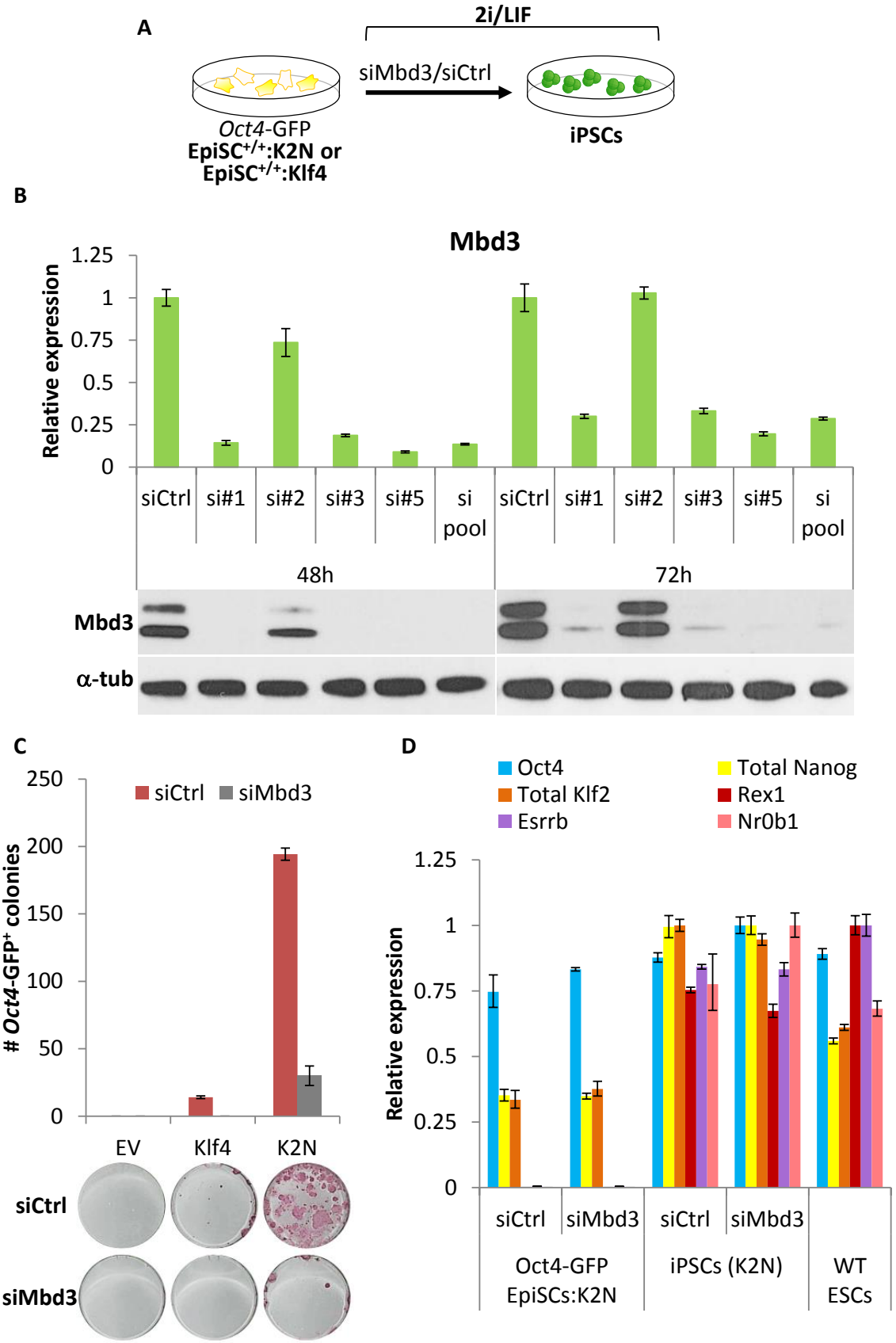
Next, I tested the impact of *Mbd3* deletion during EpiSC reprogramming to naïve pluripotency. For this, two approaches were used: siRNA mediated *Mbd3* depletion or genetic deletion.

To study the impact of *Mbd3* knockdown (KD) in the reprogramming of EpiSCs, wild type EpiSCs carrying an *Oct4-GFP* reporter were stably transfected with piggyBac (PB) vectors constitutively expressing *Klf2* and *Nanog* (K2N) or *Klf4* and transfected with either small interference RNA (siRNA) against *Mbd3* or control siRNA (siCtrl) (Figure 4.2.9A). To test which siRNA resulted in higher *Mbd3* silencing, EpiSCs were transfected with four different siRNAs, independently or pooled, and both *Mbd3* transcript and protein expression were analysed 48 or 72 hours latter (Figure 4.2.9B). Since the highest KD efficiency was achieved using siRNAs #1, #3 and #5, these were pooled and transfected together into EpiSCs for reprogramming experiments. Strikingly, *Mbd3* KD led to a complete impairment of Klf4-mediated reprogramming and up to a 6 fold reduction in the reprogramming ability of K2N (Figure 4.2.9C). Although with less efficiency, iPSCs could be generated after *Mbd3* KD, exhibiting a gene expressing profile similar to wild type ESCs (Figure 4.2.9D).

(figure on next page)

Figure 4.2.9 – Mbd3 knockdown impairs Epiblast Stem Cell reprogramming.

(A) Experimental design used to analyse the effect of Mbd3 knockdown (KD) on EpiSC reprogramming. Wild type EpiSCs (carrying an *Oct4*-GFP cassette), stably transfected with pPB-CAG-Klf2.2A.Nanog (K2N) or pPB-CAG-Klf4, were transfected with either siMbd3 or siControl (siCtrl) and after 24h plated in 2i/LIF for 12 days. **(B)** qRT-PCR and western blot analysis of *Mbd3* transcript and protein levels, respectively, 48h and 72h after siRNA-mediated KD. αtub indicates α-Tubulin. siRNAs were tested independently or pooled. A pool of si#1, #3 and #5 was used for future experiments. **(C)** The efficiency of EpiSC reprogramming after *Mbd3* KD assessed by counting *Oct4*-GFP^{+} iPSC colonies. Representative AP stained plates are also indicated. Colony number is per 1.0 x 10^{4} EpiSCs. **(D)** qRT-PCR analysis of *Oct4*, total *Nanog*, *Rex1*, *Esrrβ*, total *Klf2* and *Nr0b1* expression in Klf2-2A-Nanog transfected *Oct4*-GFP EpiSCs and derived iPSCs with siCtrl and siMbd3 transfection in 2i/LIF. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.


A
2i/LIF
siMbd3/siCtrl
Oct4-GFP
EpiSC+/+:K2N or
EpiSC+/+:Klf4
iPSCs
B
Mbd3
Relative expression
1.25
1
0.75
0.5
0.25
0
siCtrl
si#1
si#2
si#3
si#5
si pool
48h
72h
Mbd3
α-tub
C
siCtrl
siMbd3
Oct4-GFP+ colonies
250
200
150
100
50
0
EV
Klf4
K2N
siCtrl
siMbd3
D
Oct4
Total Klf2
Esrrb
Total Nanog
Rex1
Nr0b1
Relative expression
1.25
1
0.75
0.5
0.25
0
siCtrl
siMbd3
siCtrl
siMbd3
Oct4-GFP
EpiSCs:K2N
iPSCs (K2N)
WT
ESCs

Since *Mbd3* KD decreased the reprogramming efficiency of EpiSCs, I decided to test if the same finding could be achieved after genetic ablation of *Mbd3*. As described in the section 1.4.2, *Mbd3*$^{-/-}$ blastocysts die early post-implantation (Kaji et al., 2007), and since EpiSCs are derived from the post-implantation epiblast, *Mbd3*$^{-/-}$ EpiSCs could not be derived *in vivo*. However, EpiSCs can be derived *in vitro* by the culture of ESCs in Fgf2 and Activin A conditions (Fgf2/Act.A) (Guo et al., 2009; Nichols and Smith, 2009). I tested if *Mbd3*$^{-/-}$ ESCs could be differentiated into *Mbd3*$^{-/-}$ EpiSCs by *in vitro* culture in Fgf2/Act.A (Figure 4.2.10A). To follow differentiation into an EpiSC state, ESC lines stably transfected with the naïve pluripotency reporter *EOS*-GiP were used. As described above, the *EOS* promoter is only active in naïve pluripotent cells, being inactive in primed pluripotent cells, EpiSCs. Surprisingly, cells with EpiSC morphology could be seen after 2-3 passages in Fgf2/Act.A of *Mbd3*$^{-/-}$ ESCs (Figure 4.2.10B). Flow cytometry analysis revealed that less than 5% of both *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ cells were GFP^{+} after culture in Fgf2/Act.A conditions for 10 passages (Figure 4.2.10C). Moreover, both *Mbd3*$^{-/-}$ and *Mbd3*$^{fl/-}$ EpiSCs showed a downregulation of naïve pluripotency-associated genes (*Nanog*, *Rex1*, *Klf2*, *Klf4* and *Nr0b1*) and an up regulation of primed pluripotency-associated genes (*Lefty1* and *T brachyury*) (Figure 4.2.10D). Although a functional NuRD complex is required for the generation of the post-implantation epiblast *in vivo*, *Mbd3*-null EpiSCs can be generated *in vitro* by induced differentiation of naïve pluripotent *Mbd3*$^{-/-}$ ESCs (Kaji et al., 2007). This indicates that the NuRD complex is not required for this transition between naïve and primed pluripotency *in vitro*.

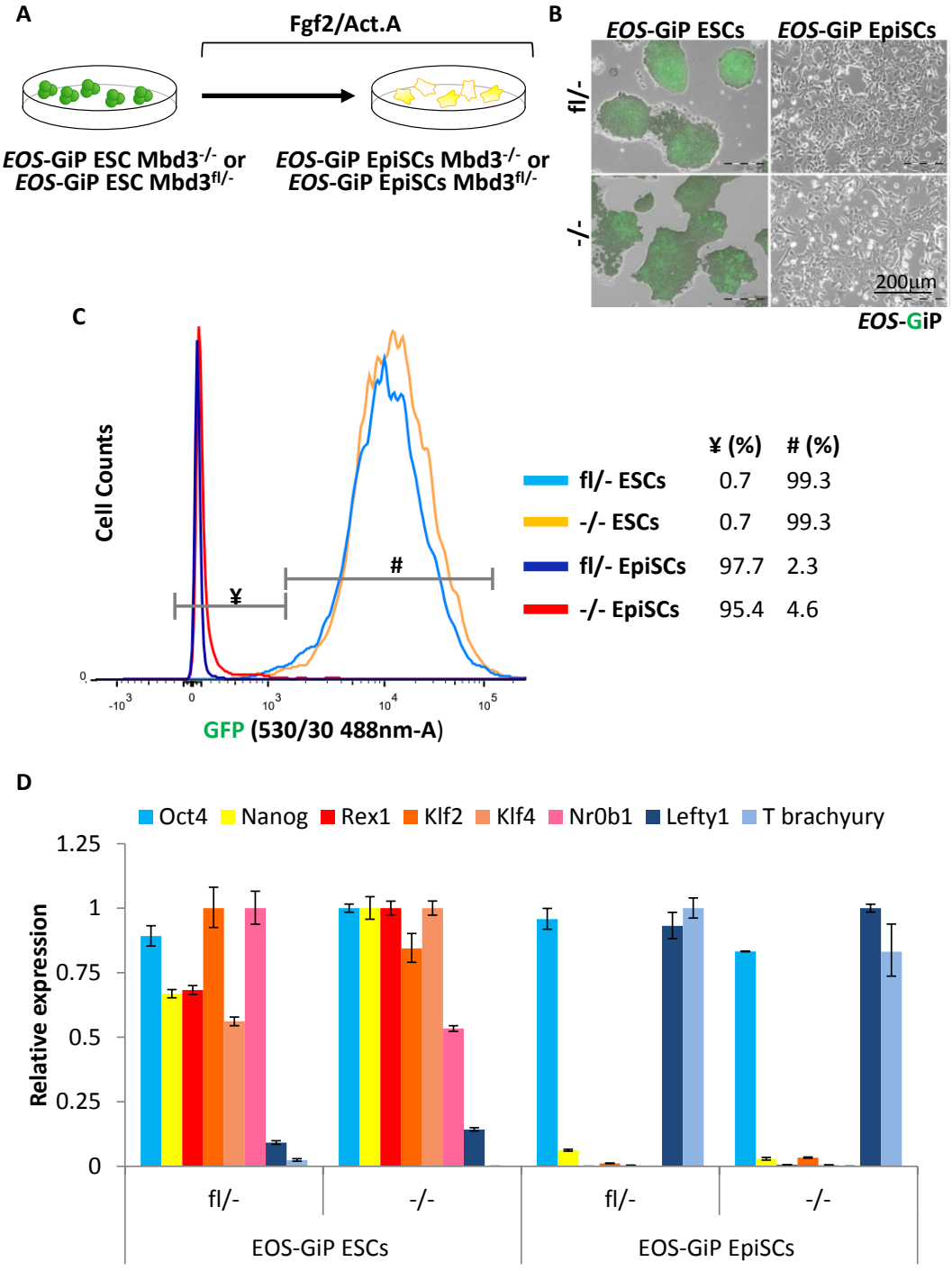


Figure 4.2.10 – Lack of NuRD complex does not impair Epiblast Stem Cell differentiation.

(A) Experimental design used to generate *Mbd3*$^{-/-}$ EpiSCs from *Mbd3*$^{-/-}$ ESCs by continuous culture in Fgf2 + Activin A (Fgf2/Act.A) conditions. The *EOS*-GiP reporter is only active in naïve pluripotent cells, becoming silent during EpiSCs differentiation. **(B)** Phase and GFP (merged) images of *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs and EpiSCs derived from them. **(C)** *EOS*-GiP levels in *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs and corresponding EpiSC derived after 10 passages in Fgf2/Act.A conditions. **(D)** qRT-PCR analysis of *Oct4*, *Nanog*, *Rex1*, *Klf2*, *Klf4*, *Nr0b1*, *Lefty1* and *T brachyury* expression in parental *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ *EOS*-GiP ESCs and corresponding *EOS*-GiP EpiSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.

After establishing the above described EpiSC lines with a Klf2-Nanog construct (K2N) these were plated in Fgf2/Act.A and switched next day to 2i/LIF conditions to minimize the possible impact of differences in cell proliferation (Figure 4.2.11A). Colonies showing reactivation of *EOS*-GiP transgene were observed 5-6 days after 2i/LIF application in cells from both genotypes, indicating that these ESC-derived EpiSCs could be reprogrammed back to full pluripotency (Figure 4.2.11B). Similarly to *Mbd3* KD, *Mbd3* genetic knockout (KO) led to a 20 fold reduction in K2N-mediated EpiSC reprogramming (Figure 4.2.11C).

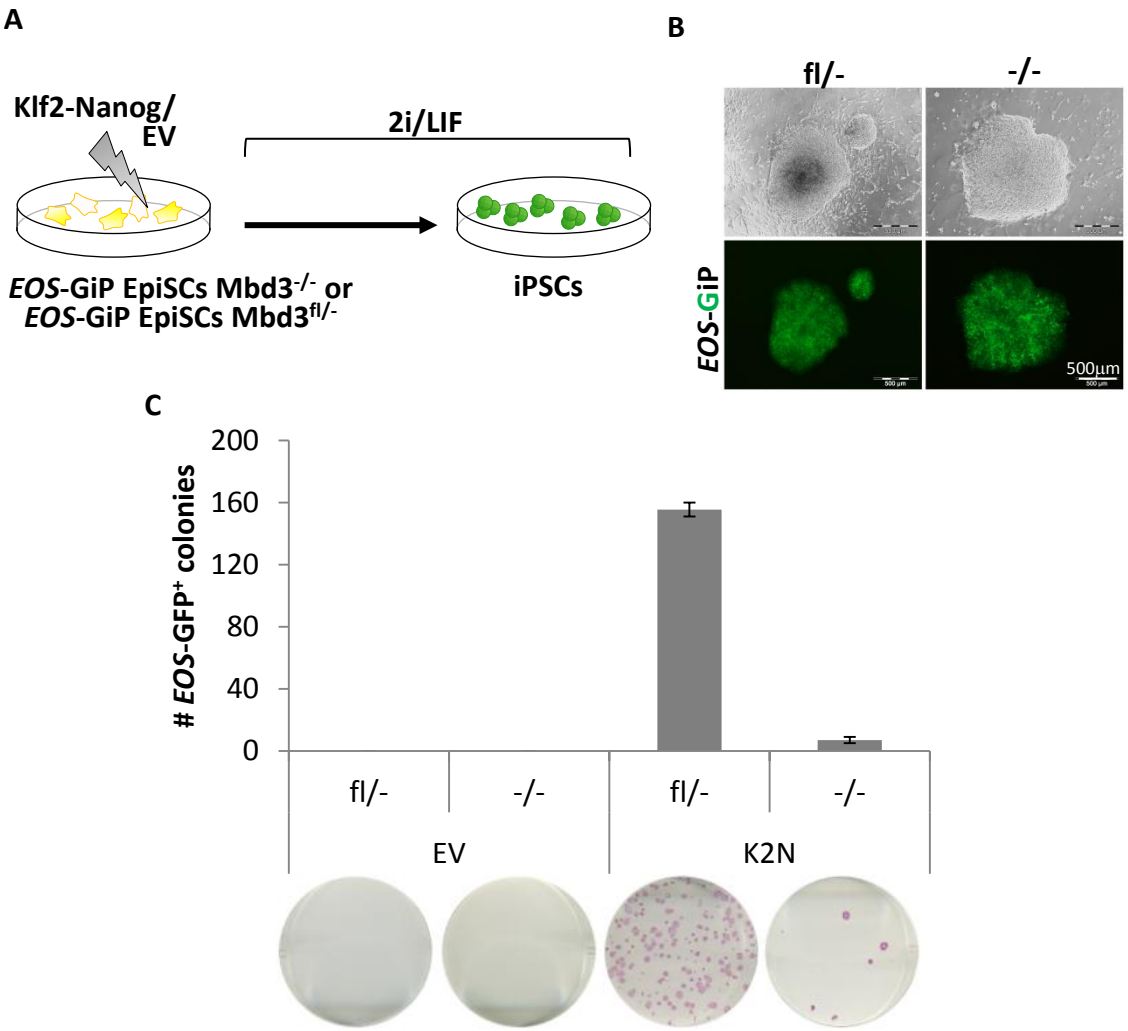


Figure 4.2.11 – Mbd3 knockout impairs Epiblast Stem Cell reprogramming.

(A) Experimental design used to analyse the effect of Mbd3 knockout (KO) on EpiSC reprogramming. *Mbd3*$^{-/-}$ or *Mbd3*$^{fl/-}$ EpiSCs carrying an Oct4-GFP reporter (*EOS*-GiP), stably transfected with pPB-CAG-Klf2.2A.Nanog (K2N) or empty vector (EV), were plated in 2i/LIF for 12 days. **(B)** Phase and *EOS*-GiP images of *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ iPSCs generated from EpiSCs. **(C)** The efficiency of EpiSC reprogramming after *Mbd3* KO was assessed by counting *EOS*-GFP^{+} iPSC colonies. Representative AP stained plates are also indicated. Colony number is per 1.5 x 10^4 EpiSCs. The error bars indicate STDEV.

The generated *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ iPSCs maintained continuous GFP expression, allowing their culture in the presence of puromycin, a consequence of complete reactivation of the *EOS*-GiP reporter (Figure 4.2.12A-B). IPSCs from both genotypes reactivated the endogenous naïve pluripotency transcriptional program, exhibiting a gene expression profile similar to the original *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs (Figure 4.2.12C).

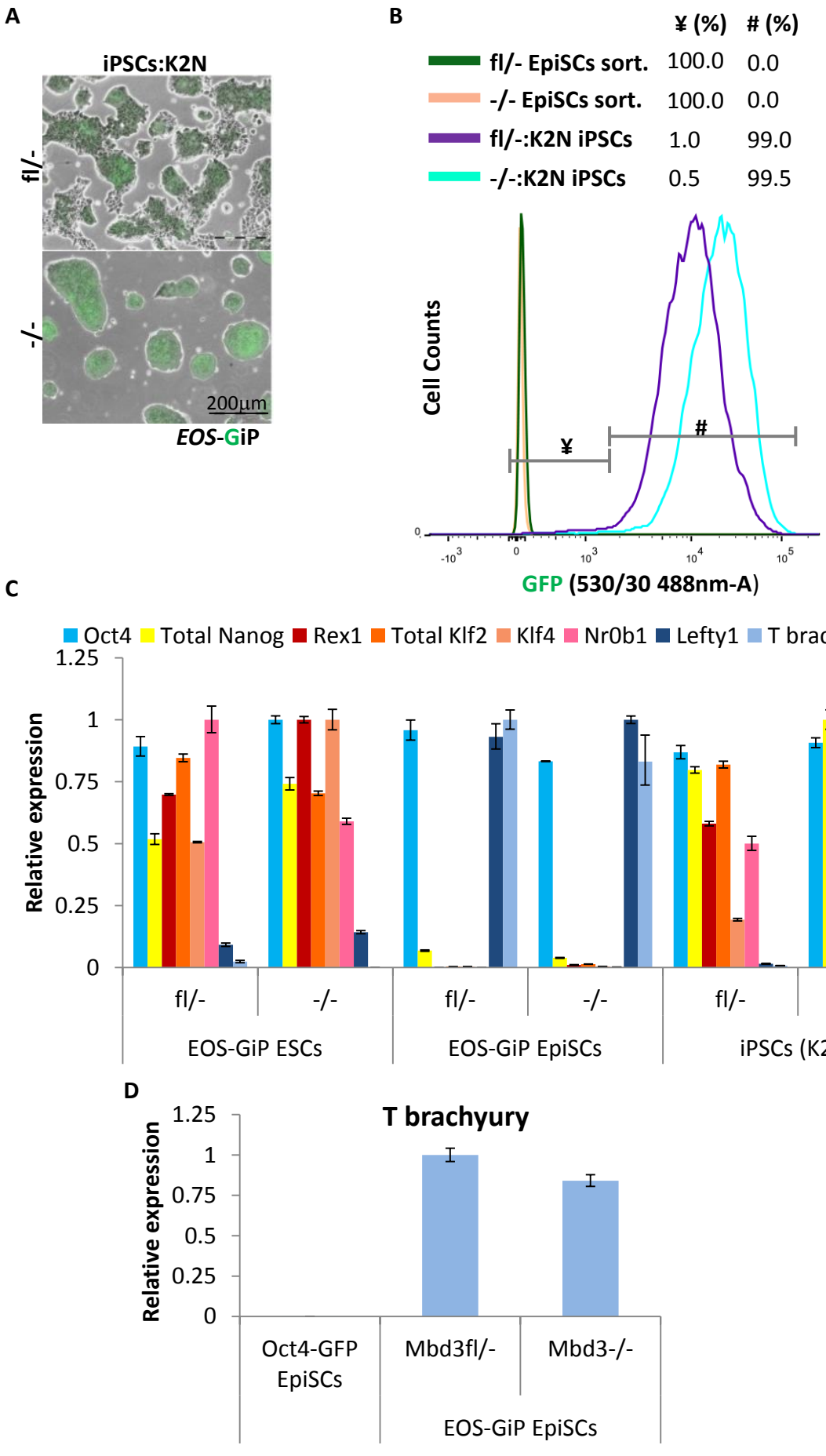
It has been reported that different populations of EpiSCs can be derived, exhibiting antagonistic levels of *Oct4* and *T brachyury* (Han et al., 2010). Those different EpiSC lines display different behaviours. *Oct4*^{high}/*T brachyury*^{low} are able to spontaneously convert to naïve pluripotency by culture in ESC media, without the use of exogenous transgenes (Han et al., 2010). To eliminate the possibility that the impaired EpiSC reprogramming after *Mbd3* deletion might be due to *T brachyury* levels, I used two independent EpiSC lines which differed in the expression levels of *T Brachyury* (Figure 4.2.12D). For the knockdown studies, the wild type EpiSC line chosen (OEC2) expresses very low levels of *T brachyury*, whereas for the knockout experiments I used *Mbd3*$^{-/-}$, and their counterpart *Mbd3*$^{fl/-}$, EpiSCs which express high levels of *T-Brachyury*. Moreover, I have never observed spontaneous reprogramming to naïve pluripotency of any EpiSC line (wild type or *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ transfected with empty vector) after culture in 2i/LIF conditions alone.

Together, these results indicate that lack of a functional NuRD complex not only impairs NSC reprogramming but also transcription factor-mediated EpiSC reprograming.

(figure on next page)

Figure 4.2.12 – Characterization of the iPSCs generated from *Mbd3*$^{-/-}$ EpiSCs.

(A) Phase and GFP (merged) images of *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ iPSCs generated from corresponding EpiSCs. **(B)** *EOS*-GFP levels in *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ EpiSCs after sorting for *EOS*-GFP^{-} EpiSCs and corresponding iPSCs. **(C)** qRT-PCR analysis of *Oct4*, total *Nanog*, *Rex1*, total *Klf2*, *Klf4*, *Nr0b1*, *Lefty1* and *T (brachyury)* expression in parental *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ *EOS*-GiP ESCs, *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ *EOS*-GiP EpiSCs obtained from them and EpiSC-derived iPSCs. **(D)** qRT-PCR analysis of *T brachyury* in the EpiSCs lines used in this study: *Oct4*-GFP EpiSCs used for knockdown experiments and *EOS*-GiP EpiSCs used for knockout experiments. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value.


A
iPSCs:K2N
fl/-
-/-
200µm
EOS-GiP
B
¥ (%)
(%)
fl/- EpiSCs sort. 100.0 0.0
-/- EpiSCs sort. 100.0 0.0
fl/-:K2N iPSCs 1.0 99.0
-/-:K2N iPSCs 0.5 99.5
Cell Counts
¥
#
GFP (530/30 488nm-A)
C
Oct4
Total Nanog
Rex1
Total Klf2
Klf4
Nr0b1
Lefty1
T brachyury
Relative expression
1.25
1
0.75
0.5
0.25
0
fl/-
-/-
fl/-
-/-
fl/-
-/-
EOS-GiP ESCs
EOS-GiP EpiSCs
iPSCs (K2N)
D
T brachyury
Relative expression
1.25
1
0.75
0.5
0.25
0
Oct4-GFP EpiSCs
Mbd3fl/-
Mbd3-/-
EOS-GiP EpiSCs

4.2.4 – The NuRD complex does not impact Mouse Embryonic Fibroblast reprogramming

All the above described results indicate that *Mbd3* is critical for efficient reprogramming. To check if similar results are observed when different reprogramming approaches are used, we performed piggyBac-mediated reprogramming of mouse embryonic fibroblasts (MEFs) combined with either *Mbd3* depletion by KD or KO. The MEF piggyBac-mediated reprogramming experiments presented below were performed in collaboration with Luca Tosti and Keisuke Kaji from the MRC Centre for Regenerative Medicine, University of Edinburgh.

First, we used an *Mbd3* knockdown system (Figure 4.2.13A). For this, *Nanog*-GFP MEFs were treated with doxycycline (DOX) for the induction of the iMKOS polycistronic cassette which had been previously transfected into those MEFs (O'Malley et al., 2013). This polycistronic cassette encodes all four Yamanaka factors which are expressed from a single promoter and are linked by self-processing 2A peptides (Carey et al., 2009), and is only active when DOX is added to the culture (O'Malley et al., 2013). After addition of DOX, MEFs were cultured in S+LIF medium supplemented with vitamin C and Alki (Tgfβ signalling inhibitor). 24h afterwards they were transduced with lentiviruses expressing shRNA against *Mbd3*. 80-90% *Mbd3* KD was observed 6 days after transduction of the MEFs with shMbd3, which corresponds to day 7 of reprogramming (Figure 4.2.13B). Reactivation of the *Nanog*-GFP reporter was observed 10-12 days after the induction of the iMOKS cassette (Figure 4.2.13C), and the *Nanog*-GFP^{+} iPSC colonies generated were scored at day 13 of reprogramming (Figure 4.2.13D). No significant differences were found between MEFs transduced with two different shRNAs against *Mbd3* or shControls (around 100 colonies were scored in all cases). To test if this difference could be due to the reprogramming cassette used, we tested two reprogramming cassettes side-by-side: the iMKOS used above (O'Malley et al., 2013) and the iOKSM cassette (O'Malley et al., 2013; Sommer et al., 2009). The only difference between these two cassettes is the order in which the factors were cloned. Once again, no impact on the amount of *Nanog*-GFP^{+} iPSC colonies generated using either cassette was observed (Figure 4.2.13E). Interestingly, we observed a 2 fold increase in the amount of *Nanog*-GFP^{-} colonies, which corresponds to partially reprogrammed colonies, after reprogramming of the MEFs transduced with shMbd3.

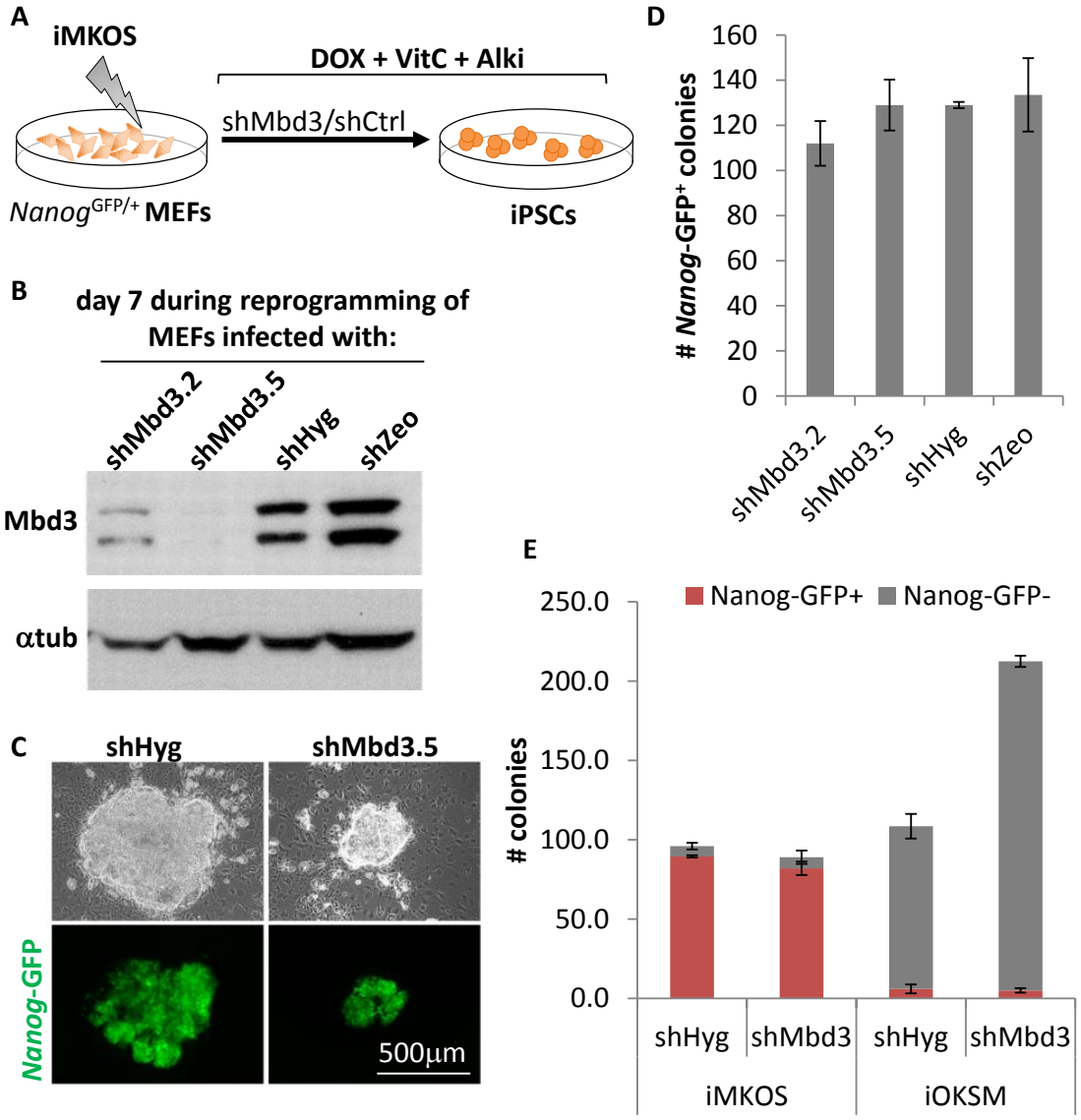


Figure 4.2.13 – Mbd3 knockdown does not impact Mouse Embryonic Fibroblasts piggyBac-mediated reprogramming.

(A) Experimental designs used to analyse the effect of Mbd3 KD. *Nanog*-GFP MEFs transfected with doxycycline-inducible MKOS piggyback transposon (iMKOS) were cultured in S+LIF + DOX + vitamin C (vitC) + Alki 24 hours before lentiviral infections of shMbd3 or shControls. **(B)** Western blot analysis of Mbd3 and α-Tubulin (αtub) protein levels at day 7 of reprogramming MEFs infected with shMbd3 or control shRNA against Hygromycin or Zeocin resistant genes (shHyg or shZeo). Over 90% of Mbd3 knockdown was observed. **(C)** Phase and *Nanog*-GFP images of iPSCs derived from MEFs transduced with lentiviruses encoding shHyg or shMbd3.5. **(D)** Number of *Nanog*-GFP^{+} iPSC colonies at day 13 of reprogramming upon infection of indicated shRNAs. **(E)** Quantification of *Nanog*-GFP^{+} iPSCs colonies and Nanog-GFP^{-} colonies generated by piggybac MEF reprogramming using two different reprogramming cassettes, iMKOS and iOKSM. *Mbd3* KD using shRNA was carried out 24h after induction of reprogramming cassettes. shMbd3.5 was used for this experiment. Typical iMKOS positive cell number at day 2 of reprogramming is 1.0-3.0 x 10^4 cells per well, providing 1-2% reprogramming efficiency. The error bars indicate STDEV.

We also tested Mbd3 depletion by KO, by treating Cre-ERt2 transduced *Mbd3*$^{ex1fl/ex1fl}$ MEFs with 4-OHT at 0h or 48hrs after the induction of reprogramming factor expression (Figure 4.2.14A). *Mbd3* exon 1 excision results in a loss of Mbd3a and Mbd3b isoforms, but a hypomorphic Mbd3c residual expression is observed (Figure 4.2.12B and Figure 3.2.1D). Since *Mbd3*$^{ex1fl/ex1fl}$ MEFs do not contain a reprogramming reporter, the resulting colonies were stained for Nanog protein 13 days after induction of the iMKOS cassette (Figure 4.2.14C). Although Mbd3 protein was found to be down-regulated, no impact on piggyBac-mediated MEF reprogramming was observed after *Mbd3* exon 1 deletion (Figure 4.2.14D), recapitulating the result obtained from *Mbd3* KD during MEF reprogramming.

Together, the above results show that Mbd3 depletion after the induction of reprogramming cassettes in a piggyBac-mediated MEF reprogramming system leads to no effect on transcription factor-induced naïve pluripotency.

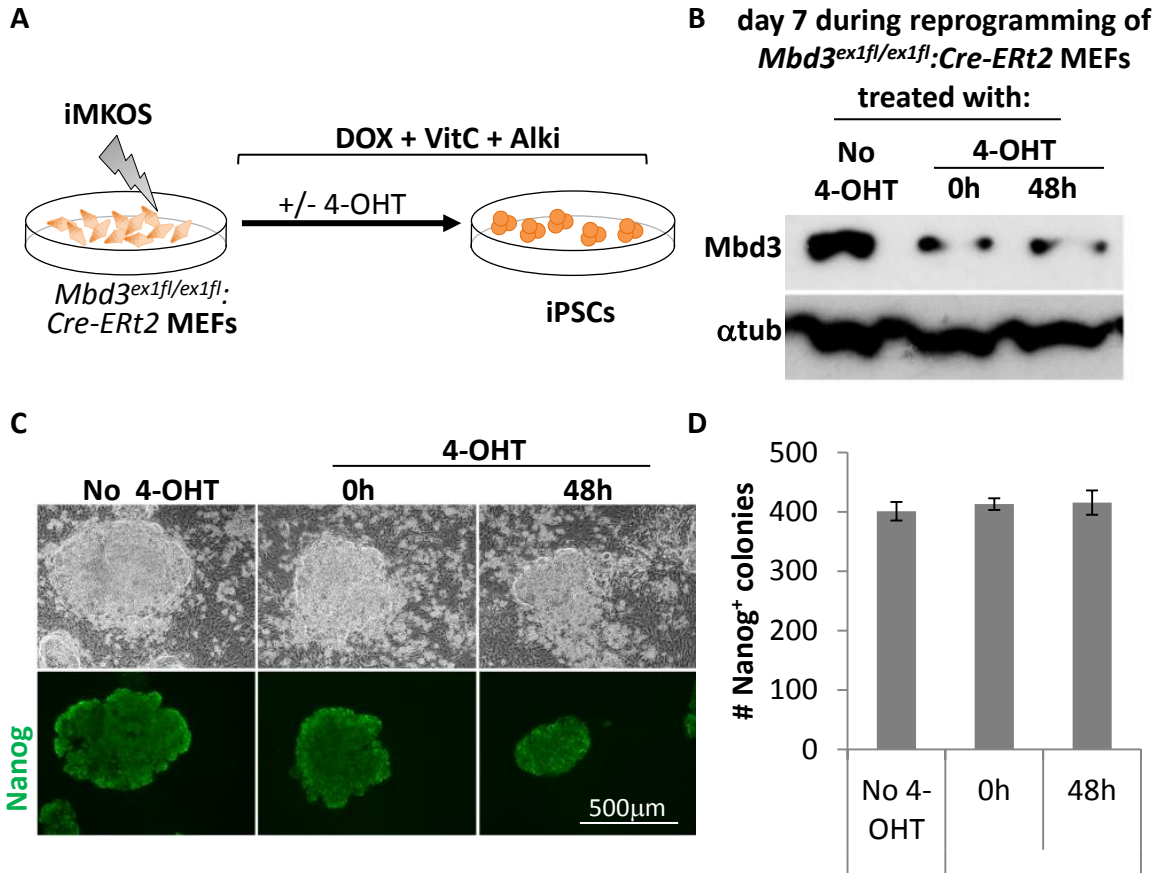


Figure 4.2.14 – Mbd3 knockout does not impact Mouse Embryonic Fibroblasts piggyBac-mediated reprogramming.

(A) Experimental designs used to analyse the effect of *Mbd3* exon 1 KO. *Mbd3*$^{ex1fl/ex1fl}$ MEFs transfected with iMKOS were infected with pMX-Cre-ERt2. Reprogramming was carried out in S+LIF + DOX + vitC + Alki, and 4-OHT was added either at the time of DOX administration (0h) or 48h later (48h). **(B)** Western blot analysis of Mbd3 and αtub protein at day 7 during reprogramming of *Mbd3*$^{ex1fl/ex1fl:}$*Cre-ERt2* MEFs treated 4-OHT 0h or 48h after induction of reprogramming cassettes. About 80% of Mbd3 knockdown was observed. **(C)** Phase and Nanog immunofluorescence positive colonies images of iPSCs derived from *Mbd3*$^{ex1fl/ex1fl:}$*Cre-ERt2* treated 4-OHT 0h or 48h after induction of reprogramming cassettes. **(D)** Number of Nanog positive colony numbers determined by immunofluorescence after 13 days of reprogramming of *Mbd3*$^{ex1fl/ex1fl}$*:Cre-ERt2* MEFs. Typical iMKOS positive cell number at day 2 of reprogramming is 1.0-3.0 x 10^4 cells per well, providing 1-2% reprogramming efficiency. The error bars indicate STDEV.

4.3 – Discussion

4.3.1 – Mbd3 is required for efficient iPSC generation

In this chapter I dissected the requirement for the Mbd3/ NuRD complex during iPSC generation. Using genetic or siRNA mediated *Mbd3* depletion, as well as different cell systems and experimental designs, I observed an impairment of reprogramming which is proportional to the amount of time cells express *Mbd3* during the initiation of reprogramming. I observed that depending on the reprogramming context, Mbd3/ NuRD depletion leads to two possible outcomes: no effect, in case of piggyBac-mediated MEF reprogramming; or a significant impairment in transcription factor-induced naïve pluripotency, in the case of NSC or EpiSC reprogramming. PiggyBac-mediated MEF reprogramming might not be affected due to the fact that Mbd3 is only deleted after the induction of the reprogramming factors. Since the loss of Mbd3 protein requires a few days from the time 4-OHT is administrated, and shMbd3 is transduced, it is possible that in this MEF reprogramming system cells go through the critical stage of reprogramming with Mbd3 protein still present. Indeed, the treatment of human fibroblasts with shRNAs against Mbd3 prior to OSKM expression led to a 2 fold reduction in reprogramming efficiency (Onder et al., 2012), reinforcing the idea that the induction of the reprogramming factors before *Mbd3* deletion during MEF reprogramming may overcome the requirement for the NuRD complex. Importantly, regardless of the system used, *Mbd3* deletion never resulted in the enhancement of reprogramming efficiency. My results identify a key role of the NuRD complex in reprogramming and show that, in a context dependent manner, a chromatin modifying complex required for cell differentiation also promotes reversion back to a naïve pluripotent cell state, acting as a facilitator of transcription factor-mediated induced pluripotency (Figure 4.3.1)

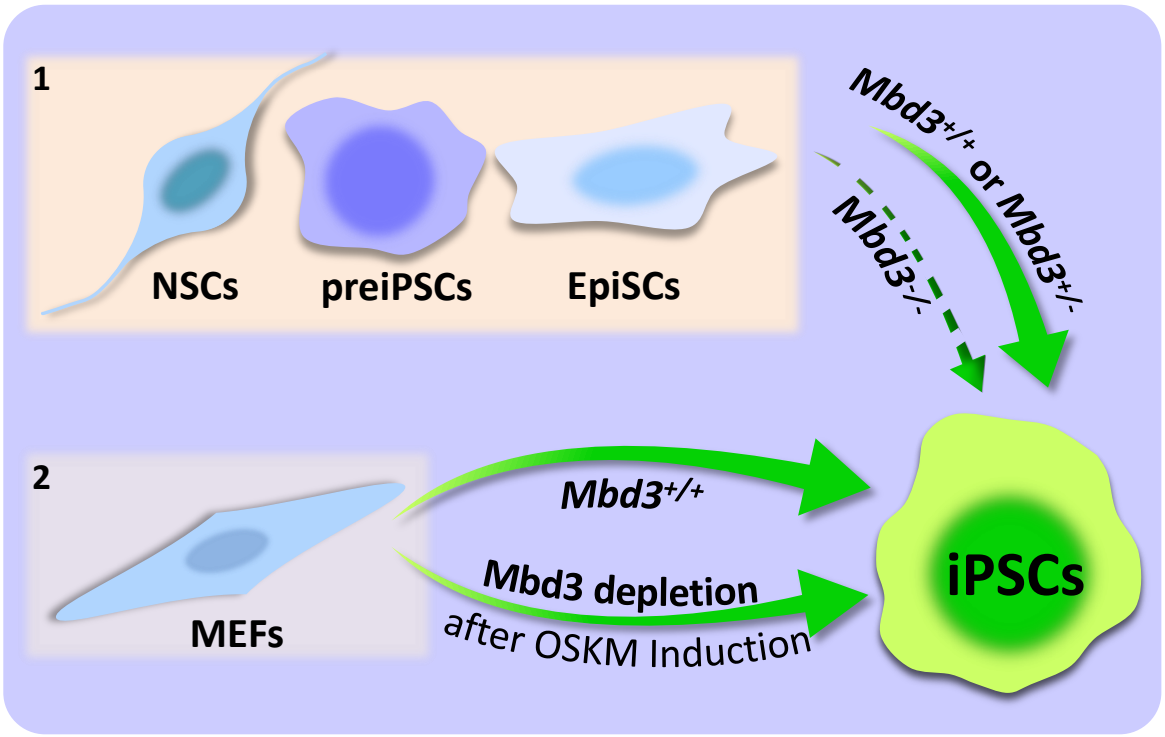


Figure 4.3.1 – Role of Mbd3/NuRD during the reprogramming.

The NuRD complex is essential for embryonic development. Its absence causes early embryo lethality (Hendrich et al., 2001; Kaji et al., 2007) and inability of ESCs to differentiate *in vitro* (Kaji et al., 2006). As demonstrated here, the NuRD complex is also critical for efficient reprogramming of NSCs, preiPSCs and EpiSCs (1), highlighting the complex as an important facilitator of cell state transitions. In my experimental setting where Mbd3 is depleted only after MKOS induction, no effect on reprogramming was observed (2).

My results are in apparent disagreement with the two reports which were published during the execution of this project (Luo et al., 2013; Rais et al., 2013). In both studies it was reported that Mbd3/ NuRD constitutes a barrier for somatic cell reprogramming. Particularly, the abrogation of Mbd3 expression led to higher reprogramming efficiencies, but the extent of this increase was different between those two studies (Luo et al., 2013; Rais et al., 2013).

In the Luo *et al* report, the system used to address the impact of the NuRD complex in reprogramming was *Mbd3* KD in MEF reprogramming system based on lentiviral-MKOS delivery (Luo et al., 2013). It was observed the shMbd3 transduction during MEF reprogramming leads to a 4 fold increase in the *Oct4*-GFP^{+} colonies generated. In our hands, while both *Mbd3* KD and KO during MEF reprogramming demonstrated about 80% downregulation of Mbd3 protein levels, neither impacted on the efficiency of MEF reprogramming. Potential explanations for the differences between Luo *et al* and our MEF

reprogramming results could be the means of reprogramming factor delivery used (4 viruses vs MKOS piggybac) and/or the reprogramming readout (alkaline phosphatase staining or *Oct4*-GFP vs *Nanog*-GFP). We addressed how reprogramming cassettes or the readout affect the results by performing *Mbd3* KD during piggyBac reprogramming using a different reprogramming cassette, iOKSM, which was reported to generate iPSC lines with abnormal imprinting patterns due to suboptimal stoichiometry of the 4 factors, unless Vitamin C is present (Carey et al., 2011; Stadtfeld et al., 2012). This iOKSM cassette was used in Rais *et al* study as well (Rais et al., 2013). Interestingly, while *Nanog*-GFP^{+} colony number was not affected by *Mbd3* KD when using either iMKOS or iOKSM reprogramming cassettes, the total colony number was increased up to 2 fold in the presence of shMbd3 in the case of the iOKSM cassette (Figure 4.2.13E). Previous reports have demonstrated that endogenous *Oct4* expression occurs in the early stages of reprogramming and is not a good predictor of successful reprogramming (Buganim et al., 2012; O'Malley et al., 2013; Silva et al., 2008; Theunissen et al., 2011b), especially if a stringent culture medium such as serum-free, KSR (knockout serum replacement)-free, 2i/LIF is not used. This is of relevance since for iPSC induction Luo *et al* used medium supplemented with 10% KSR. While further investigation is necessary, differences between Luo *et al* and our work might be due to different stoichiometry of the 4 factors and the criteria used for scoring iPSC colonies.

More recently, another report was published claiming that low levels or deletion of *Mbd3* leads to rapid deterministic reprogramming with 100% efficiency (Rais et al., 2013). The conclusions by Rais *et al* were founded on two key findings. First, they identified *Mbd3* as a barrier to reprogramming by an RNAi screen in EpiSCs. Second, they showed that MEFs derived from *Mbd3*$^{fl/-}$ ESCs reportedly expressing 20% of wild type Mbd3 levels, exhibit 100% reprogramming efficiency judging by the upregulation of an *Oct4*-GFP (genomic Oct4 fragment 18 kb, GOF-18) reporter transgene. There are, however, issues with Rais *et al's* experiments. Firstly, the EpiSCs used exhibit high expression of the naïve pluripotency marker *Klf4*, which is usually negligible in EpiSCs compared to ESCs cultured in 2i/LIF (Guo et al., 2009; Silva et al., 2009; Yang et al., 2010) (Figure 4.2.10D) and ectopic expression of Klf4 induces reprogramming to naïve pluripotency (Guo et al., 2009). This questions the reliability of their EpiSCs as an RNAi screening system for reprogramming. Secondly, the *Mbd3*$^{fl/-}$ heterozygous ESC line, which was generated by the Hendrich laboratory (University of Cambridge) and used in the study by Rais *et al* to derive *Mbd3*$^{fl/-}$ MEFs, actually express Mbd3 protein at nearly wild type levels (Reynolds et al., 2012b) (Figure 3.2.1B), in contrast to the 20% shown in Rais *et al*. Thus, MEFs derived from this line cannot be used to describe the effects of lower levels of *Mbd3* in reprogramming, nor is it clear why the cells used by

Rais *et al* show such low levels of *Mbd3*. Surprisingly, analysis of whole cell extracts input (WCE) ChIP-seq (chromatin immunoprecipitation followed by massively parallel sequencing) tracks of *Mbd3*$^{fl/-}$ MEFs and derived iPS cells published by Rais *et al* revealed a striking over-representation of *Pou5f1/Oct4* regulatory sequences, indicative of a large number of copies of the *Oct4* reporter transgene (GOF-18) in the genome (Figure 4.3.2, top panel). For this reason, reactivation of the reporter transgene could occur more readily than endogenous pluripotency-associated gene activation, leading to inappropriately inflated measures of reprogramming efficiency. Furthermore, closer inspection of the same input ChIP-seq tracks at the *Pou5f1* regulatory sequences also revealed that the 18kb Oct4 reporter transgene used by Rais *et al* in control *Mbd3*$^{+/+}$ MEFs was different from that in *Mbd3*$^{fl/-}$ MEFs. The key *Pou5f1* regulatory sequence, the proximal enhancer (PE), responsible for driving GFP expression, is found to be missing in the *Oct4-GFP* reporter of control MEFs, indicative of GOF-18ΔPE (Yeom et al., 1996; Yoshimizu et al., 1999) (Figure 4.3.2, middle and bottom panels). Thus, the intact versus the defective *Oct4-GFP* reporter transgene could easily explain the difference in *Oct4*-GFP^{+} colony numbers observed between *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ in Rais *et al's* study. While the impact of the loss of *Mbd3* on reprogramming may differ depending on cell type or reprogramming strategy (e.g. viral vs transposon, culture conditions, etc…), the claim made in the Rais *et al* study of deterministic 100% reprogramming efficiency upon *Mbd3* KO or KD is likely to be an experimental artefact.

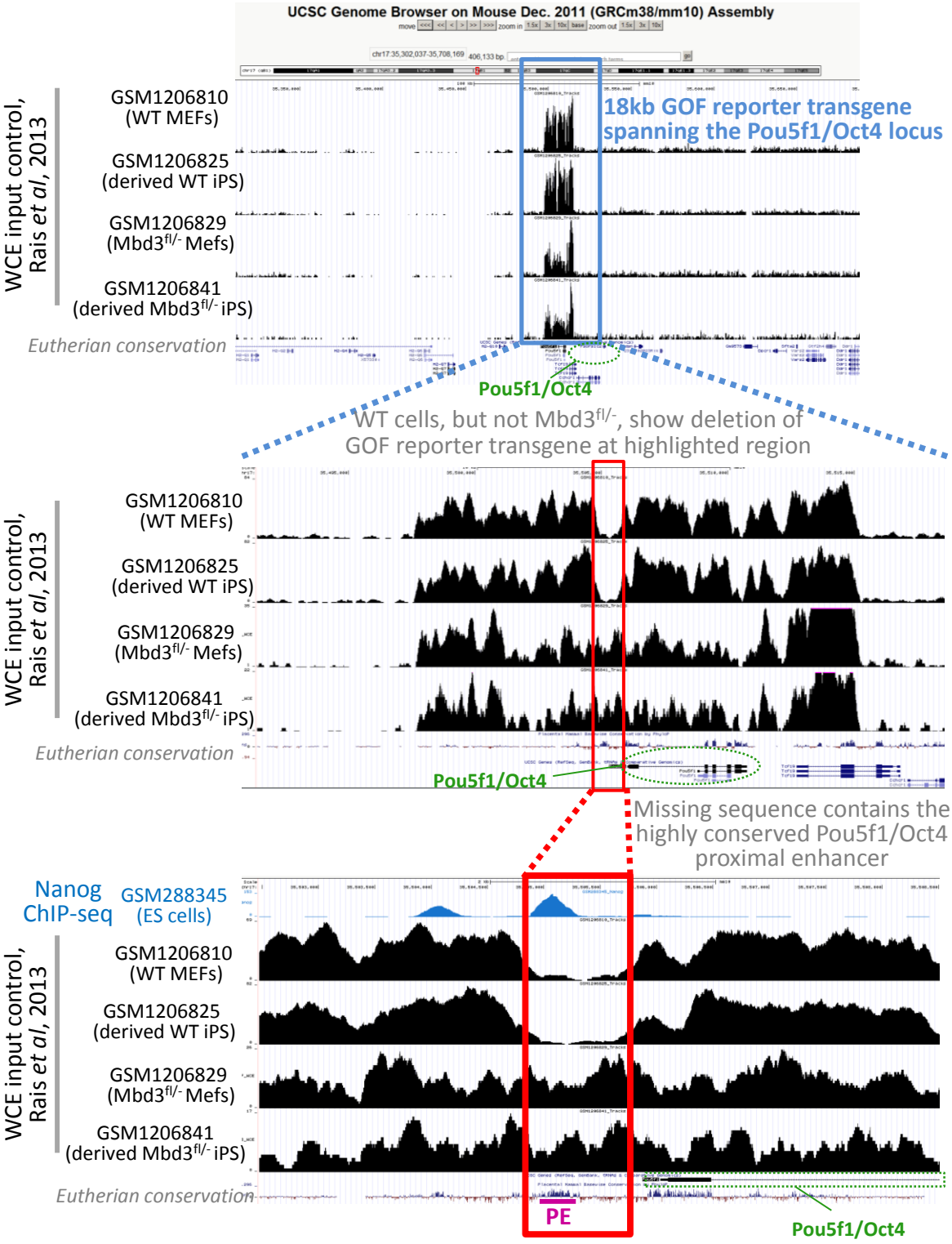
Despite observing an important role for *Mbd3* in various reprogramming contexts, I do not exclude the possibility that these opposing results could be due to some other technical differences between the studies, such as culture media, reprogramming cassettes and pluripotency reporters used.

In this chapter I also described that both *Mbd3*-null ESCs and iPSCs show globally reduced 5-hmC levels, once again raising the hypothesis that the NuRD complex might be controlling 5-hmC levels in the genome, by direct shielding of this mark from further oxidation by recruitment of Tet enzymes for 5-hmC generation. Regardless of the mechanism, this reduction in hydroxymethylation might be responsible for the decreased efficiency of *Mbd3*-null cell reprogramming.

(figure on next page)

Figure 4.3.2 – WT cells, but not Mbd3$^{fl/-}$, show deletion of GOF reporter transgene.

Shown is Rais et al whole-cell extracts (WCE) input control ChIP-seq data of wild type (WT) MEFs, WT iPSCs, *Mbd3*$^{fl/-}$ MEFs and *Mbd3*$^{fl/-}$ iPSCs. Upper panel shows a 400Mb region of chromosome 17 which contains the *Pou5f1* (*Oct4*) locus. Please note the strong signal surrounding the *Pou5f1* locus. This potentially indicates the presence of multiple genomic integrations of the 18Kb Oct4-GFP (GOF-18) (Yeom et al., 1996) reporter transgene in the genome, which could lead to unspecific reporter activation during reprogramming. Middle panel shows a closer magnification of the region exhibiting the strong signal at the *Pou5f1* locus. This reveals a deletion (red box) within the GOF-18 reporter transgene in WT MEFs and derived from them iPSCs. However, this deletion seems not to be present in *Mbd3*$^{fl/-}$ cells. Lower panel indicates that deleted region corresponds to the highly conserved *Pou5f1* regulatory sequence known as the proximal enhancer (PE), lack of which makes the expression of the reporter exclusive to the naïve pluripotent and germ cells (Yoshimizu et al., 1999). ChIP-seq data for Nanog (GSM288345) is indicated as an example for transcription factor occupancy at the *Pou5f1* regulatory sequences (peak on the left corresponds to the distal enhancer, peak on the right – to the proximal enhancer).


UCSC Genome Browser on Mouse Dec. 2011 (GRCm38/mm10) Assembly
WCE input control, Rais et al, 2013
GSM1206810 (WT MEFs)
GSM1206825 (derived WT iPS)
GSM1206829 (Mbd3fl/- Mefs)
GSM1206841 (derived Mbd3fl/- iPS)
Eutherian conservation
18kb GOF reporter transgene spanning the Pou5f1/Oct4 locus
Pou5f1/Oct4
WT cells, but not Mbd3fl/-, show deletion of GOF reporter transgene at highlighted region
Missing sequence contains the highly conserved Pou5f1/Oct4 proximal enhancer
Nanog ChIP-seq
GSM288345 (ES cells)
PE

4.3.2 – Conclusions

In this chapter, by manipulating *Mbd3* levels using genetics or shRNA/siRNA in a number of biological systems, I demonstrated that the NuRD complex facilitates the ability of cells to reprogram to naïve pluripotency. This leads to the hypothesis that in a context dependent manner, the NuRD complex might act as a gene expression activator, and that NuRD's function expands beyond simple gene repression as previously claimed.

CHAPTER 5 – Functional investigation of the effect of *Mbd3* overexpression during the induction of pluripotency

5.1 – Introduction

5.1.1 – Mbd3/ NuRD acting as facilitator of genome-wide reprogramming

In the previous chapters I observed that the complete removal of *Mbd3* or a decrease in its expression can significantly impair the generation of both preiPSCs (chapter 3) and iPSCs (chapter 4). Those observations made us postulate that Mbd3/ NuRD might act as a facilitator of transcription-factor induced pluripotency, probably by interacting with transcription factors to assist genome-wide reprogramming. I hypothesized that the NuRD complex might be recruited to key genes during nuclear reprogramming through interactions with reprogramming factors, facilitating reprogramming, perhaps by enabling the repression of lineage specific genes or by alleviating the repression of key pluripotency genes.

Since loss of NuRD function reduces reprogramming efficiency, I set out to determine the effect of higher Mbd3 and/ or NuRD complex levels on reprogramming (Figure 5.1.1).

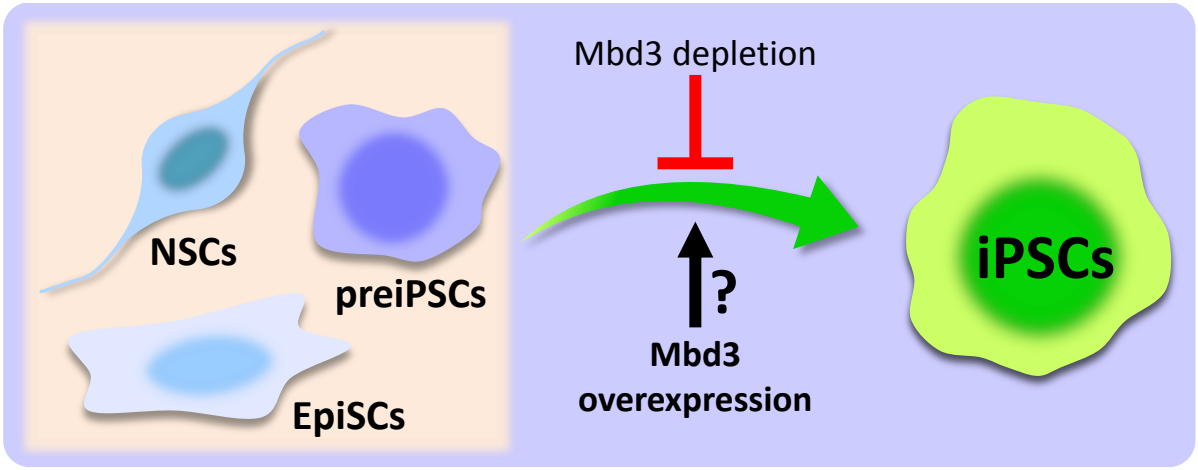


Figure 5.1.1 – Effect of *Mbd3* overexpression on reprograming.

In the previous chapters we described that Mbd3/ NuRD depletion decreases reprogramming efficiency of NSCs, preiPSCs and EpiSCs. In this chapter 5 we assessed the effect of *Mbd3* overexpression reprogramming of these cell types.

5.1.2 – Aim of the chapter

Previously I observed that *Mbd3* depletion can significantly impair nuclear reprogramming. In this chapter I aimed to test whether *Mbd3* levels are limiting for reprogramming, and weather *Mbd3* might enhance transcription-factor mediated reprogramming. For that, I used different reprogramming systems, namely NSCs, preiPSCs and EpiSCs.

5.2 – Results

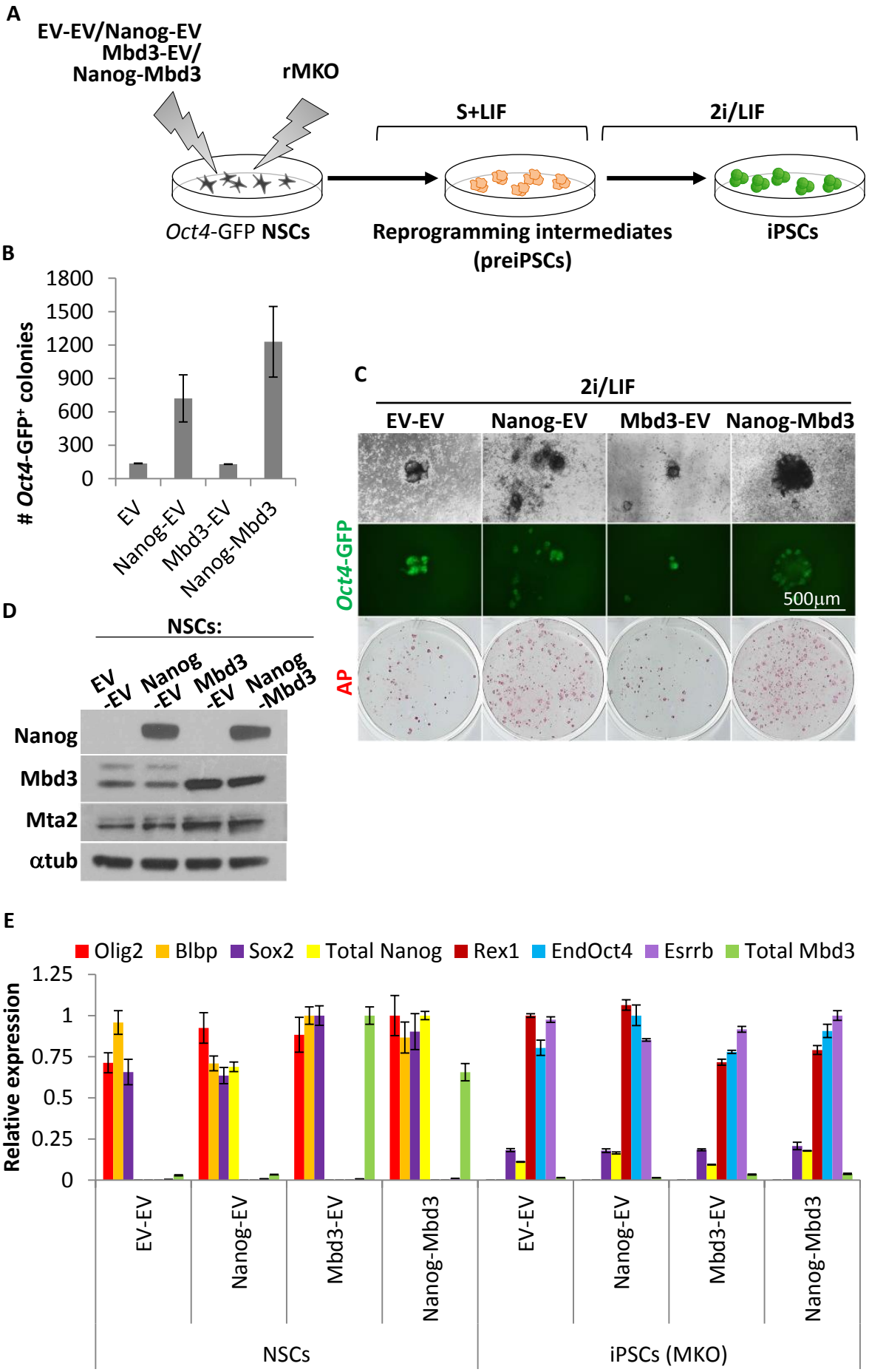
5.2.1 – Overexpression of Mbd3/NuRD facilitates Nanog-mediated reprogramming

To test whether *Mbd3* levels are limiting for reprogramming, *Oct4*-GFP reporter NSCs were stably transfected with *Mbd3*. Stable lines were subsequently transduced with rOKM, cultured in Egf+Fgf2 medium for three days, cultured in S+LIF medium for 6 days and finally switched to 2i/LIF conditions (Figure 5.2.1A). *Mbd3* overexpression had neither a positive nor a detrimental effect on the efficiency of *Oct4*-GFP$^+$ iPSC formation during NSC reprogramming (Figure 5.2.1B-C). Our laboratory has previously reported that *Nanog* overexpression enhances reprogramming efficiency in combination with 2i/LIF culture (Theunissen et al., 2011b). I used *Nanog* overexpression as a control for the enhancement of reprogramming, observing a 5 fold increase in NSC reprogramming. Interestingly, combined overexpression of *Nanog* and *Mbd3* synergistically induced a 10 fold increase in reprogramming efficiency, twice the efficiency of Nanog alone (Figure 5.2.1B). Interestingly, *Mbd3* overexpression in these cell lines caused an increase in the protein levels of Mta2, a core subunit of the NuRD complex which is degraded in the absence of *Mbd3* (Figure 5.2.1B) (Kaji et al., 2006). This suggests that the effects of *Mbd3* overexpression are potentially attributable to the total amount, and, subsequently, the total activity, of the NuRD complex. iPSCs generated from NSCs transfected with any transgene combination showed a gene expression signature typical of naïve pluripotent cells (Figure 5.2.1E).

(figure on next page)

Figure 5.2.1 – Mbd3 overexpression does not affect Neural Stem Cell reprogramming.

(A) Experimental design used to address the effect of Mbd3 overexpression on NSC reprogramming. NSCs carrying an *Oct4*-GFP cassette were stably transfected with pPB-CAG-Nanog and pPB-CAG-Mbd3b or pPB-CAG-empty controls, transduced with rOKM, cultured in Egf+Fgf2 medium for three days, switched to S+LIF medium for 6 days and then switched to 2i/LIF conditions. **(B)** Quantification of *Oct4*-GFP$^+$ colonies after 12 days in 2i/LIF conditions. Colony number is per 1.0 x 10^5 NSCs. **(C)** Phase and GFP images and AP staining of the iPSCs obtained from NSCs overexpressing respective transgenes. **(D)** Western blot analysis of Nanog, Mbd3, Mta2 and α-Tubulin (αtub) protein expression in NSCs overexpressing the indicated transgene combinations. **(E)** qRT-PCR analysis of NSC- and pluripotency-associated genes, and *Mbd3* in transgenic NSCs and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.


A
EV-EV/Nanog-EV
Mbd3-EV/
Nanog-Mbd3
rMKO
S+LIF
2i/LIF
Oct4-GFP NSCs
Reprogramming intermediates
(preiPSCs)
iPSCs
B
Oct4-GFP+ colonies
1800
1500
1200
900
600
300
0
EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
C
2i/LIF
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Oct4-GFP
AP
500µm
D
NSCs:
EV -EV
Nanog -EV
Mbd3 -EV
Nanog -Mbd3
Nanog
Mbd3
Mta2
αtub
E
Olig2
Blbp
Sox2
Total Nanog
Rex1
EndOct4
Esrrb
Total Mbd3
Relative expression
1.25
1
0.75
0.5
0.25
0
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
NSCs
iPSCs (MKO)

I then assessed the impact of Mbd3 overexpression in other reprograming systems. First, I tested if *Mbd3* expression could overcome the reprogramming block of preiPSCs. For that, *Nanog*-GFP preiPSCs were stably transfected with the same transgene combinations as the NSCs above (Figure 5.2.2A). Since I wanted to evaluate the ability of *Mbd3* to enhance reprogramming efficiency, I chose a particular preiPSC line which was derived from MEFs and that shows inefficient conversion rate to naïve pluripotency in 2i/LIF conditions. This is a suitable cell system for gain-of-function studies and this cell line has previously been used in our laboratory to dissect the Nanog-Tet1/2 synergistic effect on reprogramming (Costa et al., 2013). As observed for NSCs, Mbd3 overexpression in preiPSCs also caused an increase in the protein levels of Mta2 (Figure 5.2.2B).

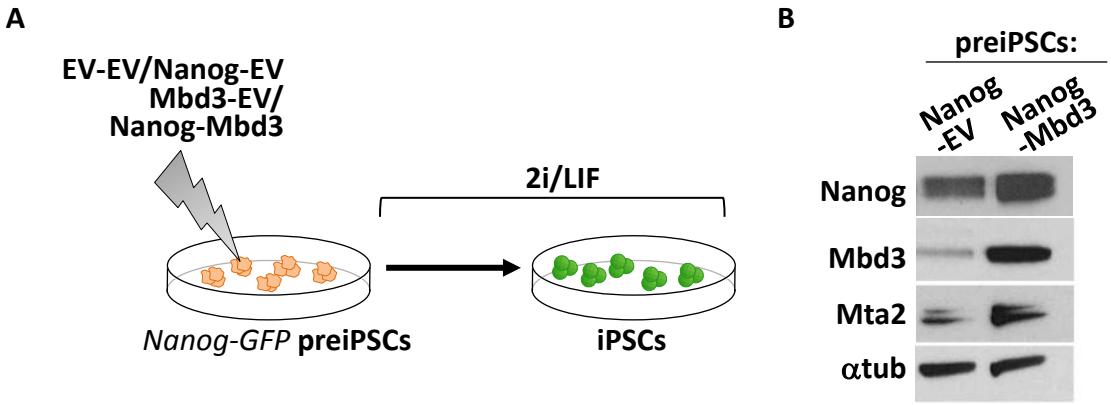


Figure 5.2.2 – Mbd3 overexpression in preiPSCs leads to increased levels of other NuRD subunits.

(A) Experimental design used to address the effect of Mbd3/NuRD overexpression on the conversion of preiPSCs to iPSCs. PreiPSCs (carrying a *Nanog*-GFP) were stably transfected with pPB-CAG-Nanog and pPB-CAG-Mbd3b or pPB-CAG-empty controls and plated in 2i/LIF conditions for 12 days. **(B)** Western blot analysis of Nanog, Mbd3, Mta2 and α-Tubulin (αtub) protein expression in preiPSCs overexpressing Nanog or Nanog and Mbd3.

To study the conversion of the established transgenic preiPSC lines to naïve pluripotency, these were plated in S+LIF medium, and switched to 2i/LIF conditions 2-3 days afterwards. Interestingly, the combined overexpression of *Mbd3* and *Nanog* in MEF-derived preiPSCs led to accelerated reprogramming kinetics, and *Nanog*-GFP^{+} colonies were visible as early as 24h after the beginning of 2i/LIF culture (Figure 5.2.3).

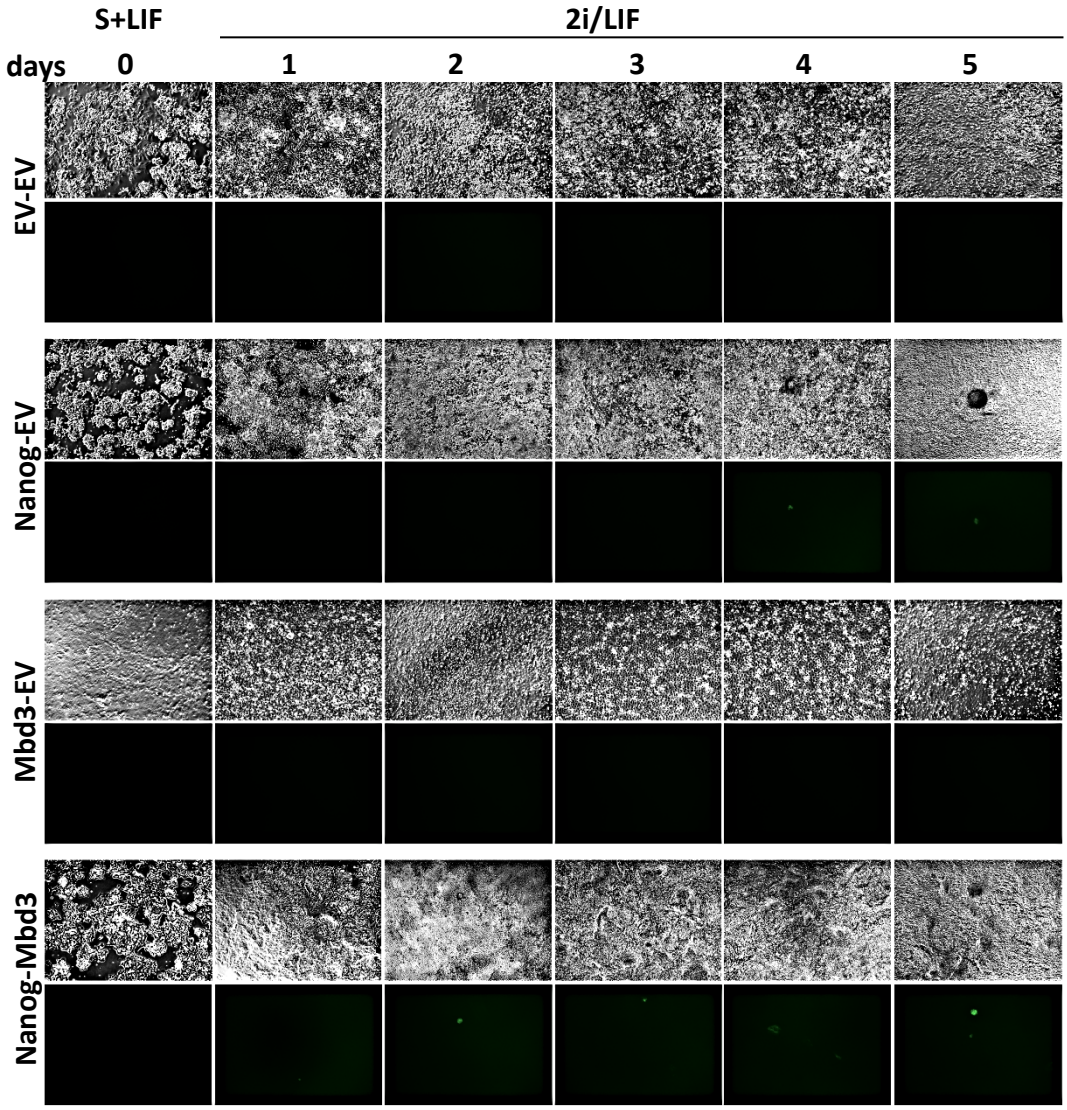
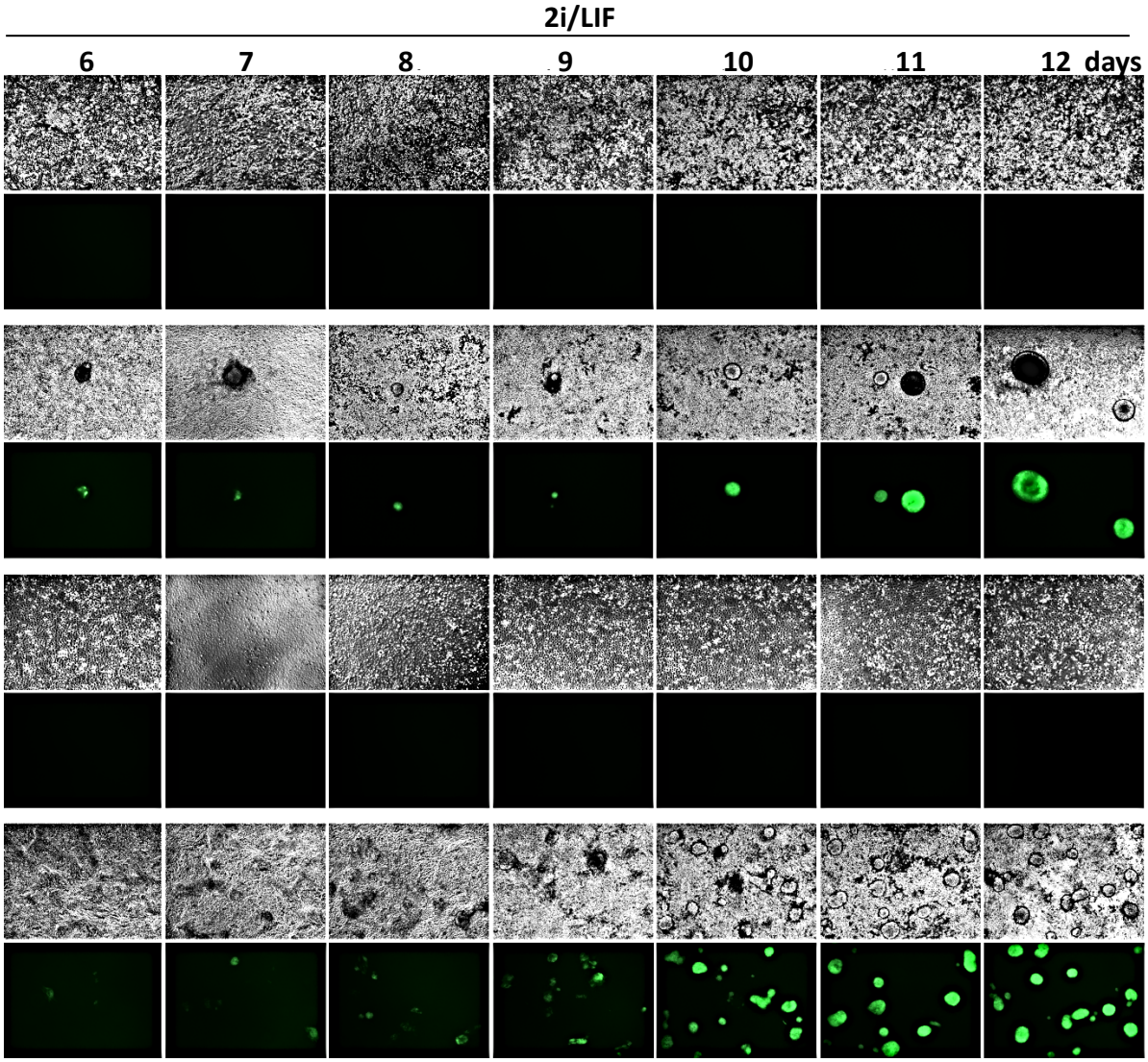


Figure 5.2.3 – Nanog-Mbd3 preiPSCs show increased reprogramming.

Phase and GFP images acquired before and after medium switch to 2i/LIF of the different transgenic *Nanog*-GFP preiPSCs lines. Images were taken every day during the 12 days of the experiment. The field of vision is not the same in the different images correspondent to the same transgenic line.

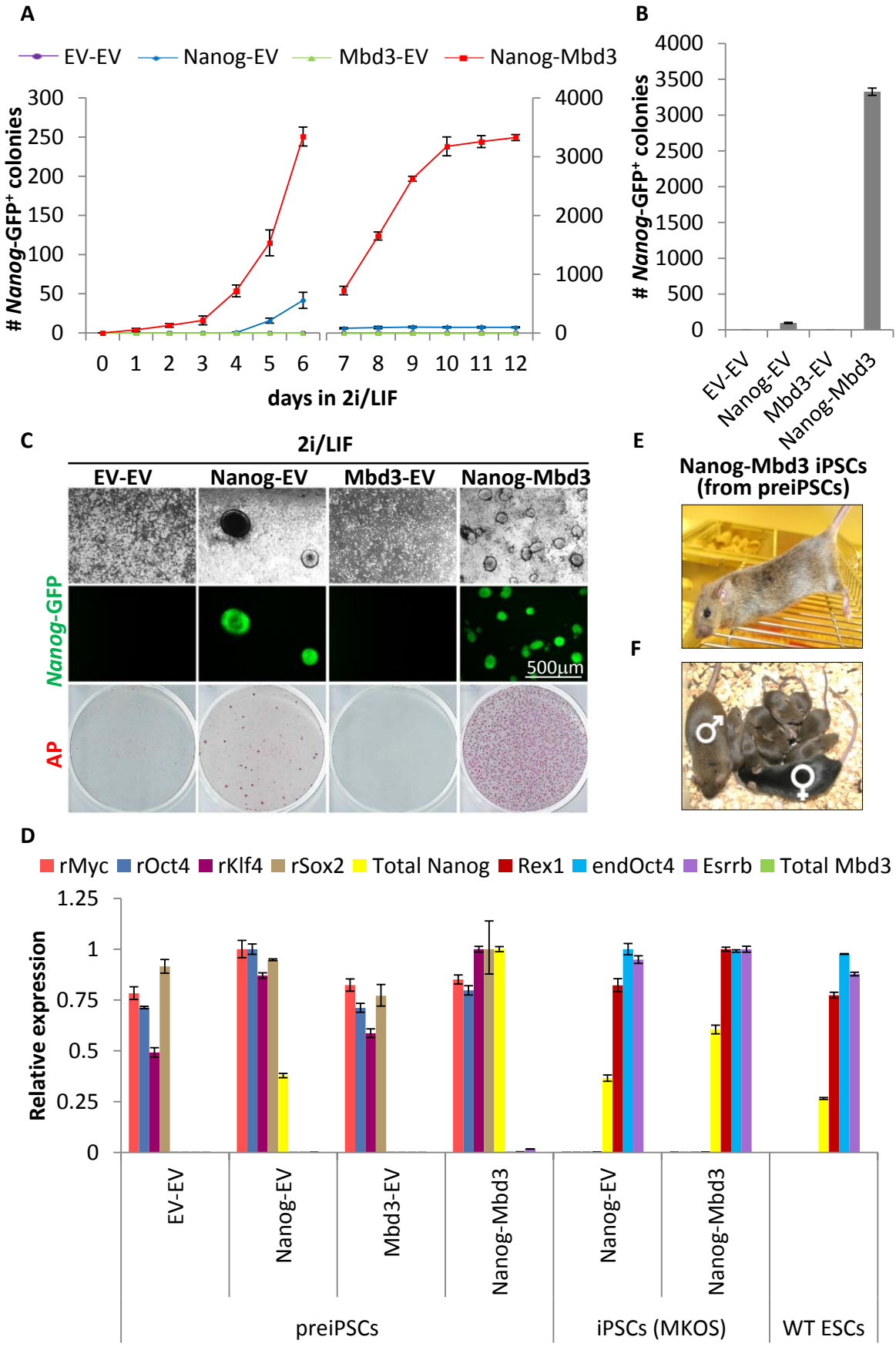

2i/LIF
6
7
8
9
10
11
12 days

I observed that the combined forced expression of *Nanog-Mbd3* not only facilitated preiPSC's conversion to naïve pluripotency in rate but also in extent (Figure 5.2.4A), and reprogramming efficiency was up to 30 fold higher compared to the *Nanog-Empty vector* (EV) control (Figure 5.2.4B). As expected, this preiPSC line, which was recalcitrant to reprogramming, failed to convert to naïve pluripotency after its transfection with EV controls (Figure 5.2.4C), since preiPSC were not allowed to expand for more than 3 days before 2i/LIF switch. Moreover, the overexpression of *Mbd3* alone was not sufficient to overcome this reprogramming block, showing that NuRD alone is not sufficient to induce reprogramming and only acts as a facilitator of Nanog-mediated reprogramming. Importantly, the iPSCs generated from preiPSCs by the overexpression of *Nanog* and *Mbd3*, exhibited the molecular properties of naïve pluripotent cells (Figure 5.2.4D), as well as the capacity for chimera contribution (Figure 5.2.4E) and germline competence (Figure 5.2.4F) after the excision of the reprogramming transgenes.

(figure on next page)

Figure 5.2.4 – Overexpression of Mbd3/NuRD facilitates Nanog-mediated preiPSC reprogramming.

(A) The kinetics of the emergence of *Nanog*-GFP$^+$ colonies from the transgenic preiPSCs during a 12-day culture in 2i/LIF conditions (y-axis scale changes at day 7). Colony number is per 1.0 x 10^5 preiPSCs. Representative field of vision of each time point can be seen in Figure 5.2.2. **(B)** Quantification of *Nanog*-GFP$^+$ colonies after 12 days of 2i/LIF culture. Colony numbers are per 1.0 x 10^5 preiPSCs. **(C)** Phase and GFP images and AP staining of the iPSCs formed from preiPSCs overexpressing respective transgenes. **(D)** qRT-PCR analysis of retroviral transgenes, pluripotency-associated, and *Mbd3* in preiPSCs and corresponding derived iPSCs. **(E)** Male chimera showing contribution from Nanog-Mbd3 iPSCs generated from preiPSCs (brown colour). **(F)** Germline contribution of Nanog-Mbd3 iPSCs. Cells were treated with TAT-Cre for reprogramming transgene excision prior to blastocyst injection. Chimeric father, C57BL/6 mother and pups resulting from cross can be viewed.


A
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Nanog-GFP+ colonies
days in 2i/LIF
B
Nanog-GFP+ colonies
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
C
2i/LIF
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Nanog-GFP
AP
500µm
E
Nanog-Mbd3 iPSCs
(from preiPSCs)
F
D
rMyc
rOct4
rKlf4
rSox2
Total Nanog
Rex1
endOct4
Esrrb
Total Mbd3
Relative expression
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Nanog-EV
Nanog-Mbd3
preiPSCs
iPSCs (MKOS)
WT ESCs

After careful analysis of the *Nanog* transcript (Figure 5.2.4D) and protein levels (Figure 5.2.2B) in the different transgenic preiPSC lines, I detected that *Nanog* transgenic levels were 2 fold higher in *Nanog-Mbd3* preiPSCs compared to *Nanog-EV* preiPSCs. To exclude the possibility of this affecting the results, I generated new *Nanog-EV* and *Nanog-Mbd3* preiPSC lines with similar transgenic *Nanog* levels (Figure 5.2.5A). As with the previous preiPSC lines, higher amounts of Mta2 were observed in the *Nanog-Mbd3* preiPSC line (Figure 5.2.5A), and the synergistic gain-of-function reprogramming phenotype after combined forced expression of *Nanog* and *Mbd3* was again observed (Figures 5.2.5B-C). The iPSCs generated from these new preiPSC lines expressed similar levels of transgenic *Nanog* were molecularly indistinguishable from the iPSCs generated in Figure 5.2.4, displaying the expected molecular properties of naïve pluripotent cells (Figures 5.2.5D).

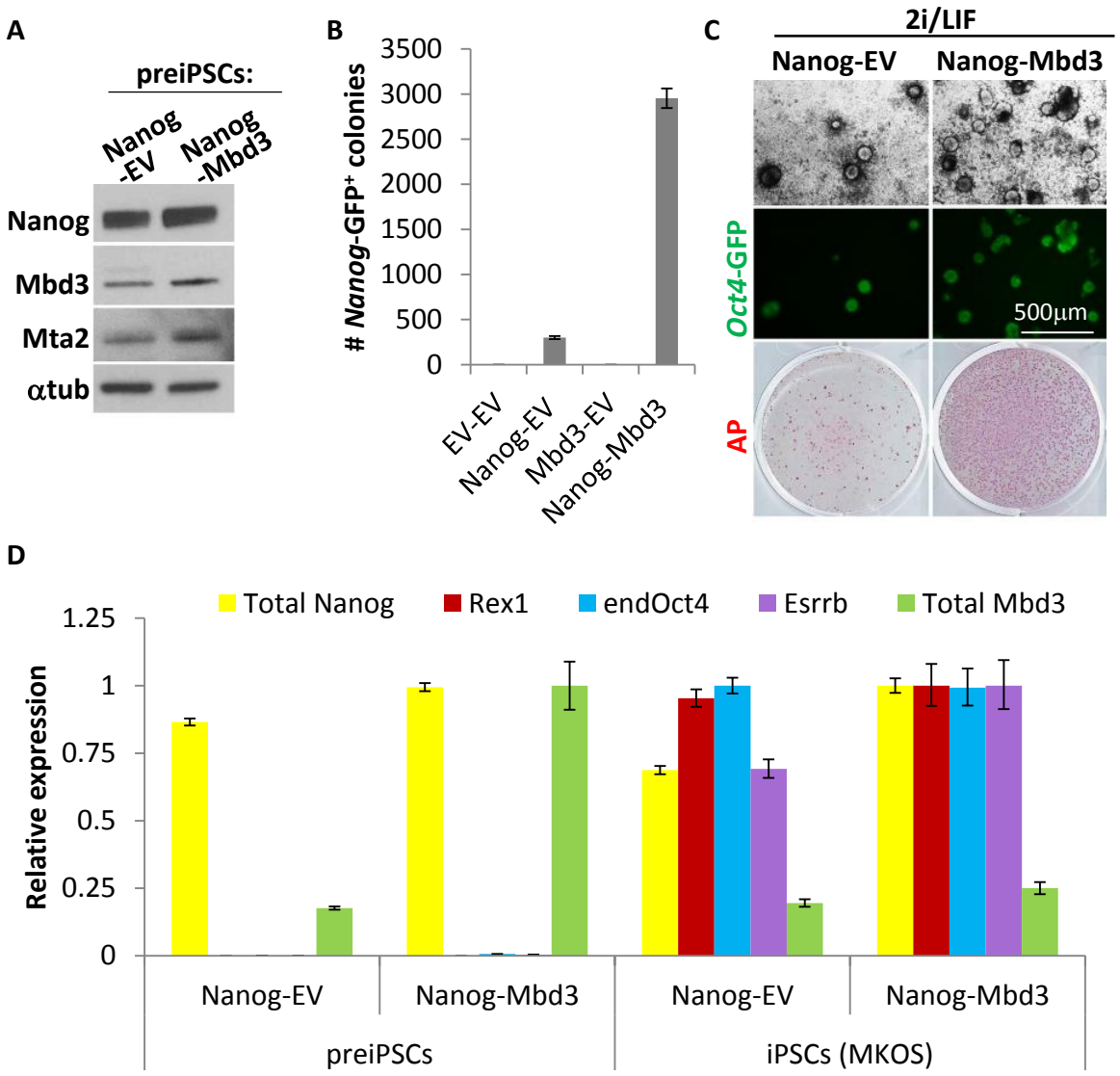


Figure 5.2.5 – Nanog/Mbd3 synergy is independent of Nanog transgenic levels.

(A) Western blot analysis of Nanog, Mbd3, Mta2 and α-Tubulin (αtub) protein expression in preiPSCs overexpressing Nanog or Nanog and Mbd3. It is notable that expression level of Nanog transgene in the cell lines are similar (refer to Figure 5.2.5 D). **(B)** Quantification of *Nanog*-GFP$^+$ after 12 days of 2i/LIF culture. Colony numbers are per 1.0 x 10^5 preiPSCs. **(C)** Phase and GFP images and AP staining of the iPSCs formed from preiPSCs overexpressing respective transgenes. **(D)** qRT-PCR analysis of pluripotency-associated genes and *Mbd3* in transgenic preiPSCs and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.

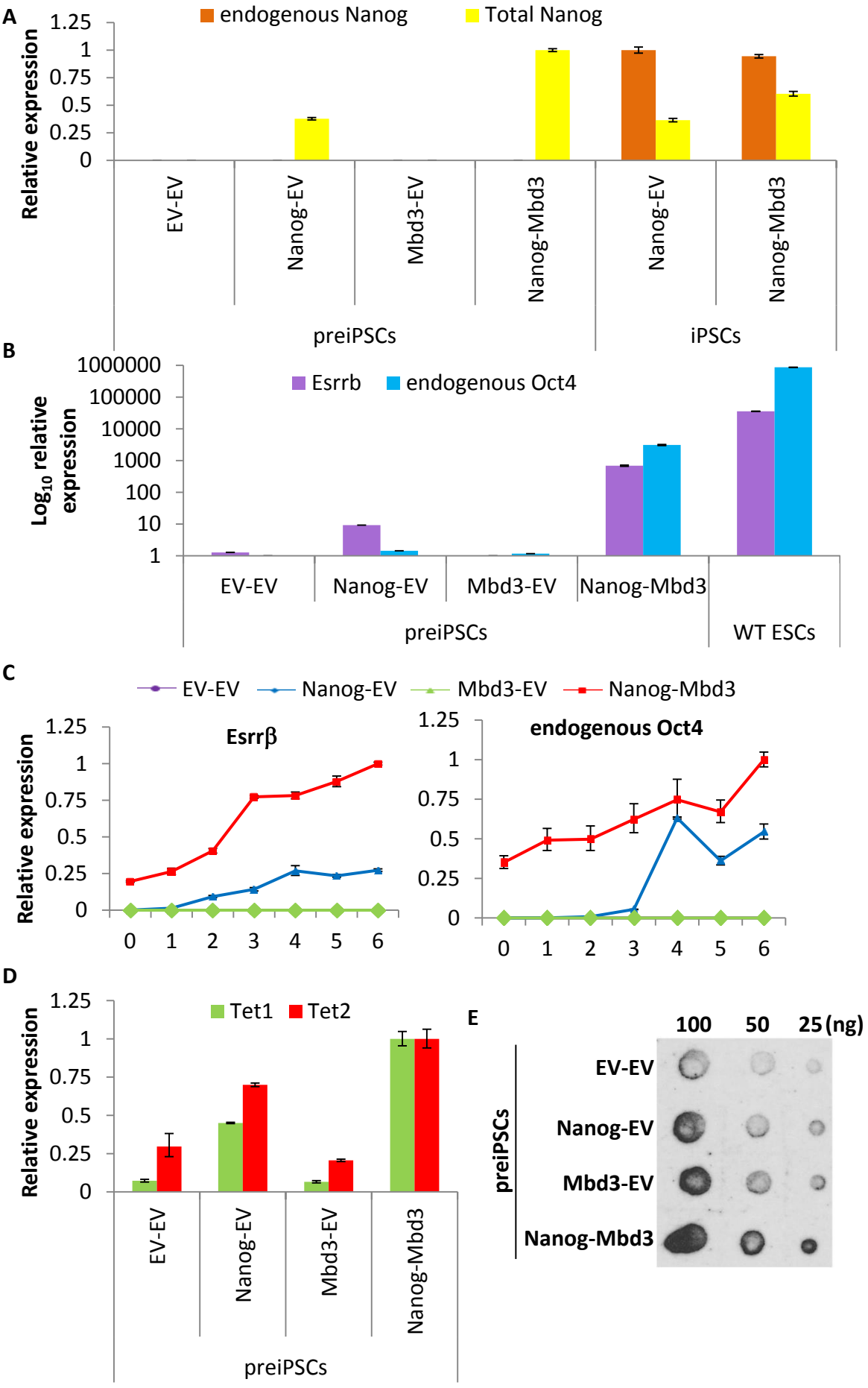
To gain some insights into the mechanism of the synergy between *Nanog* and *Mbd3*, I looked for signs of the expression of naïve pluripotency genes in these established preiPSC lines, which could explain the gain-of-function reprogramming phenotype. These analyses of primed expression were carried out 12 days after stable transfection of the preiPSCs which are cultured in S+LIF. Using TaqMan probes that distinguish between total *Nanog* (endogenous

plus transgenic) and endogenous *Nanog*, I observed that all the transgenic preiPSC lines express only background levels of endogenous *Nanog* prior to the induction of pluripotency by 2i/LIF medium switch (Figure 5.2.6A), indicating no priming of endogenous *Nanog* expression. However, I observed a significant primed expression of the key pluripotency genes *Esrrβ* and endogenous *Oct4* (Figure 5.2.6B). I noticed that the expression levels of *Esrrβ* and endogenous *Oct4* in the *Nanog-Mbd3* preiPSCs are already 5% and 3% of wild type ESC levels in 2i/LIF, respectively. The expression level of *Esrrβ* is also approximately 80 times greater than that of Nanog-EV preiPSCs and 700 times greater than that of *Mbd3-EV* and *EV-EV* preiPSCs. The expression level of endogenous *Oct4* is approximately 2000 times greater than that of *Nanog-EV* preiPSCs and 3000 times greater than that of *Mbd3*-EV and *EV-EV* preiPSCs. The *Esrrβ* and endogenous *Oct4* expression levels continued to be higher in *Nanog-Mbd3* than *Nanog-EV* preiPSCs after the medium switch to 2i/LIF (Figure 5.2.6C). After only 4 days of 2i/LIF culture the levels of *Esrrβ* and endogenous *Oct4* in *Nanog-EV* transgenic lines matched the levels of these genes in the *Nanog-Mbd3* preiPSCs cultured in S+LIF, indicating clear priming for conversion to naïve pluripotency. I have also observed a 2 fold increase in *Tet1/2* expression in *Nanog-Mbd3* preiPSCs in comparison to *Nanog-EV* preiPSCs, and 4 fold increase compared to *Mbd3-EV* or *EV-EV* preiPSC lines (Figure 5.2.6D). This is of extreme importance, since both *Oct4* and *Esrrβ* are genes are known to mark cells that are on a route to become fully reprogrammed (Buganim et al., 2012). Strikingly, *Nanog-Mbd3* preiPSCs exhibited profoundly higher levels of 5-hmC than *Nanog-EV* or *Mbd3-EV* preiPSCs (Figure 5.2.6E), a phenomena previously observed in preiPSCs overexpressing *Nano*g and *Tet1/2* (Costa et al., 2013). This genome-wide increase in 5-hmC, a marker assumed to be associated with gene activation, could explain *Esrrβ* and endogenous *Oct4* primed expression with the concomitant gain-of-function reprogramming phenotype of *Nanog-Mbd3* preiPSCs.

(figure on next page)

Figure 5.2.6 – Nanog/Mbd3 overexpression primes preiPSCs to reprogram to naïve pluripotency.

(A) qRT-PCR analysis of endogenous Nanog and Total Nanog in transgenic preiPSCs and respective generated iPSCs. **(B)** qRT-PCR analysis of *Esrrβ* and endogenous (end) *Oct4* expression in preiPSC 12 days after stable transgene transfection and culture in S+LIF (*y*-axes in $\log_{10}$ scale). The expression levels of *Esrrβ* and endogenous *Oct4* in these Nanog-Mbd3 preiPSC are 5% and 3%, respectively, of the expression levels of WT ESCs in 2i/LIF. **(C)** Time course qRT-PCR analysis of *Esrrβ* and endogenous *Oct4* expression after medium switch to 2i/LIF of the indicated preiPSCs lines. **(D)** qRT-PCR analysis of *Tet1/2* expression levels in transgenic preiPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV. **(E)** DNA dot blot analysis of 5-hydroxymethylation (5-hmC) bulk levels in transgenic preiPSCs.

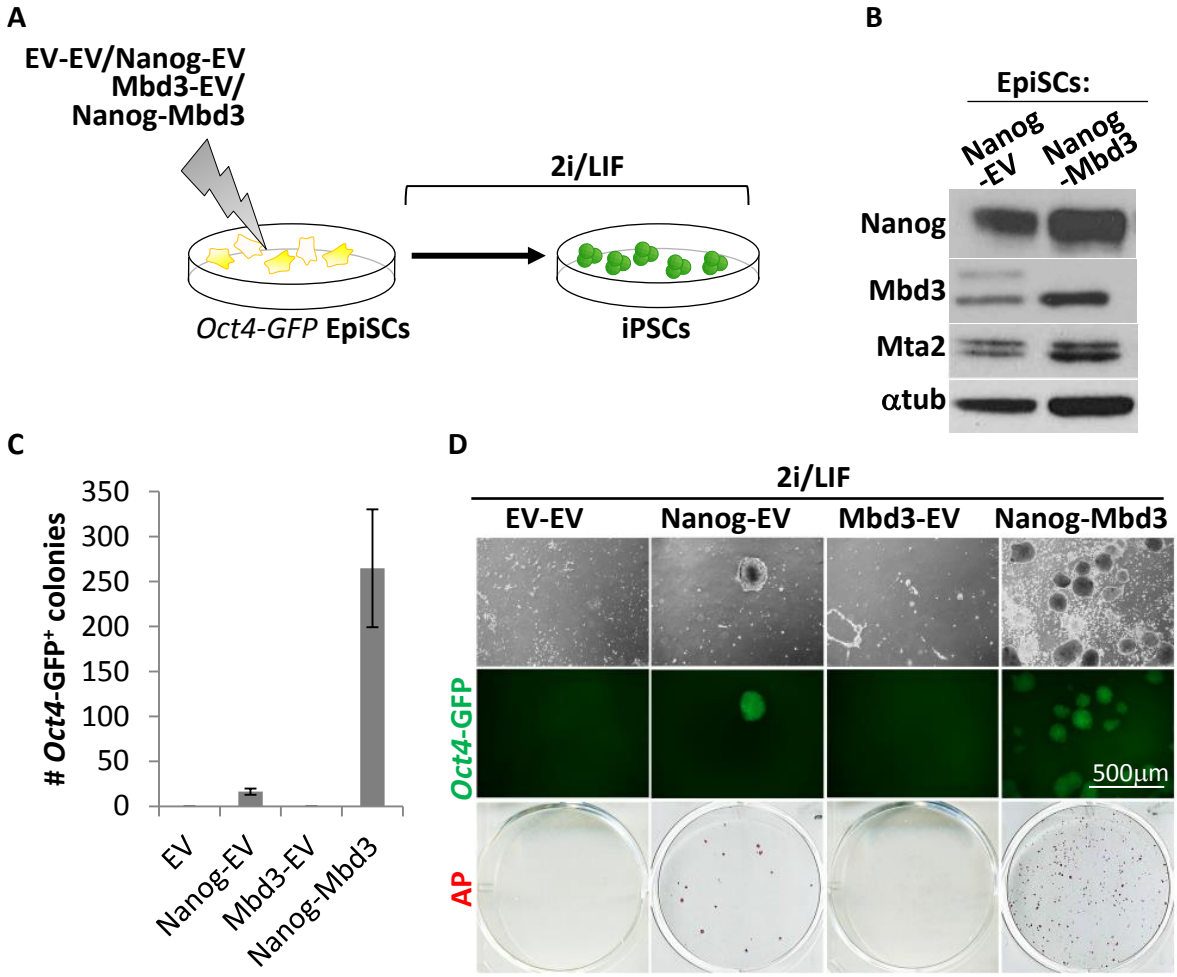

A
endogenous Nanog
Total Nanog
Relative expression
1.25
1
0.75
0.5
0.25
0
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Nanog-EV
Nanog-Mbd3
preiPSCs
iPSCs
B
Esrrb
endogenous Oct4
Log10 relative expression
1000000
100000
10000
1000
100
10
1
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
preiPSCs
WT ESCs
C
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Esrrβ
endogenous Oct4
Relative expression
1.25
1
0.75
0.5
0.25
0
0 1 2 3 4 5 6
D
Tet1
Tet2
Relative expression
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
preiPSCs
E
100
50
25 (ng)
preiPSCs
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3

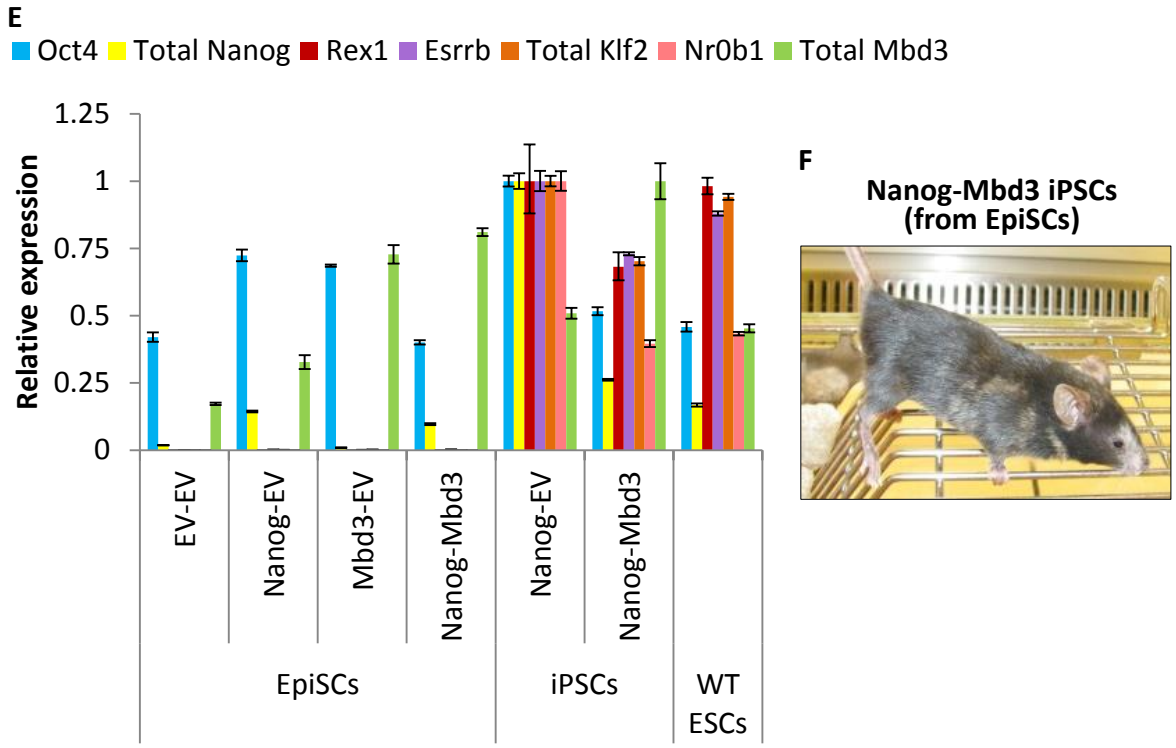
Above I showed that preiPSC reprogramming to naïve pluripotency is greatly enhanced by *Nanog-Mbd3* overexpression. I then decided to test if similar synergy could be observed in EpiSC reprogramming. For that, *Oct4*-GFP EpiSCs were stably transfected with the same transgene combinations as the NSCs and preiPSCs above (Figure 5.2.7A). Similarly to preiPSCs, EpiSCs overexpressing both *Nanog* and *Mbd3* showed also higher levels of the NuRD subunit Mta2 (Figure 5.2.7B), and a 30 fold increase in the ability to generate iPSCs relative to a *Nanog-EV* control (Figures 5.2.7C-D). iPSCs generated from EpiSCs reactivated the naïve pluripotency-associated transcriptional program (Figure 5.2.7E), and the iPSCs obtained by the overexpression of *Nanog* and *Mbd3* were able to contribute to chimeras upon blastocyst injection (Figure 5.2.7F).

(figure on next page)

Figure 5.2.7 – Overexpression of Mbd3/NuRD facilitates Nanog-mediated EpiSC reprogramming.

(A) Experimental design used to address the effect of Mbd3/NuRD overexpression on EpiSC reprogramming. EpiSCs (carrying an *Oct4*-GFP) were stably transfected with pPB-CAG-Nanog and pPB-CAG-Mbd3b or pPB-CAG-empty controls and plated in 2i/LIF conditions for 12 days. **(B)** Western blot analysis of Nanog, Mbd3, Mta2 and α-Tubulin (αtub) protein expression in EpiSCs overexpressing Nanog or Nanog and Mbd3. **(C)** Quantification of *Oct4*-GFP^{+} colonies after 12 days of 2i/LIF culture. Colony numbers are per 2.0 x 10^{4} EpiSCs. **(D)** Phase and GFP images and AP staining of the iPSCs formed from EpiSCs overexpressing respective transgenes. **(E)** qRT-PCR analysis of pluripotency-associated genes and *Mbd3* in transgenic EpiSCs and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV. **(F)** Chimera of Nanog-Mbd3 iPSCs generated from EpiSCs (brown colour).


A
EV-EV/Nanog-EV
Mbd3-EV/
Nanog-Mbd3
2i/LIF
Oct4-GFP EpiSCs
iPSCs
B
EpiSCs:
Nanog -EV
Nanog -Mbd3
Nanog
Mbd3
Mta2
αtub
C
Oct4-GFP+ colonies
350
300
250
200
150
100
50
0
EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
D
2i/LIF
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Oct4-GFP
AP
500µm


E
Oct4
Total Nanog
Rex1
Esrrb
Total Klf2
Nr0b1
Total Mbd3
Relative expression
1.25
1
0.75
0.5
0.25
0
EV-EV
Nanog-EV
Mbd3-EV
Nanog-Mbd3
Nanog-EV
Nanog-Mbd3
EpiSCs
iPSCs
WT ESCs
F
Nanog-Mbd3 iPSCs
(from EpiSCs)

5.2.2 – Synergy is specific to Nanog-driven reprogramming

I observed that increased Mbd3/NuRD levels can enhance reprogramming efficiency when co-overexpressed with the reprogramming factor *Nanog*. But is this synergy unique to Nanog? I showed that *Mbd3* overexpression alone failed to alleviate the reprogramming block of MEF-derived preiPSCs that express high amounts of retroviral *Oct4*, *Klf4*, *cMyc* and *Sox2* (Figures 5.2.4A-C). This is suggestive that increased levels of *Mbd3* do not synergise with these reprogramming factors. I checked if *Mbd3* overexpression could enhance reprogramming efficiency during *Klf4* and *Nr5a2* induced reprogramming. These are two well characterized EpiSC reprogramming factors (Guo and Smith, 2010; Guo et al., 2009). For that, *Oct4*-GFP EpiSCs were stably transfected with those transcription-factors alone or in combination with *Mbd3*, and plated in 2i/LIF conditions for conversion to naïve pluripotency. I found that combined overexpression of *Mbd3* with *Klf4* or *Nr5a2* in EpiSCs does not confer an increase in reprogramming efficiency (Figures 5.2.8A-B), although the reprogramming factors are expressed under all conditions at similar levels (Figure 5.2.8C). Thus, the observed synergy with *Mbd3* overexpression seemed to be specific to Nanog-mediated reprogramming.

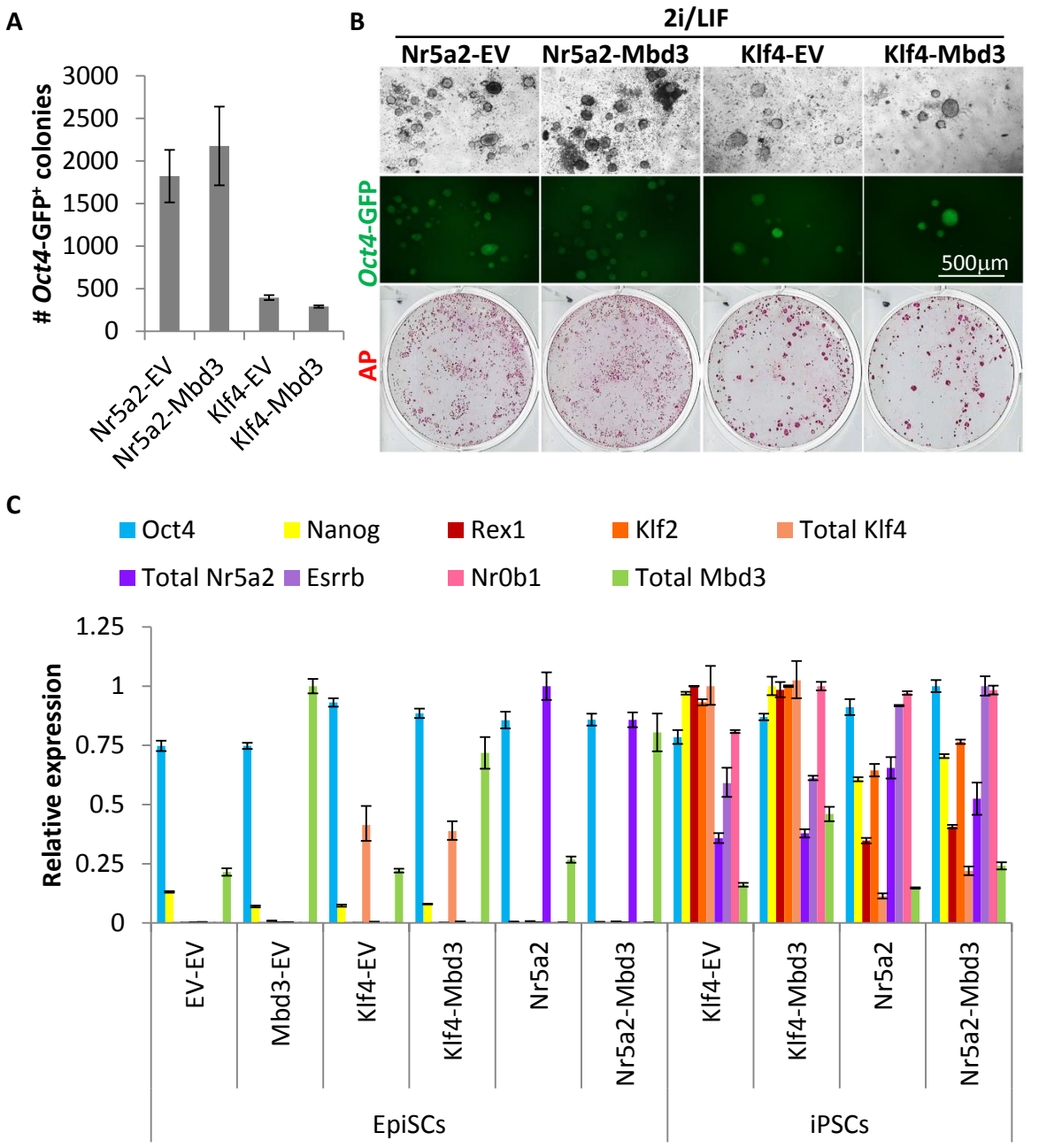


Figure 5.2.8 – The reprogramming synergy with the NuRD complex is specific to Nanog.

(A) Quantification of *Oct4*-GFP^{+} colonies after 12 days of 2i/LIF culture, generated from EpiSCs transfected with the indicated transgene combinations. Colony numbers are per 2.0 x 10^{4} EpiSCs. **(B)** Phase and GFP images and AP staining of the iPSCs formed from EpiSCs overexpressing respective transgenes. **(C)** qRT-PCR analysis of pluripotency-associated genes and *Mbd3* in transgenic EpiSCs and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.

Having identified *Nanog* as a unique factor that synergises with *Mbd3* during reprogramming, I next aimed to understand how this synergy occurs. I performed some structure-function experiments to understand which domains of *Mbd3* are important for the observed reprograming synergy with *Nanog*. It has been previously suggested that the N-terminus of *Mbd3* is required for protein-protein interaction (Aguilera et al., 2011). The *Mbd3* isoform used for the rescue and overexpression experiments was *Mbd3b*, the most abundant isoform by protein levels (Figure 3.2.1B). As shown in Figure 3.2.1F, due to alternative start sites, there are two more Mbd3 isoforms. To test if the N-terminus of *Mbd3b* is required for the synergy with *Nanog*, I tested whether overexpression of the *Mbd3c* isoform affected MEF-derived preiPSC reprogramming, since the *Mbd3c* isoform differs from *Mbd3b* in the first 60 N-terminal amino acids (Figure 5.2.9A). The same amount of *Nanog*-GFP^{+} colonies was scored after 12 days of 2i/LIF culture of both *Nanog-EV* and *Nanog-Mbd3c* transgenic lines (Figures 5.2.9B-C). Interestingly, Mta2 levels were not found to be elevated after the overexpression of *Mbd3c* (Figure 5.2.9D). As with the iPSCs generated from *Nanog-Mbd3b* preiPSCs, *Nanog-Mbd3c* iPSCs reactivated the core naïve pluripotency-associated transcriptional program (Figure 5.2.9E).

In conclusion, the results presented in this chapter indicate that *Mbd3* overexpression does not impair the induction of naïve pluripotency. Moreover, it specifically facilitates Nanog-mediated reprogramming. I also demonstrate that the N-terminal region of *Mbd3b* is required for *Nanog-Mbd3* reprogramming synergy.

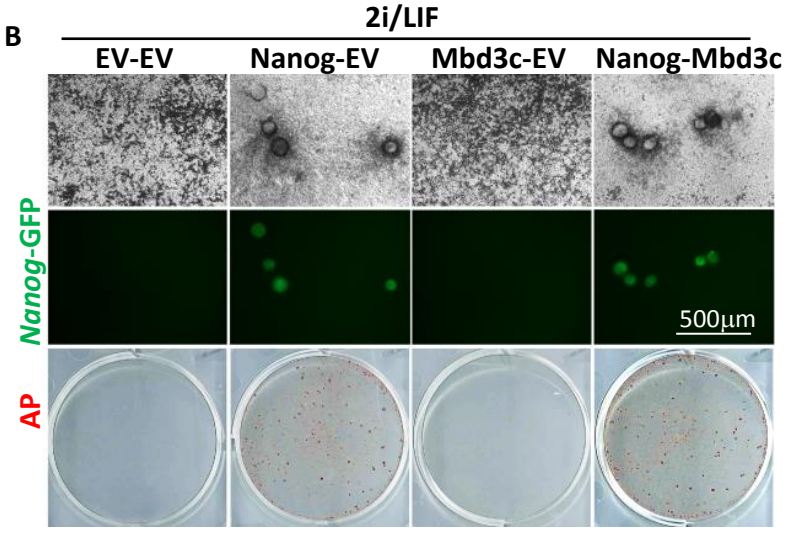
(figure on next page)

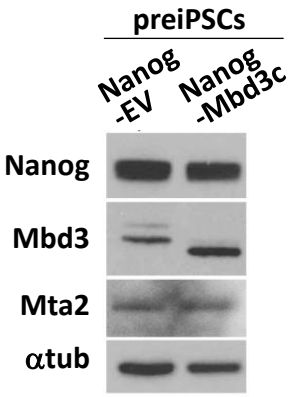
Figure 5.2.9 – The N-terminal sequence of Mbd3 is required for the synergistic effect with Nanog in reprogramming.

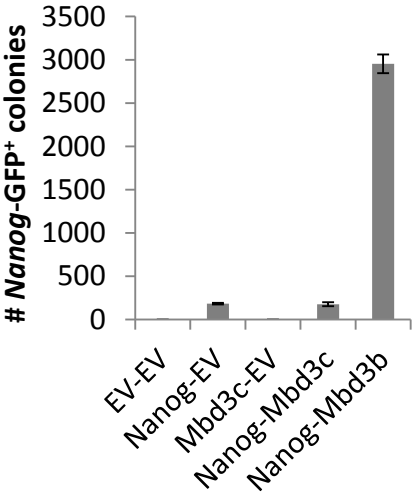
(A) ClustalW2 sequence alignment between Mbd3b isoform (cDNA used in this study unless otherwise stated) and Mbd3c isoform (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The isoforms differ in their N-terminus sequence, because of different translation start, Mbd3c being 8 amino acids shorter (dotted box). Sequence in red is hypothesized to be the domain responsible for Mbd3 protein-protein interactions. **(B)** Phase and GFP images and AP staining of the iPSCs formed from preiPSCs overexpressing respective transgenes. **(C)** Quantification of *Nanog*-GFP^{+} colonies after 12 days of 2i/LIF culture. Colony numbers are per 1.0 x 10^{5} preiPSCs. **(D)** Western blot analysis of Nanog, Mbd3, Mta2 and α-Tubulin (αtub) protein expression in preiPSCs overexpressing Nanog or Nanog and Mbd3c. **(E)** qRT-PCR analysis of pluripotency-associated genes and *Mbd3* in transgenic preiPSCs and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value. The error bars indicate STDEV.

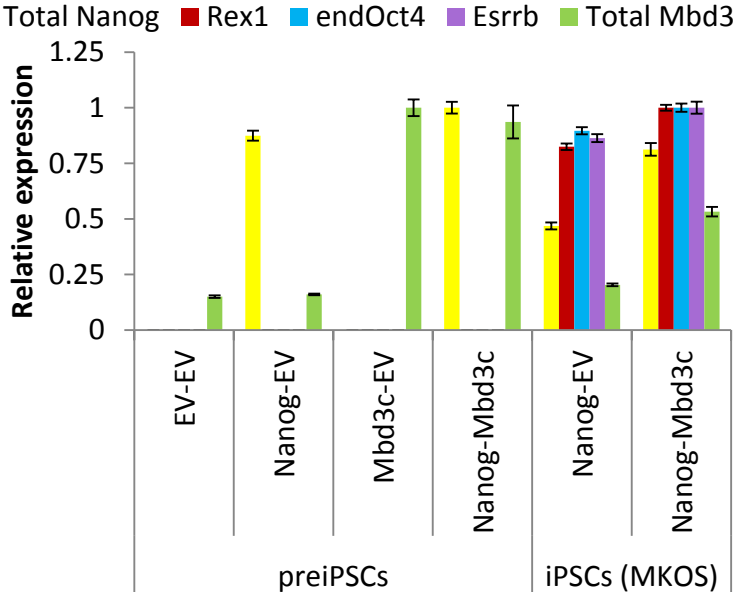
A

```
Mbd3b   MERKSPSGKKFRSKPQLARYLGGSMDLSTFDFRTGKMLMNKMNKSRQRVRYDSSNQVKGK 60
Mbd3c   MARIWFGGWEISVP---DRPPATIQALAKHLPGPSNPPWTPVGAARCRVFSP-----QGK 52
         * *   .* ::       *  .    *:..   ..:   . :. :* **         :**

Mbd3b   PDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQPRQLFWEKKLSGLSAFDIA 120
Mbd3c   PDLNTALPVRQTASIFKQPVTKITNHPSNKVKSDPQKAVDQPRQLFWEKKLSGLSAFDIA 112
        ************************************************************

Mbd3b   EELVRTMDLPKGLQGVGPGCTDETLLSAIASALHTSTLPITGQLSAAVEKNPGVWLNTAQ 180
Mbd3c   EELVRTMDLPKGLQGVGPGCTDETLLSAIASALHTSTLPITGQLSAAVEKNPGVWLNTAQ 172
        ************************************************************

Mbd3b   PLCKAFMVTDDDIRKQEELVQQVRKRLEEALMADMLAHVEELARDGEAPLDKACAEEEEE 240
Mbd3c   PLCKAFMVTDDDIRKQEELVQQVRKRLEEALMADMLAHVEELARDGEAPLDKACAEEEEE 232
        ************************************************************

Mbd3b   EEEEEEEPEPERV- 253
Mbd3c   EEEEEEEPEPERV- 245
        *************
```









5.2.3 – NuRD regulates gene expression in a context- and locus-dependent manner

Nanog expression is known to be auto-regulated. Nanog, Oct4 and Sox2 are part of a regulatory circuitry consisting of auto-regulatory and feed-forward loops that control their expression (Boyer et al., 2005; Young, 2011). However, for these key pluripotency factors to be the expressed at normal levels and be responsive to stimuli, gene repression mechanisms have to be in place. The NuRD complex has been implicated in Nanog auto-repression in ESCs, and is recruited to the Nanog locus by the transcriptional repressor Zfp281, a Nanog protein-protein interactor (Fidalgo et al., 2012). Since Nanog and the NuRD complex seemed to be co-operatively acting as facilitators of primed gene expression during reprogramming, I decided to analyse the gene expression profiles of wild type, *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs stably transfected with Nanog and/ or Mbd3 (Figure 5.2.10). Since it was recently observed that the NuRD complex promotes pluripotent gene expression heterogeneity in ESC cultured in S+LIF (Reynolds et al., 2012a), I performed my analysis in cells cultured in both 2i/LIF and S+LIF conditions. I did not observe any differences in *Oct4* or *Esrrβ* expression between transgenic lines regardless of the medium the cells were grown in. Klf4 was found to be up-regulated in *Mbd3*$^{-/-}$ ESCs compared to *Mbd3*-expressing ESCs, both in 2i/LIF and S+LIF, a mis-expression previously described and hypothesized to contribute to the differentiation defects of *Mbd3*$^{-/-}$ ESCs (Reynolds et al., 2012a). Nanog expression was found to be slightly elevated in *Mbd3*$^{-/-}$ ESCs compared to wild type and *Mbd3*$^{fl/-}$ ESCs cultured in 2i/LIF, which is in agreement with the existence of an auto-repressive Nanog mechanism involving NuRD activity (Fidalgo et al., 2012). Moreover, a lower total amount of Nanog was observed in cells transfected with *Nanog* and *Mbd3* than that in *Nanog-EV*, indicating that higher levels of the NuRD complex in ESCs might somewhat repress endogenous *Nanog* expression. I have shown that although the forced expression of *Nanog* and *Mbd3* in preiPSCs leads to primed expression of both *Esrrβ* and *Oct4*, no primed *Nanog* expression was detected (Figure 5.2.6A). This indicates that even if the NuRD complex represses the endogenous expression of *Nanog*, if an external source of *Nanog* that is not target of the endogenous repression is provided, the combined action of the exogenous *Nanog* and NuRD can enhance the expression of other pluripotency genes, such as *Esrrβ* and *Oct4*, which, in a reprogramming setting, results in enhanced efficiency.

Altogether, this data suggests that NuRD functions in a highly context- and gene-dependent manner, acting both as an activator and as a repressor of gene expression.

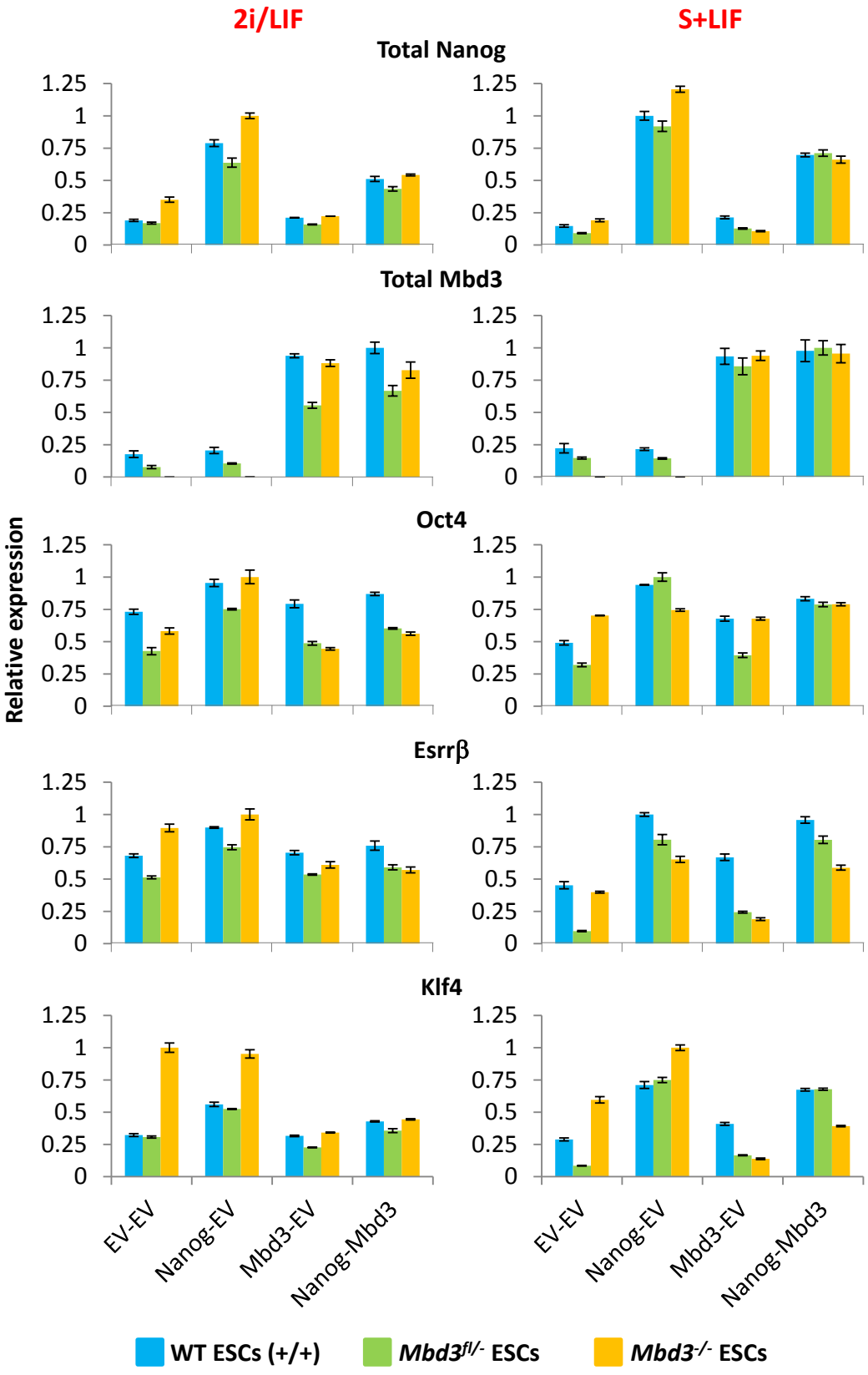


Figure 5.2.10 – Gene expression of transgenic ESCs in 2i/LIF and S+LIF conditions.

qRT-PCR analysis of pluripotency-associated genes in wild type, *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ ESCs stably transfected with pPB-CAG-Nanog and pPB-CAG-Mbd3b or pPB-CAG-empty controls and plated in either 2i/LIF or S+LIF conditions for 12 days. qRT-PCR values are normalized to Gapdh value and shown as relative to the highest value. The error bars indicate STDEV.

5.3 – Discussion

5.3.1 – Overexpression of Mbd3/NuRD can facilitate reprogramming

I have demonstrated that *Mbd3* overexpression, with resulting higher levels of the NuRD complex, can facilitate reprogramming in combination with Nanog expression (Figure 5.3.1). It has been previously shown by our laboratory that Nanog is able to overcome the reprogramming block of preiPSCs and is sufficient to reprogram EpiSCs to naïve pluripotency (Theunissen et al., 2011b). Is this chapter I demonstrate that this effect can be increased up to 30 fold by the forced expression of Mbd3 together with Nanog overexpression. Taking into account that Mbd3 alone failed to alleviate the reprogramming block of preiPSCs (Figures 5.2.4A-C), and that combined overexpression of *Mbd3* with *Klf4* or *Nr5a2* does not enhance reprogramming efficiency (Figures 5.2.8A-B), the observed synergistic effect is specific to Nanog.

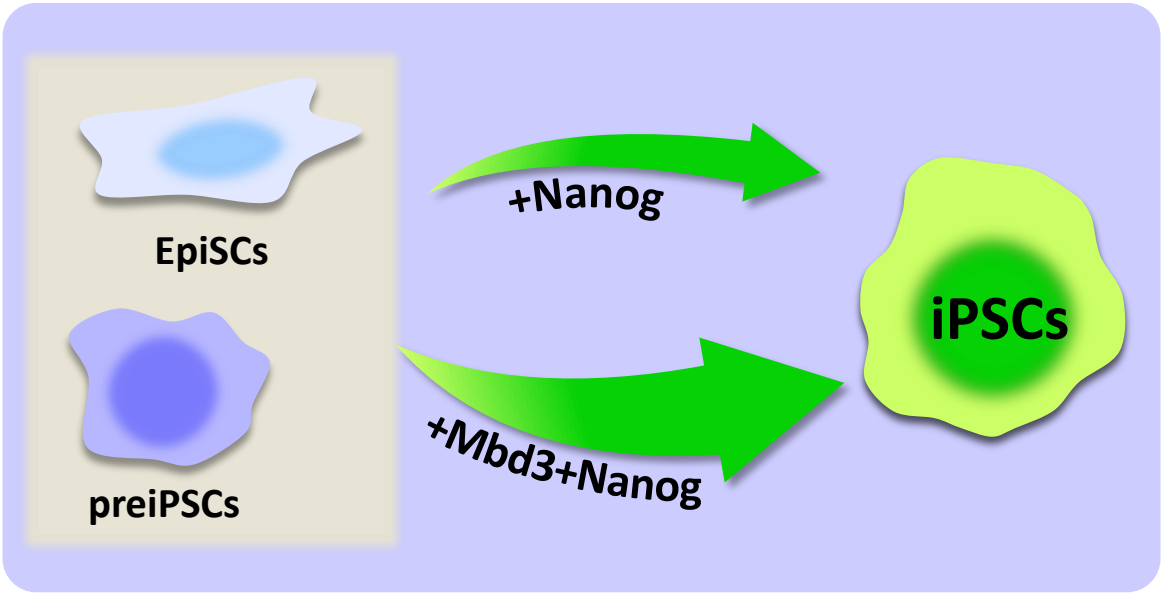


Figure 5.3.1 – Overexpression of Mbd3/NuRD facilitates Nanog-mediated reprogramming.

Increased Mbd3/NuRD levels enhance reprogramming efficiency of preiPSCs and EpiSCs when co-expressed with the reprogramming factor Nanog. The synergy between *Nanog* and *Mbd3* results in a 30 fold increase in reprogramming efficiency.

An analogous synergy was recently reported in our laboratory, where the interaction between the transcription factor *Nanog* and the epigenetic regulators *Tet1/2* facilitate reprogramming, in a *Tet1/2* catalytic activity-dependent manner (Costa et al., 2013). In that study it was observed that *Nanog* recruits *Tet1* to a subset of key pluripotency genes, such as *Esrrβ* and *Oct4*, which leads to increased hydroxymethylation of those genes, resulting in

primed expression before reprogramming to naïve pluripotency. This primed expression prior to the establishment of the pluripotency network is thought to be the reason behind the gain-of-efficiency during preiPSC reprogramming, when *Nanog-Tet1/2* are co-overexpressed compared to *Nanog-alone* (Costa et al., 2013). Interestingly, *Nanog-Tet1* stably transfected MEF-derived preiPSCs reprogram 20-30 fold more efficiently than *Nanog-EV* ones, an effect in the same order of magnitude as that seen in *Nanog-Mbd3* preiPSCs (30 fold). Since the synergies between *Nanog-Tet1/2* and *Nanog-Mbd3* during reprogramming are very similar, I looked at global hydroxymethylation levels in *Nanog-Mbd3* preiPSCs lines. Surprisingly, higher levels of 5-hmC are observed in *Nanog-Mbd3* preiPSCs compared to *Nanog-EV* or *Mbd3-EV* preiPCs. Moreover, as in the *Nanog-Tet1* preiPSCs, primed *Esrrβ* and *Oct4* expression prior to the establishment of naïve pluripotency was also observed only in *Nanog-Mbd3* preiPSCs, indicating a direct relationship between these three major players *Nanog*, *Tet1/2* and the Mbd3/ NuRD complex during reprogramming. In fact, a link between Mbd3/ NuRD and hydroxymethylation levels has been recently described in ESCs (Yildirim et al., 2011). It was observed that *Mbd3* controls 5-hmC levels in the chromatin and that Mbd3 recruitment is dependent on 5-hmC (Yildirim et al., 2011). Since *Mbd3* fails to bind methylated DNA (Hendrich and Bird, 1998; Zhang et al., 1999), how Mbd3 interacts with chromatin to regulate transcription is not known. The correlation between *Mbd3* and 5-hmC levels led the authors to hypothesise that NuRD complex might bind chromatin through an interaction of the mbd of Mbd3 and 5-hmC. However, and as discussed in section 3.3.1, this does not seem to be the case since an Mbd3 isoform, Mbd3b, which lacks the domains believed to interact with 5-hmC, is able to rescue all the *Mbd3*$^{-/-}$ phenotypes. From the above described data, and taking into account that NuRD complex subunits, like Mbd3, are high confidence protein interactors of Nanog (Costa et al., 2013; Ding et al., 2012; Gagliardi et al., 2013; Liang et al., 2008), a more plausible model is where the Mbd3/ NuRD complex is recruited to its target genes by interactions with key transcription factors, such Nanog, where it executes its function. Since Tet1/2 are also high-confidence interactors of Nanog (Costa et al., 2013), it is possible that Nanog guides both NuRD and Tets to key pluripotency genes during reprogramming, enabling their activation and priming expression prior to the 2i/LIF medium switch (Figure 5.3.2).

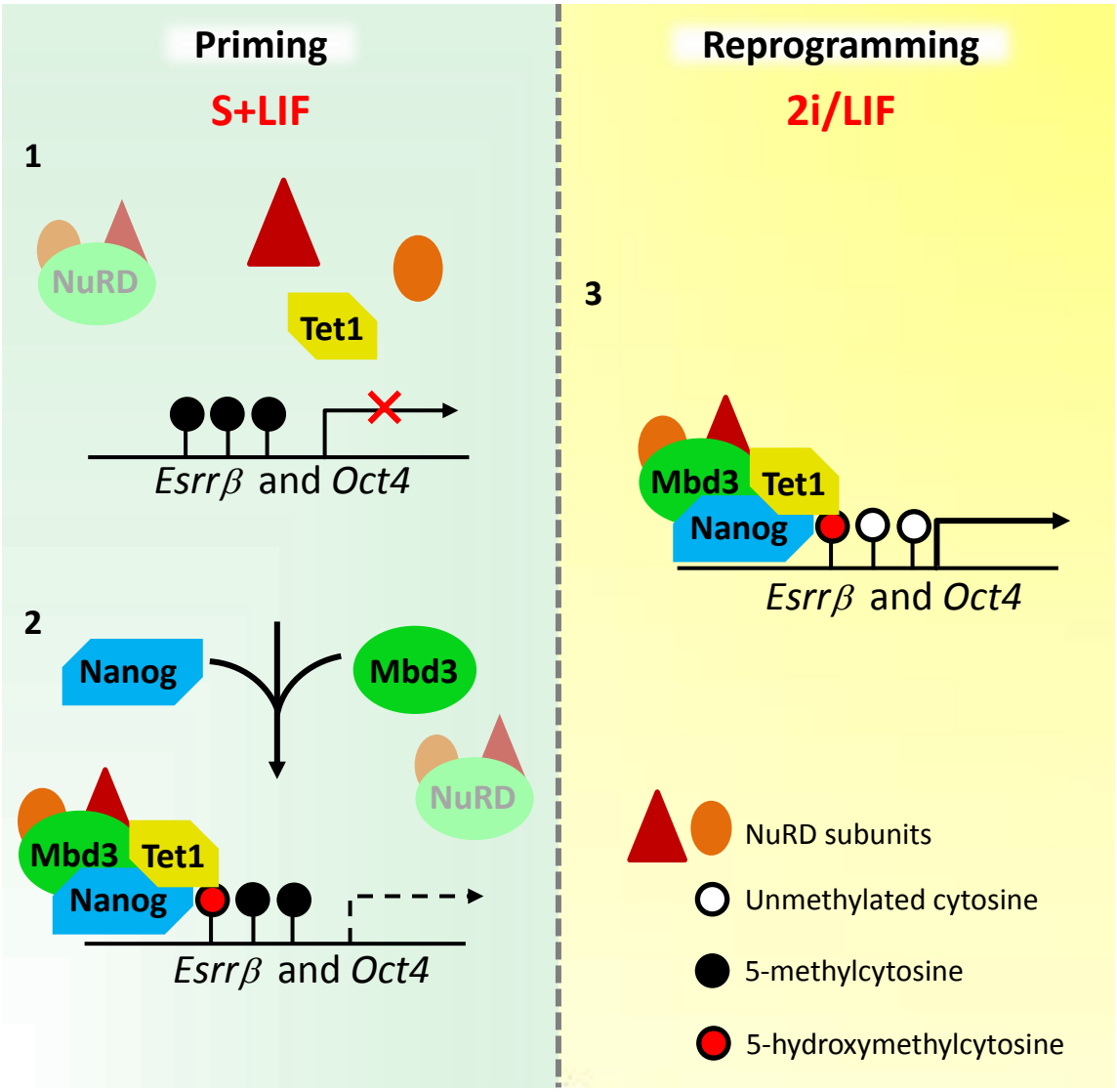


Figure 5.3.2 – Working model for Nanog-Mbd3 synergy during reprogramming.

I propose a model to explain the synergistic effect of combined forced expression of both *Nanog* and *Mbd3* during reprogramming of preiPSCs. **1-** In wild type preiPSCs, *Esrrβ* and *Oct4* promoters are hypermethylated resulting in complete repression of their expression (Costa et al., 2013; Theunissen et al., 2011b). **2-** Overexpression of Nanog and Mbd3, with resulting higher levels of the NuRD complex, results in primed expression of *Esrrβ* and *Oct4* (5% and 3%, respectively, of the expression levels of WT ESCs in 2i/LIF), and also increased global levels of 5-hmC. Since Nanog interacts with both Mbd3 and Tet1 (Costa et al., 2013; Ding et al., 2012; Gagliardi et al., 2013; Liang et al., 2008), it is possible that Nanog recruits the NuRD complex to key pluripotency genes, and in a combined action with Tet1, alters their epigenetic status, resulting in higher hydroxymethylation and lower methylation, and primed gene expression. **3-** This primed expression appears to facilitate reactivation of the core transcriptional program of naïve pluripotency after 2i/LIF culture, resulting in earlier up-regulation of gene expression of pluripotency genes, leading to increased reprograming efficiency.

Such a model assumes that if the hypothesized Nanog-Mbd3 interaction is abolished, then reprogramming synergy between these two factors would be lost. To test this assumption, I tested if reprogramming synergy during preiPSC reprogramming is maintained when an Mbd3 sequence that is thought to be responsible for protein-protein interactions is removed. It has been previously suggested that the N-terminus of Mbd3 is required for protein-protein interactions (Aguilera et al., 2011). It was observed that a 15-residue sequence (ARYLGGSMDLSTFDF – Figure 5.2.9) situated in the N-terminus of Mbd3 is necessary and sufficient for binding to c-Jun (Aguilera et al., 2011). I have also found that the N-teminus of Mbd3 is required for the observed synergistic effect with Nanog in reprogramming (Figure 5.2.9), probably due to the abolishment of the Nanog-Mbd3 interaction. Moreover, Mta2 levels were not found to be elevated after the overexpression of Mbd3c (Figure 5.2.9D), which leads to the assumption that the N-terminal sequence is required for the stabilization of the NuRD complex. Interestingly, the fact that protein-protein interactions with both Nanog or c-Jun (Aguilera et al., 2011) seem to be responsible for NuRD's targeting to chromatin and might explain how this complex interacts with DNA and why a repressor complex might act as an activator in certain conditions. Lastly, since the methyl binding domain of Mbd3 does not bind methylated DNA but is required for protein-protein interactions, this indicates that this domain may have evolved from a methyl-CpG-binding domain into a protein–protein interaction module. Future work will aim to understand how Nanog and Mbd3 work together in driving cells blocked in reprogramming, such as preiPSCs, to pluripotency.

5.3.2 – Conclusions

In this chapter I demonstrated that the reprogramming efficiency of preiPScs and EpiSCs can be boosted by the combined overexression of *Nanog* and *Mbd3*. I observed that both the rate and extent of iPSC generation from preiPSC are increased after the forced expression of *Nanog* and *Mbd3*, which seems to be due to the primed expression of *Esrrβ* and *Oct4*, possibly due to increased hydroxymethylation, with respective reduced methylation, of their loci.

CHAPTER 6 – General Discussion

6.1 – NuRD complex as a transcription modulator.

Reprogramming of somatic cells to naïve pluripotency can be robustly driven by the combined action of transcription factors and culture cues. As described in section 1.3.5, among the reprogramming transcription factors, Oct4 and Nanog play a central role in the induction of pluripotent cells. The Oct4 and Nanog interactome studies in ESCs revealed members of the NuRD complex as its highest confidence interactors (Costa et al., 2013; Ding et al., 2012; Gagliardi et al., 2013; Pardo et al., 2010; van den Berg et al., 2010). NuRD is composed of six core subunits with at least two enzymatic activities involved in gene regulation: the histone deacetylase activity of HDAC1/2 subunits and the ATP-dependent chromatin remodelling activity of Mi-2a/β subunits (section 1.4.1). As described in section 1.4.2, embryos lacking Mbd3, an essential scaffold protein in the NuRD complex, in the absence of which the complex is not assembled, die shortly after implantation (Hendrich et al., 2001; Kaji et al., 2007), and *Mbd3*-null ESCs are viable but exhibit limited differentiation capacity (Kaji et al., 2006; Reynolds et al., 2012a; Reynolds et al., 2012b). Taking into account that chromatin remodelling plays an important role in reprogramming to naïve pluripotency (section 1.3.4), and that the chromatin remodeller NuRD is a high confidence interactor of Oct4 and Nanog and a key regulator of developmental cell state transitions, I have investigated its involvement in the induction of pluripotency.

To study the role of the NuRD complex during reprogramming, the levels of Mbd3 were manipulated in somatic cells which were induced to pluripotency using different systems. In the chapter 3, I demonstrated that Mbd3 facilitates the initiation of neural stem cell reprogramming, and that the amount of preiPSC colonies formed correlates with the amount of time cells express Mbd3 during this transition. In the chapter 4, using genetics or shRNA/siRNA to manipulate *Mbd3* levels, I demonstrated that Mbd3/NuRD is also required for efficient iPSC generation from EpiSCs and preiPSCs. I also showed that Mbd3 deletion after OSKM induction does not impact MEF reprogramming. Finally, in the chapter 5, I present undisputable evidence that the overexpression of Mbd3/NuRD can boost Nanog-mediated reprogramming (Figure 6.1.1).

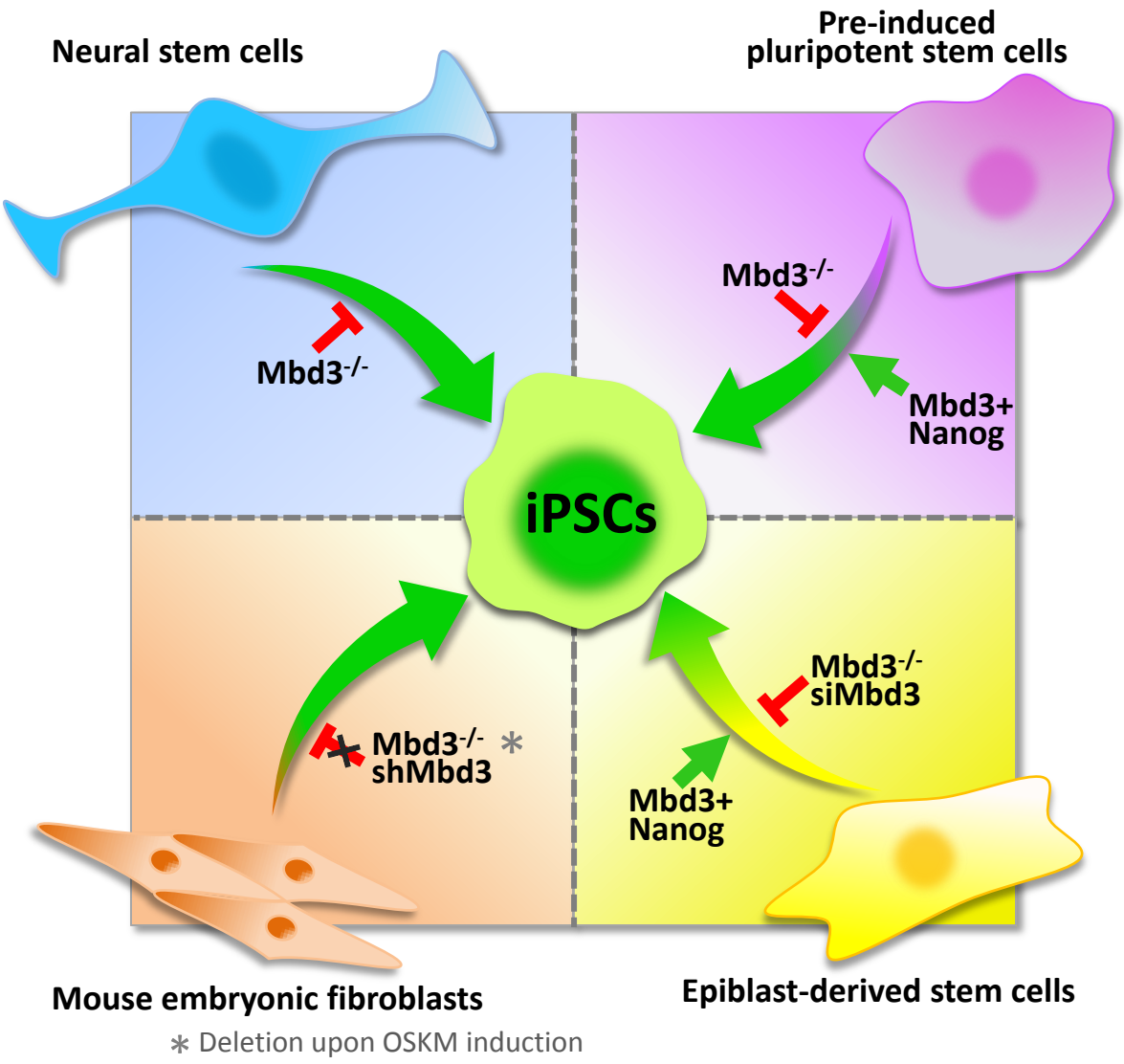


Figure 6.1.1 – The requirement of Mbd3/NuRD in induced pluripotency.

Schematic representation of the main findings of this thesis. By manipulation of the Mbd3 levels, I observed that Mbd3 depletion reduces efficiency of NSCs, EpiSCs and preiPSCs reprogramming to iPSCs. Mbd3 depletion does not seem to impact MEF reprogramming. However, it is important to notice that Mbd3 depletion in MEFs was only be carried out after induction of OSKM, not prior it, as in the other cells types. Combined forced expression of Nanog and Mbd3 enhance reprogramming efficiency of preiPSCs and EpiSCs.

Due to its subunit composition, the NuRD complex has been regarded as a transcriptional repressor. However, the data presented in this thesis suggests that the NuRD complex might be facilitating gene activation under certain conditions. In agreement, recent literature has started to re-evaluate the role of the NuRD complex in gene regulation, providing clear proof that the NuRD complex regulates gene activation as well as gene repression. The first report involving the NuRD complex in gene activation came from the analysis of T cell development

(Williams et al., 2004). The authors reported that the NuRD component Mi-2β is required for T cell maturation by directly promoting CD4 expression. It was shown that Mi-2β binds to the *CD4* enhancer and contributes to the recruitment of the transcriptional machinery for *CD4* transcription (Williams et al., 2004). Further studies by the same laboratory showed that the NuRD complex binds to many active genes involved in lymphoid maturation, and that its targeting is mediated by Ikaros, a key regulator of lymphocyte development (Georgopoulos, 2002; Zhang et al., 2011). When Ikaros interaction with the NuRD complex is disrupted, NuRD is not targeted to lymphocyte development genes, but it is still bound to other genes, where it can act both as repressor and activator of expression (Zhang et al., 2011). In addition, it was also reported that the NuRD complex, dependent on cellular and genetic context, can mediate both the activating and the repressive transcriptional functions of FOG-1 and Gata-1 during erythropoiesis (Miccio et al., 2010).

Recent genome-wide studies revealed that Mbd3 is localized at both active and inactive promoters, marked by H3K4me3 and H3K37me3, respectively (Shimbo et al., 2013), and was found to be enriched at active demethylated promoters (Gunther et al., 2013). Mbd3 was also found to be enriched at active enhancers (Shimbo et al., 2013), just as Mi-2β (Whyte et al., 2012), indicating that the NuRD complex is associated with active transcription. Surprisingly, Mbd3 was found to be enriched at super-enhancers in ESCs, which are enhancers associated with highly expressed genes (Hnisz et al., 2013). All these results suggest that the NuRD complex is not a classical "transcriptional repressor". Rather than controlling "all-or-nothing" gene expression changes, the NuRD complex might function as a transcriptional modulator, fine tuning the expression of both active and repressed genes. In agreement with this notion, genome-wide expression analysis revealed that 61% of differentially expressed genes (839 genes) are down-regulated upon Mbd3 deletion in ESCs (Reynolds et al., 2012b). Although a share of this decrease in transcription might be due to indirect effects, this reveals that the NuRD complex acts on enhancers and promoters as a mediator of transcription factor-induced gene activation.

A model where NuRD acts to fine-tune expression levels is consistent with the results presented in this thesis. As described in section 1.3.4, the OSKM factors bind to chromatin to induce both the repression and the activation of gene expression. It is possible that the different outcomes of OSKM binding to the different genes might be controlled, at least in part, by the NuRD complex. The reprogramming factors, by interacting with the NuRD complex, might guide it to the key genes that need to be transcriptionally modulated in the original somatic cell. Lack of a functional complex is likely to result in a failure of the transcriptional modulation induced by OSKM, which impedes the switch between different

transcriptional programmes characteristic of different cell types. This means that, in a probabilistic manner, only rare NuRD-null somatic cells would engage in a fate switching and be reprogrammed. Such a view of the role of the NuRD complex in fate switch assumes that the rare cells that reprogram without NuRD exhibit delayed reprogramming kinetics, as active transcriptional modulation is not taking place. Indeed, the data presented in the chapter 3 shows that *Mbd3*-null NSCs reprogram very inefficiently, and the ones which do reprogram take longer to do so. This role of the NuRD complex might also explain why no effect is observed if Mbd3 is depleted after OSKM induction during MEF reprogramming. Recent evidence implicates NuRD in the early stages of MEF reprogramming by transiently driving mTOR downregulation with subsequent activation of autophagy (Wang et al., 2013c). Moreover, the NuRD complex might also play a role in MET, a key event during early MEF reprogramming, which is thought to occur early during reprogramming. Thus, it is possible that Mbd3 deletion after this early stage of reprogramming has no effect on subsequent reprogramming stages because cells have already passed the critical stage where the NuRD complex might be particularly important.

The specific interactions established by individual NuRD subunits are thought to be the main mechanism of control of the different functions of the complex. In fact, NuRD's interaction with the reprogramming factors is a possible mechanism to describe how OSKM drive the transcriptional changes required for induced pluripotency. It is then possible to infer that NuRD activity on a particular set of genes, such as pluripotent genes, might be enhanced if a hypothetical transcription factor that targets the complex to those genes is ectopically expressed. Indeed, experimental evidence presented in this thesis supports such statement. Forced Nanog expression in preiPSCs and EpiSCs, together with elevated levels of NuRD, increases the efficiency of the establishment of naïve pluripotency by enhancing the expression of the key pluripotency genes.

In conclusion, the NuRD complex plays a key role during induced pluripotency, potentially by interacting with transcription factors to assist genome-wide reprogramming. Further insights into the function of the NuRD complex during different cell state transitions will help us understand the process of induced pluripotency as well as embryonic development.

6.2 – Future perspectives on the role of the NuRD complex in induced pluripotency.

In this thesis I have shown that the NuRD complex facilitates induced pluripotency. In the future it will be important to understand how NuRD modulates this transition. To assess if NuRD's function during OSKM-mediated reprogramming is dependent on its interaction with the reprogramming factors, it will be interesting to assess if members of the NuRD complex physically interact with any of the OSKM factors, which could be analysed by a pull-down assay after overexpression of these in somatic cells. Then, to further dissect how this hypothetical interaction might contribute to the transcriptional changes observed during induced pluripotency, high-throughput techniques such as RNA sequencing (RNA-seq) and chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) can be employed. Mbd3 ChIP-seq together with RNA-seq prior to and after OSKM transduction, using wild type and *Mbd3*-null somatic cells, such as NSCs, will reveal if OSKM binding to genome or if OSKM-induced transcriptional changes are dependent on the NuRD function. These experiments will help us to understand if the reduced reprogramming efficiency of *Mbd3*$^{-/-}$ NSCs is due to aberrant transcriptional silencing and/ or activation after OSKM induction. Moreover, by coupling ChIP-seq with RNA-seq it will be able possible to understand why NuRD acts as a facilitator during induced pluripotency, since the effect of NuRD binding to the genome can be correlated with gene expression profiles. Since in this thesis I showed that Mbd3 depletion after OSKM induction during MEF reprogramming does not affect reprogramming, it will be of interest to test the impact of Mbd3 depletion prior to the induction of expression of the reprogramming factors. Also, taking into account that MET is a key event during early MEF reprogramming, it will be important to test if the NuRD complex also plays a part in this transition, which would explain why its disruption a couple of days after induction does not impact induced pluripotency.

ChIP-seq with RNA-seq approaches can also be used to dissect how Nanog and Mbd3 work to induce the reprogramming of preiPSCs to naïve pluripotency. By comparing Mbd3 and Nanog ChIP-seq tracks obtained from different preiPSC lines stably transfected with Nanog and/ or Mbd3 transgenes, NuRD changes in localization after the ectopic expression of Nanog can be assessed, and what genome-wide transcriptional alterations occur when Nanog and Mbd3 are overexpressed can be revealed. Since global 5-hmC levels are increased in *Nanog-Mbd3* preiPSCs compared to *Nanog-EV* or *Mbd3-EV* preiPSCs, it will be interesting to understand how and why this occurs. It will be important to assess if Nanog, NuRD and Tet1/2 interact with each other, and whether the disruption of those interactions would affect Nanog-Mbd3 synergy during reprogramming. A comparison of Nanog, Mbd3 and Tet1/2

ChIP-seq tracks using *Nanog-Mbd3* preiPSCs, together with 5-hmC mapping throughout the genome, will provide a compressive and integrative view of how transcription factors and epigenetic remodellers work together to modulate gene expression changes and induce fate switching.

Such integrative approaches will help us to understand how reprogramming factors work, how the somatic epigenome is erased and how the naïve pluripotent one is established. Once the molecular mechanisms underlying induced pluripotency are understood, then the full potential of iPSCs for personalized medicine and disease modelling can be explored. This research will lead to the creation of better and safer iPSCs, bringing the applications in regenerative medicine a step closer.

CHAPTER 7 – References

Aguilera, C., Nakagawa, K., Sancho, R., Chakraborty, A., Hendrich, B., and Behrens, A. (2011). c-Jun N-terminal phosphorylation antagonises recruitment of the Mbd3/NuRD repressor complex. Nature *469*, 231-235.

Ahringer, J. (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends Genet *16*, 351-356.

Allen, H.F., Wade, P.A., and Kutateladze, T.G. (2013). The NuRD architecture. Cell Mol Life Sci *70*, 3513-3524.

Amaya, M., Desai, M., Gnanapragasam, M.N., Wang, S.Z., Zu Zhu, S., Williams, D.C., Jr., and Ginder, G.D. (2013). Mi2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells. Blood *121*, 3493-3501.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., *et al.* (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183-197.

Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y.Z., Yang, W.L., Gruber, P.J., Epstein, J.A., *et al.* (2011). Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. Cell Stem Cell *8*, 376-388.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science *321*, 699-702.

Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., *et al.* (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell *12*, 699-712.

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462-471.

Armache, K.J., Garlick, J.D., Canzio, D., Narlikar, G.J., and Kingston, R.E. (2011). Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 A resolution. Science *334*, 977-982.

Artus, J., Piliszek, A., and Hadjantonakis, A.K. (2011). The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. Dev Biol *350*, 393-404.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Batlle-Morera, L., Smith, A., and Nichols, J. (2008). Parameters influencing derivation of embryonic stem cells from murine embryos. Genesis *46*, 758-767.

Baudino, T.A., McKay, C., Pendeville-Samain, H., Nilsson, J.A., Maclean, K.H., White, E.L., Davis, A.C., Ihle, J.N., and Cleveland, J.L. (2002). c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. Genes Dev *16*, 2530-2543.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669-681.

Betschinger, J., Nichols, J., Dietmann, S., Corrin, P.D., Paddison, P.J., and Smith, A. (2013). Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. Cell *153*, 335-347.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., *et al.* (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222-226.

Blau, H.M., Chiu, C.P., and Webster, C. (1983). Cytoplasmic Activation of Human Nuclear Genes in Stable Heterocaryons. Cell *32*, 1171-1180.

Blau, H.M., Pavlath, G.K., Hardeman, E.C., Chiu, C.P., Silberstein, L., Webster, S.G., Miller, S.C., and Webster, C. (1985). Pasticity of the differentiated state. Science *230*, 758-766.

Boland, M.J., Hazen, J.L., Nazor, K.L., Rodriguez, A.R., Gifford, W., Martin, G., Kupriyanov, S., and Baldwin, K.K. (2009). Adult mice generated from induced pluripotent stem cells. Nature *461*, 91-U94.

Bouazoune, K., Mitterweger, A., Langst, G., Imhof, A., Akhtar, A., Becker, P.B., and Brehm, A. (2002). The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. EMBO J *21*, 2430-2440.

Bowen, N.J., Fujita, N., Kajita, M., and Wade, P.A. (2004). Mi-2/NuRD: multiple complexes for many purposes. Biochim Biophys Acta *15*, 1-3.

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

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Brackertz, M., Gong, Z., Leers, J., and Renkawitz, R. (2006). p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. Nucleic Acids Res *34*, 397-406.

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

Briggs, R., and King, T.J. (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc Natl Acad Sci U S A *38*, 455-463.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B.W., Lopes, S., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., *et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q.L., and Smith, A. (2008). Capture of authentic embryonic stem cells from rat blastocysts. Cell *135*, 1287-1298.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209-1222.

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat Rev Genet *14*, 427-439.

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

Campbell, K.H.S., McWhir, J., Ritchie, W.A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature *380*, 64-66.

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., and Jaenisch, R. (2009). Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc Natl Acad Sci U S A *106*, 157-162.

Carey, B.W., Markoulaki, S., Hanna, J.H., Faddah, D.A., Buganim, Y., Kim, J., Ganz, K., Steine, E.J., Cassady, J.P., Creyghton, M.P., *et al.* (2011). Reprogramming Factor Stoichiometry Influences the Epigenetic State and Biological Properties of Induced Pluripotent Stem Cells. Cell Stem Cell *9*, 588-598.

Carter, A.C., Davis-Dusenbery, B.N., Koszka, K., Ichida, J.K., and Eggan, K. (2014). Nanog-Independent Reprogramming to iPSCs with Canonical Factors. Stem Cell Reports *2*, 119-126.

Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells *26*, 1496-1505.

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

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-U1238.

Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311-2322.

Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, Kevin Andrew U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., and Ng, H.-H. (2013). Induction of a Human Pluripotent State with Distinct Regulatory Circuitry that Resembles Preimplantation Epiblast. Cell Stem Cell *13*, 663-675.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell *10*, 615-624.

Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., *et al.* (2013a). Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet *45*, 1504-1509.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Wu, Y., Guo, L., *et al.* (2013b). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet *45*, 34-42.

Christophorou, M.A., Castelo-Branco, G., Halley-Stott, R.P., Oliveira, C.S., Loos, R., Radzisheuskaya, A., Mowen, K.A., Bertone, P., Silva, J.C., Zernicka-Goetz, M., *et al.* (2014). Citrullination regulates pluripotency and histone H1 binding to chromatin. Nature *507*, 104-108.

Chung, Y.G., Eum, J.H., Lee, J.E., Shim, S.H., Sepilian, V., Hong, S.W., Lee, Y., Treff, N.R., Choi, Y.H., Kimbrel, E.A., *et al.* (2014). Human Somatic Cell Nuclear Transfer Using Adult Cells. Cell Stem Cell *15*, 00137-00134.

Climent, M., Alonso-Martin, S., Perez-Palacios, R., Guallar, D., Benito, A.A., Larraga, A., Fernandez-Juan, M., Sanz, M., de Diego, A., Seisdedos, M.T., *et al.* (2013). Functional analysis of Rex1 during preimplantation development. Stem Cells Dev *22*, 459-472.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., *et al.* (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370-374.

Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science *309*, 1369-1373.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936.

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

Davenport, T.G., Jerome-Majewska, L.A., and Papaioannou, V.E. (2003). Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. Development *130*, 2263-2273.

Davidson, R.L., Ephrussi, B., and Yamamoto, K. (1966). Regulation of pigment synthesis in mammalian cells, as studied by somatic hybridization. Proc Natl Acad Sci U S A *56*, 1437-&.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a Single Transfected cDNA Converts Fibroblasts to Myoblasts. Cell *51*, 987-1000.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., *et al.* (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell *24*, 310-323.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.-C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.-W., Page, D.C., *et al.* (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. Cell Stem Cell *9*, 166-175.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433-5438.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235-239.

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., and Wang, J. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res *22*, 155-167.

Do, D.V., Ueda, J., Messerschmidt, D.M., Lorthongpanich, C., Zhou, Y., Feng, B., Guo, G., Lin, P.J., Hossain, M.Z., Zhang, W., *et al.* (2013). A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. Genes Dev *27*, 1378-1390.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., *et al.* (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652-655.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J Embryol Exp Morphol *87*, 27-45.

Dutta, D., Ray, S., Home, P., Larson, M., Wolfe, M.W., and Paul, S. (2011). Self-renewal versus lineage commitment of embryonic stem cells: protein kinase C signaling shifts the balance. Stem Cells *29*, 618-628.

Ema, M., Mori, D., Niwa, H., Hasegawa, Y., Yamanaka, Y., Hitoshi, S., Mimura, J., Kawabe, Y., Hosoya, T., Morita, M., *et al.* (2008). Kruppel-like factor 5 is essential for

blastocyst development and the normal self-renewal of mouse ESCs. Cell Stem Cell *3*, 555-567.

Eminli, S., Foudi, A., Stadtfeld, M., Maherali, N., Ahfeldt, T., Mostoslavsky, G., Hock, H., and Hochedlinger, K. (2009). Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. Nat Genet *41*, 968-976.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., Han, D., Drexler, H.C., Arauzo-Bravo, M.J., Ng, C.K., *et al.* (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol *15*, 295-301.

Esteban, M.A., Wang, T., Qin, B.M., Yang, J.Y., Qin, D.J., Cai, J.L., Li, W., Weng, Z.H., Chen, J.K., Ni, S., *et al.* (2010). Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. Cell Stem Cell *6*, 71-79.

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

Feng, B., Jiang, J.M., Kraus, P., Ng, J.H., Heng, J.C.D., Chan, Y.S., Yaw, L.P., Zhang, W.W., Loh, Y.H., Han, J.Y., *et al.* (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol *11*, 197-U193.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477-490.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., *et al.* (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351-359.

Fidalgo, M., Faiola, F., Pereira, C.F., Ding, J., Saunders, A., Gingold, J., Schaniel, C., Lemischka, I.R., Silva, J.C., and Wang, J. (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. Proc Natl Acad Sci U S A *109*, 16202-16207.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA Repair Complex Functions as an Oct4/Sox2 Coactivator in Embryonic Stem Cells. Cell *147*, 120-131.

Fragola, G., Germain, P.L., Laise, P., Cuomo, A., Blasimme, A., Gross, F., Signaroldi, E., Bucci, G., Sommer, C., Pruneri, G., *et al.* (2013). Cell reprogramming requires silencing of a core subset of polycomb targets. PLoS Genet *9*, 28.

Fujita, N., Jaye, D.L., Geigerman, C., Akyildiz, A., Mooney, M.R., Boss, J.M., and Wade, P.A. (2004). MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. Cell *119*, 75-86.

Fujita, N., Jaye, D.L., Kajita, M., Geigerman, C., Moreno, C.S., and Wade, P.A. (2003). MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. Cell *113*, 207-219.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., *et al.* (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282-286.

Gagliardi, A., Mullin, N.P., Ying Tan, Z., Colby, D., Kousa, A.I., Halbritter, F., Weiss, J.T., Felker, A., Bezstarosti, K., Favaro, R., *et al.* (2013). A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. EMBO J *32*, 2231-2247.

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., *et al.* (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell *12*, 453-469.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., *et al.* (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863-U897.

Gaspar-Maia, A., Qadeer, Z.A., Hasson, D., Ratnakumar, K., Leu, N.A., Leroy, G., Liu, S., Costanzi, C., Valle-Garcia, D., Schaniel, C., *et al.* (2013). MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. Nat Commun *4*.

Ge, Q., Nilasena, D.S., O'Brien, C.A., Frank, M.B., and Targoff, I.N. (1995). Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen. J Clin Invest *96*, 1730-1737.

Gehring, W.J. (1996). The master control gene for morphogenesis and evolution of the eye. Genes Cells *1*, 11-15.

Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K., and Daley, G.Q. (2004). Derivation of embryonic germ cells and male gametes from embryonic stem cells. Nature *427*, 148-154.

Georgopoulos, K. (2002). Haematopoietic cell-fate decisions, chromatin regulation and ikaros. Nat Rev Immunol *2*, 162-174.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., *et al.* (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149-1163.

Gong, Z., Brackertz, M., and Renkawitz, R. (2006). SUMO modification enhances p66-mediated transcriptional repression of the Mi-2/NuRD complex. Mol Cell Biol *26*, 4519-4528.

Gonzalez, F., Georgieva, D., Vanoli, F., Shi, Z.D., Stadtfeld, M., Ludwig, T., Jasin, M., and Huangfu, D. (2013). Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state. Cell Rep *3*, 651-660.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587-594.

Gu, P., Goodwin, B., Chung, A.C., Xu, X., Wheeler, D.A., Price, R.R., Galardi, C., Peng, L., Latour, A.M., Koller, B.H., *et al.* (2005). Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol Cell Biol *25*, 3492-3505.

Gunther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., Bartkuhn, M., and Renkawitz, R. (2013). Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. Nucleic Acids Res *41*, 3010-3021.

Guo, G., Huang, Y., Humphreys, P., Wang, X.Z., and Smith, A. (2011). A PiggyBac-Based Recessive Screening Method to Identify Pluripotency Regulators. PLoS ONE *6*.

Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development *137*, 3185-3192.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063-1069.

Guo, S., Zi, X., Schulz, V.P., Cheng, J., Zhong, M., Koochaki, S.H., Megyola, C.M., Pan, X., Heydari, K., Weissman, S.M., *et al.* (2014). Nonstochastic reprogramming from a privileged somatic cell state. Cell *156*, 649-662.

Gurdon, J.B. (1962a). Adult Frogs Derived from the Nuclei of Single Somatic Cells. Dev Biol *4*, 256-&.

Gurdon, J.B. (1962b). The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. J Embryol Exp Morphol *10*, 622-&.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature *182*, 64-65.

Habib, O., Habib, G., Do, J.T., Moon, S.H., and Chung, H.M. (2014). Activation-induced deaminase-coupled DNA demethylation is not crucial for the generation of induced pluripotent stem cells. Stem Cells Dev *23*, 209-218.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., *et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360-369.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., *et al.* (2009). Oct4 and LIF/Stat3 Additively Induce Kruppel Factors to Sustain Embryonic Stem Cell Self-Renewal. Cell Stem Cell *5*, 597-609.

Han, D.W., Tapia, N., Joo, J.Y., Greber, B., Arauzo-Bravo, M.J., Bernemann, C., Ko, K., Wu, G.M., Stehling, M., Do, J.T., *et al.* (2010). Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. Cell *143*, 617-627.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010a). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc Natl Acad Sci U S A *107*, 9222-9227.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R., *et al.* (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell *133*, 250-264.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595-U563.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010b). Pluripotency and Cellular Reprogramming: Facts, Hypotheses, Unresolved Issues. Cell *143*, 508-525.

Harris, H., Miller, O.J., Klein, G., Worst, P., and Tachiban.T (1969). Suppression of Malignancy by Cell Fusion. Nature *223*, 363-&.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res *40*, 4841-4849.

Hayashi, K., and Surani, M.A. (2009). Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. Development *136*, 3549-3556.

Hemmings, B.A., and Restuccia, D.F. (2012). PI3K-PKB/Akt pathway. Cold Spring Harb Perspect Biol *4*.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol *18*, 6538-6547.

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

Heng, J.C.D., Feng, B., Han, J.Y., Jiang, J.M., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., *et al.* (2010). The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells. Cell Stem Cell *6*, 167-174.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A *106*, 5187-5191.

Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R., and Crabtree, G.R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. Nat Cell Biol *13*, 903-913.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling

complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A *106*, 5181-5186.

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

Hogan, B.L. (1976). Changes in the behaviour of teratocarcinoma cells cultivated in vitro. Nature *263*, 136-137.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature *460*, 1132-U1195.

Horton, J.R., Elgar, S.J., Khan, S.I., Zhang, X., Wade, P.A., and Cheng, X. (2007). Structure of the SANT domain from the Xenopus chromatin remodeling factor ISWI. Proteins *67*, 1198-1202.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., *et al.* (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science *341*, 651-654.

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

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W., *et al.* (2014). Tet and TDG Mediate DNA Demethylation Essential for Mesenchymal-to-Epithelial Transition in Somatic Cell Reprogramming. Cell Stem Cell *14*, 512-522.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell Rep *2*, 1571-1578.

Huangfu, D.W., Maehr, R., Guo, W.J., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol *26*, 795-797.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., *et al.* (2009). A Small-Molecule Inhibitor of Tgf-beta Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. Cell Stem Cell *5*, 491-503.

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

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567-582.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-360.

Judson, R.L., Greve, T.S., Parchem, R.J., and Blelloch, R. (2013). MicroRNA-based discovery of barriers to dedifferentiation of fibroblasts to pluripotent stem cells. Nat Struct Mol Biol *20*, 1227-1235.

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

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development *134*, 1123-1132.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771-U112.

Kang, E., Wu, G., Ma, H., Li, Y., Tippner-Hedges, R., Tachibana, M., Sparman, M., Wolf, D.P., Scholer, H.R., and Mitalipov, S. (2014). Nuclear reprogramming by interphase cytoplasm of two-cell mouse embryos. Nature *509*, 101-104.

Kang, L., Wang, J.L., Zhang, Y., Kou, Z.H., and Gao, S.R. (2009). iPS Cells Can Support Full-Term Development of Tetraploid Blastocyst-Complemented Embryos. Cell Stem Cell *5*, 135-138.

Karwacki-Neisius, V., Goke, J., Osorno, R., Halbritter, F., Ng, J.H., Weisse, A.Y., Wong, F.C., Gagliardi, A., Mullin, N.P., Festuccia, N., *et al.* (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell *12*, 531-545.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Belmonte, J.C.I. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140-U1107.

Khalfallah, O., Rouleau, M., Barbry, P., Bardoni, B., and Lalli, E. (2009). Dax-1 knockdown in mouse embryonic stem cells induces loss of pluripotency and multilineage differentiation. Stem Cells *27*, 1529-1537.

Kim, J.B., Greber, B., Arauzo-Bravo, M.J., Meyer, J., Park, K.I., Zaehres, H., and Scholer, H.R. (2009a). Direct reprogramming of human neural stem cells by OCT4. Nature *461*, 649-643.

Kim, J.B., Sebastiano, V., Wu, G., Arauzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., *et al.* (2009b). Oct4-induced pluripotency in adult neural stem cells. Cell *136*, 411-419.

Kim, J.B., Zaehres, H., Wu, G., Gentile, L., Ko, K., Sebastiano, V., Arauzo-Bravo, M.J., Ruau, D., Han, D.W., Zenke, M., *et al.* (2008). Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. Nature *454*, 646-650.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

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

Kumar, R., DiMenna, L., Schrode, N., Liu, T.C., Franck, P., Munoz-Descalzo, S., Hadjantonakis, A.K., Zarrin, A.A., Chaudhuri, J., Elemento, O., *et al.* (2013). AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. Nature *500*, 89-92.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895-2902.

Kuo, C.T., Veselits, M.L., and Leiden, J.M. (1997). LKLF: A transcriptional regulator of single-positive T cell quiescence and survival. Science *277*, 1986-1990.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nature Reviews Cancer *11*, 588-596.

Latos, P.A., Helliwell, C., Mosaku, O., Dudzinska, D.A., Stubbs, B., Berdasco, M., Esteller, M., and Hendrich, B. (2012). NuRD-dependent DNA methylation prevents ES cells from accessing a trophectoderm fate. Biol Open *1*, 341-352.

Lauberth, S.M., and Rauchman, M. (2006). A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. J Biol Chem *281*, 23922-23931.

Le Bin, G.C., Munoz-Descalzo, S., Kurowski, A., Leitch, H., Lou, X., Mansfield, W., Etienne-Dumeau, C., Grabole, N., Mulas, C., Niwa, H., *et al.* (2014). Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. Development *141*, 1001-1010.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol *26*, 843-851.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev *24*, 265-276.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Leitch, H.G., Blair, K., Mansfield, W., Ayetey, H., Humphreys, P., Nichols, J., Surani, M.A., and Smith, A. (2010). Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. Development *137*, 2279-2287.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., *et al.* (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol *20*, 311-316.

Lejon, S., Thong, S.Y., Murthy, A., AlQarni, S., Murzina, N.V., Blobel, G.A., Laue, E.D., and Mackay, J.P. (2011). Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48.FOG-1 complex. J Biol Chem *286*, 1196-1203.

Lenardo, M.J., Staudt, L., Robbins, P., Kuang, A., Mulligan, R.C., and Baltimore, D. (1989). Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. Science *243*, 544-546.

Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Canamero, M., Blasco, M.A., and Serrano, M. (2009a). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature *460*, 1136-U1101.

Li, P., Tong, C., Mehrian-Shai, R., Jia, L., Wu, N., Yan, Y., Maxson, R.E., Schulze, E.N., Song, H., Hsieh, C.L., *et al.* (2008). Germline competent embryonic stem cells derived from rat blastocysts. Cell *135*, 1299-1310.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., *et al.* (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51-63.

Li, W., Li, K., Wei, W., and Ding, S. (2013). Chemical approaches to stem cell biology and therapeutics. Cell Stem Cell *13*, 270-283.

Li, W.L., Wei, W., Zhu, S.Y., Zhu, J.L., Shi, Y., Lin, T.X., Hao, E.G., Hayek, A., Deng, H.K., and Ding, S. (2009b). Generation of Rat and Human Induced Pluripotent Stem Cells by Combining Genetic Reprogramming and Chemical Inhibitors. Cell Stem Cell *4*, 16-19.

Liang, G., He, J., and Zhang, Y. (2012). Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. Nat Cell Biol *14*, 457-U445.

Liang, J., Wan, M., Zhang, Y., Gu, P.L., Xin, H.W., Jung, S.Y., Qin, J., Wong, J.M., Cooney, A.J., Liu, D., *et al.* (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol *10*, 731-739.

Liew, C.K., Simpson, R.J., Kwan, A.H., Crofts, L.A., Loughlin, F.E., Matthews, J.M., Crossley, M., and Mackay, J.P. (2005). Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. Proc Natl Acad Sci U S A *102*, 583-588.

Lin, C.Y., Loven, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56-67.

Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K., *et al.* (2008). Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell Stem Cell *3*, 587-590.

Liu, X., Sun, H., Qi, J., Wang, L., He, S., Liu, J., Feng, C., Chen, C., Li, W., Guo, Y., *et al.* (2013). Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. Nat Cell Biol *15*, 829-838.

Lu, J., Jeong, H.W., Kong, N., Yang, Y., Carroll, J., Luo, H.R., Silberstein, L.E., Yupoma, and Chai, L. (2009). Stem cell factor SALL4 represses the transcriptions of PTEN and SALL1 through an epigenetic repressor complex. PLoS ONE *4*, 18.

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

Luo, M., Ling, T., Xie, W., Sun, H., Zhou, Y., Zhu, Q., Shen, M., Zong, L., Lyu, G., Zhao, Y., *et al.* (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells *31*, 1278-1286.

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., and Yamanaka, S. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. Nature *474*, 225-229.

Maherali, N., and Hochedlinger, K. (2009). Tgf beta Signal Inhibition Cooperates in the Induction of iPSCs and Replaces Sox2 and cMyc. Curr Biol *19*, 1718-1723.

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

Mali, P., Chou, B.K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R.A., Ohm, J.E., Yu, W., Baylin, S.B., *et al.* (2010). Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells *28*, 713-720.

Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., and Mackay, J.P. (2011). Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. J Biol Chem *286*, 11779-11791.

Mansour, A.A., Gafni, O., Weinberger, L., Zviran, A., Ayyash, M., Rais, Y., Krupalnik, V., Zerbib, M., Amann-Zalcenstein, D., Maza, I., *et al.* (2012). The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature *488*, 409-413.

Marion, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009a). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature *460*, 1149-U1119.

Marion, R.M., Strati, K., Li, H., Tejera, A., Schoeftner, S., Ortega, S., Serrano, M., and Blasco, M.A. (2009b). Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. Cell Stem Cell *4*, 141-154.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., *et al.* (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell *149*, 590-604.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J *32*, 2561-2574.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Gottgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491-504.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *78*, 7634-7638.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., *et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-635.

Masui, S., Ohtsuka, S., Yagi, R., Takahashi, K., Ko, M.S., and Niwa, H. (2008). Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. BMC Dev Biol *8*, 8-45.

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

Mattout, A., Biran, A., and Meshorer, E. (2011). Global epigenetic changes during somatic cell reprogramming to iPS cells. J Mol Cell Biol *3*, 341-350.

McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: Roles of NuRD and Sin3 co-repressor complexes during mammalian development. Int J Biochem Cell Biol *41*, 108-116.

Miccio, A., Wang, Y., Hong, W., Gregory, G.D., Wang, H., Yu, X., Choi, J.K., Shelat, S., Tong, W., Poncz, M., *et al.* (2010). NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. EMBO J *29*, 442-456.

Mikkelsen, T.S., Hanna, J., Zhang, X.L., Ku, M.C., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-U41.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D.L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F., *et al.* (2011). Reprogramming of mouse and human cells to pluripotency using mature microRNAs. Cell Stem Cell *8*, 633-638.

Montserrat, N., Nivet, E., Sancho-Martinez, I., Hishida, T., Kumar, S., Miquel, L., Cortina, C., Hishida, Y., Xia, Y., Esteban, C.R., *et al.* (2013). Reprogramming of human fibroblasts to pluripotency with lineage specifiers. Cell Stem Cell *13*, 341-350.

Munsie, M.J., Michalska, A.E., O'Brien, C.M., Trounson, A.O., Pera, M.F., and Mountford, P.S. (2000). Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. Curr Biol *10*, 989-992.

Murzina, N.V., Pei, X.Y., Zhang, W., Sparkes, M., Vicente-Garcia, J., Pratap, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F., *et al.* (2008). Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure *16*, 1077-1085.

Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P., and Kutateladze, T.G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. Biochem J *423*, 179-187.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proc Natl Acad Sci U S A *90*, 8424-8428.

Najm, F.J., Chenoweth, J.G., Anderson, P.D., Nadeau, J.H., Redline, R.W., McKay, R.D.G., and Tesar, P.J. (2011). Isolation of Epiblast Stem Cells from Preimplantation Mouse Embryos. Cell Stem Cell *8*, 318-325.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

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

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

Nichols, J., Jones, K., Phillips, J.M., Newland, S.A., Roode, M., Mansfield, W., Smith, A., and Cooke, A. (2009). Validated germline-competent embryonic stem cell lines from nonobese diabetic mice. Nat Med *15*, 814-U135.

Nichols, J., and Smith, A. (2009). Naive and Primed Pluripotent States. Cell Stem Cell *4*, 487-492.

Nichols, J., and Smith, A. (2011). The origin and identity of embryonic stem cells. Development *138*, 3-8.

Nichols, J., and Smith, A. (2012). Pluripotency in the embryo and in culture. Cold Spring Harb Perspect Biol *4*.

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

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., *et al.* (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell *151*, 68-79.

Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., *et al.* (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Dev Cell *16*, 398-410.

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

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

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

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118-122.

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

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88-91.

O'Malley, J., Woltjen, K., and Kaji, K. (2009). New strategies to generate induced pluripotent stem cells. Curr Opin Biotechnol *20*, 516-521.

O'Shaughnessy, A., and Hendrich, B. (2013). CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. Biochem Soc Trans *41*, 777-782.

Ogas, J., Kaufmann, S., Henderson, J., and Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proc Natl Acad Sci U S A *96*, 13839-13844.

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

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Mancarci, B.O., Unternaehrer, J., Gupta, P.B., *et al.* (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598-U119.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. Development *139*, 2288-2298.

Papp, B., and Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell *152*, 1324-1343.

Parchem, R.J., Ye, J., Judson, R.L., Larussa, M.F., Krishnakumar, R., Blelloch, A., Oldham, M.C., and Blelloch, R. (2014). Two miRNA Clusters Reveal Alternative Paths in Late-Stage Reprogramming. Cell Stem Cell *14*, 617-631.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. Cell Stem Cell *6*, 382-395.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H.G., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141-U141.

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

Pasque, V., Radzisheuskaya, A., Gillich, A., Halley-Stott, R.P., Panamarova, M., Zernicka-Goetz, M., Surani, M.A., and Silva, J.C. (2012). Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. J Cell Sci *125*, 6094-6104.

Pawlak, M., and Jaenisch, R. (2011). De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. Genes Dev *25*, 1035-1040.

Pease, S., Braghetta, P., Gearing, D., Grail, D., and Williams, R.L. (1990). Isolation of embryonic stem (ES) cells in media supplemented with recombinant leukemia inhibitory factor (LIF). Dev Biol *141*, 344-352.

Pegoraro, G., Kubben, N., Wickert, U., Gohler, H., Hoffmann, K., and Misteli, T. (2009). Ageing-related chromatin defects through loss of the NURD complex. Nat Cell Biol *11*, 1261-1267.

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

Pijnappel, W.W., Esch, D., Baltissen, M.P., Wu, G., Mischerikow, N., Bergsma, A.J., van der Wal, E., Han, D.W., Bruch, H., Moritz, S., *et al.* (2013). A central role for TFIID in the pluripotent transcription circuitry. Nature *495*, 516-519.

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

Pollard, S.M., Benchoua, A., and Lowell, S. (2006). Neural stem cells, neurons, and glia. In Embryonic Stem Cells, I.L.R. Klimanskaya, ed., pp. 151-169.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., *et al.* (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617-1632.

Qian, Y.W., Wang, Y.C., Hollingsworth, R.E., Jr., Jones, D., Ling, N., and Lee, E.Y. (1993). A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. Nature *364*, 648-652.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Radzisheuskaya, A., Chia Gle, B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol *15*, 579-590.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., *et al.* (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65-70.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development *137*, 395-403.

Ramirez-Solis, R., Liu, P., and Bradley, A. (1995). Chromosome engineering in mice. Nature *378*, 720-724.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., *et al.* (2012a). NuRD Suppresses Pluripotency Gene Expression to Promote Transcriptional Heterogeneity and Lineage Commitment. Cell Stem Cell *10*, 583-594.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012b). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J *31*, 593-605.

Rizzino, A. (2009). Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells. Wiley Interdiscip Rev Syst Biol Med *1*, 228-236.

Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J Biol Chem *277*, 35434-35439.

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

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. Cell Stem Cell *7*, 64-77.

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

Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816-818.

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

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

Schrode, N., Xenopoulos, P., Piliszek, A., Frankenberg, S., Plusa, B., and Hadjantonakis, A.K. (2013). Anatomy of a blastocyst: cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. Genesis *51*, 219-233.

Schwarz, B.A., Bar-Nur, O., Silva, J.C., and Hochedlinger, K. (2014). Nanog is dispensable for the generation of induced pluripotent stem cells. Curr Biol *24*, 347-350.

Scotland, K.B., Chen, S., Sylvester, R., and Gudas, L.J. (2009). Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. Dev Dyn *238*, 1863-1877.

Seelig, H.P., Moosbrugger, I., Ehrfeld, H., Fink, T., Renz, M., and Genth, E. (1995). The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation. Arthritis Rheum *38*, 1389-1399.

Segre, J.A., Bauer, C., and Fuchs, E. (1999). Klf4 is a transcription factor required for establishing the barrier function of the skin. Nat Genet *22*, 356-360.

Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell *32*, 491-502.

Shimamoto, R., Amano, N., Ichisaka, T., Watanabe, A., Yamanaka, S., and Okita, K. (2014). Generation and characterization of induced pluripotent stem cells from aid-deficient mice. PLoS ONE *9*.

Shimbo, T., Du, Y., Grimm, S.A., Dhasarathy, A., Mav, D., Shah, R.R., Shi, H., and Wade, P.A. (2013). MBD3 localizes at promoters, gene bodies and enhancers of active genes. PLoS Genet *9*, 26.

Shinagawa, T., Takagi, T., Tsukamoto, D., Tomaru, C., Huynh, L.M., Sivaraman, P., Kumarevel, T., Inoue, K., Nakato, R., Katou, Y., *et al.* (2014). Histone variants enriched in oocytes enhance reprogramming to induced pluripotent stem cells. Cell Stem Cell *14*, 217-227.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., *et al.* (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell *153*, 963-975.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. PLoS Biol *6*, 2237-2247.

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

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722-737.

Sims, J.K., and Wade, P.A. (2011). Mi-2/NuRD complex function is required for normal S phase progression and assembly of pericentric heterochromatin. Mol Biol Cell *22*, 3094-3102.

Singhal, N., Graumann, J., Wu, G., Arauzo-Bravo, M.J., Han, D.W., Greber, B., Gentile, L., Mann, M., and Scholer, H.R. (2010). Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. Cell *141*, 943-955.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688-690.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543-549.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., *et al.* (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. Nat Cell Biol *15*, 872-882.

Sridharan, R., and Plath, K. (2011). Small RNAs loom large during reprogramming. Cell Stem Cell *8*, 599-601.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the Murine Reprogramming Factors in the Induction of Pluripotency. Cell *136*, 364-377.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S., Kim, S.Y., Bestor, T.H., Shioda, T., *et al.* (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat Genet *44*, 398-405.

Stadtfeld, M., Brennand, K., and Hochedlinger, K. (2008a). Reprogramming of pancreatic beta cells into induced pluripotent stem cells. Curr Biol *18*, 890-894.

Stadtfeld, M., and Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. Genes Dev *24*, 2239-2263.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008b). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230-240.

Steingrimsson, E., Tessarollo, L., Pathak, B., Hou, L., Arnheiter, H., Copeland, N.G., and Jenkins, N.A. (2002). Mitf and Tfe3, two members of the Mitf-Tfe family of bHLH-Zip transcription factors, have important but functionally redundant roles in osteoclast development. Proc Natl Acad Sci U S A *99*, 4477-4482.

Stuart, H.T., van Oosten, A.L., Radzisheuskaya, A., Martello, G., Miller, A., Dietmann, S., Nichols, J., and Silva, J.C. (2014). NANOG Amplifies STAT3 Activation and They Synergistically Induce the Naive Pluripotent Program. Curr Biol *24*, 340-346.

Subramanyam, D., Lamouille, S., Judson, R.L., Liu, J.Y., Bucay, N., Derynck, R., and Blelloch, R. (2011). Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. Nat Biotechnol *29*, 443-448.

Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.S., Sritanaudomchai, H., *et al.* (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. Cell *153*, 1228-1238.

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

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

Tai, C.I., and Ying, Q.L. (2013). Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state. J Cell Sci *126*, 1093-1098.

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

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

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T., and Akira, S. (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proc Natl Acad Sci U S A *94*, 3801-3804.

Tapia, N., Reinhardt, P., Duemmler, A., Wu, G., Arauzo-Bravo, M.J., Esch, D., Greber, B., Cojocaru, V., Rascon, C.A., Tazaki, A., *et al.* (2012). Reprogramming to pluripotency is an ancient trait of vertebrate Oct4 and Pou2 proteins. Nat Commun *3*.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Theunissen, T.W., Costa, Y., Radzisheuskaya, A., van Oosten, A.L., Lavial, F., Pain, B., Castro, L.F.C., and Silva, J.C.R. (2011a). Reprogramming capacity of Nanog is functionally conserved in vertebrates and resides in a unique homeodomain. Development *138*, 4853-4865.

Theunissen, T.W., van Oosten, A.L., Castelo-Branco, G., Hall, J., Smith, A., and Silva, J.C.R. (2011b). Nanog Overcomes Reprogramming Barriers and Induces Pluripotency in Minimal Conditions. Curr Biol *21*, 65-71.

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

Toh, Y., Pencil, S.D., and Nicolson, G.L. (1994). A novel candidate metastasis-associated gene, mta1, differentially expressed in highly metastatic mammary adenocarcinoma cell lines. cDNA cloning, expression, and protein analyses. J Biol Chem *269*, 22958-22963.

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

Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., and Schreiber, S.L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature *395*, 917-921.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. Development *135*, 909-918.

Toyooka, Y., Tsunekawa, N., Akasu, R., and Noce, T. (2003). Embryonic stem cells can form germ cells in vitro. Proc Natl Acad Sci U S A *100*, 11457-11462.

Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M., and Yamanaka, S. (2009). Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. Genes Cells *14*, 683-694.

Tweedie, S., Ng, H.H., Barlow, A.L., Turner, B.M., Hendrich, B., and Bird, A. (1999). Vestiges of a DNA methylation system in Drosophila melanogaster? Nat Genet *23*, 389-390.

Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature *460*, 1145-U1112.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. Cell Stem Cell *6*, 369-381.

van Oosten, A.L., Costa, Y., Smith, A., and Silva, J.C.R. (2012). JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. Nat Commun *3*.

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet *23*, 62-66.

Wade, P.A., Jones, P.L., Vermaak, D., and Wolffe, A.P. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. Curr Biol *8*, 843-846.

Wakayama, T., Perry, A.C.F., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature *394*, 369-374.

Wang, G., Guo, X., Hong, W., Liu, Q., Wei, T., Lu, C., Gao, L., Ye, D., Zhou, Y., Chen, J., *et al.* (2013a). Critical regulation of miR-200/ZEB2 pathway in Oct4/Sox2-induced mesenchymal-to-epithelial transition and induced pluripotent stem cell generation. Proc Natl Acad Sci U S A *110*, 2858-2863.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013b). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910-918.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y.L., Zhou, B., Han, L., Fargo, D.C., *et al.* (2014). INO80 Facilitates Pluripotency Gene Activation in Embryonic Stem Cell Self-Renewal, Reprogramming, and Blastocyst Development. Cell Stem Cell *14*, 575-591.

Wang, S., Xia, P., Ye, B., Huang, G., Liu, J., and Fan, Z. (2013c). Transient Activation of Autophagy via Sox2-Mediated Suppression of mTOR Is an Important Early Step in Reprogramming to Pluripotency. Cell Stem Cell *13*, 617-625.

Wang, T., Chen, K., Zeng, X., Yang, J., Wu, Y., Shi, X., Qin, B., Zeng, L., Esteban, M.A., Pan, G., *et al.* (2011). The Histone Demethylases Jhdm1a/1b Enhance Somatic Cell Reprogramming in a Vitamin-C-Dependent Manner. Cell Stem Cell *9*, 575-587.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al.* (2010). Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. Cell Stem Cell *7*, 618-630.

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. Cell Stem Cell *13*, 36-47.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc Natl Acad Sci U S A *86*, 5434-5438.

Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell *2*, 10-12.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M.C., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-U312.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.

Williams, C.J., Naito, T., Arco, P.G., Seavitt, J.R., Cashman, S.M., De Souza, B., Qi, X., Keables, P., Von Andrian, U.H., and Georgopoulos, K. (2004). The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. Immunity *20*, 719-733.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684-687.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P.T., Gertsenstein, M., *et al.* (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766-U106.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-U246.

Wu, G., Han, D., Gong, Y., Sebastiano, V., Gentile, L., Singhal, N., Adachi, K., Fischedick, G., Ortmeier, C., Sinn, M., *et al.* (2013). Establishment of totipotency does not depend on Oct4A. Nat Cell Biol *15*, 1089-1097.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., *et al.* (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134-1148.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell *2*, 851-861.

Yamada, M., Johannesson, B., Sagi, I., Burnett, L.C., Kort, D.H., Prosser, R.W., Paull, D., Nestor, M.W., Freeby, M., Greenberg, E., *et al.* (2014). Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells. Nature *28*.

Yamaguchi, Y., Ogura, S., Ishida, M., Karasawa, M., and Takada, S. (2005). Gene trap screening as an effective approach for identification of Wnt-responsive genes in the mouse embryo. Dev Dyn *233*, 484-495.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell Stem Cell *12*, 368-382.

Yang, J.A., van Oosten, A.L., Theunissen, T.W., Guo, G., Silva, J.C.R., and Smith, A. (2010). Stat3 Activation Is Limiting for Reprogramming to Ground State Pluripotency. Cell Stem Cell *7*, 319-328.

Ye, S., Li, P., Tong, C., and Ying, Q.L. (2013). Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. EMBO J *32*, 2548-2560.

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

Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell *147*, 1498-1510.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-U515.

Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y.I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Scholer, H.R., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Dev Growth Differ *41*, 675-684.

Young, R.A. (2011). Control of the Embryonic Stem Cell State. Cell *144*, 940-954.

Yu, J.Y., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

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

Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y., Aburatani, H., Ura, K., Koseki, H., Niwa, H., *et al.* (2009). Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression. Stem Cells *27*, 796-805.

Zhang, H., Jiao, W., Sun, L., Fan, J., Chen, M., Wang, H., Xu, X., Shen, A., Li, T., Niu, B., *et al.* (2013). Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. Cell Stem Cell *13*, 30-35.

Zhang, J., Jackson, A.F., Naito, T., Dose, M., Seavitt, J., Liu, F., Heller, E.J., Kashiwagi, M., Yoshida, T., Gounari, F., *et al.* (2011). Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. Nat Immunol *13*, 86-94.

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

Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell *95*, 279-289.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev *13*, 1924-1935.

Zhao, W., Li, Q., Ayers, S., Gu, Y., Shi, Z., Zhu, Q., Chen, Y., Wang, H.Y., and Wang, R.F. (2013). Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. Cell *152*, 1037-1050.

Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L., *et al.* (2009). iPS cells produce viable mice through tetraploid complementation. Nature *461*, 86-U88.

Zheng, B., Mills, A.A., and Bradley, A. (2001). Introducing defined chromosomal rearrangements into the mouse genome. Methods *24*, 81-94.

Zhou, H.Y., Wu, S.L., Joo, J.Y., Zhu, S.Y., Han, D.W., Lin, T.X., Trauger, S., Bien, G., Yao, S., Zhu, Y., *et al.* (2009). Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. Cell Stem Cell *4*, 381-384.

Zhu, D., Fang, J., Li, Y., and Zhang, J. (2009). Mbd3, a component of NuRD/Mi-2 complex, helps maintain pluripotency of mouse embryonic stem cells by repressing trophectoderm differentiation. PLoS ONE *4*, 0007684.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., *et al.* (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642-654.

Appendix

CellPress

Cell Stem Cell
Short Article

MBD3/NuRD Facilitates Induction of Pluripotency in a Context-Dependent Manner

Rodrigo L. dos Santos,[1,2] Luca Tosti,[3] Aliaksandra Radzisheuskaya,[1] Isabel M. Caballero,[3,5] Keisuke Kaji,[3,4] Brian Hendrich,[1,4] and José C.R. Silva[1,4,*]

[1]Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

[2]Doctoral Programme in Experimental Biology and Biomedicine, Centre for Neuroscience and Cell Biology and Institute for Interdisciplinary Research, University of Coimbra, 3030-789 Coimbra, Portugal

[3]MRC Centre for Regenerative Medicine, University of Edinburgh, Edinburgh BioQuarter, 5 Little France Drive, Edinburgh EH16 4UU, UK

[4]Co-senior author

[5]Present address: Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, 171 77 Sweden

*Correspondence: jcs64@cscr.cam.ac.uk

http://dx.doi.org/10.1016/j.stem.2014.04.019

SUMMARY

The Nucleosome Remodeling and Deacetylase (NuRD) complex is essential for embryonic development and pluripotent stem cell differentiation. In this study, we investigated whether NuRD is also involved in the reverse biological process of induction of pluripotency in neural stem cells. By knocking out MBD3, an essential scaffold subunit of the NuRD complex, at different time points in reprogramming, we found that efficient formation of reprogramming intermediates and induced pluripotent stem cells from neural stem cells requires NuRD activity. We also show that reprogramming of epiblast-derived stem cells to naive pluripotency requires NuRD complex function and that increased MBD3/NuRD levels can enhance reprogramming efficiency when coexpressed with the reprogramming factor NANOG. Our results therefore show that the MBD3/NuRD complex plays a key role in reprogramming in certain contexts and that a chromatin complex required for cell differentiation can also promote reversion back to a naive pluripotent cell state.

INTRODUCTION

Reprogramming of somatic cells to naive pluripotency can be robustly driven by the combined action of transcription factors and culture cues. Among the reprogramming transcription factors, OCT4 plays a central role, as it is sufficient and essential for the induction of pluripotent cells (Kim et al., 2009; Radzisheuskaya et al., 2013; Radzisheuskaya and Silva, 2014). OCT4 interactome studies in embryonic stem cells (ESCs) revealed members of the Nucleosome Remodeling and Deacetylase (NuRD) complex as its highest confidence interactors (Ding et al., 2012; Liang et al., 2008; Pardo et al., 2010; van den Berg et al., 2010). NuRD is composed of six core subunits with at least two enzymatic activities involved in gene regulation: histone deacetylase activity of HDAC1/2 subunits and ATP-dependent chromatin remodeling activity of Mi-2a/β subunits (Lai and Wade, 2011; McDonel et al., 2009). Methyl-CpG binding domain protein 3 (MBD3) is an essential scaffold protein of the NuRD complex, in the absence of which the complex is not assembled (Kaji et al., 2006; Zhang et al., 1999). Embryos lacking MBD3 die shortly after implantation (Hendrich et al., 2001; Kaji et al., 2007) and *Mbd3*-null ESCs are viable but show severely impaired lineage commitment and exhibit limited differentiation capacity (Kaji et al., 2006; Reynolds et al., 2012a, 2012b). Chromatin remodeling plays an important role in reprogramming to naive pluripotency (Apostolou and Hochedlinger, 2013; Papp and Plath, 2013). Because the NuRD complex is a high confidence interactor of Oct4 and a key regulator of developmental cell state transitions, we have investigated its involvement in the induction of pluripotency.

RESULTS

MBD3 Facilitates the Initiation of Reprogramming from Neural Stem Cells

To address the requirement of the NuRD complex in the reprogramming process, we established an *Mbd3*$^{-/-}$ clonal neural stem cell (NSC) line from *Mbd3*$^{fl/-}$ NSCs and an *Mbd3*$^{-/-}$ rescue NSC line (*Mbd3*$^{-/-}$*:Mbd3*) by stable transfection of an *Mbd3* transgene (Figures S1A–S1C available online). These NSC lines were transduced with retroviruses encoding *cMyc*, *Klf4*, and *Oct4* (rMKO) to initiate their reprogramming and were then switched to serum plus LIF (S+LIF) conditions (Figure 1A), which typically results in the formation of highly proliferative reprogramming intermediates, or preiPSCs (Silva et al., 2008). When we used retroviruses encoding *GFP* (rGFP), equal percentages of GFP^{+} cells were observed 72 hr after transduction of *Mbd3*$^{fl/-}$ or *Mbd3*$^{-/-}$ NSCs, indicating that *Mbd3* deletion does not affect transduction efficiency (Figures S1D and S1E). Strikingly, the kinetics of preiPSC emergence was markedly delayed in the *Mbd3*$^{-/-}$ cells. While *Mbd3*-expressing preiPSCs dominated the culture by day 4 posttransduction (d.p.t.), *Mbd3*$^{-/-}$

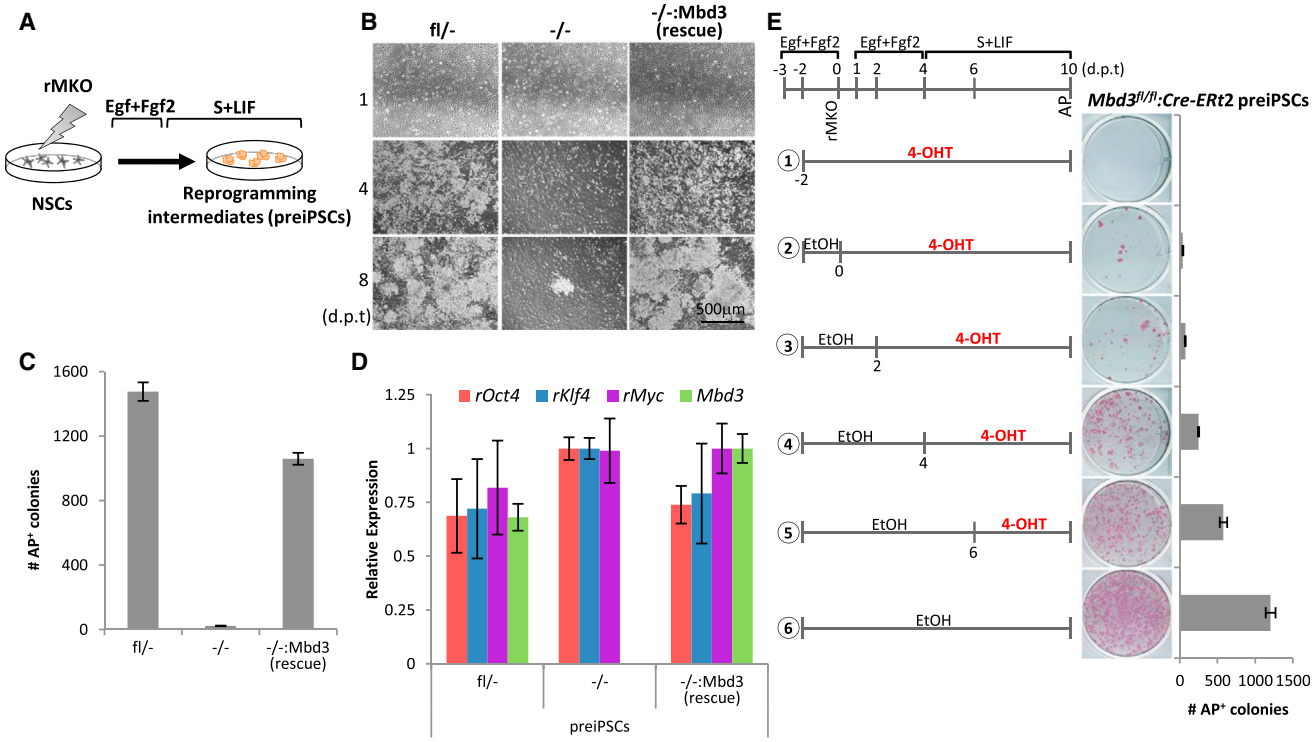


Figure 1. MBD3 Facilitates the Initiation of Reprogramming

(A) Experimental design used to address the kinetics and efficiency of initiation of reprogramming in NSCs with different *Mbd3* genotypes. NSCs were transduced with retroviruses encoding *cMyc*, *Klf4*, and *Oct4* (rMKO), maintained in Egf+Fgf2 medium for 3 days, and then switched to S+LIF medium.

(B) Phase images of the reprogramming intermediates (preiPSCs) emerging from $Mbd3^{fl/-}$, $Mbd3^{-/-}$, and $Mbd3^{-/-}$:*Mbd3* (rescue) NSCs at different days posttransduction (d.p.t.).

(C) Efficiency of preiPSC colony formation per 2.5 × 10^5 NSCs as assessed by alkaline phosphatase (AP) staining at day 9 posttransduction.

(D) qRT-PCR analysis of retroviral transgenes (*rOct4*, *rKlf4*, and *rMyc*) and *Mbd3* expression in the obtained preiPSCs maintained in S+LIF. Three independent NSCs transductions were carried out and gene expression was assessed 12 days after transduction. Values are normalized to *Gapdh* value and shown as relative to the highest value.

(E) Time course of MBD3 requirement during preiPSC formation. $Mbd3^{fl/fl}$ NSCs were stably transfected with pCAG-CreERt2 transgene, transduced with retroviral transgenes, and treated with 4-OHT at indicated time points to induce Cre-mediated deletion of the floxed alleles during reprogramming. Ethanol (EtOH) was used as a control. The encircled numbers correspond to different conditions. PreiPSC colony formation was assessed by AP staining at day 10 post-transduction and is presented as the number of colonies per 7.5 × 10^4 NSCs. The error bars indicate STDEV.

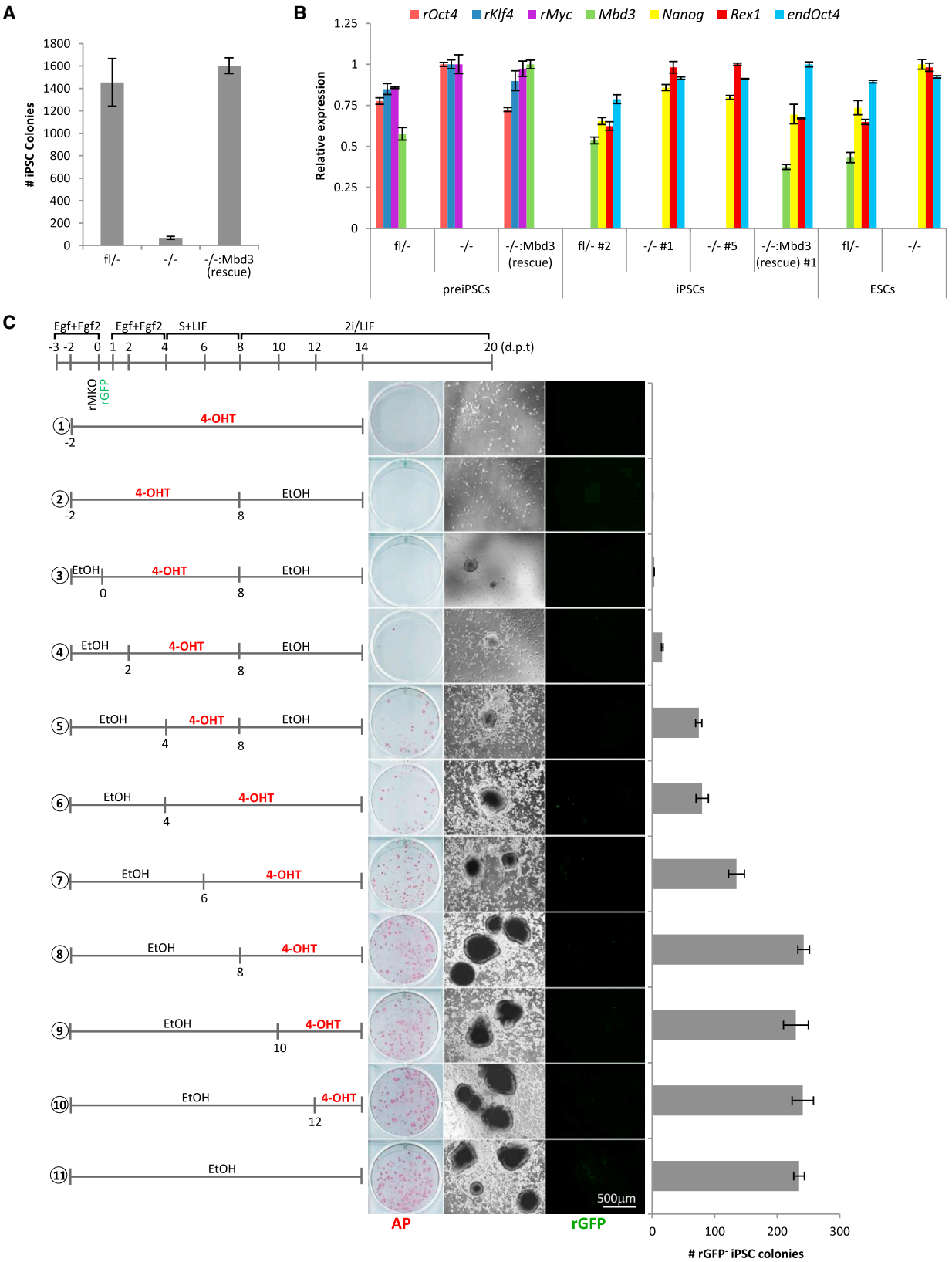
preiPSCs emerged only by 7–8 d.p.t. (Figure 1B). In addition, the number of emerging alkaline-phosphatase-positive (AP^+) $Mbd3^{-/-}$ preiPSC colonies was significantly reduced compared to parental and rescue cell lines (Figure 1C and Figure S1F). Nevertheless, it was possible to establish and expand $Mbd3^{-/-}$ preiPSCs, although less efficiently and with delayed kinetics. Both $Mbd3^{-/-}$ NSCs and *Mbd3*-null preiPSCs derived from them exhibited slower proliferation, consistent with previous reports of $Mbd3^{-/-}$ ESCs (Kaji et al., 2006; Sims and Wade, 2011) (Figures S1G and S1H). $Mbd3^{-/-}$ preiPSCs expressed slightly higher levels of retroviral transgenes compared to control cells (Figure 1D), suggesting that dosage of reprogramming factors is not the reason for the reduced efficiency of reprogramming initiation that we observed.

To further dissect the requirement for MBD3 in the initiation of reprogramming, we analyzed the effect of *Mbd3* deletion at different experimental time points. For this experiment, we stably transfected $Mbd3^{fl/fl}$ NSCs with *Cre-ERt2*, which enabled Cre-mediated excision of the floxed *Mbd3* alleles upon addition of 4-hydroxytamoxifen (4-OHT) (Figures S1I–S1L). We found that earlier removal of *Mbd3* reduced the number of preiPSC colonies formed (Figure 1E). We also obtained similar results after conditional deletion of *Mbd3* exon 1 (ex1fl) which removes all but a small amount of a truncated MBD3 protein isoform (MBD3C) (Aguilera et al., 2011; Kaji et al., 2006) (Figures S1M and S1N).

Taken together, these results demonstrate that lack of a functional NuRD complex strongly impairs the initiation of reprogramming from NSCs.

MBD3 Is Required for Efficient iPSC Generation from NSCs, preiPSCs, and EpiSCs

We then evaluated the role of MBD3 in later stages of reprogramming. To induce completion of the reprogramming process, $Mbd3^{fl/-}$, $Mbd3^{-/-}$, and rescued $Mbd3^{-/-}$:*Mbd3* preiPSCs were switched to serum-free medium containing LIF and inhibitors of both mitogen-activated protein kinase and glycogen synthase kinase-3 signaling (2i/LIF) (Silva et al., 2008), and the resulting iPSC colonies were scored 12 days later. We observed that the efficiency of conversion to naive pluripotency of $Mbd3^{-/-}$ preiPSCs is strongly reduced compared to



(legend on next page)

Mbd3$^{fl/-}$ and *Mbd3*$^{-/-}$*:Mbd3* preiPSCs (Figure 2A). The *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$, and *Mbd3*$^{-/-}$*:Mbd3* iPSCs that were obtained could be expanded clonally in 2i/LIF, and they exhibited reactivation of the pluripotency transcriptional program and silencing of the retroviral reprogramming transgenes as expected (Figure 2B). The *Mbd3*$^{-/-}$ iPSCs were phenotypically similar to previously reported *Mbd3*-null ESCs (Kaji et al., 2006), exhibiting impaired embryoid body differentiation and slower proliferation (Figures S2A and S2B). We also observed that *Mbd3* deletion in an established preiPSC line before the 2i/LIF medium switch impaired reprogramming to naive pluripotency (Figures S2C and S2D).

Next we performed a time course experiment to define the time frame during reprogramming for which MBD3 is required. For this analysis, we transduced *Mbd3*$^{fl/fl}$*:Cre-ERt2* NSCs with rMKO and rGFP and treated them with 4-OHT at different experimental time points (Figure 2C). The growth medium was changed to S+LIF 4 days after transduction and subsequently, 4 days later, to 2i/LIF. The number of iPSC colonies exhibiting silencing of retroviral GFP expression was assessed 12 days after 2i/LIF medium switch (Figure 2C and Figure S2E). We observed that the number of iPSC colonies formed was proportional to the amount of time cells expressed MBD3 during the initiation phase of reprogramming (prior to 2i/LIF culture). We observed neither a reduction nor a gain of reprogramming efficiency when *Mbd3* was deleted at the 2i/LIF stage. Regardless of the stage of *Mbd3* deletion, the resulting iPSCs displayed a pluripotency-associated transcriptional signature (Figure S2F). Thus, our data suggest that MBD3 is specifically required for the initiation and intermediate stage of NSC reprogramming rather than establishment of pluripotency.

Epiblast stem cells (EpiSCs) can be reprogrammed to naive pluripotency by a combination of overexpression of at least one transcription factor, such as KLF4, KLF2, or NANOG, and the use of serum-free 2i/LIF medium, which not only promotes reprogramming of EpiSCs but also blocks their self-renewal (Guo et al., 2009; Silva et al., 2009). To examine the role of MBD3 in reprogramming in this context, we stably transfected wild-type EpiSCs carrying an *Oct4*-GFP reporter with piggyBac (PB) vectors constitutively expressing *Klf2* and *Nanog* (K2N) or *Klf4*, and we then transfected these with either small interfering RNA (siRNA) against *Mbd3* or control siRNA (Figure 3A). Strikingly, *Mbd3* knockdown led to a complete impairment of KLF4-mediated reprogramming and to a 6-fold reduction in the reprogramming ability of K2N (Figure 3B and Figures S3A and S3B). Similar results were obtained when EpiSCs with *Mbd3* genetic knockout were used (Figure 3C and Figures S3C and S3D).

All the results described above indicate that MBD3 is critical for efficient reprogramming in the contexts that we examined, contrasting with previous reports (Luo et al., 2013; Rais et al., 2013). To examine whether this difference is a reflection of the specific reprogramming systems that we used, we performed PB-mediated reprogramming of mouse embryonic fibroblasts (MEFs) combined with *Mbd3* depletion using two different approaches (Figure 3D). First, we used an *Mbd3* knockdown system in which *Nanog*-GFP MEFs were treated with doxycycline (DOX) for the induction of the MKOS or STEMCCA reprogramming cassettes (Kaji et al., 2009; Sommer et al., 2009) and cultured in S+LIF medium supplemented with vitamin C and Alki (Tgfβ signaling inhibitor). Twenty-four hours after induction they were transduced with lentiviruses expressing shRNA against *Mbd3* (Figure 3E and Figures S3E–S3G). Second, we depleted *Mbd3* by treating *Cre-ERt2*-transduced *Mbd3*$^{ex1fl/ex1fl}$ MEFs with 4-OHT at 0 hr or 48 hrs after induction of reprogramming factor expression (Figure 3F and Figures S3H and S3I). While both systems demonstrated about 80% downregulation of MBD3 protein, neither impacted on the efficiency of MEF reprogramming. However, depletion of MBD3 protein would take a few days from the time of 4-OHT administration, so it is possible that in this system cells go through the most critical stage of reprogramming with MBD3 protein still present at a sufficient level.

From our experiments we therefore found that, depending on the reprogramming context, MBD3/NuRD depletion can either have no apparent effect on reprogramming or significantly impair the transition to naive pluripotency.

Overexpression of MBD3/NuRD Can Facilitate Reprogramming

Because the complete removal of *Mbd3* or a decrease in its expression can significantly impair the generation of both pre-iPSCs and iPSCs from NSCs, and iPSCs from preiPSCs and EpiSCs, we tested whether MBD3 levels are limiting for reprogramming. For that, *Oct4*-GFP reporter NSCs and *Nanog*-GFP MEF-derived preiPSCs were stably transfected with *Mbd3* (Figures 4A and 4E). MBD3 overexpression had neither a positive nor detrimental effect on the efficiency of iPSC formation in both systems (Figures 4B, 4C, 4F and 4G). However, combined overexpression of MBD3 and NANOG in MEF-derived preiPSCs led to accelerated reprogramming kinetics and an up to 30-fold increase in reprogramming efficiency compared to *Nanog-Empty* vector (*EV*) control (Figures 4F and 4G and Figures S4A and S4B). This synergistic effect correlated with the upregulation of both *Esrrβ* and endogenous *Oct4* expression, to 5% and 3% of the expression levels of wild-type ESCs, respectively, prior to induction of pluripotency by 2i/LIF medium switch (Figure 4I). Interestingly, MBD3 overexpression in these cell lines caused an increase in protein levels of MTA2, a core subunit of the NuRD

Figure 2. MBD3 Is Required for Efficient iPSC Generation

(A) Quantification of iPSC colonies generated from *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$, and *Mbd3*$^{-/-}$*:Mbd3* (rescue) preiPSCs after 2i/LIF culture for 12 days. Colony number is per 1.0 × 10^5 preiPSCs.

(B) qRT-PCR analysis of retroviral transgenes, *Mbd3*, and pluripotency-associated factors in preiPSCs and corresponding derived iPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value.

(C) *Mbd3*$^{fl/fl}$*:Cre-ERt2* NSCs were transduced with rMKO and rGFP, maintained in Egf+Fgf2 medium for 3 days, switched to S+LIF for 4 more days to allow preiPSC emergence, and then switched to 2i/LIF conditions to induce iPSC formation. 4-OHT was added at different time points (before or after preiPSC emergence) to induce *Mbd3*-floxed alleles excision. The encircled numbers correspond to different conditions. At day 20 after transfection, GFP$^-$ iPSC colonies were counted and subsequently stained for AP. The number of colonies is presented per 7.5 × 10^4 NSCs. The error bars indicate STDEV.

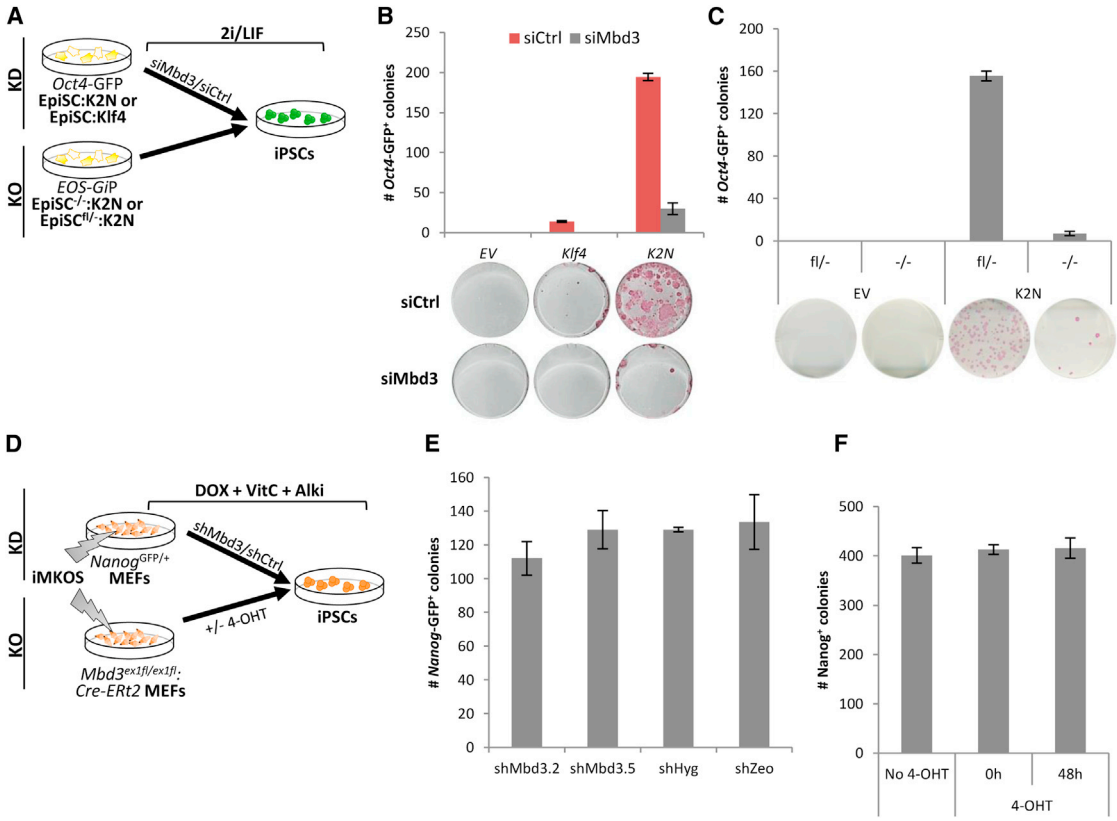


Figure 3. Requirement of MBD3 in Other Reprogramming Systems

(A) Experimental designs used to analyze the effect of *Mbd3* KD and KO on EpiSC reprogramming efficiency. For the KD experiments, wild-type EpiSCs (carrying an *Oct4*-GFP cassette), stably transfected with pPB-CAG-Klf2.2A.Nanog (K2N) or pPB-CAG-Klf4, were transfected with either siMbd3 or siControl (siCtrl) and, after 24 hr, were plated in 2i/LIF for 12 days. For the KO experiments, $Mbd3^{-/-}$ or $Mbd3^{fl-}$ EpiSCs carrying an *Oct4*-GFP reporter (*EOS*-GiP), stably transfected with K2N (or empty vector control, EV), were plated in 2i/LIF for 12 days.

(B and C) The efficiency of EpiSC reprogramming after *Mbd3* removal, either by KD (B) or KO (C), was assessed by counting *Oct4*-GFP$^+$ colonies. Representative AP stained plates are also indicated. 1.0×10^4 EpiSCs were plated in (B). 1.5×10^4 EpiSCs were plated in (C).

(D) Experimental designs used to analyze the effect of *Mbd3* KD and *Mbd3* exon 1 KO. For the KD experiments, *Nanog*-GFP MEFs transfected with doxycycline-inducible MKOS piggyBac transposon (iMKOS) were cultured in S+LIF + DOX + vitamin C (vitC) + Alki 24 hr before lentiviral infections of shMbd3 or shControls. For the KO experiments, $Mbd3^{ex1fl/ex1fl}$ MEFs transfected with iMKOS were infected with pMX-Cre-ERt2. Reprogramming was carried out in S+LIF + DOX + vitC + Alki, and 4-OHT was added either at the time of DOX administration (0h) or 48 hr later (48h).

(E) Number of *Nanog*-GFP$^+$ iPSC colonies at day 13 of reprogramming upon infection of indicated shRNAs.

(F) Number of Nanog$^+$ colony numbers determined by immunofluorescence after 13 days of reprogramming of $Mbd3^{ex1fl/ex1fl}$:*Cre*-ERt2 MEFs. The error bars indicate STDEV. Typical iMKOS positive cell number at day 2 of reprogramming is $1.0–3.0 \times 10^4$ cells per well, providing 1%–2% reprogramming efficiency.

complex degraded in the absence of MBD3 (Figures 4D and 4H) (Kaji et al., 2006). This suggests that the effects of MBD3 overexpression are potentially attributable to the total amount, and subsequently total activity, of the NuRD complex. We also found that EpiSCs overexpressing both Nanog and MBD3 showed a 30-fold increase in the ability to generate iPSCs relative to a *Nanog* only control (Figures 4K–4N). Importantly, all iPSCs generated by the overexpression of NANOG and MBD3 exhibited the molecular properties expected for naive pluripotent cells (Figures S4C–S4F) as well as chimera and germline competence after the excision of reprogramming transgenes (Figures 4J and 4O). We did not see reprogramming synergy with the NuRD complex for two other known reprogramming factors, KLF4 and NR5A2 (Figure 4P and Figure S4G).

The MBD3 isoform that we used for the rescue and overexpression experiments was MBD3B, the most abundant isoform by protein levels in ESCs (Figure S4K). In contrast to MBD3B, we found that MBD3C did not synergize with NANOG (Figure 4Q and Figures S4H–S4J). The two isoforms differ in only the first 60 N-terminal amino acids (Figure S4H), indicating that this region is of importance for NANOG-dependent MBD3 ability to facilitate reprogramming.

These results demonstrate that MBD3 overexpression does not impair induction of naive pluripotency and that it can in fact facilitate reprogramming in conjunction with enhanced NANOG expression.

DISCUSSION

In this study we have identified a positive facilitator role for MBD3/NuRD in transcription-factor-mediated reprogramming of NSCs and EpiSCs. In our analyses, we found that genetic

or siRNA-mediated depletion of *Mbd3* led to a reduction in the efficiency of reprogramming in these contexts, but not in reprogramming of MEFs. More specifically, we found through time course experiments that MBD3/NuRD function is particularly important during the initiation phase of reprogramming of NSCs and is more dispensable in the later stages when the pluripotency network is becoming more stably established.

We also found that MBD3 overexpression, with resulting higher levels of the NuRD complex, facilitates reprogramming of MEF-derived preiPSCs and EpiSCs when combined with expression of NANOG, but not with other tested reprogramming factors, which is consistent with previous observations that NuRD complex subunits are high confidence protein interactors of both OCT4 and NANOG (Costa et al., 2013; Ding et al., 2012; Gagliardi et al., 2013; Liang et al., 2008; Pardo et al., 2010; van den Berg et al., 2010). In our experiments the N teminus of the MBD3B isoform, which has previously been suggested to be required for protein-protein interactions (Aguilera et al., 2011), appeared to be required for the observed synergistic effect with NANOG in reprogramming (Figure 4Q). In the future we will aim to understand how NANOG and MBD3 work together to drive cells (preiPSCs) that are arrested in the reprogramming process toward pluripotency.

Our data suggest that the NuRD complex might be facilitating gene activation during reprogramming. Interestingly, MBD3 was recently shown to localize to the regulatory sequences of active genes (Günther et al., 2013; Reynolds et al., 2013; Shimbo et al., 2013), including ESC super-enhancers (Hnisz et al., 2013). Moreover, genome-wide expression analysis revealed that 61% of differentially expressed genes are downregulated after *Mbd3* deletion in ESCs (Reynolds et al., 2012b). Although some of this decrease in transcription might be due to indirect effects, it seems likely that the NuRD complex acts at enhancers as a mediator of transcription-factor-induced gene activation and thus could also interact with pluripotency factors such as NANOG to support genome-wide reprogramming. In addition, NuRD has been proposed to mediate transient mTOR downregulation and subsequent activation of autophagy, a key step during early stages of reprogramming (Wang et al., 2013).

Our results are in apparent disagreement with two recent reports that suggested an inhibitory role for MBD3 in reprogramming (Luo et al., 2013; Rais et al., 2013), including one (Rais et al., 2013) that argued that reduction or deletion of *Mbd3* leads to rapid deterministic reprogramming with 100% efficiency. There are a number of differences between our study and these two previous reports, including the choice of reprogramming cassettes and the reprogramming culture conditions. In contrast to our study, Rais et al. (2013) used a secondary system for somatic cell reprogramming and lentiviral cassette delivery, and it has been reported that both of these factors influence iPSC generation efficiency (Stadtfeld and Hochedlinger, 2010). Moreover, distinct reprogramming factor stoichiometry can provide varying intracellular environments, which may show different dependencies on MBD3 activity for reprogramming. In addition, in our hands heterozygous $Mbd3^{fl/-}$ ESCs express MBD3 at nearly wild-type levels (Figures S4K–S4M; Reynolds et al., 2012b), but Rais et al. (2013) reported that their $Mbd3^{fl/-}$ ESCs expressed MBD3 at 20% of wild-type levels. Further examination of these and other practical and procedural differences between our study and the previous work should help clarify the basis of the apparent differences seen.

Overall, taking into account the results that we report here and previous studies, our conclusion is that at least in some contexts MBD3/NuRD plays a positive role in reprogramming, and that loss of MBD3 expression leads to a reduction in the efficiency of the reprogramming process.

NuRD plays well-documented roles in controlling gene expression and developmental transitions in a wide variety of different metazoan systems (Ahringer, 2000; McDonel et al., 2009; Reynolds et al., 2013). MBD3 is known to be required for embryonic development and pluripotent cell differentiation (Kaji et al., 2006, 2007), and the composition of the complex or specific interactions of its individual subunits may regulate different aspects of its function (Allen et al., 2013; Reynolds et al., 2013). Further insights into the function of the NuRD complex during different cell state transitions will help us understand the process of induced pluripotency as well as embryonic development.

EXPERIMENTAL PROCEDURES

Cell Culture

Platinum-E, preiPSCs, and MEFs were cultured in GMEM (Sigma-Aldrich) supplemented with 10% FCS, 1× NEAA, 1× Pen/Strep, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, 2 mM L-glutamine, and 20 ng/ml of LIF (homemade), indicated as *S*+LIF medium throughout. ESCs and iPSCs were maintained in N2B27-based medium (DMEM/F12 and Neurobasal [both Life Technologies] in 1:1 ratio, 1× Pen/Strep, 0.1 mM 2-mercaptoethanol, 2 mM L-glutamine, 1:200 N2 [PAA], and 1:100 B27 [Life Technologies]) supplemented 20 ng/ml of LIF and 2i inhibitors: CHIR99021 (3 μM) and PD0325901 (1 μM), indicated as 2i/LIF throughout (Ying et al., 2008). NSCs were cultured in DMEM/F12 (GIBCO) supplemented with 1× NEAA, 0.1 mM 2-mercaptoethanol, 1× Pen/Strep, 1:100 B27, 1:200 N2 supplement, 4.5 μM HEPES, 0.03 M glucose, 120 μg/ml BSA, 10 ng/ml of Egf (Peprotech), and 20 ng/ml of Fgf2 (homemade), indicated as Egf+Fgf2 medium throughout. EpiSCs were maintained in N2B27-based medium containing 12 ng/ml of Fgf2 and 20 ng/ml of Activin A (homemade), indicated as Fgf2/Act.A medium throughout. EpiSCs and NSCs were cultured on plastic coated with fibronectin (10 μg/ml, Millipore) or laminin (10 μg/ml, Sigma-Aldrich), respectively. All other cell types were grown on gelatine. All cell types were maintained at 7% CO_2. For Cre-mediated transgene excision, cells were treated with 500 nM of 4-OHT.

Derivation of Cell Lines

NSCs

Brains from $Mbd3^{fl/fl}$ and $Mbd3^{ex1fl/ex1fl}$ E13.5 embryos were dissected, dissociated in Egf+Fgf2 medium, and plated onto the laminin-coated cell culture flasks. $Mbd3^{fl/-}$ NSCs were derived from ESCs as described (Pollard et al., 2006). Briefly, ESCs were seeded on gelatinized 10 cm dishes in N2B27 medium for 7 days. After this period, cells were trypsinized and plated on nongelatinized dishes for 3 days in Egf+Fgf2 medium. The emergent neurospheres were then seeded on gelatinized plates and maintained in monolayer in Egf+Fgf2 medium. For Cre-excision of the *Mbd3* floxed allele, $Mbd3^{fl/-}$ NSCs were nucleofected with a pCAG-Cre-ires-Puro plasmid and clonal lines of $Mbd3^{-/-}$ NSCs were expanded.

MEFs

Organ-deprived carcasses from E12.5 or E13.5 embryos were dissociated into small pieces, trypsinized, and plated in S+LIF medium.

EpiSCs

$Mbd3^{fl/-}$ and $Mbd3^{-/-}$ EpiSCs were derived from ESCs as previously described (Guo et al., 2009). Briefly, ESCs transfected with pPB-EOS-GFP-ires-Puro (*EOS*-GiP; GFPiresPuro under the control of early transposon promoter and *Oct4* and *Sox2* enhancers) were cultured in Fgf2/Act.A medium

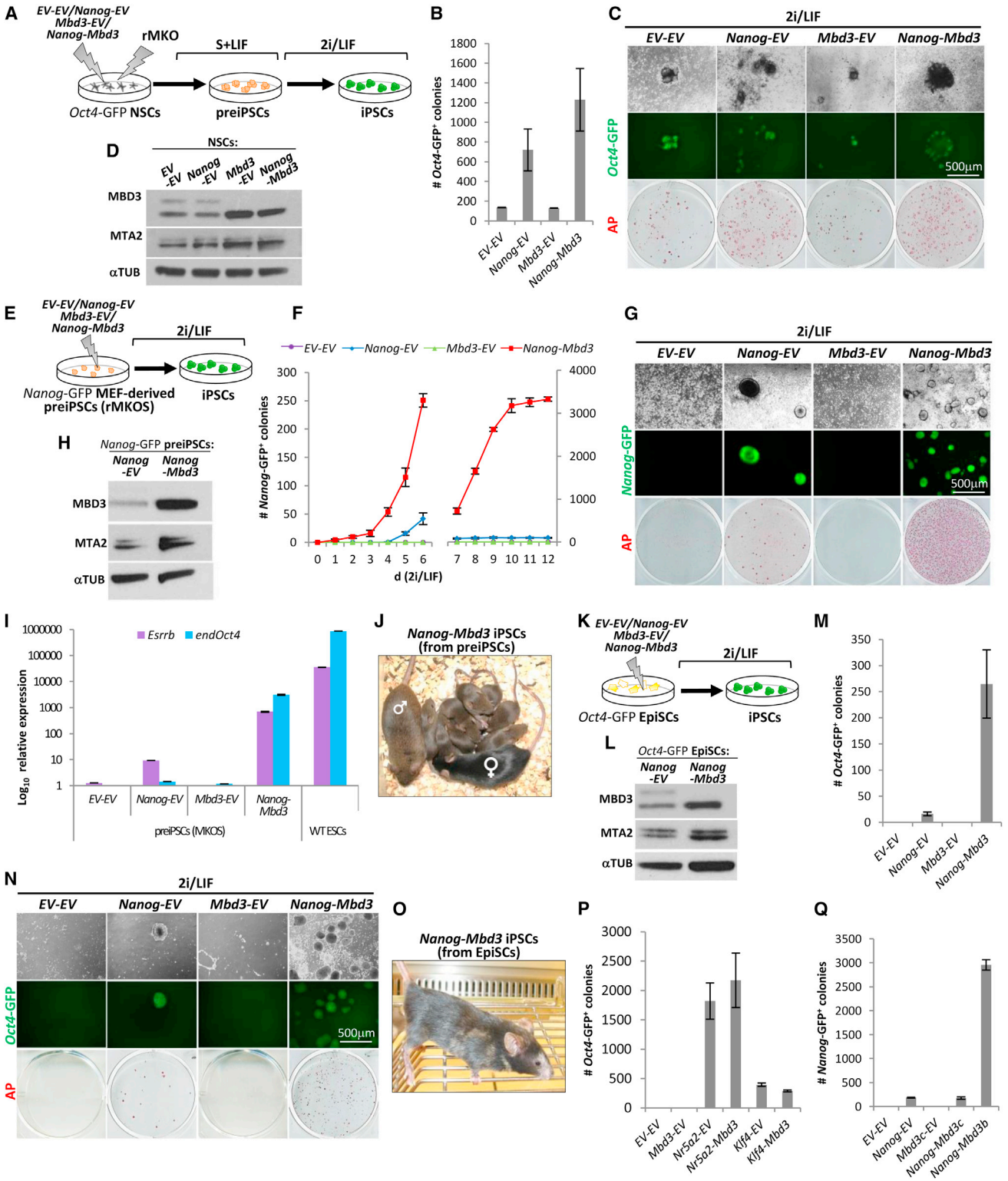


Figure 4. Overexpression of MBD3/NuRD Facilitates NANOG-Mediated Reprogramming

(A) Experimental design used to address the effect of MBD3 overexpression on NSC reprogramming. NSCs carrying an *Oct4*-GFP cassette were stably transfected with pPB-CAG-Nanog and pPB-CAG-Mbd3b or pPB-CAG-empty controls, transduced with rOKM, cultured in Egf+Fgf2 medium for 3 days, switched to S+LIF medium for 6 days, and then switched to 2i/LIF conditions.

(B) Quantification of *Oct4*-GFP$^+$ colonies after 12 days in 2i/LIF conditions. Colony number is per 1.0 × 10^5 NSCs.

(C) Phase and GFP images and AP staining of the iPSCs obtained from NSCs overexpressing respective transgenes.

(legend continued on next page)

for at least 10 passages before analysis. To obtain a pure EpiSC culture, GFP$^+$ cells were removed by FACS.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes four figures and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2014.04.019.

ACKNOWLEDGMENTS

We acknowledge N. Reynolds and J. Ramalho-Santos for advice and helpful discussions. We are also grateful to Y. Costa for critical reading of the manuscript. This study was supported by a Wellcome Trust Fellowship (WT101861), an ERC starting grant, and the Anne Rowling Clinic. J.C.R.S. and B.H. are Wellcome Trust Senior Research Fellows in the Basic Biomedical Sciences. R.L.S. is a recipient of a Ph.D. fellowship from the Portuguese Foundation for Sciences and Technology, FCT (SFRH/BD/51198/2010). L.T. is a recipient of a Ph.D. fellowship from The College of Medicine and Veterinary Medicine, University of Edinburgh.

Received: January 8, 2014
Revised: March 27, 2014
Accepted: April 24, 2014
Published: May 15, 2014

REFERENCES

Aguilera, C., Nakagawa, K., Sancho, R., Chakraborty, A., Hendrich, B., and Behrens, A. (2011). c-Jun N-terminal phosphorylation antagonises recruitment of the Mbd3/NuRD repressor complex. Nature *469*, 231–235.

Ahringer, J. (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends Genet. *16*, 351–356.

Allen, H.F., Wade, P.A., and Kutateladze, T.G. (2013). The NuRD architecture. Cell. Mol. Life Sci. *70*, 3513–3524.

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462–471.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., and Wang, J. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res. *22*, 155–167.

Gagliardi, A., Mullin, N.P., Ying Tan, Z., Colby, D., Kousa, A.I., Halbritter, F., Weiss, J.T., Felker, A., Bezstarosti, K., Favaro, R., et al. (2013). A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. EMBO J. *32*, 2231–2247.

Günther, K., Rust, M., Leers, J., Boettger, T., Scharfe, M., Jarek, M., Bartkuhn, M., and Renkawitz, R. (2013). Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. Nucleic Acids Res. *41*, 3010–3021.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.

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

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

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

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development *134*, 1123–1132.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771–775.

Kim, J.B., Sebastiano, V., Wu, G., Araúzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., et al. (2009). Oct4-induced pluripotency in adult neural stem cells. Cell *136*, 411–419.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat. Rev. Cancer *11*, 588–596.

Liang, J., Wan, M., Zhang, Y., Gu, P.L., Xin, H.W., Jung, S.Y., Qin, J., Wong, J.M., Cooney, A.J., Liu, D., and Songyang, Z. (2008). Nanog and Oct4

(D) Western blot analysis of MBD3, MTA2, and TUBULIN (TUB) protein expression in NSCs overexpressing the indicated transgene combinations.
(E) Experimental design used to address the effect of MBD3/NuRD overexpression on the conversion of preiPSCs to iPSCs. PreiPSCs (carrying a *Nanog*-GFP) were stably transfected with the same transgene combinations as in (A) and plated in 2i/LIF conditions for 12 days.
(F) The kinetics of the emergence of *Nanog*-GFP$^+$ colonies from the transgenic preiPSCs during a 12 day culture in 2i/LIF conditions (y axis scale changes at day 7). Colony number is per 1.0 × 10^5 preiPSCs.
(G) Phase and GFP images and AP staining of the iPSCs formed from preiPSCs overexpressing respective transgenes.
(H) Western blot analysis of MBD3, MTA2, and TUBULIN (TUB) protein expression in preiPSCs overexpressing NANOG or NANOG and MBD3.
(I) qRT-PCR analysis of *Esrrβ* and endogenous (end) *Oct4* expression in preiPSCs 12 days after stable transgene transfection and culture in S+LIF (y axes in $\log_{10}$ scale). The expression levels of *Esrrβ* and endogenous *Oct4* in these *Nanog-Mbd3* preiPSCs are 5% and 3%, respectively, of the expression levels of WT ESCs in 2i/LIF. The *Esrrβ* expression level is also approximately 80 times greater than that of *Nanog-EV* preiPSCs and 700 times greater than that of *Mbd3-EV* and *EV-EV* preiPSCs. The endogenous *Oct4* expression level is approximately 2,000 times greater than that of *Nanog-EV* preiPSCs and 3,000 times greater than that of *Mbd3-EV* and *EV-EV* preiPSCs. qRT-PCR values are normalized to *Gapdh* value and shown as relative to the highest value.
(J) Germ line contribution of *Nanog-Mbd3* iPSCs generated from preiPSCs (brown color). Cells were treated with TAT-Cre for reprogramming transgene excision prior to blastocyst injection. Chimeric father, C57BL/6 mother, and pups resulting from cross can be viewed.
(K) Experimental design used to address the effect of MBD3/NuRD overexpression on EpiSC reprogramming. EpiSCs (carrying an *Oct4*-GFP) were stably transfected with the same transgene combinations as in (A) and (E) and plated in 2i/LIF conditions for 12 days.
(L) Western blot analysis of MBD3, MTA2, and TUBULIN (TUB) protein expression in EpiSCs overexpressing NANOG or NANOG and MBD3.
(M) Quantification of *Oct4*-GFP$^+$ colonies after 12 days of 2i/LIF culture. Colony numbers are per 2.0 × 10^4 EpiSCs.
(N) Phase and GFP images and AP staining of the iPSCs formed from EpiSCs overexpressing respective transgenes.
(O) Chimera of *Nanog-Mbd3* iPSCs generated from EpiSCs (brown color). Cells were treated with TAT-Cre for reprogramming transgene excision prior to blastocyst injection.
(P) Quantification of *Oct4*-GFP$^+$ colonies after 12 days of 2i/LIF culture, generated from EpiSCs transfected with *Klf4* or *Nr5a2* (together or not with *Mbd3*). Colony numbers are per 2.0 × 10^4 EpiSCs.
(Q) Quantification of *Nanog*-GFP$^+$ colonies after 12 days of 2i/LIF culture generated from MEF-derived preiPSCs stably transfected with *Nanog* alone, or *Nanog* together with *Mbd3b* or *Mbd3c*. Colony numbers are per 1.0 × 10^5 preiPSCs. The error bars indicate STDEV.

associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell Biol. *10*, 731–739.

Luo, M., Ling, T., Xie, W., Sun, H., Zhou, Y., Zhu, Q., Shen, M., Zong, L., Lyu, G., Zhao, Y., et al. (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells *31*, 1278–1286.

McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. Int. J. Biochem. Cell Biol. *41*, 108–116.

Papp, B., and Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell *152*, 1324–1343.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

Pollard, S.M., Benchoua, A., and Lowell, S. (2006). Neural stem cells, neurons, and glia. In Embryonic Stem Cells, I.L.R. Klimanskaya, ed. (Boston: Elsevier Academic Press), pp. 151–169.

Radzisheuskaya, A., and Silva, J.C. (2014). Do all roads lead to Oct4? The emerging concepts of induced pluripotency. Trends Cell Biol. *24*, 275–284.

Radzisheuskaya, A., Chia, Gle.B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat. Cell Biol. *15*, 579–590.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012a). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012b). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J. *31*, 593–605.

Reynolds, N., O'Shaughnessy, A., and Hendrich, B. (2013). Transcriptional repressors: multifaceted regulators of gene expression. Development *140*, 505–512.

Shimbo, T., Du, Y., Grimm, S.A., Dhasarathy, A., Mav, D., Shah, R.R., Shi, H., and Wade, P.A. (2013). MBD3 localizes at promoters, gene bodies and enhancers of active genes. PLoS Genet. *9*, e1004028.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol. *6*, e253.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Sims, J.K., and Wade, P.A. (2011). Mi-2/NuRD complex function is required for normal S phase progression and assembly of pericentric heterochromatin. Mol. Biol. Cell *22*, 3094–3102.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543–549.

Stadtfeld, M., and Hochedlinger, K. (2010). Induced pluripotency: history, mechanisms, and applications. Genes Dev. *24*, 2239–2263.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

Wang, S., Xia, P., Ye, B., Huang, G., Liu, J., and Fan, Z. (2013). Transient activation of autophagy via Sox2-mediated suppression of mTOR is an important early step in reprogramming to pluripotency. Cell Stem Cell *13*, 617–625.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924–1935.

nature cell biology

ARTICLES

A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages

Aliaksandra Radzisheuskaya[1], Gloryn Le Bin Chia[2], Rodrigo L. dos Santos[1,3], Thorold W. Theunissen[1,5], L. Filipe C. Castro[4], Jennifer Nichols[2] and José C. R. Silva[1,6]

Oct4 is considered a master transcription factor for pluripotent cell self-renewal, but its biology remains poorly understood. Here, we investigated the role of Oct4 using the process of induced pluripotency. We found that a defined embryonic stem cell (ESC) level of Oct4 is required for pluripotency entry. However, once pluripotency is established, the Oct4 level can be decreased up to sevenfold without loss of self-renewal. Unexpectedly, cells constitutively expressing Oct4 at an ESC level robustly differentiated into all embryonic lineages and germline. In contrast, cells with low Oct4 levels were deficient in differentiation, exhibiting expression of naive pluripotency genes in the absence of pluripotency culture requisites. The restoration of Oct4 expression to an ESC level rescued the ability of these to restrict naive pluripotent gene expression and to differentiate. In conclusion, a defined Oct4 level controls the establishment of naive pluripotency as well as commitment to all embryonic lineages.

Naive pluripotency characterizes the cells that can give rise to all cell types of an organism except extraembryonic tissues. In mouse embryos these cells arise during pre-implantation development in the naive epiblast. This transient cell population can be captured *in vitro* as ESCs. In addition to its developmental potential, the naive pluripotent state is characterized by a unique set of properties, including the lack of an inactive X chromosome in female cells, self-renewing response to Mek/Erk signalling inhibition, and simultaneous expression of Esrrb, Nanog, Rex1, Klf2 and Klf4 (ref. 1).

Oct4 plays a fundamental role in mammalian development as a master transcriptional regulator of naive pluripotency maintenance. It belongs to the POU family of transcription factors and possesses the POU DNA-binding domain characteristic of this family[2–4]. Oct4 is expressed in oocytes, blastomeres, inner cell mass (ICM), naive and post-implantation epiblast, germ cells, and in pluripotent cells *in vitro*[2,3,5]. Its knockout causes pre-implantation lethality of mouse embryos due to failure to form a pluripotent ICM (ref. 6). Moreover, both cessation and overexpression of Oct4 cause exit from ESC self-renewal[7,8]. Oct4 is also sufficient to trigger reprogramming of mouse and human somatic cells in the absence of other reprogramming transgenes, albeit with decreased efficiency and delayed kinetics[9–12].

Here, we established a system using $Oct4^{-/-}$ somatic cells and 2i/LIF culture medium, containing LIF and inhibitors of mitogen-activated protein kinase signalling and glycogen synthase kinase-3β. Using this, we uncovered the existence of biological roles of Oct4 that have a critical impact on pluripotency acquisition, self-renewal and on *in vitro* and *in vivo* cell differentiation.

RESULTS

An ESC level of Oct4 marks acquisition of naive pluripotency

To investigate Oct4 function during induced pluripotency we generated $Oct4^{-/-}$ neural stem cells (NSCs; Supplementary Fig. S1a–e). $Oct4^{-/-}$ and control $Oct4^{+/-}$ NSCs were transduced with retroviruses expressing c-Myc and Klf4 (rMK) and transfected with a piggyBac (PB) vector containing a ubiquitous promoter (CAG) driving Oct4 expression (PB-Oct4; Fig. 1a). The CAG promoter, unlike the retroviral promoter, does not undergo silencing during reprogramming. This strategy produced reprogramming intermediates in serum/LIF conditions that, on medium switch to 2i/LIF, formed induced

[1]Wellcome Trust—Medical Research Council Cambridge Stem Cell Institute and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. [2]Wellcome Trust—Medical Research Council Cambridge Stem Cell Institute and Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. [3]Doctoral Programme in Experimental Biology and Biomedicine, Centre for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal. [4]CIMAR/CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas, Porto 4050-123, Portugal. [5]Present address: Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. [6]Correspondence should be addressed to J.C.R.S. (e-mail: jcs64@cscr.cam.ac.uk)

Received 19 September 2012; accepted 26 March 2013; published online 30 April 2013; DOI: 10.1038/ncb2742

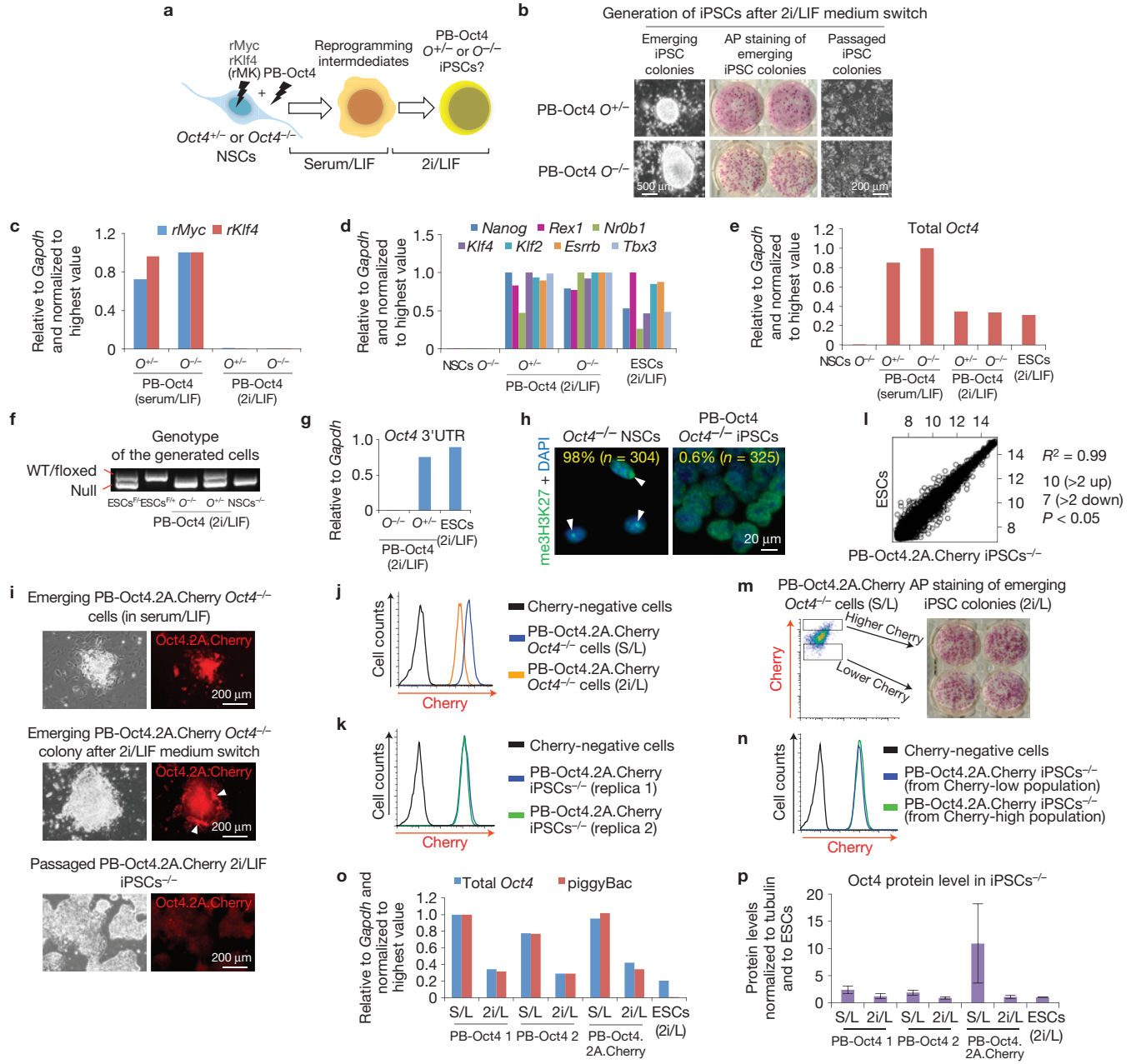


Figure 1 An ESC level of Oct4 marks pluripotency acquisition. (**a**) Generation of rMK+PB-Oct4 iPSCs$^{-/-}$. Reprogramming intermediates are represented in orange, and iPSCs in yellow. (**b**) Phase images and alkaline phosphatase (AP) staining of rMK+PB-Oct4 iPSCs$^{+/-}$ and iPSCs$^{-/-}$. (**c**–**e**) Quantitative PCR with reverse transcription (qRT-PCR) analysis of retroviral transgenes (**c**), pluripotency markers (**d**) and total Oct4 (**e**) expression in rMK+PB-Oct4 cells before and after 2i/LIF induction. Serum/LIF indicates reprogramming intermediates; 2i/LIF indicates iPSCs. ESCs were grown in 2i/LIF. Data shown are the mean of 3 replicates and are from 1 of 3 representative experiments. (**f**) *Oct4* locus genotyping for rMK+PB-Oct4 iPSCs$^{+/-}$ and iPSCs$^{-/-}$ and control *Oct4*$^{-/-}$ NSCs (NSCs$^{-/-}$), *Oct4*$^{flox/-}$ (ESCs$^{F/-}$) and *Oct4*$^{flox/+}$ (ESCs$^{F/+}$) ESCs. (**g**) qRT–PCR analysis of *Oct4* 3′UTR expression in rMK+PB-Oct4 iPSCs$^{+/-}$ and iPSCs$^{-/-}$. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**h**) me3H3K27 immunostaining of NSCs$^{-/-}$ and rMK+PB-Oct4 iPSCs$^{-/-}$. White arrowheads indicate representative inactive X chromosomes. (**i**) Phase and Cherry images of rMK+PB-Oct4.2A.Cherry reprogramming intermediates, the rMK+PB-Oct4.2A.Cherry iPSC$^{-/-}$ colony with surrounding Cherry-high reprogramming intermediates (white arrowheads) in 2i/LIF and of the established rMK+PB-Oct4.2A.Cherry iPSC$^{-/-}$ line. (**j**,**k**) Cherry flow cytometry analysis of rMK+PB-Oct4.2A.Cherry *Oct4*$^{-/-}$ cells before and after 2i/LIF induction (**j**) and of two independently derived pools of rMK+PB-Oct4.2A.Cherry iPSCs$^{-/-}$ (**k**). Serum/LIF indicates reprogramming intermediates; 2i/LIF indicates iPSCs. (**l**) Scatter plot comparing global gene expression profiles of ESCs and rMK+PB-Oct4.2A.Cherry iPSCs$^{-/-}$. (**m**) Alkaline phosphatase staining demonstrating the comparable ability of Cherry-high and -low reprogramming intermediates to acquire pluripotency. (**n**) Cherry flow cytometry analysis of rMK+PB-Oct4.2A.Cherry iPSCs$^{-/-}$ derived from Cherry-high and -low reprogramming intermediates. (**o**) qRT–PCR analysis of piggyBac and total Oct4 expression in rMK+PB-Oct4 *Oct4*$^{-/-}$ cells before and after 2i/LIF induction. Serum/LIF (S/L) indicates reprogramming intermediates; 2i/LIF (2i/L) indicates iPSCs. PB-Oct4 1 and 2 indicate biological replicates. ESCs were grown in 2i/LIF. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**p**) Oct4 protein quantification in rMK+PB-Oct4 cells before and after 2i/LIF induction. Data shown are the mean of 3 independent experiments; error bars represent ± s.d. See Supplementary Fig. S9 for uncropped data.

pluripotent stem cell (iPSC) colonies (Fig. 1b). PB-Oct4 iPSCs$^{-/-}$ exhibited silencing of retroviral transgenes and upregulation of naive pluripotency markers (Fig. 1c,d). Strikingly, the total Oct4 level was similar between iPSCs$^{+/-}$, iPSCs$^{-/-}$ and ESCs (Fig. 1e). The absence of endogenous Oct4 expression in iPSCs$^{-/-}$ was confirmed by genotyping (Fig. 1f) and undetected expression of the *Oct4* 3′UTR, which is absent in the knockout loci and PB-Oct4 transgene (Fig. 1g). Acquisition of a naive pluripotent cell state was further confirmed by the loss of the trimethyl(me3)H3K27 nuclear focus indicative of X chromosome reactivation (Fig. 1h). Thus, we generated and maintained iPSCs$^{-/-}$dependent exclusively on constitutively expressed PB-Oct4 transgene. We also observed that, independently of the source of Oct4 expression, iPSCs exhibit an ESC level of Oct4 transcript on pluripotency establishment.

To monitor PB transgene expression at the single-cell level during reprogramming, we used a PB-Oct4.2A.Cherry construct. Consistent with gene expression data (Fig. 1e), reprogramming intermediates showed a strong Cherry signal, whereas iPSCs$^{-/-}$ obtained after 2i/LIF induction demonstrated a lower Cherry expression level (Fig. 1i,j and Supplementary Fig. S1f). Notably, in each experiment PB-Oct4.2A.Cherry iPSCs$^{-/-}$ represented a pool of hundreds of colonies formed as a result of multiple independent reprogramming events. Importantly, the pools of iPSCs$^{-/-}$ obtained in independent experiments demonstrated a similar small range of Cherry and, thereby, Oct4 expression (Fig. 1k). This indicates selection and/or modulation of Oct4 transgene expression during reprogramming. The PB-Oct4.2A.Cherry iPSCs$^{-/-}$ had a global gene expression profile similar to ESCs, with only 17 genes differentially expressed by more than twofold (Fig. 1l). To assess whether a particular level of Oct4 transgene expression facilitates reprogramming intermediates to transit into naive pluripotency, we sorted the highest and lowest Cherry-expressing cells and plated these in 2i/LIF. We did not observe any difference in the ability of Oct4 high- and low-expressing cell fractions to undergo reprogramming (Fig. 1m). Importantly, obtained iPSCs$^{-/-}$ exhibited similar Cherry expression profiles (Fig. 1n). Combined gene expression and western blot analysis of independently derived iPSC$^{-/-}$ lines further confirmed that these exhibit an ESC level of Oct4 on entry into the pluripotent cell state (Fig. 1o,p and Supplementary Fig. S1g). We also generated *Oct4*$^{-/-}$ iPSCs from an independent somatic cell type, mouse embryonic fibroblasts, and, consistently, these exhibited an ESC level of Oct4 expression (Supplementary Fig. S1h). This shows that pluripotent cell state acquisition and/or maintenance requires modulation of Oct4 transgene expression to an ESC level.

Low Oct4 expression sustains self-renewal

It is established that abolishment of Oct4 expression in ESCs in serum/LIF leads to differentiation towards trophectoderm[7]. As the Oct4 transgene in the PB vector is flanked by *loxP* sites and our iPSCs$^{-/-}$ express 4-hydroxytamoxifen (4OHT)-inducible Cre recombinase, we tested these for the capacity to undergo trophectoderm differentiation on Oct4 deletion. Consistent with previous reports, 4OHT treatment in serum/LIF resulted in some trophectoderm differentiation judging by morphology and expression of trophectoderm marker Pl-1 (Fig. 2a,b). However, Oct4 expression was not completely abolished (Fig. 2c). As cells probably contain multiple PB transgene integrations, this indicates that not all of the inserts were excised. Surprisingly, an average 12-fold reduction in Oct4 expression level did not affect the average expression of naive pluripotency markers Nanog and Rex1 (Fig. 2c). When the same cells were treated with 4OHT in 2i/LIF conditions (Fig. 2d), we observed a sevenfold reduction in the Oct4 level (Fig. 2e). Again, both Rex1 and Nanog expression remained unchanged, indicating that these cells maintain a naive pluripotent cell state (Fig. 2e). Oct4 and Nanog immunocytochemistry revealed a wide range of Oct4 expression in 4OHT-treated cells, with some of the Oct4-low cells showing strong Nanog signal, above that of control cells (Fig. 2f and Supplementary Fig. S2a).

We picked a colony of 4OHT-treated iPSCs$^{-/-}$ and established a cell line with reduced Oct4 transcript and protein levels (Fig. 2h,i). It retained ESC morphology (Fig. 2g), Nanog expression (Fig. 2h) and could be serially passaged without any signs of differentiation. To validate this, we 4OHT-treated an independent cell line, PB-Oct4.2A.Cherry iPSCs$^{-/-}$, and subsequently single-cell sorted them for low Cherry-expressing cells. This allowed establishment of several Oct4-low iPSC$^{-/-}$ lines (Fig. 2j). These clones exhibited a similar decrease in Oct4 transcript and protein levels (Fig. 2k,l) and retained expression of naive pluripotency genes at levels comparable to ESCs (Fig. 2m). To further demonstrate that, despite reduced Oct4 expression, Oct4-low iPSCs$^{-/-}$ match the molecular criteria of the naive pluripotent state, we performed global gene expression analysis of these in 2i/LIF. This revealed a very similar profile to parental iPSCs$^{-/-}$ with an ESC Oct4 level (Oct4-WT iPSCs$^{-/-}$; Fig. 2n), and to ESCs (data not shown). We also confirmed that Oct4-low cells do not express epiblast stem cell markers (Supplementary Fig. S2b). Moreover, Oct4-low iPSCs$^{-/-}$ contained unmethylated Nanog and Oct4 regulatory regions (Supplementary Fig. S2c), and exhibited absence of the H3K27me3 nuclear focus (Supplementary Fig. S2d).

To determine the highest Oct4 expression level permissive of pluripotent cell self-renewal, we transfected wild-type ESCs with PB-Oct4.2A.Cherry construct. Consistent with previous findings[7], we observed that the cells expressing the highest level of Cherry had differentiated morphology and were lost on passaging (Supplementary Fig. S2e). Self-renewing cells exhibited lower Cherry levels than cells transfected with PB-Cherry alone, further indicating that high Oct4 expression is detrimental for the pluripotent state (Supplementary Fig. S2f). We picked and expanded the highest Cherry-expressing ESC-like clones. All demonstrated similar total Oct4 levels but lower endogenous Oct4 when compared with parental ESCs (Supplementary Fig. S2g). They also had lower Cherry expression levels than PB-Oct4.2A.Cherry iPSCs$^{-/-}$ (Supplementary Fig. S2h). This indicates that Oct4 transgene expression is compatible with self-renewal only if ESCs$^{+/+}$ compensate by equally downregulating endogenous Oct4 expression. We also overexpressed Oct4 episomally[13] in ESCs to determine which differentiation genes become upregulated. An average 11-fold Oct4 overexpression led to significant upregulation of differentiation markers representing all three embryonic lineages: Gata4; Zeb2 and Snai2; and Sox1 (Supplementary Fig. S2i).

In conclusion, expression of up to fivefold lower levels of Oct4 sustains naive pluripotent cell self-renewal. This signifies that an ESC level of Oct4 expression is a requirement for pluripotency entry but not self-renewal.

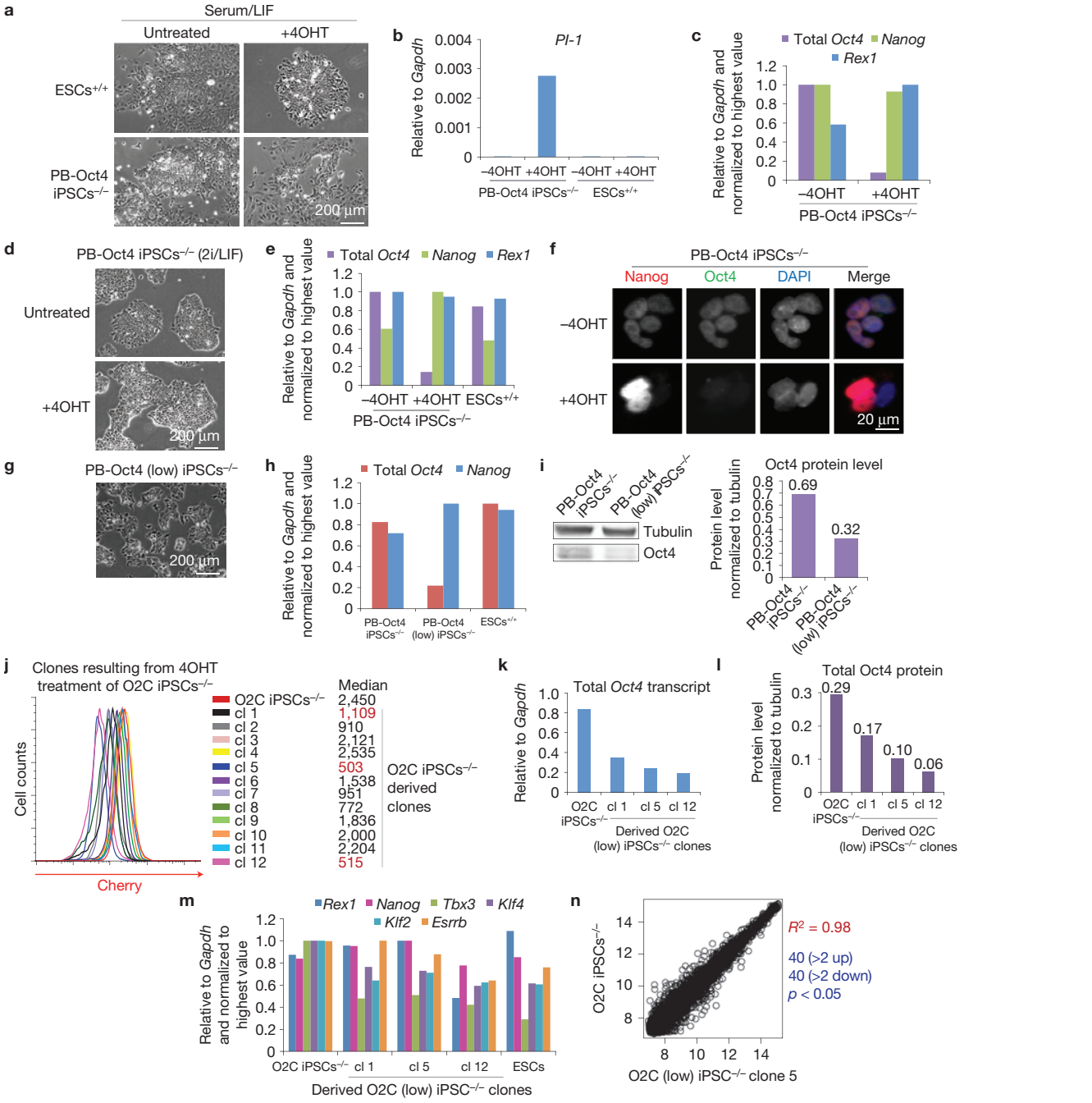


Figure 2 Low levels of Oct4 expression sustain self-renewal. (**a**) Phase images of ESCs and rMK+PB-Oct4 iPSCs$^{-/-}$ treated and untreated with 4OHT in serum/LIF culture conditions. (**b**,**c**) qRT–PCR analysis of *Pl-1* (**b**) and pluripotency markers (**c**) expression in ESCs and rMK+PB-Oct4 iPSCs$^{-/-}$ treated and untreated with 4OHT in serum/LIF culture conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**d**) Phase images of rMK+PB-Oct4 iPSCs$^{-/-}$ treated and untreated with 4OHT in 2i/LIF culture conditions. (**e**) qRT–PCR analysis of total Oct4, Nanog and Rex1 expression in rMK+PB-Oct4 iPSCs$^{-/-}$ treated and untreated with 4OHT in 2i/LIF culture conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**f**) Nanog and Oct4 immunocytochemistry in rMK+PB-Oct4 iPSCs$^{-/-}$ treated and untreated with 4OHT in 2i/LIF culture conditions. (**g**) Phase image of rMK+PB-Oct4 (low) iPSCs$^{-/-}$ in 2i/LIF conditions. (**h**) qRT–PCR analysis of total Oct4 and Nanog expression in PB-Oct4, PB-Oct4 (low) iPSCs$^{-/-}$ and ESCs in 2i/LIF culture conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**i**) Western blot analysis of Oct4 protein levels in PB-Oct4 and PB-Oct4 (low) iPSCs$^{-/-}$ in 2i/LIF culture conditions. See Supplementary Fig. S9 for uncropped data. (**j**) Flow cytometry analysis of PB-Oct4.2A.Cherry (low) iPSC$^{-/-}$ clones obtained as a result of 4OHT treatment of PB-Oct4.2A.Cherry iPSCs$^{-/-}$ and subsequent single-cell sorting. PB-Oct4.2A.Cherry is indicated as O2C. Clones for which the median is indicated in red were chosen for subsequent analysis. (**k**) qRT–PCR analysis of total *Oct4* expression in PB-Oct4.2A.Cherry (low) iPSC$^{-/-}$ clones 1, 5 and 12. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**l**) Oct4 protein levels in PB-Oct4.2A.Cherry (low) iPSC$^{-/-}$ clones 1, 5 and 12. (**m**) qRT–PCR analysis of pluripotency gene expression in PB-Oct4.2A.Cherry (low) iPSC$^{-/-}$ clones 1, 5 and 12. Data shown are the mean of 3 replicates and are from 1 of 3 representative experiments. (**n**) Scatter plot comparing global gene expression profiles of PB-Oct4.2A.Cherry iPSCs$^{-/-}$ and PB-Oct4.2A.Cherry (low) iPSC$^{-/-}$ clone 5.

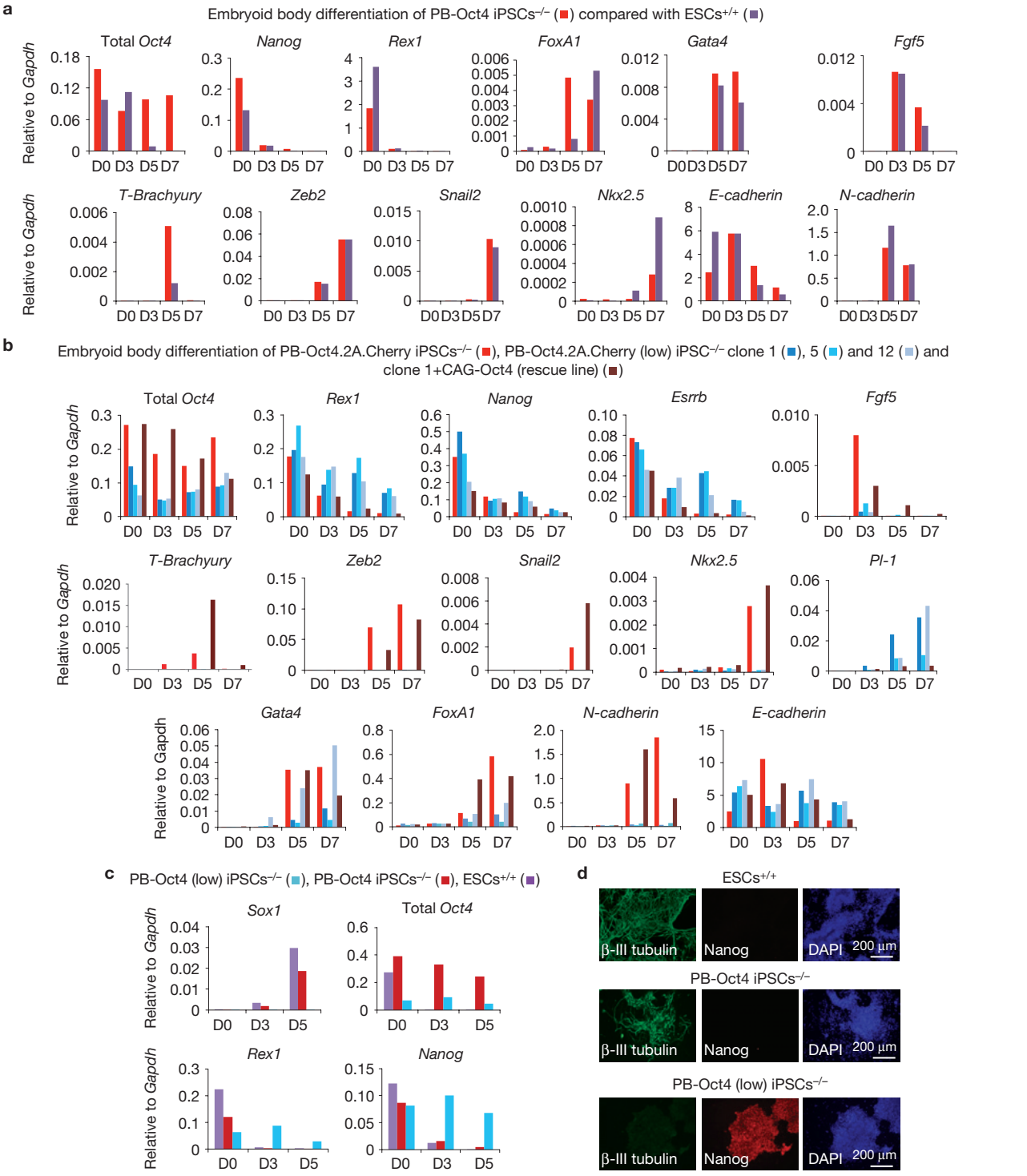


Figure 3 Oct4 expression at an ESC level is required for *in vitro* differentiation. (**a**) qRT–PCR analysis of pluripotency (total *Oct4, Nanog, Rex1*), endoderm (*FoxA1, Gata4*), ectoderm (*Fgf5*) and mesoderm (*T-Brachyury, Zeb2, Snai2, Nkx2.5, N-cadherin*) markers and *E-cadherin* expression during embryoid body differentiation of ESCs$^{+/+}$ and rMK+PB-Oct4 iPSCs$^{-/-}$. D0–D7 indicate the number of days of differentiation. Data shown are the mean of 3 replicates and are from 1 of 4 representative experiments. (**b**) qRT–PCR analysis of pluripotency (total *Oct4, Nanog, Rex1, Esrrb*), endoderm (*FoxA1, Gata4*), ectoderm (*Fgf5*), mesoderm (*T-Brachyury, Zeb2, Snai2, Nkx2.5, N-cadherin*) and trophectoderm (*Pl-1*) markers and *E-cadherin* expression during embryoid body differentiation of PB-Oct4.2A.Cherry iPSCs$^{-/-}$, PB-Oct4.2A.Cherry (low) iPSC$^{-/-}$ clone 1, 5 and 12 iPS$^{-/-}$ and PB-Oct4.2A.Cherry (low) clone 1 + CAG-Oct4 (rescue line). Data shown are the mean of 3 replicates and are from 1 of 3 representative experiments. (**c**) qRT–PCR analysis of *Sox1*, total *Oct4, Rex1* and *Nanog* expression during neural induction by monolayer culture of control ESCs$^{+/+}$, PB-Oct4 iPSCs$^{-/-}$ and PB-Oct4 (low) iPSCs$^{-/-}$. D0–D5 indicate the number of days of differentiation. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**d**) Immunocytochemistry detection of βIII-tubulin and Nanog expression after 7 days of neural induction by monolayer culture of control ESCs$^{+/+}$, PB-Oct4 iPSCs$^{-/-}$ and PB-Oct4 (low) iPSCs$^{-/-}$.

In vitro differentiation requires an ESC level of Oct4

To further define the properties of Oct4-low and Oct4-WT iPSCs$^{-/-}$ we analysed their ability to differentiate *in vitro*. Despite constitutive expression of the Oct4 transgene, Oct4-WT iPSCs$^{-/-}$ underwent efficient embryoid body differentiation into all three germ layers, as judged by the upregulation of mesoderm, endoderm and ectoderm markers (Fig. 3a). Efficient E-cadherin downregulation demonstrated successful epithelial-to-mesenchymal transition (Fig. 3a). These also efficiently formed beating heart cells on embryoid body outgrowth (Supplementary Video S1). Strikingly, Oct4-low iPSCs$^{-/-}$ differentiated poorly in embryoid body assays, exhibiting failure to downregulate pluripotent gene expression and to upregulate differentiation markers (Fig. 3b and Supplementary Fig. S3a). They also showed trophectoderm marker upregulation (Fig. 3b and Supplementary Fig. S3a) and did not form beating heart cells. Importantly, restoration of Oct4 expression to an ESC level in Oct4-low iPSCs$^{-/-}$ by the introduction of a CAG-Oct4 transgene rescued their differentiation defect (Fig. 3b). In addition, whereas Oct4-WT iPSCs$^{-/-}$ differentiated into the neural lineage in the monolayer protocol[14], judging by Sox1 upregulation and the formation of β-III tubulin-positive neurons, Oct4-low iPSCs$^{-/-}$ failed to downregulate Nanog and Rex1 and to upregulate neural lineage markers (Fig. 3c,d and Supplementary Fig. S3b,c). Importantly, restoration of Oct4 expression to an ESC level in Oct4-low cells reinstated their ability to form neural lineage (Supplementary Fig. S3b,c).

These data show that Oct4 expression at an ESC level is required for efficient *in vitro* differentiation.

In vivo differentiation requires an ESC level of Oct4

To address the ability of PB-Oct4.2A.Cherry and PB-Oct4.2A.Cherry (low) iPSCs$^{-/-}$ to contribute to mouse development we performed morula aggregations. Consistent with their naive pluripotent state, by the blastocyst stage both iPSC$^{-/-}$ lines efficiently incorporated into the pre-implantation epiblast (Fig. 4a). Despite constitutive Oct4 expression, embryonic day (E)6.5 PB-Oct4.2A.Cherry chimaeric embryos appeared normal with the whole epiblast consisting of Cherry-positive cells (Fig. 4b). Strikingly, despite efficiently incorporating into the naive epiblast, PB-Oct4.2A.Cherry (low) iPSCs$^{-/-}$ failed to proceed in development. Only 4/30 E6.5 embryos showed some contribution to the post-implantation epiblast (Fig. 4b and Supplementary Fig. S4a). The remaining embryos were either Cherry-negative or contained Cherry-positive cells between the epiblast and trophectoderm (Fig. 4b,c and Supplementary Fig. S4a,b). These remaining embryos were lost on *in vitro* culture of the embryos (Supplementary Fig. S4c), potentially reflecting cell exclusion from development due to failure to differentiate. This is supported by the fact that when compared with the blastocyst stage, few E6.5 embryos show the presence of Cherry-positive cells. To validate this, we dissociated PB-Oct4.2A.Cherry (low) E6.5 chimaeric embryos into single-cell suspension and plated them in 2i/LIF, which would select for cells remaining in a naive pluripotent state. As a result, we observed the emergence of numerous Cherry-positive colonies (Supplementary Fig. S4d), which retained Nanog, Klf4 and Rex1 expression (Supplementary Fig. S4e). This indicates that most, if not all, of the PB-Oct4.2A.Cherry (low) cells retained a naive pluripotent state. Together, these results highlight a requirement for an ESC level of Oct4 expression for post-implantation development. In agreement with this, restoration of Oct4 expression to an ESC level rescued the ability of PB-Oct4.2A.Cherry (low) iPSCs$^{-/-}$ to efficiently incorporate into the post-implantation embryo (Fig. 4d).

Subsequently, we analysed PB-Oct4.2A.Cherry iPSC$^{-/-}$ E7.5 chimaeric embryos. Strikingly, constitutive Oct4 expression at an ESC level was compatible with contribution to all embryonic lineages (Fig. 4e,f and Supplementary Fig. S4f), specifically, early mesoderm marked by T-Brachyury[15] and presumptive neuroectoderm marked by Sox2 (ref. 16; Fig. 4f). Foxa2, marking early progenitors of all germ layers at the anterior primitive streak[17], was largely co-expressed with Cherry in the chimaeric embryos (Fig. 4f). Moreover, Cherry was co-expressed with Sox17 exclusively in the inner layer of Sox17-positive cells, representing nascent definitive endoderm[18] (Fig. 4f).

At E8.5, PB-Oct4.2A.Cherry embryos appeared morphologically normal and, judging by Cherry and Oct4 protein expression, virtually all embryonic tissues were of iPSC$^{-/-}$ origin (Fig. 4g,h). However, chimaeric embryos inefficiently proceeded in development after this stage (Supplementary Fig. S4g). This is also a developmental stage when Oct4 expression becomes restricted to the germ lineage[5,19]. As obtained embryos demonstrated high if not 100% chimaerism, we performed blastocyst injections with 1–5 cells to determine whether chimaeric embryos can develop further in the presence of host embryo cells. Widespread iPSC$^{-/-}$ contribution was observed in E12.5 chimaeric embryos, which was the latest time point analysed (Supplementary Fig. S4h). Analysis of genital ridges at E12.5 demonstrated contribution of PB-Oct4.2A.Cherry iPSCs$^{-/-}$ to the germline (Fig. 4i). We also observed that PB-Oct4.2A.Cherry iPSCs$^{-/-}$ are more efficient at entering embryonic development than control ESCs (Fig. 4j,k). This suggests that premature loss of Oct4 expression in ESCs on embryo injection leads to decreased chimaerism and further shows that Oct4 facilitates cell state transitions during pluripotent state exit. We also performed teratoma assays for Oct4-WT and Oct4-low iPSCs$^{-/-}$. Whereas teratomas derived from Oct4-WT and rescue iPSCs$^{-/-}$ contained various tissues representing all three embryonic lineages, teratomas derived from the two Oct4-low iPSC$^{-/-}$ lines had only discernible areas of trophectoderm-like and undifferentiated cells (Supplementary Fig. S5).

In summary, Oct4 expression at an ESC level is required for efficient *in vivo* differentiation into all three germ layers.

Oct4 genomic binding is converse to Nanog and is linked to downregulation of naive pluripotency genes

Pluripotent cell cultures in 2i/LIF are homogeneous and do not contain differentiated cells or cells primed for differentiation. However, pluripotent cultures in serum/LIF have subpopulations of cells with variable levels of naive pluripotency marker expression[20–24]. We analysed these in Oct4-WT and Oct4-low iPSCs$^{-/-}$ in 2i/LIF compared with serum/LIF in the presence of selection for *Oct4* promoter activity to eliminate differentiated cells (Supplementary Fig. S1b). There were no major differences in pluripotency marker expression, at both transcript and protein levels, between analysed cell lines in 2i/LIF (Fig. 5a,c). However, in serum/LIF, we observed higher expression levels of naive pluripotency markers in all of the Oct4-low iPSC$^{-/-}$ lines, at both transcript and protein levels, when compared with parental iPS$^{-/-}$ and rescued Oct4-low iPSCs$^{-/-}$ (Fig. 5b,d). The observed expression differences could not be attributed to differentiation (Supplementary Fig. S6a,b). To gain insight

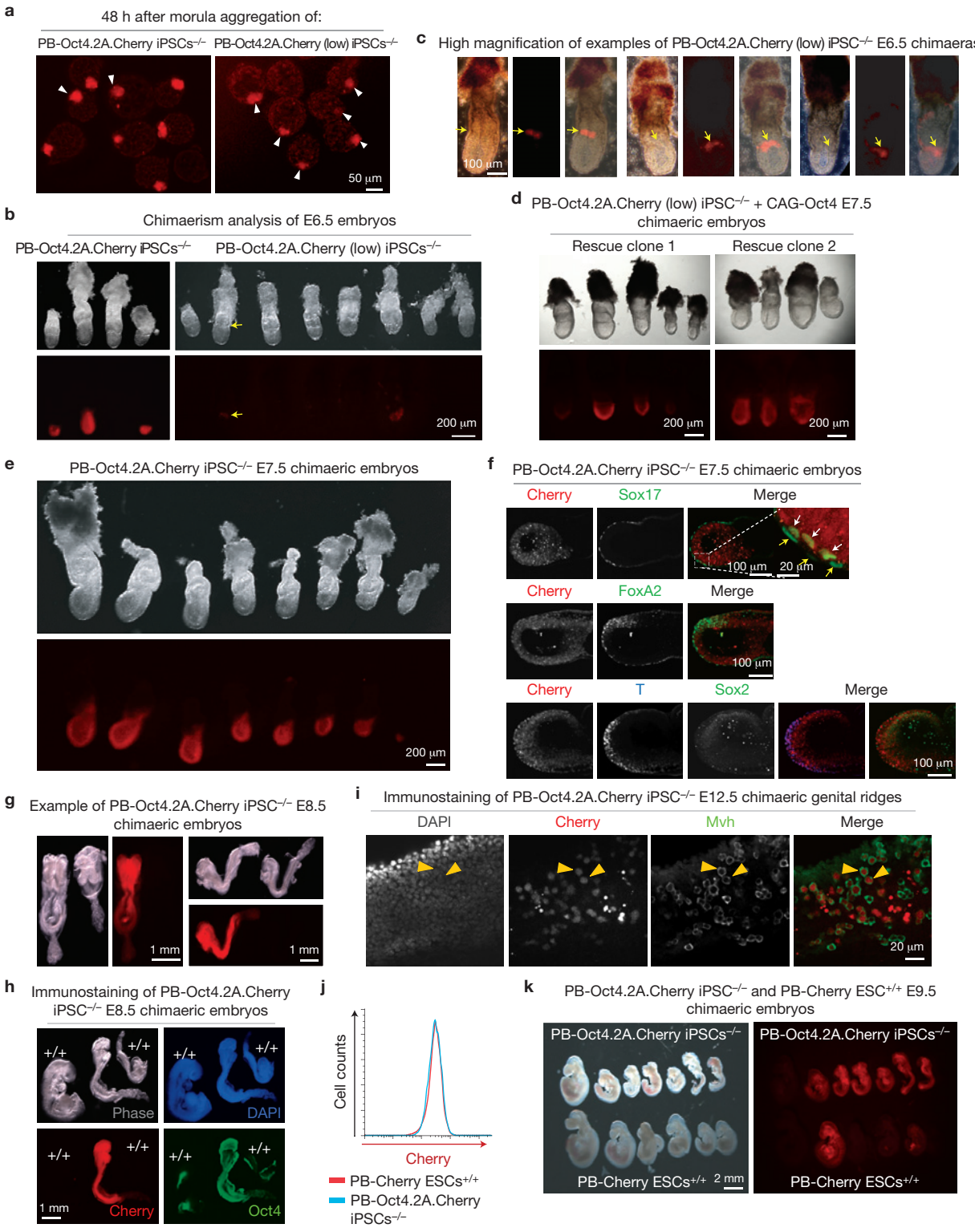


Figure 4 Oct4 expression at an ESC level is required for *in vivo* differentiation. (**a**) Phase and Cherry images of embryos obtained 48 h after morula aggregations of PB-Oct4.2A.Cherry or PB-Oct4.2A.Cherry (low) clone 1 iPSCs$^{-/-}$. Representative contribution to naive epiblast is indicated with white arrowheads. (**b**) Phase and Cherry images of E6.5 embryos obtained as a result of PB-Oct4.2A.Cherry or PB-Oct4.2A.Cherry (low) clone 1 iPSC$^{-/-}$ morula aggregations. (**c**) High-magnification images of E6.5 PB-Oct4.2A.Cherry (low) clone 1 chimaeric embryos. (**d**) Phase and Cherry images of E7.5 embryos obtained after rescuing PB-Oct4.2A.Cherry (low) clone 1 iPSCs$^{-/-}$ with CAG-Oct4. (**e**) Phase and Cherry images of E7.5 PB-Oct4.2A.Cherry chimaeric embryos. (**f**) E7.5 PB-Oct4.2A.Cherry chimaeric embryos immunostained for Sox17, FoxA2, Sox2 and T-Brachyury. (**g**,**h**) E8.5 PB-Oct4.2A.Cherry chimaeric embryos. Phase and Cherry images (**g**) and immunostaining for Oct4 (**h**). (**i**) Confocal images of the genital ridges from E12.5 PB-Oct4.2A.Cherry chimaeric embryos immunostained for Mvh. (**j**) Flow cytometry analysis of PB-Oct4.2A.Cherry iPSCs$^{-/-}$ and ESCs$^{+/+}$ with CAG-Cherry reporter. (**k**) Representative Phase and Cherry images of litters of PB-Oct4.2A.Cherry iPSC$^{-/-}$ and CAG-Cherry ESC$^{+/+}$ chimaeric embryos.

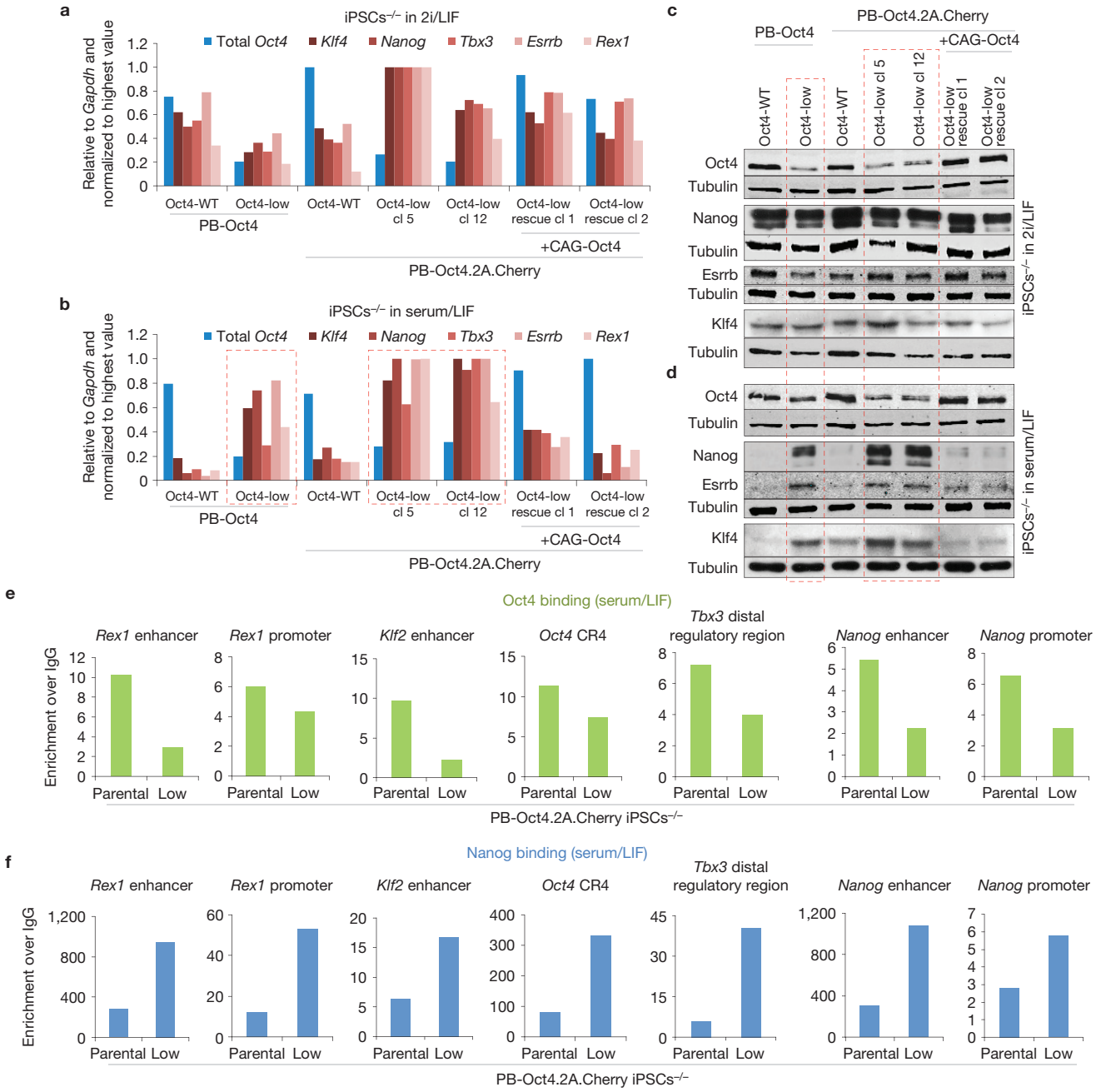


Figure 5 Oct4 binding is converse to Nanog and is linked to downregulation of naive pluripotency genes. (**a**,**b**) qRT–PCR analysis of total *Oct4, Klf4, Nanog, Tbx3, Esrrb* and *Rex1* expression in PB-Oct4 (Oct4-WT and -low), PB-Oct4.2A.Cherry (Oct4-WT and -low) iPSCs$^{-/-}$ and PB-Oct4.2A.Cherry (low) + CAG-Oct4 iPSC$^{-/-}$ rescue clones in 2i/LIF (**a**) and serum/LIF with selection for geneticin (G418) resistance driven by the *Oct4* promoter (Supplementary Fig. S1b; **b**). Oct4-low lines are indicated with red rectangles. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**c**,**d**) Western blot analysis of Oct4, Nanog, Esrrb and Klf4 protein expression in PB-Oct4 (Oct4-WT and -low), PB-Oct4.2A.Cherry (Oct4-WT and -low) iPSCs$^{-/-}$ and PB-Oct4.2A.Cherry (low) + CAG-Oct4 iPSC$^{-/-}$ rescue clones in 2i/LIF (**c**) and serum/LIF with selection for geneticin (G418) resistance driven by *Oct4* promoter (**d**). Oct4-low lines are indicated with red rectangles. See Supplementary Fig. S9 for uncropped data. (**e**,**f**) ChIP analysis of Oct4 (**e**) and Nanog (**f**) binding at the regulatory regions of pluripotency genes in PB-Oct4.2A.Cherry and PB-Oct4.2A.Cherry (low) iPSCs$^{-/-}$. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments.

into why Oct4-low cells fail to downregulate naive pluripotency gene expression, we performed chromatin immunoprecipitation (ChIP) for Oct4, Nanog and Esrrb in serum/LIF conditions in Oct4-low and Oct4-WT cells at regulatory sequences of key naive pluripotency genes. Strikingly, Oct4 genomic occupancy was markedly reduced and Nanog and Esrrb significantly increased at these targets in Oct4-low cells (Fig. 5e,f and Supplementary Fig. S6c). This suggests that in suboptimal self-renewing culture conditions and on induction of cell differentiation Oct4 can act as a repressor of key naive pluripotency genes and that a reduction in the Oct4 level leads to an incapacity to downregulate these.

To assess whether the phenotype of Oct4-low cells could be attributed to the high levels of Nanog, we performed Nanog knockdown in

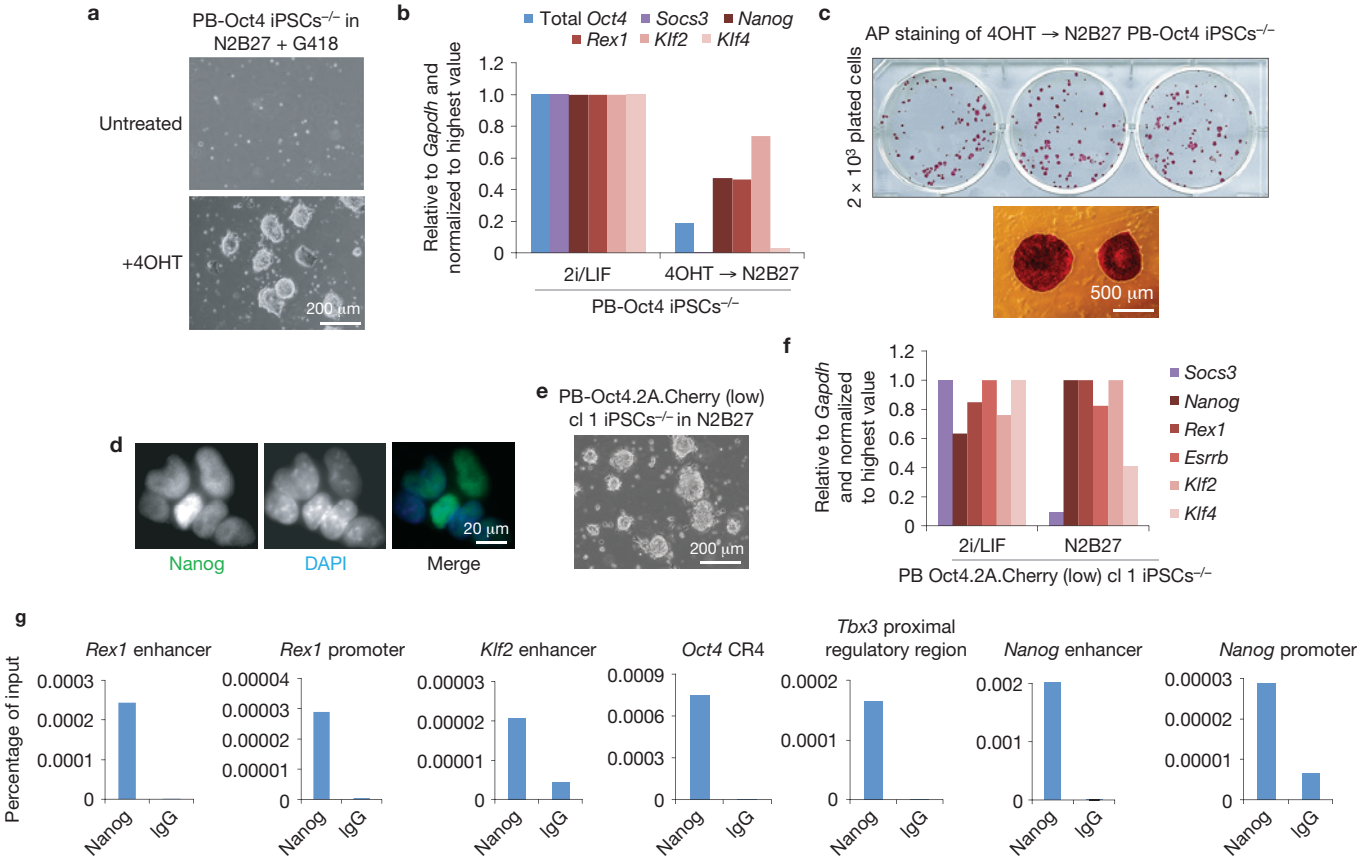


Figure 6 Oct4-low iPSCs self-renew in the absence of pluripotent culture requisites. (**a**) Phase images of PB-Oct4 iPSCs$^{-/-}$ treated or untreated with 4OHT for 24 h and subsequently cultured in N2B27 conditions with selection for geneticin (G418) resistance driven by *Oct4* promoter (Supplementary Fig. S1b). Untreated cells differentiated and died after the first passage, whereas treated cells self-renewed indefinitely. (**b**) qRT–PCR analysis of total *Oct4, Socs3, Nanog, Rex1, Klf2* and *Klf4* expression in PB-Oct4 iPSCs$^{-/-}$ cultured in 2i/LIF and 4OHT-treated iPSCs$^{-/-}$ cultured for 15 passages in N2B27 conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**c**) Alkaline phosphatase (AP) staining of 4OHT-treated PB-Oct4 iPSCs$^{-/-}$ (passage 15 in N2B27) plated at clonal density. Two thousand cells were plated per well and cultured for 8 days in N2B27. (**d**) Immunocytochemistry detection of Nanog in 4OHT-treated PB-Oct4 iPSCs$^{-/-}$ cultured in N2B27 conditions. (**e**) Phase image of the previously established (Fig. 2j–l) PB-Oct4.2A.Cherry (low) clone 1 iPSCs$^{-/-}$ cultured in N2B27 conditions. (**f**) qRT–PCR detection of *Socs3, Nanog, Rex1, Esrrb, Klf2* and *Klf4* in PB-Oct4.2A.Cherry (low) clone 1 iPSCs$^{-/-}$ cultured in 2i/LIF and N2B27 conditions. Data shown are the mean of 3 replicates and are from 1 of 3 representative experiments. (**g**) ChIP analysis of Nanog binding at main target genes in PB-Oct4.2A.Cherry (low) clone 1 iPSCs$^{-/-}$ cultured in N2B27 conditions for 15 passages. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments.

Oct4-low cells and subjected these to either a self-renewal assay at clonal density in serum minus LIF or to embryoid body differentiation. We confirmed Nanog knockdown and observed decreased expression of Klf4 and Esrrb, known Nanog targets[25]. However, the expression of these and other naive pluripotency genes remained higher than in self-renewing Oct4-WT cells (Supplementary Fig. S6d). Oct4-low cells with Nanog knockdown did not differentiate (Supplementary Fig. S6e,f), suggesting that ongoing high-level expression of other naive pluripotency genes such as *Klf2* and *Tbx3*, which can confer LIF-independent self-renewal[26,27], prevents differentiation. Together, these data further confirm that Oct4-low cells are robustly locked in a self-renewing state.

Low Oct4 is sufficient to sustain self-renewal in the absence of pluripotent culture requisites

Failure to differentiate together with robust expression of naive pluripotency genes led us to investigate whether Oct4-low iPSCs$^{-/-}$ can self-renew after removal of 2i and LIF from the serum-free medium (N2B27). Under these conditions, pluripotent cells undergo differentiation[14,28]. Oct4-WT iPSCs$^{-/-}$ were treated with 4OHT for 24 h, which induced partial Oct4 transgene loss (Fig. 2e), and then switched to N2B27 with selection for *Oct4* promoter activity to eliminate differentiated cells. In contrast to untreated cells, which differentiated and died after the first passage, we were able to maintain 4OHT-treated cells indefinitely (Fig. 6a). These cells expressed naive pluripotency markers but lost Socs3 and downregulated Klf4 expression, indicating an absence of LIF/STAT3 signalling (Fig. 6b and Supplementary Fig. S7a). Oct4 transgene expression was decreased by at least 2.5-fold (Fig. 6b and Supplementary Fig. S7a,b). These cells self-renewed at clonal density (Fig. 6c and Supplementary Fig. S7c) and expressed Nanog protein (Fig. 6d and Supplementary Fig. S7d). We were also able to maintain previously established (Fig. 2j–l) Oct4-low iPSC$^{-/-}$ clones indefinitely in N2B27 (Fig. 6e,f). Nanog binding at key genomic targets in these cells was also maintained (Fig. 6g).

To confirm our observations in an independent system we established ESCs with low Oct4 expression levels. To derive these, we co-transfected *Oct4*$^{flox/-}$ ESCs with CAG-Oct4 and CAG-CreERT2

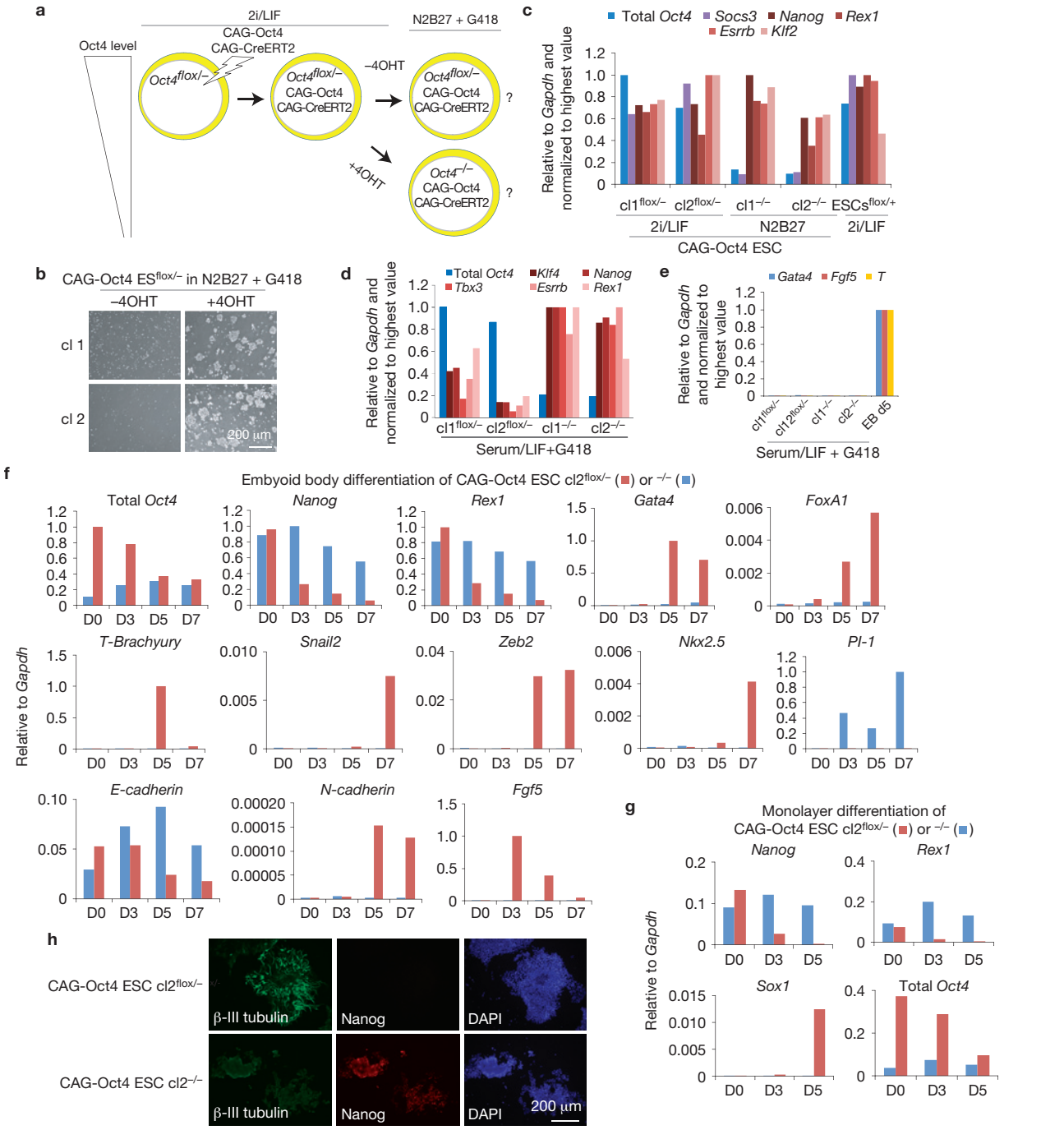


Figure 7 A defined Oct4 level is also required for downregulation of key naive pluripotency genes in ESCs. (**a**) Experimental design used to obtain Oct4-low ESCs. ESCs$^{flox/-}$ were stably transfected with CAG-Oct4 and CAG-CreERT2 transgenes and subsequently treated with 4OHT to induce Cre-mediated excision of the floxed *Oct4* allele but not the CAG-Oct4 transgene, which is not flanked by *loxP* sites. (**b**) 4OHT-treated and untreated CAG-Oct4 ESCs$^{flox/-}$ after the first passage in N2B27 culture conditions with selection (G418) for *Oct4* promoter activity. (**c**) qRT–PCR detection of total *Oct4, Socs3, Nanog, Rex1, Esrrb, Klf2* and *Klf4* in Oct4-low ESCs (CAG-Oct4 ESCs$^{-/-}$) cultured in N2B27 conditions for 15 passages in comparison with 4OHT-untreated cells and wild-type ESCs cultured in 2i/LIF conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**d**) qRT–PCR detection of total *Oct4, Klf4, Nanog, Tbx3, Esrrb* and *Rex1* in CAG-Oct4 ESCs$^{flox/-}$ and ESCs$^{-/-}$ cultured in serum/LIF conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**e**) qRT–PCR detection of *Gata4, T-Brachyury (T)* and *Fgf5* in CAG-Oct4 ESCs$^{flox/-}$ and ESCs$^{-/-}$ cultured in serum/LIF conditions. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. EB, embryoid body. (**f**) qRT–PCR detection of *Oct4, Nanog, Rex1, Gata4, FoxA1, T-Brachyury, Snai2, Zeb2, Nkx2.5, Pl-1, E-cadherin, N-cadherin* and *Fgf5* expression during embryoid body differentiation of CAG-Oct4 ESCs$^{flox/-}$ and ESCs$^{-/-}$ clone 2. Data shown are the mean of 3 replicates and are from 1 of 3 representative experiments. (**g**) qRT–PCR analysis of *Sox1*, total *Oct4, Rex1* and *Nanog* expression during 5 days of neural induction in monolayer culture of CAG-Oct4 ESCs$^{flox/-}$ and ESCs$^{-/-}$ clone 2. D0–D5 indicate the number of days of differentiation. Data shown are the mean of 3 replicates and are from 1 of 2 representative experiments. (**h**) Immunocytochemistry detection of βIII-tubulin and Nanog expression after 7 days of neural induction in monolayer culture of CAG-Oct4 ESCs$^{flox/-}$ and ESCs$^{-/-}$.

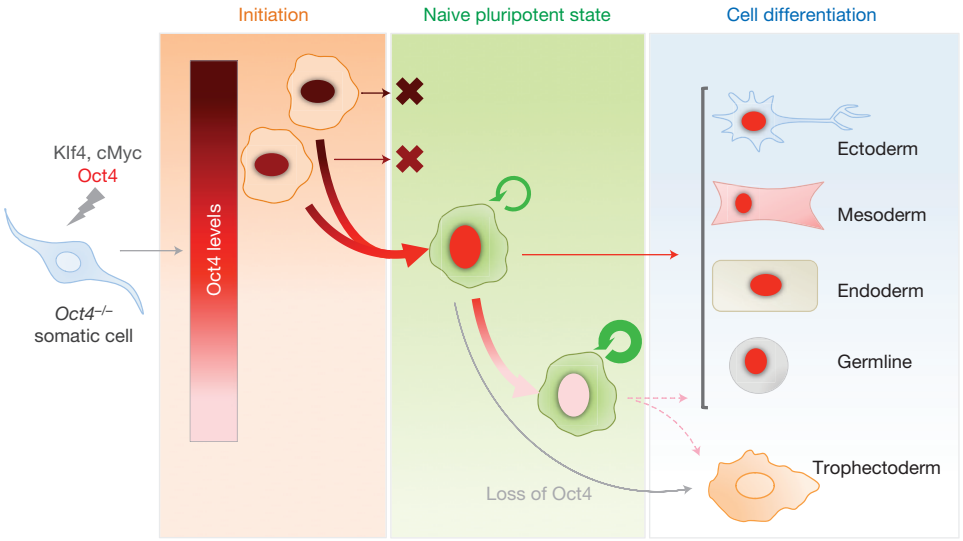


Figure 8 A defined Oct4 level controls cell state transitions around pluripotency. The combined transduction and transfection of reprogramming transgenes into *Oct4*$^{-/-}$ somatic cells results in the generation of highly proliferative reprogramming intermediates. On exposure to 2i/LIF culture conditions some of these undergo conversion to a pluripotent cell state. Remarkably, independently of the Oct4 expression level in reprogramming intermediates, generated iPSCs$^{-/-}$ always show an invariable ESC level of Oct4 expression. Once cells have entered a pluripotent state they can be maintained within a range of Oct4 expression from an ESC level to up to sevenfold less without loss of self-renewing capacity. This indicates a specific requirement for a defined Oct4 level for the acquisition rather than maintenance of the naive pluripotent state. As shown before for ESCs, complete abolishment of Oct4 expression in iPSCs$^{-/-}$ leads to differentiation towards the trophectoderm lineage. Surprisingly, pluripotent cells with a constitutive ESC level of Oct4 can efficiently differentiate into the three germ layers and germline on the provision of appropriate signalling cues. At the same time ESCs/iPSCs with low Oct4 levels demonstrate enhanced self-renewing capabilities independently of culture conditions and fail to exit the pluripotent state on the induction of differentiation. Overall these data demonstrate that Oct4 actively controls cell state transitions taking place during the entry into and exit from the naive pluripotent cell state.

transgenes. Similarly to the above results (Supplementary Fig. S2g), the obtained self-renewing cells exhibited an ESC level of Oct4 (Fig. 7c). Subsequently, we treated these with 4OHT to induce Cre-mediated excision of the floxed *Oct4* allele but not the CAG-Oct4 transgene, as the latter was not *loxP*-flanked (Fig. 7a). The cells were then switched to N2B27 with selection to eliminate non-self-renewing cells. Untreated cells differentiated and died after the first passage. Consistent with iPSCs$^{-/-}$, 4OHT-treated ESCs (CAG-Oct4 ESCs$^{-/-}$) exhibited 5–7-fold lower than ESC Oct4 levels and could be maintained in N2B27 indefinitely (Fig. 7b,c). Moreover, when placed in serum/LIF, they exhibited higher expression levels of naive pluripotency markers than parental cells (Fig. 7d,e). Furthermore, CAG-Oct4 ESCs$^{-/-}$ failed to undergo embryoid body and neural differentiation, instead remaining locked in a naive pluripotent state (Fig. 7f–h and Supplementary Fig. S7e).

Next we investigated whether the DNA-binding capacity of Oct4 is important for its function in cell differentiation. We used the Oct4-267VP mutant, which cannot bind DNA or sustain ESC self-renewal in doxycycline-repressible ZHBTc4.1 cells[29,30] (Supplementary Fig. S7f). We transfected Oct4-low ESCs$^{-/-}$ with the PB-Oct4-267VP transgene and established lines with an ESC level of total Oct4 (Supplementary Fig. S7g). Importantly, mutant Oct4 could not rescue the differentiation defect of Oct4-low ESCs$^{-/-}$ (Supplementary Fig. S7g), demonstrating that the capacity of Oct4 to bind DNA is required for differentiation.

In summary, the naive pluripotent state can be maintained in the absence of pluripotency medium requisites, if the Oct4 expression level is decreased.

DISCUSSION

In this study, using optimized culture conditions for the induction and maintenance of pluripotent cells[31,32] and *Oct4*$^{-/-}$ somatic cells as a starting point to provide a stringent functional assay, we demonstrated that a defined Oct4 level is critical for naive pluripotency acquisition. Once pluripotency is established, Oct4 levels can be decreased without loss of self-renewal. However, an ESC level of Oct4 is then required for the efficient downregulation of the naive pluripotent program during cell differentiation (Fig. 8).

It has previously been proposed that more than 50% Oct4 downregulation would lead to trophectoderm differentiation in doxycycline-repressible ZHBTc4.1 ESCs (ref. 7). This apparently differs from our observations. However, we found that by titrating the doxycycline concentration self-renewing Oct4-low ZHBTc4.1 ESCs with around 30% of the starting Oct4 level could be established (Supplementary Fig. S8).

Several studies reported the involvement of Oct4 in ESC differentiation *in vitro*. Ref. 33 claimed that Oct4 suppresses neuroectoderm differentiation and promotes mesendoderm differentiation of mouse ESCs *in vitro*. Ref. 34 claimed that Oct4 suppresses human ESC differentiation into definitive endoderm. Both studies, however, are based on Oct4 overexpression and knockdown experiments in ESCs. In contrast, our system allowed investigation of the effect of biological Oct4 levels on pluripotent cell differentiation. Thus, and in apparent disagreement with what was expected from the published literature, constitutive Oct4 expression at an ESC level led to a robust contribution to nascent ectoderm, mesoderm, endoderm and germline in the embryo. This result is however consistent with the observed Oct4 expression in the progeny of all germ layers until the late somite stage[5,19].

It was reported that episomal overexpression of an Oct4 DNA-binding mutant (Oct4-267VP; ref. 30) causes spontaneous ESC differentiation[29]. In our system, this did not rescue differentiation defects of Oct4-low cells. As the transactivation activity of Oct4-267VP depends on the recruitment by a functional POU factor[30], we believe that Oct4-267VP induces differentiation only on strong overexpression and in the presence of an ESC level of wild-type Oct4.

In conclusion, this study redefines our previous understanding of the biological roles of Oct4 from a factor known to be important for reprogramming and self-renewal[6,7] to one also actively controlling cell state transitions during entry into and exit from the naive pluripotent state. The future challenge will be to define the molecular mechanisms underlying this dual function of Oct4. □

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank W. Mansfield and C-E. Dumeau for blastocyst injections and morula aggregations, R. Walker for flow cytometry, and M. McLeish and H. Skelton for histological processing of teratomas. We are grateful to H. Niwa for providing mice with different *Oct4* genotypes and A. Smith and J. Betschinger for providing plasmids. We are also grateful to Y. Costa and P. Shliaha for technical assistance and H. Stuart for critical reading of the manuscript. The study was supported by Wellcome Trust Fellowship WT086692MA. J.C.R.S. is a Wellcome Trust Career Development Fellow. A.R. is a recipient of the Darwin Trust of Edinburgh Postgraduate Scholarship.

AUTHOR CONTRIBUTIONS

A.R. performed and designed the experiments, analysed the data and wrote the manuscript. R.L.S. and G.L.B.C. performed experiments. T.W.T., L.F.C., A.R. and J.S. designed the study. J.N. analysed data. J.S. supervised the study, designed the experiments, analysed the data, and wrote and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at www.nature.com/doifinder/10.1038/ncb2742
Reprints and permissions information is available online at www.nature.com/reprints

1. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem. Cell* **4,** 487–492 (2009).
2. Rosner, M. H. *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345,** 686–692 (1990).
3. Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.* **9,** 2185–2195 (1990).
4. Okamoto, K. *et al.* A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60,** 461–472 (1990).
5. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122,** 881–894 (1996).
6. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379–391 (1998).
7. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet* **24,** 372–376 (2000).
8. Theunissen, T. W. *et al.* Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr. Biol.* **21,** 65–71 (2011).
9. Kim, J. B. *et al.* Direct reprogramming of human neural stem cells by OCT4. *Nature* **461,** 649–643 (2009).
10. Yuan, X. *et al.* Combined chemical treatment enables Oct4-induced reprogramming from mouse embryonic fibroblasts. *Stem Cells* **29,** 549–553 (2011).
11. Li, Y. *et al.* Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Res.* **21,** 196–204 (2011).
12. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136,** 411–419 (2009).
13. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643–655 (2003).
14. Ying, Q. L., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21,** 183–186 (2003).
15. Wilkinson, D. G., Bhatt, S. & Herrmann, B. G. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* **343,** 657–659 (1990).
16. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126–140 (2003).
17. Monaghan, A. P., Kaestner, K. H., Grau, E. & Schutz, G. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3α,β and γ genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* **119,** 567–578 (1993).
18. Kanai-Azuma, M. *et al.* Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129,** 2367–2379 (2002).
19. Downs, K. M. Systematic localization of Oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. *Dev. Dyn.* **237,** 464–475 (2008).
20. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
21. Hayashi, K., Lopes, S. M., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem. Cell* **3,** 391–401 (2008).
22. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135,** 909–918 (2008).
23. Kobayashi, T. *et al.* The cyclic gene Hes1 contributes to diverse differentiation responses of embryonic stem cells. *Genes Dev.* **23,** 1870–1875 (2009).
24. Canham, M. A., Sharov, A. A., Ko, M. S. & Brickman, J. M. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* **8,** e1000379 (2010).
25. Festuccia, N. *et al.* Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem. Cell* **11,** 477–490 (2012).
26. Hall, J. *et al.* Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem. Cell* **5,** 597–609 (2009).
27. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460,** 118–122 (2009).
28. Van Oosten, A. L., Costa, Y., Smith, A. & Silva, J. C. JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. *Nat. Commun.* **3,** 817 (2012).
29. Niwa, H., Masui, S., Chambers, I., Smith, A. G. & Miyazaki, J. Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Mol. Cell Biol.* **22,** 1526–1536 (2002).
30. Vigano, M. A. & Staudt, L. M. Transcriptional activation by Oct-3: evidence for a specific role of the POU-specific domain in mediating functional interaction with Oct-1. *Nucleic Acids Res.* **24,** 2112–2118 (1996).
31. Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6,** e253 (2008).
32. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
33. Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145,** 875–889 (2011).
34. Teo, A. K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* **25,** 238–250 (2011).