Linking open chromatin to pluripotency in ES cells

Relacionando cromatina aberta com pluripotência em células estaminais embrionárias

The role of the chromatin remodeler Chd1

O papel do remodelador da cromatina Chd1

Alexandre Gaspar Maia

Universidade de Coimbra

2009

Dissertação de candidatura à obtenção do grau de Doutor em Biologia, especialização em Biologia Molecular, apresentada à Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

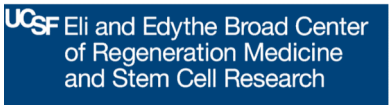
Dissertation submitted for the degree of Doctor of Philosophy in Biology, specialization in Molecular Biology, presented to Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

The work presented in this dissertation was carried out in the scope of the Doctoral Programme in Experimental Biology and Biomedicine (PDBEB), Center for Neuroscience and Cell Biology (CNC), University of Coimbra, Portugal. The first six months of the PhD were spent at the Center for Neuroscience and Cell Biology, Coimbra, Portugal, attending the advanced courses organized by PDBEB for the class of 2003-2007.

The research work presented in this dissertation was conducted at the University of California San Francisco, under the supervision of Dr. Miguel Ramalho-Santos (Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at University of California San Francisco) and co-supervision of Dr. João Ramalho-Santos (CNC, University of Coimbra, Portugal).

This work was financially supported by Fundação Para a Ciência e Tecnologia, Portugal (fellowship SFRH/BD/11818/2003; co-funded by the Programa Operacional Ciência e Inovação 2010 [POCI 2010], Governo da República Portuguesa and Fundo Social Europeu [FSE], União Europeia).

Work in the laboratory of Dr. Miguel Ramalho-Santos was supported by grants from NIH Director's New Innovator Award, California Institute for Regenerative Medicine and Juvenile Diabetes Research Foundation.



Financiamento no âmbito do III Quadro Comunitário de Apoio, comparticipado pelo Fundo Social Europeu.

À minha mãe, ao meu pai e ao meu irmão,
que mais do que ninguém acreditam em mim

Acknowledgments

Este doutoramento foi um processo muito enriquecedor, com tantas etapas diferentes, umas de muito entusiasmo, outras de muita incerteza, e algumas de frustração. E felizmente tive em meu redor um grupo de pessoas fantástico, com tantas coisas para me dar, com tanta energia para me transmitir, e com saber e experiência. Enumerar todos aqueles que me ajudaram é por si só uma missão. Mas não gostava de terminar esta aventura sem deixar marcado o meu apreço por todos os que me ajudaram e fizeram esta aventura uma realidade. A todos um grande obrigado.

Gostava especialmente de agradecer ao Miguel, o meu orientador, com o qual cresci para a ciência. Foi sempre uma pessoa com extrema sensibilidade para nos momentos menos bons ter as palavras certas para me dirigir no bom caminho. Aprendi imenso com os meus erros, mas sobretudo com o modo como ele lidou com os meus erros. Foi uma lição de vida que ele me deu.

E uma palavra também para todos os membros do meu laboratório, com os quais cresci e aprendi muito. Fizeram com que cada dia no laboratório não fosse apenas mais um dia. Diverti-me imenso com todos. Mas queria especialmente agradecer à Marica, que foi sem dúvida uma segunda mentora para mim. Ajudou-me a cada passo, a cada experiência, com toda a sua paciência e energia.

(A few words for all the members of the Santos lab, with whom I learned and grew up a lot. They made every day in lab different and enjoyable. I had a lot of fun with everyone. But I would like to thank especially Marica, who was like a second mentor for me. She helped me in every step I took, with every experiment, with all her patience and energy.)

And then I would like to thank everyone in the Diabetes Center, in the IRM, and at UCSF that made my PhD a bit richer everyday. And to our collaborators, that definitely had a big impact in my project, especially to Michael McManus, Eran Meshorer and Kathrin Plath.

Gostava de agradecer aos BEBs, especialmente aos BEB2, com quem vivi o ínicio desta aventura, e com quem partilhei a magia de ir atrás do desconhecido. E uma palavra de apreço também ao meu co-orientador João Ramalho Santos.

Tudo isto não seria possível, se não tivesse passado quase dois anos no laboratório de Leonor Teles-Grilo, com quem comecei a dar os primeiros passos, e que me ensinou muito mais do que ciência, uma atitude de vida.

Finalmente resta-me agradecer à minha familia, que todos os anos esperou por mim no aeroporto, de cada vez que voltava a Portugal. E que me acompanhou ao longe, com toda a força e alegria. Espero que no final todos saibam que valeu a pena.

E aos meus amigos, por me ouvirem falar de ciência, sem sequer quererem muito saber. E por me ouvir dizer coisas sem sentido a toda a hora, sem sequer perceberem muito bem o que digo. E à Sophia, que leu e releu esta tese vezes sem conta e me apoiou muito especialmente nesta fase final.

(And to all my friends, for listening to me talking about science, without really wanting to know about it. And for listening to my nonsense all the time, without really understanding quite well what I say. And to Sophia, that read this thesis so many times and supported me a lot, specially in this last period.)

Thank you.

Obrigado.


Rik Derynck
Benoit Bruneau
Masayo Yumoto
Silvia Curado
Didier Stainer
Marco Conti
Mary Mohrin
Renee Reijo Pera
Amander Clark
Connie Wong
Tatiana Hochgreb
Xiaowei Liu
Brian King
Paolo Rinaudo
Emmanuele Passagué
Cori Nicholas
Kee Kehkooi
Diana Laird
Mark Fox
Kelly Haston
Marianne Santaguida
Damien Reynaud
Stacey Morikawa-Wan
Margaret Rossi-Holman
Cassandra Belair
Michael Wong
June Silathong
Archana Shenoy
Collin Melton
Emmanuele Huillard
Erica Stidham
Robert Blelloch
Yangming Wang
Joshua Babiarz
Monica Venere
Barbara Panning
Rui Galvão
Minoree Kohwi
Rostislav Medvid
Tom Fazzio
Daniel Lim
Hirohide Takebayashi
Stephen Noctor
Erica Jackson
Holger Willenbring
Ashkan Javaherian
Florian Merkle
Ricardo Romero
Amar Sharma
Arnold Kriegstein
Young Goo Han
Jan Lui
David Castaneda
Arturo Alvarez Buylla
Cori Harwell
Zaman Mirzadeh
Derek Southwell
Marta Gaglia
Melissa Lezameta
David Cristina
Kathryn Blaschke
Cynthia Yaschine
Salma Salmi
Miro Brajenovic
Marica Grskovic
Varvara Zeldovich
Marta Wegorzewska
Clara Chiu
Yuki Ohi
Fanny Polesso
Andy Newman
Michael McManus
Miguel Ramalho Santos
Veena Singla
Amy Heidersbach
Michael Sachs
Jonathan Yen
Julie Gaulden
Trinna Cuellar
Christina Chaivorapol
Jeremy Reiter
Victoria Wu
Kevin Corbitt
Alison Brincat
Amy Shyer
Leonor Teles Grilo
Grace Wei
Emre Turer
Eran Meshorer
Limor Landsman
Barbara Malynn
Adi Alajem
Patrick Heisser
John Morris IV
Rita Tavares
Averil Ma
David Cano
Laurence Clement
David Boone
Sara Gozalo
Mathias Hebrok
Helene Favre
Sophia Chai
Rupa Sridhran
Sara Cervantes-Roldan
Sapna Puri
Michael Mason
Christian Vaisse
Geraldine Bienvenu
Kathrin Plath
Matthew Spindler
João Pereira
Karyl Nakamura
Cedric Louvet
Freia Spilman
Linda Ta
Fran DeNoto
Jeff Bluestone
Marty Bigos
Fred Fermin
Richard Locksley
Fernando Botelho
Robert Perry
Valerie Stepps
Geeta Narlika
Michael Parmaceck
Yoichi Shinkai
Lee Reinhardt
Cliff McArthur
Joanna Wysocka
Ufuk Gunesdogan
Thomas Jenuwein
Silvia Vilarinho
Thea Tlsty
Tara Rambaldo
Sepideh Khorasanizadeh
Austin Smith
Shuwei Jiang
Adriana Oliveira dos Santos
Catarina Cruz
Joana Sá
Tiffany Reese
Ana Luisa Vital
Renata Brito
Catarina Grandela
Filipe Josué
Tiago Paixão
Joana Sequeira Mendes
Sandra Santos
Ana Filipa Simões
Euclides Pires
Sukalyan Chaterjee
BEB2
Manuela Ferreira
David Mendes da Silva
Sandra Pinho
Miguel Castelo Branco
Isaura Simões
Carlos Duarte
Sandrina Nóbrega
Carlos Faro
Cristina Rego
Susana Maia Lopes
João Ramalho Santos
Luis Almeida

“Se ao menos eu soubesse saber.”
Anónimo

Table of contents

Abstract

Embryonic Stem (ES) cells are the prototypical pluripotent stem cells that can self-renew indefinitely and give rise to all cell types of the body. Through a transcriptional profile using microarrays, we have discovered that the mouse embryonic germ line have high transcriptional similarities to pluripotent stem cells, including mouse ES cells. A group of about 220 genes are up-regulated in pluripotent stem cells when compared to differentiated cells. From this cluster (that we called the pluricluster) we decided to test the effect of down-regulation in ES cells self-renewal and pluripotency. We developed a lentiviral RNA-interference screen to test the role of 41 candidate genes (chosen amongst the pluricluster with a bias towards transcription factors, chromatin remodelers, DNA/RNA binding proteins, oncogenes and unknown genes). Our screen identified 18 genes with RNAi phenotype including Chd1, NFYa, NFYb and Sall4. The last three genes allowed me to follow further characterization of their role in ES cells and reprogramming, as a collaboration project. However, Chd1 was the only novel regulator that had both self-renewal defects and loss of Oct4 activity in this screen. Chd1 is a chromatin remodeler that has been shown to recognize H3K4me3 and is associated with transcriptional activation and splicing. Our work shows clear evidence that Chd1 correlates with H3K4me3 enrichment, which overlaps with RNA polymerase II localization. Down-regulation of Chd1 using RNAi in mouse ES cells, leads to increased heterochromatin foci marked by H3K9me3 and HP1. Moreover, in the absence of Chd1, ES cells have self-renewal and pluripotency defects. Chd1-deficient cells are not able to form a typical endodermal layer upon embryoid body differentiation, and have high propensity for neural differentiation. Chd1 also plays a role in the generation of iPS cells by over-expression of the four factors Oct4, Sox2, N-Myc and Klf4. Down-regulation of Chd1 decreases the efficiency of the process, suggesting that Chd1 may be necessary for the chromatin reorganization during reprogramming to pluripotency. The role of Chd1 in ES cells suggests that pluripotent stem cells exist in a dynamic state of opposing influences between euchromatin and heterochromatin. In this thesis, I also suggest that indeed open chromatin may be necessary for ES cells to differentiate into all three germ layers, linking open chromatin to pluripotency.

Sumário

As células estaminais embrionárias são a população típica de células pluripotentes, que têm uma capacidade infinita de auto-renovação e podem dar origem a todas as células do corpo. Através da análise global de expressão genética usando *microarrays*, descobrimos que células da linha germinal de ratinho são em grande parte semelhantes a células pluripotentes, incluindo células estaminais embrionárias de ratinho. Um conjunto de cerca de 220 genes estão sobre-expressos em células pluripotentes quando comparados com células diferenciadas. Neste conjunto de genes (a que chamámos *pluricluster*) decidimos testar o efeito da supressão de expressão na capacidade de auto-renovação e pluripotência das células estaminais. Desenvolvemos um ensaio de crivagem de supressão de 41 genes candidatos (em que foram escolhidos preferencialmente factores de transcrição, remodeladores da cromatina, oncogenes, moléculas de ligação a DNA e RNA, e genes com função desconhecida), usando RNA de intereferência através de uma infecção lentiviral. Desse ensaio identificámos 18 genes com fenótipo após a supressão de expressão, entre os quais os genes Chd1, NFYa, NFYb e Sall4. A função em células estaminais embrionárias e em reprogramação dos últimos três genes mencionados foi consequentemente caracterizada em projectos de colaboração. No entanto, o único novo regulador de células estaminais embrionárias que cujo fenótipo apresentava defeitos de auto-renovação e supressão de expressão do marcador Oct4 foi o gene Chd1. Chd1 é um remodelador da cromatina que reconhece a tri-metilação da lisina 4 da histona 3 (H3K4me3) e está associada a activação da transcrição e *splicing*. Aqui mostramos que Chd1 se liga a regiões do genoma enriquecidas ma marca H3K4me3 e também a RNA polimerase II. Supressão de Chd1 através de RNA de interferência em células estaminais embrionárias induz a formação de focus de heterocromatina marcados por H3K9me3 e HP1. Para além disso, as células perdem a capacidade de pluripotência e mostram defeitos na auto-renovação. Células Chd1-deficientes, após indução de diferenciação em corpos embrioides, não são capazes de formar a típica camada de endoderme, e têm uma alta propensão para neuro diferenciação. Chd1 tem também uma função na reprogramação de células diferenciadas em células pluripotentes induzidas (iPS) através dos factores Oct4, Sox2, Klf4 e N-Myc. Supressão da expressão de Chd1 reduz a eficiência do processo de reprogramação, o que sugere que Chd1 possa ser necessário para a reorganização da cromatina durante o processo de reprogramação. O papel de Chd1 em células estaminais embrionárias sugere que em células pluripotentes existe um estado dinâmico entre eucromatina e heterocromatina. Nesta tese, eu sugiro que de facto, uma cromatina aberta será necessária em células estaminais embrionárias para se diferenciarem em células de todas a linhas germinais, relacionando directamente uma cromatina aberta com pluripotência.

Abreviations

ATPase	Adenosine triphosphate dephosphorylase
BMP	Bone morphogenetic proteins
BSA	Bovine serum albumin
bp	Base pair
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CpG	Cytosine and Guanine dinucleotide
CMV	Human cytomegalovirus promoter
DAPI	4',6-diamidino-2-phenylindole
ddH_2O	Double distilled water
DMEM	Dulbeco's modified Eagle's medium
DMSO	Dimethylsulfoxide
DMP	Dimethylpimelimidate
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferases
dsRNA	Double stranded RNA
E8.5/ E10.5	Embryonic day 8.5/ Embryonic day 10.5
EB	Embryoid bodies
EC	Embryonic carcinoma cells
EG	Embryonic germ cells
EpiSCs	Epiblast stem cells
ES/ESC	Embryonic stem cells
EtOH	Ethanol
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast growth factor
FRAP	Fluorescence recovery after photobleaching
GFP	Green fluorescence protein
h/hr	Hour
H&E	Hematoxilin and Eosin
H3K4me2	Di-methyl lysine 4 histone 3
H3K4me3	Tri-methyl lysine 4 histone 3
H3K9me2	Di-methyl lysine 9 histone 3
H3K9me3	Tri-methyl lysine 9 histone 3
H3K27me3	Tri-methyl lysine 27 histone 3
H3K79	Lysine 79 histone 3
HDMase	Histone demethylase
HDAC	Histone deacetylase
hES/hESC	Human embryonic stem cells
hEG	Human germ cells
HMTase	Histone methyltransferases
HRP	Horseradish peroxidase
ICM	Inner cell mass
IF	Immunofluorescence
Ig	Immunoglobulin
IHC	Immunohistochemistry

iPS	Induced pluripotent stem cell
Kb	Kilobase
KDa	Kilodalton
LIF	Leukemia inhibitory factor
LOCKs	Large organized chromatin lysine 9 modifications
µg	Microgram
µl	Microliter
µm	Micrometer
MEFs	Mouse embryonic fibroblasts
mES	Mouse embryonic stem cells
mEG	Mouse germ cells
mGS	Multipotent germ cells
min	Minute
mm	Millimeter
mRNA	Messenger RNA
miRNA	Micro RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
N	Amino
n	Number
NOD	Non-obese diabetic mouse
NT-ES	Nuclear transfer embryonic stem cells
NuRD	Nucleosomal remodeling and histone deacetylase complex
ORF	Open reading frame
p	p-value
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBT	PBS-Tween
PCR	Polymerase chain reaction
PcG	Polycomb group
PE	Phycoerythrin
PEG	Polyethylene glycol
PGCs	Primordial germ cells
PFA	Paraformaldehyde
PI	Propidium iodide
PTM	Post-translational modification
RA	Retinoic acid
RIPA	Radioimmunoprecipitation assay buffer (modified)
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNApolII	RNA polymerase II
RT-PCR	Reverse transcriptase PCR
SDS	Sodium dodecyl sulfate
SCID	Severe combined immunodeficiency
SCNT	Somatic cell nuclear transfer
siRNA	Small interfering RNA
shRNA	Short hairpin RNA
VSV-G	Vesicular stomatitis virus G protein
WB	Western blotting

Chapter 1.

Introduction

Pluripotent stem cells

Definition of pluripotency

Stem cells are characterized by their ability to proliferate in an undifferentiated state and to give rise to differentiated progeny. There are two major kinds of stem cells: embryonic stem (ES) cells, and adult stem cells. ES cells are the prototypical pluripotent stem cell (Figure 1.1), that is, they have the capacity to generate differentiated progeny from all three embryonic germ layers (endoderm, mesoderm and ectoderm) as well as the germline (Bradley et al., 1984). They can be expanded extensively in culture because of their self-renewing capacity. Despite their proliferative capacity, they maintain their genomic stability, are diploid and kariotypically normal (Pera et al., 2000). In contrast to ES cells, adult stem cells (such as neural stem cells or hematopoietic stem cells) have a more restricted differentiation capacity and usually generate cells of the tissue from which they are derived, and are therefore called multipotent. Adult stem cells are maintained throughout the life of the organism by their ability to self-renew. In mammals, totipotency, which is the ability to generate a whole organism (including the extra-embryonic tissue necessary for the formation of the body), only exists in the zygote and in the first cleaved blastomeres (Suwińska et al., 2008).

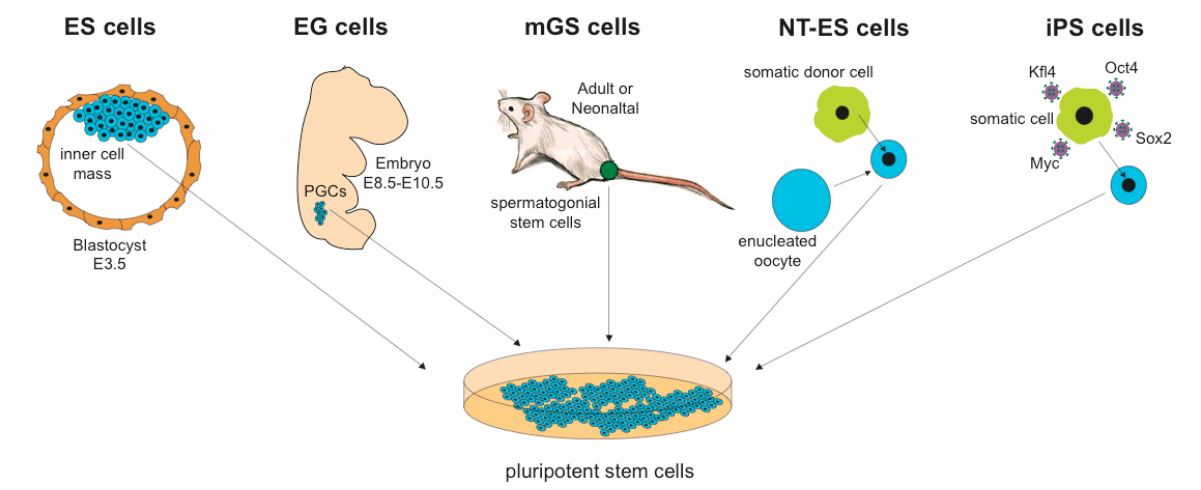


Figure 1.1 The origin of pluripotent stem cells. There are five sources of pluripotent stem cells in mouse (that according to the most stringent criteria, are able to give rise to a whole organism): ES cells, derived from the inner cell mass of a blastocyst; embryonic germ (EG) cells, that are derived from primordial germ cells (PGCs) in the early embryo (Embryonic day E8.5-E10.5); the multipotent germline stem (mGS) cells derived from spermatogonial stem cells in adult and neonatal testis; nuclear transfer embryonic stem (NT-ES) cells resultant of reprogramming of a nucleus using an enucleated oocyte; and induced pluripotent stem cells (iPS) through manipulation of somatic cells with expression of reprogramming factors (most commonly used Oct4, Sox2, Myc and Klf4).

There are several criteria to assess pluripotency (Jaenisch and Young, 2008). The first one is *in vitro* differentiation. When pluripotent stem cells are cultured without LIF and in non-attachment conditions, they form embryoid bodies (EBs),

recapitulating early differentiation expression patterns, with all the three germ layers represented (Keller, 2005). This test has the obvious limitation of not being functional. Another assay used is the teratoma formation, where pluripotent cells are injected in immunodeficient mice, to induce tumors with differentiated cells from all lineages, in an *in vivo* setting. However, this does not test for the ability of the cells to promote normal development. Chimera formation is the third assay to assess pluripotency. Chimerism is the ability to generate all cell types once introduced into a host blastocyst, giving rise to an organism with fully differentiated cells from the host blastocyst and from the injected ES cells. The disadvantage of this assay is that it does not exclude non-autonomous signaling defects that may be compensated by the host cells. If there is contribution of the donor cells of these chimeras to the germline (i.e. germ cells with the genotype of the donor cell that fertilize and give rise to a whole organism), this excludes all possible genetic defects (from the pluripotent cells) but does not exclude epigenetic defects, since these can be erased upon germ cell differentiation. Finally, the most stringent assay for pluripotency is tetraploid complementation, where donor cells are injected with tetraploid blastocyst (cells that are unable to contribute to the embryo, but help form the extraembryonic tissues), giving rise to the whole embryo.

Historic perspective

ES cells were first derived in 1981 (Figure 1.2) from the inner cell mass of the mouse blastocyst (Evans and Kaufman, 1981; Martin, 1981). Before the derivation of mouse ES cells (mES cells) it had been shown that some tumors called teratocarcinomas behaved as a pluripotent and self-renewing population *in vitro* (Kleinsmith and Pierce*, 1964)*. Cell lines derived from these tumors were called embryonal carcinoma (EC) cells (Finch and Ephrussi, 1967). ES cell lines have most of the molecular, morphological and growth characteristics of EC cell lines. However, unlike EC cells, mES cells can contribute to all tissues when injected into blastocysts, including to the germline (Bradley et al., 1984). In 1992 another pluripotent cell type was isolated, this time from mouse primordial germ cells (PGCs), in a very early stage of specification (between embryonic days E8.5 and E10.5). These cells are called embryonic germ (EG) cells and resemble both mES cells and EC cells (Matsui et al., 1992; Resnick et al., 1992).

In 1998 a major event for ES cell research took place: the derivation of human embryonic stem (hES) cell lines (Thomson et al., 1998). Human ES cells are derived from the inner cell mass of blastocysts at about one week post-fertilization. Like mES cells, hES cells are a self-renewing and pluripotent population. Injection of hES cells into immunocompromised mice results in the formation of teratomas, containing cells from the three embryonic layers. Due to obvious ethical reasons it is unclear whether these cells can contribute to a human embryo when introduced into the blastocyst. Also in 1998, human embryonic germ (hEG) cells were derived from gonadal ridges containing PGCs (5-9 weeks post-fertilization) (Shamblott et al., 1998).

The extraordinary capacity of these cells to give rise to all cell types of the body has also been explored in the context of nuclear reprogramming from a differentiated cell into a pluripotent stem cell. Typically, embryonic development is a unidirectional process, with a progressive loss of differentiation potential. It starts with the formation of the totipotent zygote, and ends with the establishment of the specialized cells (more than 200 different types of cells in mammals). To understand the reversibility of the process of specification and loss of differentiation potential, early work in frog (Gurdon and Uehlinger, 1966) has shown that a differentiated nucleus could regain pluripotency by being transferred into an enucleated oocyte (nuclear transfer). This indicated that the genome of an individual adult cell kept the integrity necessary to generate a whole new organism. Moreover, the oocytic cytoplasm contained all the information to reprogram the nucleus back to pluripotency. Since then, nuclear transfer techniques have been used in mammalian differentiated cells to give rise to whole organisms, from sheep to mouse (Wilmut et al., 1997; Wakayama et al., 1998). The ability to generate reprogrammed cells from mammals (specially mouse) has played an important role in the development of ES cell biology and our comprehension of the different cues for specific cell type differentiation (Munsie et al., 2000).

Therapeutic cloning (cloning using adult somatic cells for therapeutic purposes) was a promise for replacement of defective cells in particular diseases, and it was shown in mouse, as a proof of principle. Mutated blood cells from Rag2-/-

mice (a gene involved in the recombination of the antigen specificity machinery and critical in the development of T and B cells) were reprogrammed into nuclear transfer NT-ES cells, genetically modified to introduce the Rag2 gene, and differentiated into blood cells, repopulating successfully the Rag2 -/- mice (Rideout et al., 2002). Probably, the biggest promise of this technology was the ability to generate NT-ES cells from patients, which would allow *in vitro* studies of the development of specific diseases. However, NT-ES cells from humans were never successfully obtained, and involve several ethical issues.

The idea that a particular set of molecules would be sufficient to trigger such dramatic changes in the cell fate, turning fully differentiated cells into undifferentiated cells, was also pursued in cell fusion assays. The first observation that these pluripotent cells contained the information necessary for reprogramming came by fusing EC cells with somatic cells to form tetraploid pluripotent cells (Miller and Ruddle, 1976). Later, EG cells (Tada et al., 1997), ES cells (Tada et al., 2001) and human ES cells (Cowan et al., 2005) were also used in fusion assays. However, these are not able to integrate into the normal developmental program, when injected into blastocysts, so they do not form chimeras, as they are tetraploid. Instead, they proved to be very useful in providing insights into the molecular mechanisms that govern pluripotency and mostly were used as a tool to test and study candidate molecules (Cowan et al., 2005) (Wong et al., 2008), before new technologies became available.

Through ectopic expression of genes (using retrovirus) over-represented in ES cells, a set of four transcription factors (Oct4, Sox2, cMyc and Klf4) was shown to efficiently reprogram differentiated mouse cells to pluripotent state. These induced pluripotent cells were designated iPS cells (Takahashi and Yamanaka, 2006). The surprising effect of only four factors in inducing such a dramatic change in cell fate, initiated a whole new field of research, following the great potential of these cells, but also allowed for a better understanding of the pluripotent state. Initially, the iPS cells formed were found to be similar to ES cells, but not able to support the development of chimeric mice. Further technical improvements (namely using drug selection driven by known ES cell regulators) allowed for the isolation of fully reprogrammed iPS cells (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007), as well as derivation from various differentiated cell types, including blood (Hanna et al., 2008), liver and stomach (Aoi et al., 2008), pancreas (Stadtfeld et al., 2008a), brain (Eminli et al., 2008; Kim et al., 2008), intestine and adrenals (Wernig et al., 2008). This increasing list of cells shows that the induction of pluripotency is possible in many types of cells from all three germ layers. Moreover, human fibroblasts (Lowry et al., 2008; Park et al., 2008b; Takahashi et al., 2007) and keratinocytes (Aasen et al., 2008; Maherali et al., 2008) have been converted into iPS cells using either the described four factors, or using a different combination of factors, including OCT4, SOX2, LIN28 and NANOG (Yu et al., 2007). Finally, a proof of principle of the capacity of these cells to give rise to a whole organism was shown using tetraploid complementation, a technique where the iPS cells were injected back into a female, along with tetraploid host cells (Kang et al., 2009).

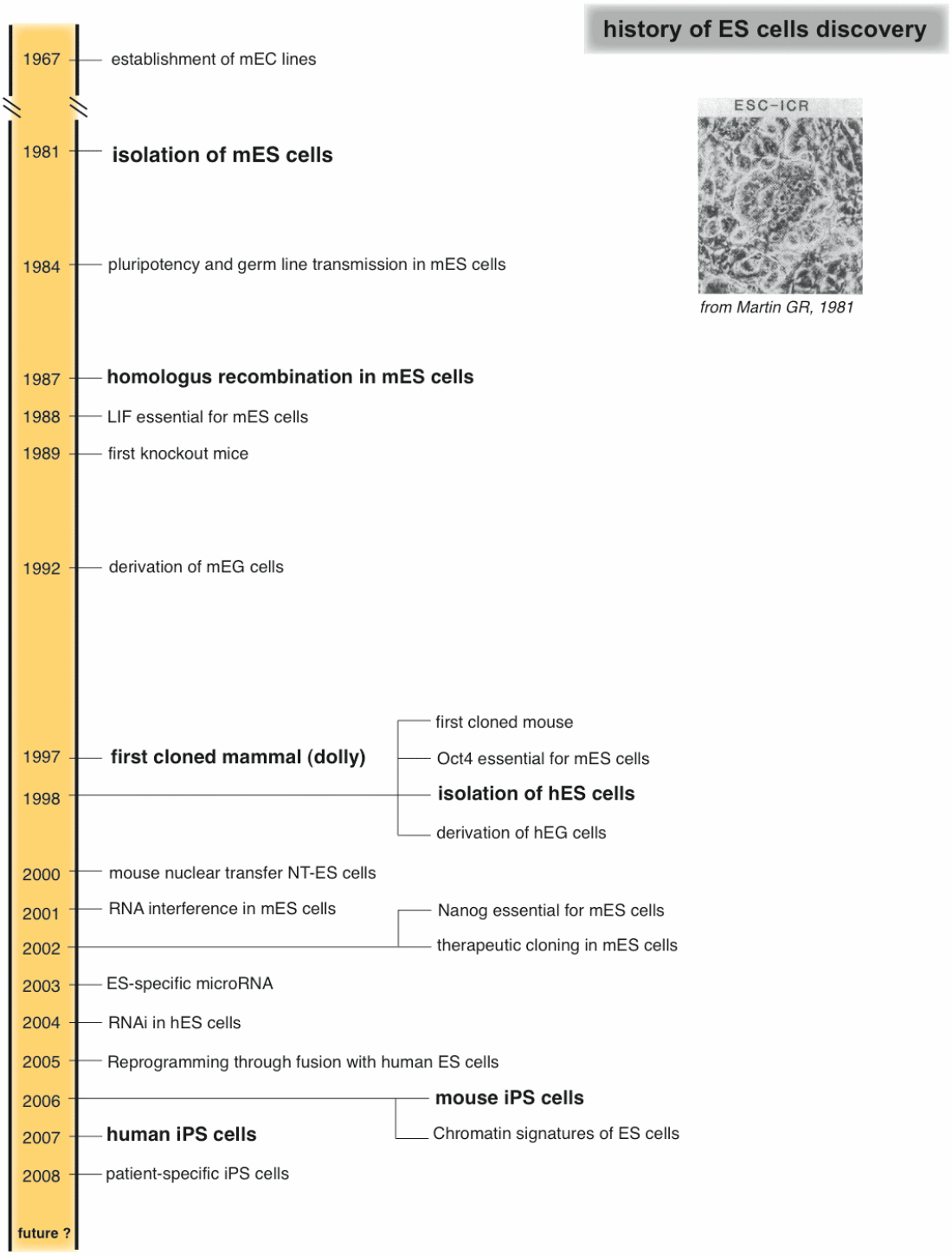


Figure 1.2 The history of pluripotent stem cells. The most significant events that followed the isolation of ES cells in mouse in 1981, are represented until our days.

Therapeutic and Clinical potential

Historically mouse ES cells have been very useful for generating genetically engineered animals (Koller et al., 1989; Thomas and Capecchi, 1990; Thompson et al., 1989) for research purposes by using homologous recombination (Doetschman et al., 1987; Thomas and Capecchi, 1987). Recently, however, the focus of ES cell research has been directed towards clinical applications, such as development of cell-replacement and gene therapies. The big challenge is to be able to apply all the knowledge on ES cell biology and to obtain well-defined protocols for differentiation for cell-based therapies, where some damaged tissues could be replaced by ES or iPS-derived cells. ES and iPS cell pluripotency enables a broader use of these cells in such therapies, when compared to adult stem cells. Besides, they are much easier to grow in culture. Moreover, the use of iPS cells avoids certain ethical issues that arise with the use of hES cells, and with hNT-ES cells. Diseases that involve the loss or damage of a single or very few types of cells are the most attractive candidates for ES/iPS cell therapies. Some attempts have been made to direct ES cells to differentiate into particular cell types, providing interesting insights on how these cells could be potentially used.

In vitro directed differentiation

Lessons learned from developmental biology have been successfully applied to establish differentiation protocols, first using mouse ES cells and more recently human ES cells (Murry and Keller, 2008). Cell types from all three major lineages, endoderm, mesoderm and ectoderm have been obtained. Defining the early signals that establish the specification of these layers was critical for such protocols. FGF signaling is known to induce neuroectoderm in the early embryo. BMP specifies the mesoendodermal lineage that is originated from the primitive streak after gastrulation. *In vitro,* both endo and mesodermal layers can be obtained using BMP4, Wnt or Activin (Nodal activator). Some examples of ES cell differentiation assays are described below.

Endodermal lineages

In the endodermal-derived lineage there are several cell types that are good candidates for cell-based therapies, such as hepatocytes for the liver and pancreatic β cells for replacement in type I diabetes. Using human ES cells, two groups have been able to generate C peptide-positive cells that are capable of releasing insulin in response to glucose, following transplantation into animal models of diabetes (Jiang et al., 2007; Shim et al., 2007). An important physiological test for β cells is precisely their ability to produce insulin in response to glucose, having the C-peptide as a bi-product.

Mesodermal lineages

The hematopoietic, vascular, cardiac, and skeletal muscle lineages develop from subpopulations of the mesodermal lineage. Several groups have described the development of human ES cell-derived populations with a limited *in vivo* hematopoietic repopulating potential, either by transplantation into immunodeficient (NOD/SCID) mice (Tian et al., 2006; Wang et al., 2005a) or into fetal sheep (Narayan et al., 2006). Hematopoietic maturation from hES cells would allow for blood cell

replacement therapies, but a robust sustained multilineage engraftment has been challenging. Cardiomyocytes have also been efficiently derived from hES cells (Laflamme et al., 2007; Yao et al., 2006). Both studies used defined media and induced differentiation without serum and by adding activin and BMP4. The efficiency of differentiation reported was high (>30% cardiomyocytes) but it could be improved by enriching the population (to 80%–90% cardiomyocytes) by using density-gradient centrifugation. The enrichment and purity of the population of differentiated cells is an obvious quality requirement for their proper use in therapy.

Neuroectodermal lineages

Neural cells and skin are the main derivatives of this lineage. Of particular interest in this area are specific types of neurons that would allow for transplantation in cases of loss in certain diseases. Some examples of these studies involve differentiation of dopamine neurons for Parkinson's disease (Yan et al., 2005), motor neurons and oligodendrocytes for motor neuron loss and spinal cord injuries (Keirstead et al., 2005). In most of these studies, the major concern is cell survival, but an obvious caveat is that human-derived cells may also not have the appropriate environment for survival when transplanted into other animal models, such as rat. Moreover, functional improvements have been hard to achieve.

The iPS cells potential

The generation of iPS cells holds great promise as they can be derived from any individual, allowing for patient-specific cell lines (Dimos et al., 2008; Park et al., 2008a) that can be used as a tool to understand the mechanisms of particular diseases as well as new drug discovery. These iPS cells lines may enable the identification of specific genes involved in the pathophysiology of the disease, as well as the characterization of the impact of mutations or allelic diversity among different groups of patients. Once differentiation is well established for a particular cell type, the etiology of a disease can be studied at the molecular and cellular level using these *in vitro* models, allowing manipulations that would otherwise be impossible. The extraordinary capacity of reprogramming cells from an 82-year-old woman diagnosed with amyotrophic lateral sclerosis (ALS) is a proof of principle of the potential of this technology (Dimos et al.). In this study, skin fibroblasts collected from an elder ALS patient were reprogrammed into iPS cells, and differentiated *in vitro* into motor neurons (or at least, cells expressing markers of mature motor neurons), the exact cell type that is destroyed in ALS patients. The ability of these iPS cells to turn into neurons recapitulates the idea that, through development, the right specification signals are followed to form neurons in these patients, and it is through time that the disease arises. With a complexity of the interaction between genetic and environmental factors, these *in vitro* models may help define potential causes, and eventually play a role in correcting or preventing the disease. Furthermore, these cells may also be used for cell-based therapies, since they circumvent the problem of immune rejection. Simply by playing the pluripotency program backwards, we can also ask several questions about the role of particular transcription regulators, and about the epigenetic mechanisms involved in the process (Meissner et al., 2008; Mikkelsen et al., 2008).

These therapies are, however, still far from being safe, especially because of

the tumorogenic potential of these highly proliferative cells. Now, with the use of iPS cells that are transformed and selected specially for their proliferative capacity, the other big challenge is to improve reprogramming efficiency with safer technology. So far, the best way to make iPS cells has been using retro (Yu et al., 2007) or lentiviral-based expression vectors (Takahashi et al., 2007a). This involves insertional mutagenesis that could disrupt gene expression, which could be masked in the undifferentiated state, but would be important further in development. Recent developments, such as the ability to generate iPS cells using non-integrating adeno-virus (Stadtfeld et al., 2008b), transient plasmid transfection (Okita et al., 2008) or the production of recombinant proteins with protein transduction domains which allow for an efficient delivery into the nucleus (Zhou et al., 2009), pave the way for new methods to produce unaltered iPS cells. These methods are still very inefficient, and will require further optimizations.

Transcriptional regulation

The undifferentiated state is a property that ES cells retained from their *in vivo* counterparts, the inner cell mass of the blastocyst (Yamanaka et al., 2006). This capacity to differentiate into all cell types of the body requires obviously a very tight and defined regulation. One way to view pluripotency is to consider it as a state maintained through the prevention of differentiation while promoting proliferation (Boiani and Schöler, 2005; Niwa, 2007a). Yet, another way to view may be to consider the pluripotent state as a ground state for a mammalian cell, which is maintained by its self-replicating capacity without any specification determinants or epigenetic constrains (Silva and Smith, 2008).

Extracellular signals

Initial studies identified extracellular signals as key players in the maintenance of ES cells *in vitro*. In mouse, the leukemia inhibitory factor (LIF) prevents ES cells from differentiating through the JAK/STAT pathway (Smith et al., 1988; Williams et al., 1988; Niwa et al., 1998). This cytokine that belongs to the IL-6 family, is sufficient to maintain mouse ES cells without the presence of feeders or serum containing medium. LIF acts via heterodimerization of two members of class I cytokines, the low affinity LIF receptor (LIFR) and the signal transducer gp130 (Davis et al., 1993; Gearing et al., 1992). However, this signaling pathway does not seem to be required *in vivo* as the presence of gp130 is not required for early development (Nichols et al., 2001). This might be due to the transient nature of the pluripotent population of cells in the blastocyst, the epiblast, and thus, disruption of the signaling pathway might be inconsequent. Moreover, it is possible to maintain undifferentiated cells by completely bypassing this cytokine signaling pathway in ES cells, which indicates that an innate program for the undifferentiated state of ES cells does not require extrinsic induction, a “ground state” for a mammalian cell (Ying et al., 2008). In human ES cells, however, the addition of recombinant LIF is not sufficient to maintain the undifferentiated state even though they express the LIF receptor and can activate the STAT pathway (Humphrey et al., 2004). Instead, they require blocking the BMP signaling and the presence of FGF sustain the pluripotent state (Xu et al., 2005).

Transcriptional network

Further research has allowed for the dissection of different pathways, especially in the transcriptional network that underlies the execution of pluripotency, or the maintenance of the undifferentiated state. The first genetic experiments revealed three major regulators that were essential for pluripotency both *in vitro* and *in vivo*: Oct4 (Nichols et al., 1998; Niwa et al., 2000), Nanog (Chambers et al., 2003; Mitsui et al., 2003) and Sox2 (Avilion et al., 2003). The POU family transcription factor Oct4 (also known as Oct3/4 or Pou5f1) is a major regulator of the undifferentiated state, and is expressed in critical and very tight levels in ES cells. In fact, it was revealed through a conditional expression system that over-expression of Oct4 causes differentiation into primitive endoderm and mesoderm, and down-regulation of Oct4 induces the trophectodermal lineage (Niwa et al., 2000; Hay et al., 2004). Additionally, the Sry-related transcription factor Sox2 acts together with Oct4,

not only to prevent the expression of the trophectodermal marker Cdx2 (Niwa et al., 2005), but also to cooperate broadly by binding the same genomic regions (Okumura-Nakanishi et al., 2005; Rodda et al., 2005). The NK2 class transcription factor Nanog is also essential for the maintenance of the undifferentiated state, by preventing the endodermal regulator Gata6, which *in vivo,* allows for the definition of the epiblast (expressing Nanog) and primitive endoderm (expressing Gata6) layers (Singh et al., 2007). Nanog is also sufficient to overcome the need of the LIF signaling pathway activation for ES cell self-renewal (Chambers et al., 2003; Mitsui et al., 2003). Nanog may play a role similar to a gatekeeper of pluripotency, which would be governed mostly by Oct4 and Sox2. In fact, even though Oct4 and Sox2 help maintaining the undifferentiated state, they induce expression of FGF4, which propels cells for differentiation through the Fgf4/Erk signaling pathway (Kunath et al., 2007). This propensity for differentiation seems to be blocked by overcoming this pathway, either by pharmacologically blocking FGF receptors, or adding LIF, or by overexpressing Nanog.

Analysis of the binding of transcriptional regulators using chromatin immunoprecipitation has helped understand how they may be connected, but opens more questions on how the regulation is established. The core of transcription factors (Oct4, Sox2 and Nanog) defines a molecular circuitry that is essential for the maintenance of the undifferentiated state. The identification of the genomic regions, both in human (Boyer et al., 2005) and in mouse (Chen et al., 2008; Loh et al., 2006), occupied by Oct4, Sox2 and Nanog revealed that these genes often bind together to target genes, including at their own promoters, forming an auto-regulatory loop. They are present not exclusively at promoter regions of active genes (mostly ES cell house-keeping genes, including proliferation genes) but also present in repressed genes (developmental genes), indicating that these three transcription factors work in different contexts to either induce or prevent expression (Liang et al., 2008). In other words, this circuitry is used to maintain the self-renewing capacity of ES cells, and to prevent differentiation.

Other transcription factors have also been connected to this network, both by acting together with the expression of the core network, such as FoxD3 (Guo et al., 2002; Hanna et al., 2002), Sall4 (Zhang et al., 2006) and Tcl3 (Tam et al., 2008; Yi et al., 2008), or as effectors of this core circuitry, such as Tcl1, Esbrr and Tbx3 (Ivanova et al., 2006).

More recently, as new technologies became available for the analysis of both transcripts and proteins, it has been possible to grasp the real complexity of the transcriptional networks. A protein interaction network was created by affinity purification of all the factors binding to the major ES cell regulators, Nanog and Oct4, as well as Rex1, another highly expressed gene in ES cells (Wang et al., 2006). It is clear from this analysis that these regulators cooperate mostly with nuclear factors, both genetic or epigenetic regulators, and that these can be part of either activating or repressive complexes, reinforcing the idea of this dual function of the ES cell network. Deep sequencing strategies also allowed for the identification of non-coding RNAs, such as microRNAs, that are either ES cell-specific or expressed upon differentiation (Houbaviy et al., 2003; Suh et al., 2004; Marson et al., 2008). The most represented family of microRNA in ES cells (miR-290-295) seems to be associated

with the tight regulation of expression of the ES cell-specific genes, allowing for a fine tuning of their levels and promoting their rapid clearance as cells transition into a differentiated state (Marson et al., 2008).

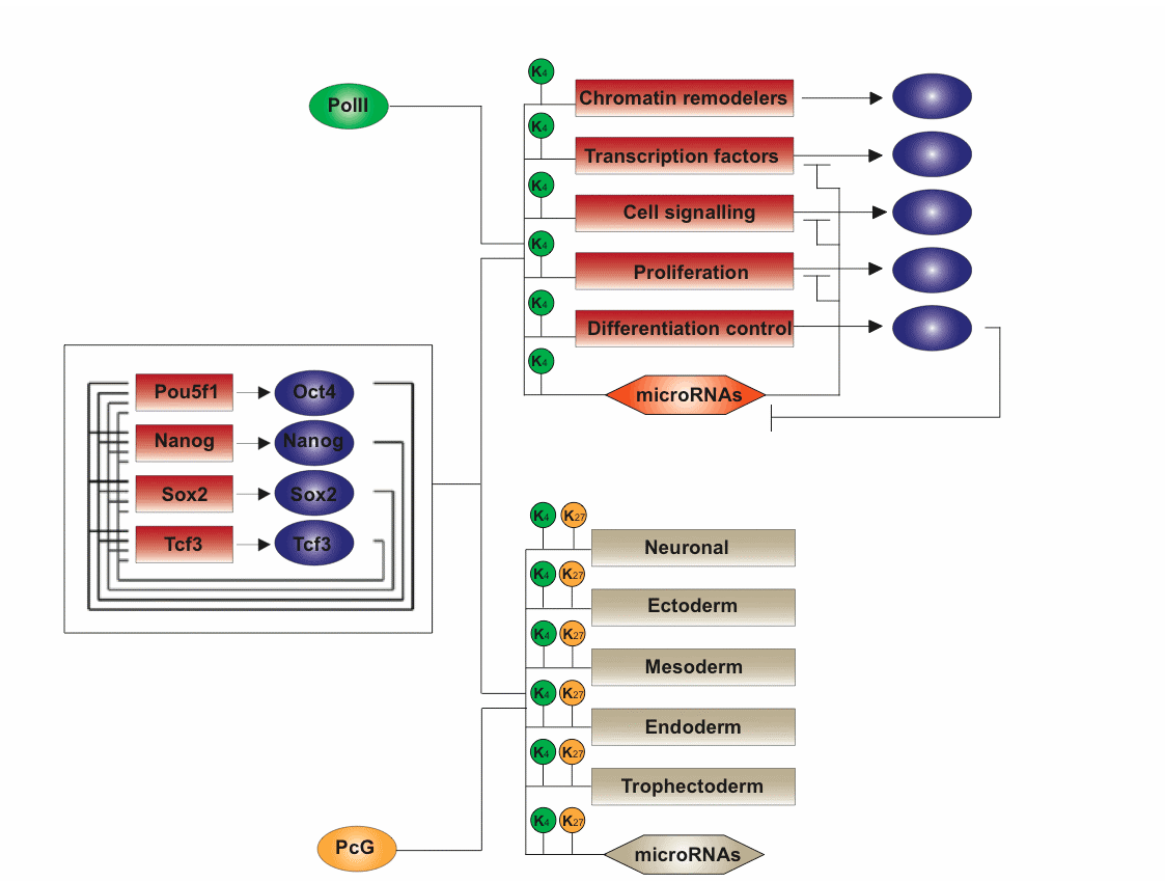


Figure 1.3 An integrative view of the ES cell transcriptional network. An interconnected auto-regulatory loop of transcription factors, where all the proteins (purple ovals) bind directly to the core genes (red box), is shown on the left side. This core of transcription factors binds to both active (red box) and inactive (gray box) genes, and also to microRNAs (orange hexagon). Active genes (including ES cell specific genes with various functions) are bound also by RNA polymerase II and are marked with the active mark H3K4me3 (green circle). Repressed developmental genes are bound also by the Polycomb protein Suz12, and marked with the inactive histone mark H3K27me3 (yellow circle). MicroRNAs in ES cells regulate levels of ES cell genes, but can also be regulated by differentiation control genes, such as Lin28. Adapted from Marson *et al.*, 2008 and Lee *et al.*, 2006.

An interesting example of the complex regulation that these non-coding RNAs allow is the case of the let-7 microRNA family. The regulators Oct4/Sox2/Nanog/Tcl3 bind to the promoters of the let-7g microRNA, activating the expression of the primary transcripts, which are abundant in ES cells. However, the maturation of this microRNA is blocked by Lin28 (Viswanathan et al., 2008), which is also activated by Oct4/Sox2/Nanog/Tcl3. This way, the Let-7 microRNA family is repressed in ES cells, being only represented in mouse embryonic fibroblasts or neuro-precursors. But at the same time, allows for a rapid activation as soon as the maturation block stops, through down-regulation of Lin28 (Marson et al., 2008). This study also shows a very interesting attempt to describe the pluripotency network, by using integration of high-resolution genomic binding data of protein-coding and miRNA genes, systematic identification of miRNA promoters through analysis of activating promoter histone

marks, such as H3K4me3 (which will be discussed further later on), and quantitative sequencing of short transcripts in multiple cell types. In other words, this study establishes a link between transcription factors (Oct4/Sox2/Nanog/Tcl3), chromatin marks (H3K4me3) and regulators (polycomb complex protein Suz12), and non-coding RNAs, which are all working together to maintain pluripotency and to allow for its execution, that is, allowing for differentiation upon the right signals (Figure 1.3).

However, it is important to keep in mind that all these studies do not take into consideration the heterogeneity of ES cell population, as they only represent the analysis of whole populations. One good example of heterogeneous expression in ES cells is Nanog (Singh et al., 2007). Nanog seems to be the limiting factor in this Oct4/Sox2/Nanog transcriptional core. In fact, over expression of Nanog overcomes the need for LIF activation by bypassing Fgf/Erk signaling. This limiting factor is indeed expressed at various levels, both in ES cells and *in vivo* prior to the establishment of the epiblast. Nanog could be a key player in maintaining this intrinsic self-replicating and undifferentiated state of the mammalian cell, supporting the theory of a ground state of the cell that, nevertheless, is very dynamic and allows for a heterogeneity of states that can be interconversible or can determine irreversible differentiation.

Mouse and human ES cells

As it has already been pointed out, there are some differences between human and mouse ES cells. These cells not only grow in different culture conditions but also exhibit different expression profiles (Ginis et al., 2004). Clearly, the core transcriptional network is similar in both (Boyer et al., 2005; Loh et al., 2006). But the downstream targets of Oct4, Sox2 and Nanog show very limited overlap, as defined by the genome-wide ChIP analysis (with the caveat that these studies used different technology platforms). Moreover, the expression pattern of several genes related to cell cycle regulation, apoptosis, cytokine expression and others is clearly different. One good example is the LIFR complex, LIFR/gp130, which is significantly increased in mouse. This is related to their signaling differences. LIF signaling does not support self-renewal of human ES cells, whereas BMP actively promotes their differentiation. A combination of FGF and activin/nodal signaling maintains self-renewal of human ES cells under serum-free conditions (Vallier et al., 2005). In terms of differentiation potential there are some differences as well. Human ES cell cultures have been reported to contain cells with trophoblast-like expression patterns (Thomson et al., 1998), which are lineages that will not form the embryo but the supporting tissue during development. Mouse ES cells do not give rise to trophoblast cells unless genetically manipulated (Niwa et al., 2000) or under certain culture conditions (Schenke-Layland et al., 2007). As already mentioned before, to evaluate the full pluripotentiality of these cells, chimeric formation and tetraploid complementation are used in the case of mouse ES cells. For obvious ethical reasons, human ES cells can only be tested *in vitro*, or in very limited *in vivo* conditions.

The differences between these two cell types raised the question of whether they were actually equivalent, with some species-specific differences, or they could actually be representative of two different states of pluripotency. Recent work by two groups simultaneously, led to the discovery of another pluripotent stem cell that may

provide interesting insights on the nature of both mouse and human ES cells. These cell lines, termed EpiSCs, were derived directly from the early post-implantation epiblast in the mouse (Brons et al., 2007; Tesar et al., 2007). These cell lines still express the main pluripotency transcription factors, Oct4, Sox2, and Nanog, but need to be cultured in the presence of FGF and activin rather than LIF. EpiSCs were able to generate tissues from all three germ layers *in vitro* and to form teratomas but they were unable to contribute to normal tissues in mouse chimeras. The similarities of these mouse EpiSCs with hES started to become more and more intriguing. As well as requiring the same growth factors for self-renewal as human ES cells, EpiSCs were also reported to show gene expression profiles and transcription factor networks closer to human ES cells than to mouse ES cells. Interestingly, like human ES cells, EpiSCs could express markers of trophoblast and primitive endoderm when treated with BMP4 (Brons et al., 2007). They also share X chromosome silencing in female cells, a feature that is not shared by mouse ES cells.

These similarities have led to suggest that hES cells might be the equivalent of the early post-implantation epiblast. However, this does not explain their propensity for trophoblastic differentiation, as this lineage in the embryo has been already specified before implantation. It is possible that these pluripotent cells may open pathways *in vitro* due to their culturing conditions that otherwise are tightly controlled in the developing embryo. On the other hand, mouse ES cells could be representing an earlier stage on the pre-implantation epiblast. Again, culturing requirements and the differentiation signaling pathways of mouse ES cells may obscure their ability to form trophoblast-like cells. As LIFR/gp130 activation is not required *in vivo* (Nichols et al., 2001), it is hard to recapitulate exactly what happens in the mouse embryo and what the full potential may be. We also know that mouse ES cells are very responsive to FGF signaling towards differentiation, which could eventually mask any trophoblastic propensity.

Epigenetic regulation

Epigenetic signature

The term “epigenetics” was first used to describe events that could not be explained merely at the light of genetic principles. The epigenetic landscape was a visual metaphor used to describe the process of cell differentiation during development (Figure 1.4). Much of the epigenetic field is converging on the study of how modifications in chromatin can influence transcriptional outcome (Goldberg et al., 2007).

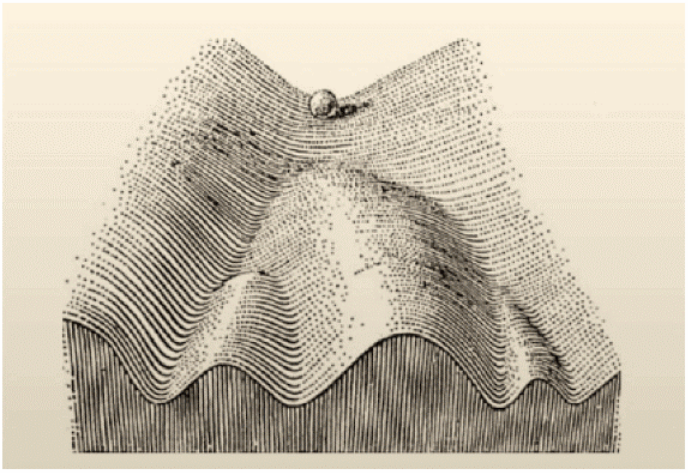


Figure 1.4 Waddington’s “epigenetic landscape” (1957). This visual diagram was used as a metaphor, showing how cells (represented as a sphere), with identical genotypes, would choose different paths (represented as valleys) during embryogenesis (Waddington, 1957).

Chromatin is a complex assemblage of DNA, histone proteins and other non-histone protein components. Histone proteins form chromatin building blocks, the nucleosomes (each nucleosome consisting of an octamer of the canonical histones H2A, H2B, H3 and H4), that establish folding into higher order structures through various mechanisms, thus determining whether genomic regions can be accessible for transcription or not (Lodén and van Steensel, 2005). Another way of regulating transcription is to actively promote the association of chromatin binding proteins, through chromatin modifications (often called epigenetic marks), thus signaling for activation or repression transcription. There are two types of chromatin modifications regulating transcription: DNA methylation and histone post-translational modifications (histone PTMs).

DNA methylation occurs at cytosine residues usually within CG dinucleotides, and generally opposes transcription (Bird, 2002). These CG dinucleotides are normally clustered in regions called CpG islands, localized in various regions of the genome, such as promoters, repetitive elements and transposons. The addition of a methyl group to the DNA is carried out by DNA methyltransferases (DNMTs) that are responsible for both the maintenance of methylation after replication (Dnmt1) and the *de novo* methylation with the establishment of new methylation patterns (Dnmt3a, Dnmt3b and Dnmt3L). Both enzymes are critical for mouse development (Lei et al., 1996; Li et al., 1992a; Okano et al., 1999) and ES cell differentiation (Jackson et al., 2004; Panning and Jaenisch, 1996). On the other hand, active DNA demethylation is still a very controversial field, and may involve players from different functions such as DNA repair like Gadd45 (Barreto et al., 2007) or could even be initiated by the

same DNA methyltransferases Dnmt3a and Dnmt3b (Kangaspeska et al., 2008; Métivier et al., 2008).

There are several post-translational histone modifications identified so far, most of them are in the amino acid residues of the histone tail (which is the structure more exposed, out of the nucleosome core): Lysine acetylation, Serine or Threonine phosphorilation, Arginine and lysine methylation, Lysine ubiquitination, Lysine sumoylation and Proline isomerization (Figure 1.5). A typical model for the role of histone PTMs is that positive-acting marks are established across promoters and open reading frames (ORFs) during gene activation through recruitment of the relevant enzymes by DNA-bound activators and RNA polymerase. Similarly, negative-acting marks are laid down across genes for repression by reversible association with repressive complexes or by irreversibly associating with heterochromatic proteins. In summary, histone modifications mark particular genomic regions by presenting chemical surfaces, which then recruit regulatory protein complexes. Simple combinations of consistently behaving marks would therefore correspond to a definable and predictable outcome. This so called histone code has been intensively dissected in the past few years (Jenuwein and Allis, 2001; Turner, 2007). However, the complexity and the dependence of the cellular context to predict a single histone PTM meaning or outcome, opens the debate on how this histone code can be defined, and whether it should be renamed more accurately, to histone signature (Sims and Reinberg, 2008). It is true that particular marks present consistent meanings, but in a broad biological perspective, most of them show a context dependence, which allows for multiple combinations of recruitment of regulatory proteins, making the predictability of a code almost impossible. Nevertheless, this epigenetic signature, or epigenome, has provided various insights on several biological processes like development, homeostasis, cancer, and, of course, ES cell biology.

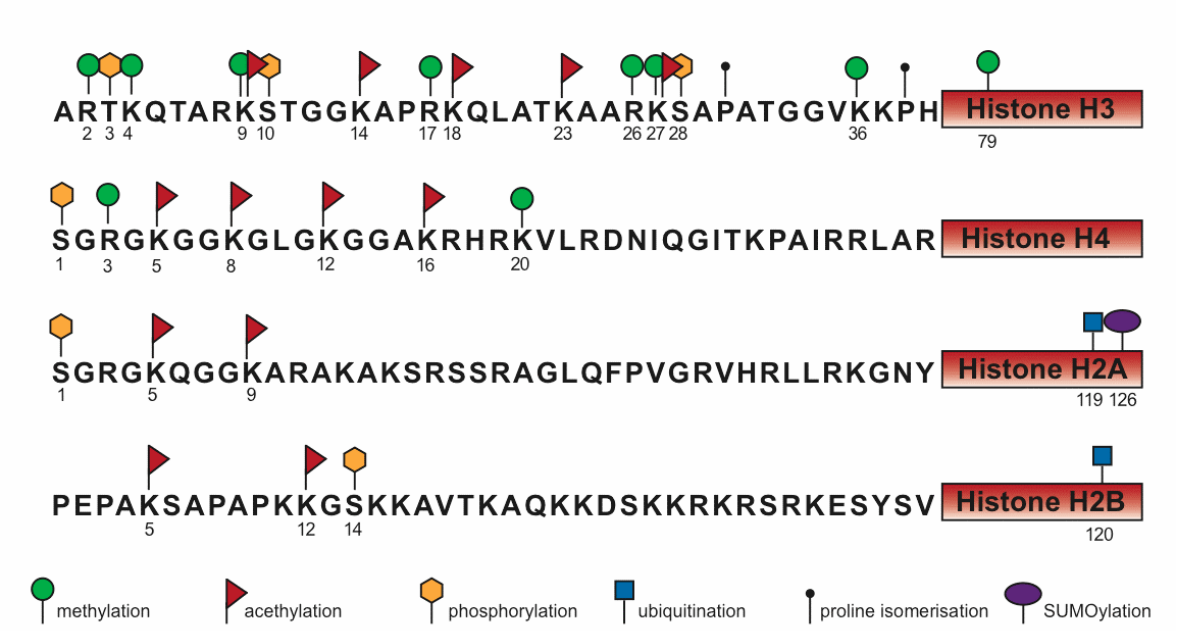


Figure 1.5 Core histone post-translation modifications. Histone tails (N terminus), here represented by the aminoacid sequence, are the most exposed regions for the modifying enzymes to act on, but some modifications can also occur in the body of histones (red box).

The pluripotency epigenome

Genome-wide mapping of chromatin marks has been of great use in defining the patterns that may govern pluripotency. Additionally, in recent years, several chromatin modifiers, from DNA or histone methyltransferases (HMTase), histone demethylases (HDMase), histone deacetylases (HDACs) and chromatin remodeling complexes, have been studied in the context of ES cells, and they have been shown to have a critical role in the regulation of the undifferentiated state and in the process of differentiation. The importance of epigenetic regulators in ES cells is evident since they are over-represented, along with transcription factors, in the transcriptome of ES cells (Efroni et al., 2008).

In general, ES cells lacking these epigenetic regulators tend to exhibit a defective differentiation, reflecting their role in determining and engaging particular lineage specific programs. Nevertheless, the re-introduction of deleted genes coding for epigenetic regulators restores the ability to differentiate, which indicates that pluripotency has remained intact. Epigenetic regulators may in fact be dispensable for the maintenance of the undifferentiated state, as none of the mutants seem to affect ES cell viability (Niwa, 2007a). It has been suggested that the epigenetic processes are not responsible for the maintenance of the pluripotency program, but rather for its execution upon differentiation (Niwa, 2007b). Alternatively, chromatin would function as a buffer for the “noise” of transcriptional variability (Chi and Bernstein, 2009), which seems to be especially important in cells with high levels of permissive transcription, like ES cells (Efroni et al., 2008; Szutorisz et al., 2006). Yet, it becomes clear that understanding epigenetic regulation and especially the epigenetic marks, this so called epigenome, is critical to draw a full picture of how undifferentiated cells maintain their state and execute the differentiation pathways (Bernstein et al., 2007).

ES cells are abundant in marks of active transcription, such as histone H3 lysine 4 tri-methylation (H3K4me3) and the acetylation of Histone 4 (H4Ac) but some mechanisms are necessary to silence developmental genes that would otherwise drive cells to specific differentiation pathways. Interestingly, replication-timing analysis suggests that lineage specification genes are more accessible in ES cells than in differentiated cells (Azuara et al., 2006). These developmental genes, despite being transcriptionally silenced, are proned for activation by having both the activating mark (H3K4me3) and a repressive mark (H3K27me3) (Bernstein et al., 2006; Pan et al., 2007). These so-called bivalent domains, even if not strictly specific in ES cells, are markedly prevalent in these cells, indicating a mechanism by which ES cells are able to activate (upon induction) specific differentiation pathways.

The repressive H3K27 methylation is regulated by the Polycomb group of proteins (PcG). Through a genome wide analysis (both in human and mouse) of the presence of several members of the PcG, it was possible to define their role in silencing developmental regulatory genes, loci that are clearly enriched in PcG members (Boyer et al., 2006; Lee et al., 2006). Moreover, presence of PcG members also overlaps with the transcription factors Oct4, Sox2 and Nanog, known to be integrated in unique repressive complexes in ES cells (Liang et al., 2008). However, PcG proteins are not essential for ES cell pluripotency. In the absence of *Eed*

(Chamberlain et al., 2008), *Suz12* (Lee et al., 2006) and *Ezh2* (Shen et al., 2008), ES cells can still self-renew and are able to differentiate into all three germ layers, indicating some possible redundancy in this complex. More recently, the histone variant H2AZ was shown to occupy regions that follow a similar pattern of the Polycomb group protein Suz12 in the promoter regions of developmental genes that are inactive (Creyghton et al., 2008). Their occupancy is interdependent and down-regulation of H2AZ shows this variant is necessary for ES cell lineage commitment.

The lack of significant repressive marks was also described in a recent study reporting that large H3K9me2 repressive domains, called LOCKs (large organized chromatin K9 modifications), cover only 4% of the ES cell genome, when compared to differentiated cells (31%). These domains are inversely related to gene expression, and dependent on the activity of the histone methyltransferase enzyme G9a. Interestingly, they are conserved between human and mouse (Wen et al., 2009).

G9a is important not only in establishing LOCKs in differentiated cells, but also in repressing ES cell specific genes upon differentiation. It has been shown that upon LIF removal, Oct4 expression is reduced until completely repressed in about 10 days. This repression is irreversible, since these cells are not able to re-express Oct4, even if stimulated with the addition of LIF. The process of silencing involves the methylation of H3K9 at the promoter region of Oct4 gene and the recruitment of DNA methylases to further signal more definite repressive state, that includes Heterochromatin Protein 1 (HP1) binding to the site. G9a plays an important role in silencing Oct4 expression since G9a null ES cells can re-express Oct4 after prolonged LIF removal (Feldman et al., 2006). G9a might have a dual role of methylating H3K9 (as a known HMTase) and recruiting DNA methyltransferases, an example of how several layers of regulation accomplish proper silencing of a particular gene.

The balance between the addition of repressive marks and their removal is also essential for the proper gene expression in ES cells. Two demethylases of H3K9, Jmjd1a and Jmjd2c, in addition to regulating global levels of H3K9me2 and H3K9m3, respectively, play a critical role as regulators of ES cell maintenance (Loh et al., 2007). Both are regulated by Oct4, and independently regulate the expression of key transcription factors, Tcl1 (Jmjd1a) and Nanog (Jmjd2c) by demethylating H3K9 at their promoter region. This is a good example of a positive feedback-loop integrating both genetic and epigenetic systems. A core regulator of ES cells, Oct4, instructs the activation of other core regulators (the transcription factors Tcl1 and Nanog) indirectly, through the induction of two demethylases of the repressive mark H3K9, that act on the promoter regions of these transcription factors.

A different layer of epigenetic regulation in ES cells is the DNA methylation of CpG islands. DNA methyltransferases (DNMTs) are responsible for these repressive marks, which are interestingly correlated with specific histone marks (Meissner et al., 2008). Methylated CpG islands are present mainly at promoter regions of repressed genes, usually correlated with unmethylated H3K4 and with tri-methylated H3K9 (Ooi et al., 2007), and represent around 30% of genes in ES cells. Cross-referencing methylation patterns to binding of Oct4, Nanog, Sox2 and polycomb proteins

revealed little overlap (Fouse et al., 2008). This suggests that these methylation patterns might represent a unique epigenetic program that complements other regulatory layers of repression, allowing for a tight regulation of specific transcriptional programs that ES cells activate upon differentiation. It may also be a safeguard against aberrant transcription that is often observed in cancer, and is associated with different patterns of DNA methylation and chromatin regulation in general. Interestingly, *de novo* DNA methylation is required for ES cell differentiation, and ES cell hypomethylation also results in hyperacetylation, showing how these repressive marks may be critical for the proper silencing of the ES cell program (Jackson et al., 2004).

In this section there should also be included chromatin remodelers, which relate directly and indirectly with histone marks, but act mostly by changing DNA-histone interactions. These proteins will be discussed further, later in the text, as the class of proteins that include Chd1, the protein that this study focused on.

Open chromatin

Stem and progenitor cells need a particular transcriptional plasticity, since they have to be able to activate several different specification pathways. These cells have been described as having a typical open chromatin conformation, from neoblast cells in planaria (Reddien and Sanchez-Alvarado, 2004), to hematopoietic stem cells in mammalian cells (Ford et al., 1992; Spangrude et al., 1988), and more recently to ES cells.

Chromatin is classified in three types, depending on the level of compactation: constitutive heterochromatin, facultative heterochromatin and euchromatin, the open chromatin. In particular, the last two domains can be very dynamic and, to some extent, interchangeable (Trojer and Reinberg, 2007). The first distinction between euchromatin and heterochromatin is credited to Heitz in 1928, based on the interphase nuclear appearance (Heitz, 1928). It has been classically accepted that heterochromatin is correlated with repressed regions (Brown, 1966) and is composed of genomic regions where nucleosomes are close together, forming a compacted chromatin domain, preventing access to the transcriptional machinery (Trojer and Reinberg, 2007). Telomeric and centromeric regions of the chromosomes are characterized by constitutive heterochromatin, and are generally gene-poor, with the presence of some transposable elements, like satellite repeats, that in most cells remain silenced. Several other features characterize heterochromatin, such as enrichment in H3K9me3, binding of HP1, and DNA methylation. Active (expressed) genes are part of euchromatin, but there are exceptions, maybe due to its dynamic nature (Sapojnikova et al., 2008). Interestingly, nucleosomes are not always regularly spaced in euchromatin, since they seem to be absent in promoter regions of active genes (Ozsolak et al., 2007), suggesting an important role in transcriptional activation.

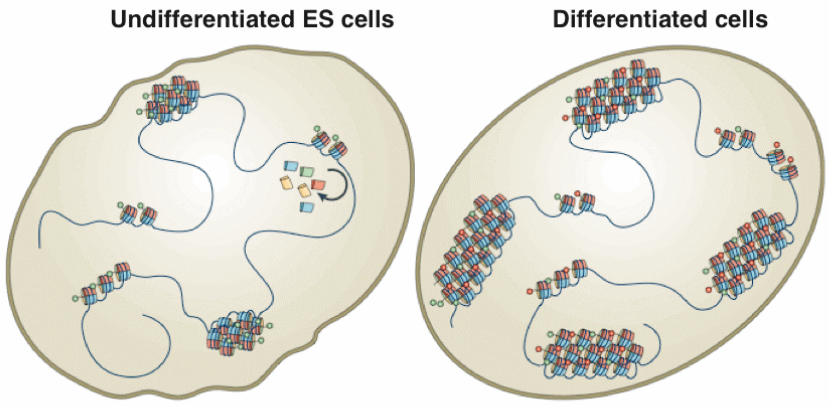


Figure 1.6 . Hyperdynamic chromatin in ES cells. In ES cells (left), chromatin is globally decondensed, enriched in active histone marks (green circular tags), and contains a fraction of loosely bound architectural chromatin proteins. As cells differentiate (right), regions of condensed heterochromatin form and silencing histone marks accumulate (red circular tags), which leads to a more stable association and tighter compaction of structural chromatin proteins with DNA (Meshorer and Misteli, 2006).

In pluripotent ES cells chromatin is globally decondensed and contains a fraction of loosely bound architectural chromatin proteins, such as core and linker histones as well as HP1 (Meshorer et al., 2006) (Figure 1.6). This fraction is a true hallmark of ES cells as it is no longer available in differentiating cells. In addition, the ES cell genome is transcriptionally hyperactive: it transcribes normally silenced repetitive elements and contains transcripts of both coding and non-coding regions,

with increased levels of total and mRNAs (Efroni et al., 2008). It has also been shown that in ES cells, the proteasome has an important role in restricting permissive transcription by proteolytic degradation of pre-initiation transcription assemblies, that form in specific regulatory regions primed for transcription (Szutorisz et al., 2006). This would allow for the maintenance of a low threshold of expression of these tissue-specific genes, in order to preserve the pluripotency network, in this transcriptional permissive context.

The way by which chromatin structure is more accessible in ES cells may be determined mostly through the enrichment of active histone marks, previously described. However, it may also be the result of the direct action of ATPase remodelers involved in the disassembly of nucleosomes and in the reorganization of the higher-order chromatin structure, as it is suggested by high levels of expression of these chromatin remodelers in ES cells (Efroni et al., 2008). These remodelers are known to effectively open specific loci and allow for the transcriptional machinery to bind, at a local level (Xella et al., 2006). It is easy to envision the abundant levels of remodelers orchestrating together with histone marks modulators, a global loosening of the chromatin. While the open chromatin state in ES cells has been described at different levels, it is still not clear whether this open chromatin is indeed essential for ES cell pluripotency. It could well be just a consequence of the abundance of such regulatory machinery. In this thesis, it is proposed that Chd1, a chromatin-remodeling enzyme, is important for the maintenance of this state.

Lessons from iPS reprogramming

The still intriguing capacity of merely four factors to induce such a dramatic change in cell fate and nuclear organization opens several questions about the process of reprogramming, but more importantly, about the pluripotent state itself. The sequence of events has been dissected, but much of what we know now is through indirect use of markers that we can recognize in this process. For instance, upon expression of the four factors, alkaline phosphatase activity and expression of the cell surface marker SSEA1 (in mouse) are already detected after 3 and 9 days, respectively. Endogenous Oct4 and Nanog can only be detected after 2 weeks and the viral expression needs to be on during all this period. However, cells can only fully reprogram upon silencing of the viral vectors (Brambrink et al., 2008). The main question that arises is: how do these transcription factors act to induce pluripotency? It is known that, at least Oct4 and Sox2 are part of this autoregulatory loop that maintains pluripotency in ES cells (Boyer et al., 2005) and that cMyc binds to several genes not bound to Oct4, Sox2 and Klf4 (Chen et al., 2008), namely genes that regulate cell cycle like E2F1 and Zfx. Interestingly, cMyc seems to be a major contributor in the early events of reprogramming, probably through its pleiotropic functions and through a stronger effect on repressing the expression of differentiated genes (Sridharan et al., 2009). It could also be detrimental to prevent cellular senescence of fibroblasts, through inactivation of pathways like Rb (Zhao and Daley, 2008) and p53 (Hong et al., 2009). However, Myc does not seem to be essential as iPS cells can be obtained, both in mouse and human, without it, even tough at a much lower efficiency (Nakagawa et al., 2008; Wernig et al., 2008; Yu et al., 2007). One could speculate that these factors allow for both the activation of the self-renewing and the pluripotency programs, but the role of each factor independently requires analysis at the single cell level, which is not yet available.

It is important to keep in mind that the process of reprogramming also involves a large reconfiguration of the chromatin structure, from DNA methylation to histone modifications and nucleosome packaging. These epigenetic barriers are generally used as repressive mechanisms in somatic cells, to prevent unwanted gene expression from other lineages. How these barriers are overcome is another central question. Several lines of evidence support the notion that the process of reprogramming involves rare stochastic epigenetic events. The first one comes from the observation that the reprogramming process is slow and gradual, with several intermediate states (Sridharan et al., 2009). More importantly, it has been shown that the expression of the four factors alone is not sufficient for induced reprogramming. In fact, the use of genetically modified fibroblasts where the four genes could be directly activated through a doxycycline-inducible system, showed a great improvement in efficiency, but nevertheless, only 3-5% of the cells gave rise to iPS cells (Hockemeyer et al., 2008; Maherali et al., 2008; Wernig et al., 2008). Moreover, reactivation of the endogenous ES cell genes like Oct4, as seen by a reporter cell line with Oct4-GFP, showed that even similar morphological iPS colonies start expressing Oct4 at different times (Meissner et al., 2007). Treatment of somatic cells during the reprogramming process with agents affecting chromatin modifications, such as DNA methylation (using DNA methyltransferase inhibitor 5-aza-cytidine), histone acetylation (using the histone deacetylase inhibitor valproic acid) and histone methylation (using a G9a methyltransferase chemical inhibitor), increases efficiency

of reprogramming and sometimes substitutes for a particular factor (Huangfu et al., 2008a; Huangfu et al., 2008b; Mikkelsen et al., 2008; Shi et al., 2008). Clearly, reprogramming is a very complex sequence of events that involves overcoming epigenetic barriers, the silencing of the somatic cell program, and resetting of the self-renewing and pluripotency programs, not necessarily by this order.

Other lessons from iPS reprogramming are related to the pluripotent state itself. Recent studies have shown that iPS cells should be considered a unique subtype of pluripotent cell, as they have a recurrent expression signature, including coding genes and non-coding RNA species, that is very specific to iPS, regardless of their origin (Chin et al., 2009; Lowry et al., 2008). Interestingly, human iPS cells become more similar to hES cells after several passages, suggesting some form of "reprogramming" upon culture, where cells are selected or acquire an identity closer to an embryonic stem cell. However, it becomes evident that depending on the origin of these cells, either from the blastocyst in the case of ES cells, or from an *in vitro* manipulation of somatic cells as in iPS cells, they acquire specific features and signatures that are not clearly understood, and might have to do with some kind of memory that we are not able to define at this point. It is known that in frog, embryos generated by nuclear transfer of muscle cells (expressing MyoD), still maintain expression of this gene even after several divisions (Ng and Gurdon, 2008). This memory has been shown to be maintained through the deposition of a histone variant H3.3 that is associated with active genes and is replication-independent (Henikoff and Ahmad, 2005). This chromatin mark could establish, through an unknown mechanism, the memory of the genes that had been previously transcribed in the somatic cell. This histone variant is less prone for H3K9 methylation, a typical repressive mark, which could be a possible way of maintaining transcription active in a particular site (McKittrick et al., 2004). The biological significance and importance of such memory is fascinating, but unclear. Yet, we may hypothesize that iPS cells may retain some epigenetic memory, which can potentially explain differences between iPS and ES cells.

Using iPS technology, it has also been possible to understand better the pluripotent state of different cell types. As pointed out previously, EpiSCs are another subtype of pluripotent cells (Brons et al., 2007; Tesar et al., 2007). Even though these cells express the known pluripotent genes, they have a developmental restriction because they are not able to contribute to the embryo when injected into a blastocyst. Using the single reprogramming factor Klf4 into EpiSCs, it was possible to convert these cells into a different pluripotent state, allowing them to contribute to chimeras with germline transmission. These cells were called Epi-iPS cells, and were similar to ES cells. In this case, Klf4 seems to be a main player in bringing back the EpiSCs into a different state, opening the possibility of conceiving not one pluripotent state, but several metastable states. Metastability refers to the interconversion of one pluripotent state into another, which seems to be, for genetic or epigenetic reasons, at different levels. Another good example comes from the analysis of the non-obese diabetic (NOD) mouse. This mouse strain is nonpermissive to the derivation of ES cells, having a genetic background different from the typical 129 or C57BL6 strains that are used to derive ES cells. However, EpiSCs have been derived (Brons et al., 2007). Moreover, NOD iPS cells were shown to be dependent on the expression of the four factors, when used in an inducible system, showing that the pluripotent state

they acquired is not stable (Hanna et al., 2009). Using constitutive expression of either Klf4 or Myc, researchers were able to generate stable pluripotent ES-like cells, which then contribute to chimeras and yielded germline transmission. Thus, it is possible to conceive two states of pluripotency, or even two levels or pluripotency: 1) the ICM derived ES cell-like state and 2) the epiblast derived EpiSC-like state. Exogenous factors such as Oct4, Sox2, Myc and Klf4 can induce the ES-like state from somatic cells, and the stability of this state could be determined by the genetic background. While iPS or ES cells from typical 129 or C57BL/6 backgrounds are stable once established in the presence of LIF, iPS and ES cells from the ‘‘nonpermissive’’ NOD background remain unstable with the maintenance of the pluripotent state depending on the continuous expression of the exogenous factors. The same happens with the EpiSCs; only after forced expression of Klf4, these cells become responsive to LIF, and acquire full pluripotentiality.

The chromatin remodeler Chd1

Chromatin remodeling

Several nuclear functions, such as the initiation of a particular transcriptional program, replication and cell division or even the DNA repair machinery, require a series of chromatin changes, both at the level of higher-order chromatin structure and chromatin organization at specific genetic loci. These chromatin changes are performed by specialized enzymes (usually structural components of large macromolecular chromatin complexes) that recognize and bind directly or indirectly to DNA altering the histone/DNA contact through an ATP-dependent helicase domain. All these ATP-dependent domains belong to the SNF2 protein family (de la Serna et al., 2006). The disruption of the histone/DNA contact itself is poorly understood, but the consequences are the exposure of DNA to regulatory proteins, and the active mobility of nucleosomes and the histones within. These ATP-dependent remodelers also recognize histone modifications allowing the interplay between the information given by the genetic sequence (DNA) and the epigenetic information (histone marks). They fall into three main classes, based on their functional domains: the SWI/SNF family that includes BRM and BRG, with a bromo domain that binds to acetylated histones (Hassan et al., 2002); the ISWI family (from imitation SWI), including SNF2H and SNF2L that contain a SANT histone-binding domain and may also recognize histone modifications; and finally the chromodomain and helicase-like domain (CHD) family, where Chd1 was the first known member. The role of these chromatin-remodeling complexes is critical for mouse development as disruption of specific SNF2 units usually leads to embryonic lethality consistent with a defect in self-renewal, pluripotency or derivation of the ICM (de la Serna et al., 2006). Examples of this are the mutations of genes encoding the chromatin remodeling complex units Brg1, Baf47/Snf5, Baf155 and Baf250a that result in embryonic lethality around implantation (Burman et al., 2000; Gao et al., 2008; Guidi et al., 2001; Kim et al., 2001; Klochendler-Yeivin et al., 2000; Roberts et al., 2000), and embryos lacking Tip60 and Trrap, two components of the Tip60-p400 histone acetyltransferase (HAT) and nucleosome remodeling complex, die before implantation (Gorrini et al., 2007; Herceg et al., 2001).

Chromatin remodeling in ES cells

With such a broad role in terms of functions within the cell, it is hard to dissect their specific effect on the maintenance of the undifferentiated state. However, some interesting correlations have been drawn through coupling these complexes with the transcriptional network of ES cells. This is the case with the SWI/SNF remodeling BAF complex that comprises a specialized subunit composition in ES cells (esBAF). This esBAF complex was shown, with proteomic and global binding analysis, to interact with the core factors Oct4, Nanog and Sox2 (Ho et al., 2009). Another example is the chromatin remodeling Tip60-p400 complex which is known to help transcription, by combining its nucleosome remodeling function with the acetylation activity (another mark for active transcription) in specific loci. In a recent study, this Tip60-p400 complex has shown to read the H3K4me3 mark along with the binding of Nanog. Interestingly, depletion of Nanog reduces p400 binding to target promoters

without affecting the histone mark, suggesting that the binding to the histone mark and Nanog are independent (Fazzio et al., 2008).

Other studies highlight how important it is to silence specific ES cell regulators upon differentiation involving the nucleosome-remodeling complex NuRD (Kaji et al., 2006; Kaji et al., 2007). This repressor complex, which includes two histone de-acetylases (Hdac1 and Hdac2), is required for proper differentiation of ES cells. ES cells lacking Mbd3 (another core element of the NuRD complex) are incapable of silencing Oct4 expression upon LIF removal, and show aberrant differentiation potential. One example is the expression of extra-embryonic trophectodermal markers, not usually derived from ES cells. This repressive complex is therefore, critical for the proper silencing of both the ES-specific genes during differentiation, and the differentiation marks on the ES cell state. Similarly, a member of the chromatin remodeling complex PBAF, Smarcc1/Baf155, is involved in the compaction of the chromatin during ES cell differentiation, necessary for the proper silencing of the ES cell transcriptional network, namely Nanog (Schaniel et al., 2009).

The CHD family

There are 9 CHD members known in mice and humans (Figure 1.7). The CHD family (Chromodomain helicase-like DNA binding family) consists in large proteins (all over 200KDa) that can be clustered into three main classes according to their domain conserved organization. The first class includes CHD1 and CHD2, the second class, CHD3 and CHD4, and the third class, the remaining CHD5 through CHD9. All of these proteins share three main components: the chromodomain, the helicase/ATPase domain, and a DNA-binding domain, which are described in detail below. The unique combination of these domains is what gives this family of proteins their name.

- The first domain is the conserved chromodomains (typically two) in the N-terminus. The chromodomain, named for its function as a chromatin organization modifier, is a highly conserved sequence motif that has been identified in a variety of animal and plant species, and recognizes specific histone tail post-translation modifications. It exists in other chromatin regulators, such as the HP1 family or Pc-like proteins (Koonin et al., 1995; Paro and Hogness, 1991). The most divergent group of chromodomain proteins is the CHD family.

- The central SNF2-like helicase-ATPase domain is the best conserved domain amongst the family, with up to 95% identity conserved, and is responsible for the ATP-dependent nucleosome remodeling activity. The helicase activity drives the rotation of nucleosomal DNA, promoting the release of the histone octamer, while the ATPase activity provides the energy necessary for the nucleosome displacement (Hall and Georgel, 2007).

-The DNA binding domain is the most divergent domain. Although it is very well defined in CHD1/2, it remains poorly characterized in the other two classes, since there is no canonical DNA-binding domain. However, these two classes contain several motifs, such as telobox, PHD or SANT domains, which may confer DNA affinity (Hall and Georgel, 2007).

The three classes

CHD1 and CHD2 are characterized by the well-defined DNA binding domain, which is not sequence specific but with a strong preference to A+T rich regions (Stokes and Perry, 1995). Most of what is known about this class comes from studies in CHD1, mostly because CHD2 only exists in mammals but not in yeast or drosophila. Despite the similarities of both sequences, some critical differences may confer alternative functions to these two molecules. For instance, CHD2 has a longer insert in its chromodomain 2, which likely interferes with the peptide binding surface, and consequently to its binding to histone marks. In fact, an *in vitro* study shows how the CHD2 fragment is 30-fold weaker at interacting with the H3K4me peptide than CHD1 (Flanagan et al., 2007). The function of CHD1 will be further discussed later on, but in general, this class seems to be associated with transcriptional activation through recognition of active histone marks, such as acetylation and H3K4 tri-methylation. Their function in development is still unclear but it has been shown that a Chd2 homozygotic gene trap leads to perinatal lethality (Marfella et al., 2006). There is no data available on Chd1 mutants. There could be some redundancy in functions between these two, but considering their differences in affinity, one interesting hypothesis could be a more complex coordination between the two during several stages of transcription. This could involve a temporal regulation of CHD2 activity through phosphorylation of the longer insert in the chromodomain 2, which would relieve its interference (Flanagan et al., 2007).

The second class includes CHD3 and CHD4, also known as Mi-2a and Mi-2b, which are part of the repressive histone deacetylase complex NuRD. This complex has been well studied, and its developmental role has been dissected in several species. It has been involved in vulval cell fate determination in *C. elegans* (Solari and Ahringer, 2000) T-cell maturation and expression of CD4 in mouse (here as an expression activator), (Williams et al., 2004), regulation of Hox genes in *Drosophila* (Kehle et al., 1998) and embryonic development in plants (Ogas et al., 1999). The NuRD complex has also been shown to be critical for repression of Oct4 during mouse ES cell differentiation (Kaji et al., 2006; Kaji et al., 2007).

The third class includes the less studied members, CHD5-CHD9. Nevertheless, they also seem to play a role in transcriptional regulation and development, some of which involving interaction between members of the CHD family, suggesting different layers of regulation within the remodeling process. One example is the interaction between Chd5 homologue Kismet and dChd1 in *Drosophila* during the transcriptional elongation process (Srinivasan et al., 2005). A recent study shows that CHD7 binding patterns are cell specific and are been tightly connected with H3K4me3. Moreover, CHD7 localization also changes during ES cell differentiation, revealing a binding pattern very similar to the one that will be discussed in this thesis, Chd1 (Schnetz et al., 2009).

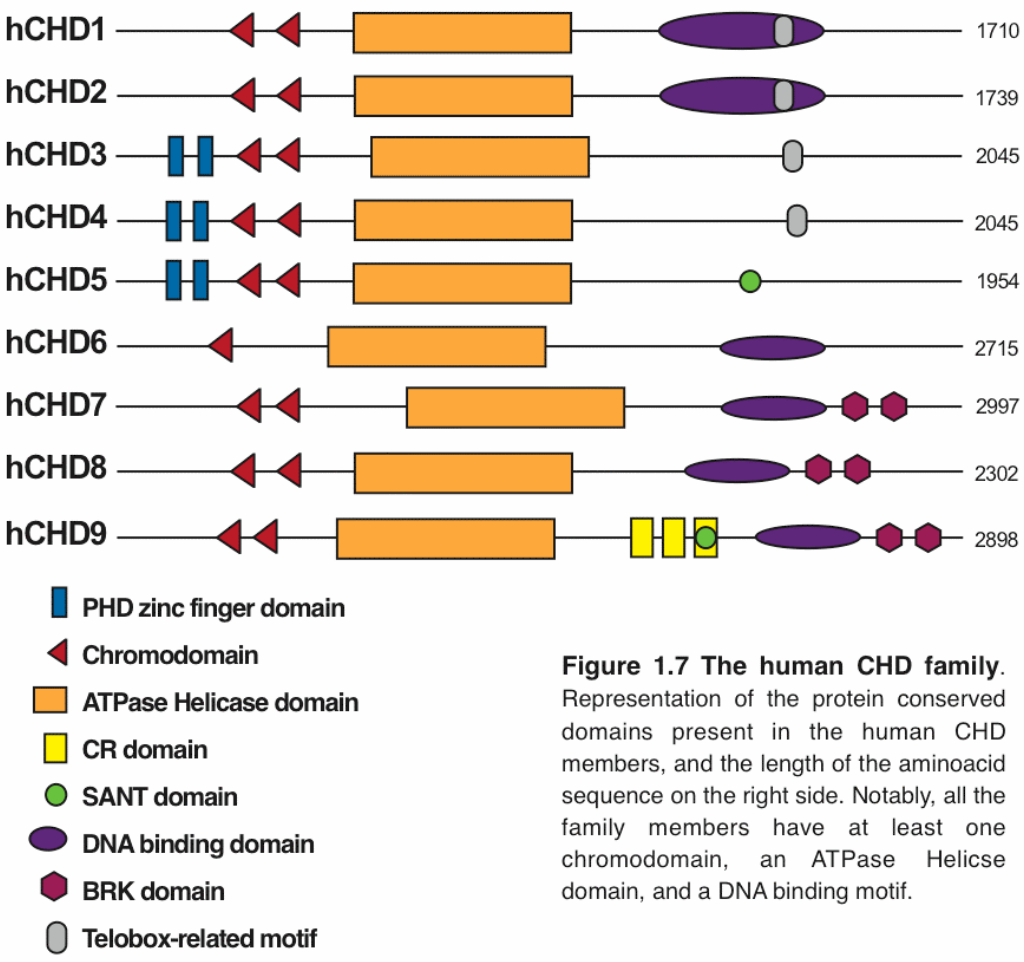


Figure 1.7 The human CHD family. Representation of the protein conserved domains present in the human CHD members, and the length of the aminoacid sequence on the right side. Notably, all the family members have at least one chromodomain, an ATPase Helicse domain, and a DNA binding motif.

Chromodomain helicase DNA binding protein 1

Chd1, the first member of the CHD family, was isolated in mouse (Delmas et al., 1993). The first descriptions of mChd1 acknowledge the presence of the two chromodomains, the helicase binding domain, and a DNA binding domain, without sequence specificity, but with an A+T rich sequence preference (Stokes and Perry, 1995). However, this sequence preference is not the only determinant for CHD1 localization, since CHD1 does not bind to centromeric DNA highly enriched with A+T tracts. In fact, early studies in yeast show a dependence of both C and N terminus for proper localization of ScChd1 (Woodage et al., 1997). Further studies unraveled the use of the two chromodomains in the N terminus as determinant for proper localization of CHD1. First in yeast, it was shown that the chromodomain was critical for the binding to the active mark H3K4me3 (Flanagan et al., 2005). However, in yeast this recognition may require more than one post-translational modification, as it was shown through the connection between the SAGA and SLIK complexes and Chd1, involving histone acetylation (Pray-Grant et al., 2005). Some differences between yeast and human CHD1 were revealed, when it was shown that hCHD1 is able to bind directly to H3K4 di and tri-methylation, using both of its chromodomains (Sims et al., 2005).

The first insights on Chd1 function were gained after finding the Chd1 homologue in fruit fly (dCHD1), where it was shown to be concentrated in the interbands of polytene chromosomes, and co-localized with the puffed bands that usually represent regions of high transcriptional activity (Stokes et al., 1996). This was later confirmed along with the association of the DNA binding protein SSRP1, both of which localized at decompacted interphase chromatin in human HeLa cells, suggesting a common role in transcription activation (Kelley et al., 1999). In yeast, ScCHD1 or Chd1p, has been more thoroughly studied and implicated in transcriptional activation, mostly through an interaction with elongation factors, such as Rtf1, member of the Paf1 complex. This complex is associated with RNA polymerase II and is known to interact with the histone methyltransferase Set1, which is responsible for methylation of the lysine 4 H3 (Simic et al., 2003; Warner et al., 2007). These data suggest that a larger complex involving Chd1 could in fact be involved in the activation or maintenance of the active mark H3K4me3 during the process of transcription. The same study shows, through yeast two hybrid experiments, other components of the elongation machinery interacting with Chd1, such as Spt4-Spt5 and the yeast SSRP1 homologue Pob3 (Simic et al., 2003). Moreover, Chd1 binding is dependent on Cdk9 activity, a kinase responsible for the activation of a RNA polymerase II large subunit. After recruitment to specific promoters, RNA polymerase II undergoes several modifications to allow for the transition from transcription initiation to the elongation phase of RNA processing. A key regulatory step is the phosphorylation of the C-terminal domain (CTD) that is regulated by Cdk9. Knockdown of Cdk9 in *Drosophila* shows a decrease of both Chd1 binding and H3K4me3 enrichment (Eissenberg et al., 2007). Recent work in human CHD1 has confirmed its role as a positive transcriptional regulator by interacting with transcription elongation factors that associate with RNA polymerase II, in an H3K4me3 dependent manner. Interestingly, Chd1 is also important in splicing, as its ablation reduced splicing efficiency, and it was shown to interact with some components of the U2 snRNP spliceosome (Sims et al., 2007).

There is also some evidence that Chd1, along with Set1 (the K4 methyltransferase) may act as a repressor, in yeast, by counteracting the transcription activator complex yFACT (Biswas et al., 2007; Biswas et al., 2008). Its role as a repressor has also been pointed out through interactions with the repressive complex NCoR, and with the histone deacetylases HDAC1 and HDAC2. It is important to add, however, that ScCHD1 is the only member of the CHD family in yeast, and that may help explain its dual role, probably controlled through a more tight context dependent regulation.

The ATPase-dependent remodeling activity of CHD1 has also been described in the literature. *In vitro* assembly assays performed with naked DNA, core histones and chaperones, show how dCHD1, existing as a monomer, is able to transfer histones from the chaperone NAP1 to the DNA in a regularly spaced fashion, as analyzed after a micrococcal nuclease digestion (Lusser et al., 2005). However, Chd1 is not able to assemble chromatin containing H1, the linker histone that is not part of the core nucleosome. Interestingly, in fission yeast, the CHD homologues Hrp1 and Hrp3 interact with Nap1, and mutation of any of these players leads to increased nucleosome density (more compact chromatin), suggesting that these CHD remodelers can stimulate both assembly and disassembly of nucleosomes (Walfridsson et al., 2007). Using the stress response gene ADH2 in yeast, CHD1 has also been shown to be critical in destabilizing the chromatin structure on the coding region and around the terminus of the ADH2 site, but not important for the chromatin reorganization that occurs upon activation of the gene (Xella et al., 2006). *In vivo*, the remodeling role of Chd1 has recently been revealed in the *Drosophila* oocyte, as a factor necessary for incorporation of the histone variant H3.3 into sperm chromatin, a step necessary for decondensation and development. This work shows how Chd1 mediates the incorporation of H3.3, mediated by the chaperone HIRA, allowing for the proper fusion of the male and the female pronucleus after fertilization (Konev et al., 2007). The H3.3 variant is known to be incorporated in a replication independent manner (Ahmad and Henikoff, 2002), and is localized at transcriptional active sites (Schwartz and Ahmad, 2005).

While the role of Chd1 in mouse or human development and disease has not been described yet, as there is no available mutant, some insights about its role have been revealed through indirect information about the expression and binding patterns. For instance, Chd1 is up-regulated both in mouse (Efroni et al., 2008; Ramalho-Santos et al., 2002) and human ES cells (Skottman et al., 2005). In mouse ES cells, recent transcription factor location data (through genome-wide binding analysis) indicate that the Chd1 gene is bound by Oct4, Sox2, Nanog, Smad1, Zfx and E2f1 (Chen et al., 2008). Being a target of multiple regulators of pluripotency and self-renewal explains the high expression levels and suggests a possible role in the maintenance of this network, as other chromatin remodelers may be key elements in the undifferentiated state (Efroni et al., 2008).

Aims

Analysis of the transcriptional profile of pluripotent stem cells, using microarrays, revealed a group of more than 200 genes that were highly represented when compared to differentiated cells, a group of genes we called pluricluster (Grskovic et al., 2007). We hypothesized that some of these up-regulated genes could be directly involved in maintaining the undifferentiated state. This is based on the fact that among the genes highly up-regulated are genes known to be essential for ES cells, such as Oct4, Nanog and Sox2. To better understand the regulation of the undifferentiated state in mouse ES cells, and to find new regulators and potentially new layers of regulation, I proposed the following specific aims:

1. To implement a versatile system for genetic analysis in pluripotent stem cells, making use of RNA interference to down-regulate candidate genes (including transcription factors, chromatin regulators, oncogenes and unknown genes);

2. To use the system described above to identify the requirements for novel candidate genes in the regulation of pluripotency;

3. To test the ability of known and novel regulators of stem cell pluripotency to reprogram somatic cells towards pluripotency.

Chapter 2.

Chd1 regulates open chromatin and pluripotency of embryonic stem cells

Alexandre Gaspar-Maia, Adi Alajem, Fanny Polesso, Rupa Sridharan, Mike J. Mason, Michael Sachs, Amy Heidersbach, João Ramalho-Santos, Michael T. McManus, Kathrin Plath, Eran Meshorer and Miguel Ramalho-Santos. (2009) Nature *460*, 863-868

doi:10.1038/nature08212

ARTICLES

Chd1 regulates open chromatin and pluripotency of embryonic stem cells

Alexandre Gaspar-Maia[1,2,3], Adi Alajem[4], Fanny Polesso[1,2], Rupa Sridharan[5], Mike J. Mason[5], Amy Heidersbach[2], João Ramalho-Santos[6], Michael T. McManus[2], Kathrin Plath[5], Eran Meshorer[4] & Miguel Ramalho-Santos[1,2]

An open chromatin largely devoid of heterochromatin is a hallmark of stem cells. It remains unknown whether an open chromatin is necessary for the differentiation potential of stem cells, and which molecules are needed to maintain open chromatin. Here we show that the chromatin remodelling factor Chd1 is required to maintain the open chromatin of pluripotent mouse embryonic stem cells. Chd1 is a euchromatin protein that associates with the promoters of active genes, and downregulation of Chd1 leads to accumulation of heterochromatin. Chd1-deficient embryonic stem cells are no longer pluripotent, because they are incapable of giving rise to primitive endoderm and have a high propensity for neural differentiation. Furthermore, Chd1 is required for efficient reprogramming of fibroblasts to the pluripotent stem cell state. Our results indicate that Chd1 is essential for open chromatin and pluripotency of embryonic stem cells, and for somatic cell reprogramming to the pluripotent state.

The genome of eukaryotic cells is organized into accessible euchromatin that is permissive for gene activation, and packaged heterochromatin that is largely silenced. Different cellular states may be defined at least in part by differential allocation of genomic regions to specific chromatin domains[1]. Several types of stem cells in organisms ranging from planarians[2] to mammals[3,4] have been reported to have an open chromatin largely devoid of heterochromatin. This phenomenon has been analysed in greater detail in pluripotent mouse embryonic stem (ES) cells. These cells have an open, 'loose' chromatin with high rates of histone protein exchange, and accumulate regions of more rigid heterochromatin after differentiation[5,6]. An open chromatin correlates with a globally permissive transcriptional state, and has been proposed to contribute to the developmental plasticity, or pluripotency, of ES cells[6]. Although there is a strong correlation between open chromatin and the undifferentiated state of stem cells, it remains unknown whether open chromatin is necessary for stem cell potential. Furthermore, little is known about the molecules that may regulate open chromatin in stem cells. Here we report the identification of the chromatin remodeller Chd1 as an essential regulator of open chromatin and pluripotency of ES cells, and of somatic cell reprogramming to pluripotency.

Chd1 regulates ES cell self-renewal

We have recently characterized the transcriptional profiles of pluripotent stem cells, including ES cells, and the cells in the mouse embryo from which they are derived (ref. 7 and G. Wei, R.-F. Yeh, M. Hebrok and M.R.-S., unpublished observations). These studies led to the identification of chromatin remodellers and transcription factors upregulated in pluripotent cells. To test the role of 41 candidate factors in the regulation of pluripotency, we carried out an RNA interference (RNAi) screen in ES cells (Supplementary Fig. 1). ES cells expressing green fluorescent protein (GFP) under the control of the *Oct4* promoter (Oct4-GiP) were infected with a short hairpin RNA (shRNA)-expressing lentiviral vector, pSicoR-mCherry[8]. Using 1–2 shRNAs per candidate target gene, we identified 18 genes that when downregulated led to defects in expansion of ES cells, and 7 that led to lower activity of the *Oct4* promoter. *Chd1* was the only gene with phenotypes in both assays that had not been previously implicated in ES cells (Supplementary Fig. 1).

Chd1 is a chromatin-remodelling enzyme that belongs to the chromodomain family of proteins and contains an ATPase SNF2-like helicase domain[9]. The two chromodomains in Chd1 are essential for recognition of histone H3 di- or tri-methylated at lysine 4 (H3K4me2/3; ref. 10), and Chd1 has been implicated in transcriptional activation in yeast[11], *Drosophila*[12] and mammalian cells[13]. Recent transcription factor location studies indicate that the *Chd1* gene is bound in mouse ES cells by Oct4 (also known as Pou5f1), Sox2, Nanog, Smad1, Zfx and E2f1, suggesting that it is a target of several regulators of pluripotency and self-renewal[14].

RNAi against *Chd1* in Oct4-GiP ES cells, using two independent shRNAs targeting different regions of the messenger RNA, led to a decrease in the expansion of ES cells and to lower Oct4–GFP levels (Fig. 1a, b and Supplementary Fig. 2a). Control cells were infected with empty pSicoR-mCherry or with pSicoR-mCherry expressing an shRNA targeting GFP (empty and GFP RNAi, respectively), and behaved like uninfected cells (Supplementary Figs 1d and 2a). Downregulation of *Chd1* mRNA after RNAi was confirmed by reverse transcription followed by quantitative PCR (qRT–PCR) (Supplementary Fig. 2b). Endogenous *Oct4* downregulation was confirmed in Chd1-deficient (*Chd1* RNAi) ES cells (Supplementary Fig. 3a). *Oct4* downregulation induced differentiation into the trophectoderm lineage[15] (marked by *Cdx2* and *Eomes*), unlike knockdown of *Chd1*, indicating that the *Chd1* RNAi phenotype is not simply one of trophectoderm differentiation due to the loss of *Oct4* (Supplementary Fig. 3).

Chd1 downregulation decreased clonogenic potential in two independent ES cell lines (Oct4-GiP and E14), but *Chd1* RNAi cells were

[1]Departments of Ob/Gyn and Pathology, Center for Reproductive Sciences and Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143-0525, USA. [2]Diabetes Center, University of California, San Francisco, California 94143-0534, USA. [3]PhD Programme in Biomedicine and Experimental Biology (BEB), Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal. [4]Department of Genetics, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. [5]Department of Biological Chemistry and Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, PO Box 951737, Los Angeles, California 90095, USA. [6]Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal.

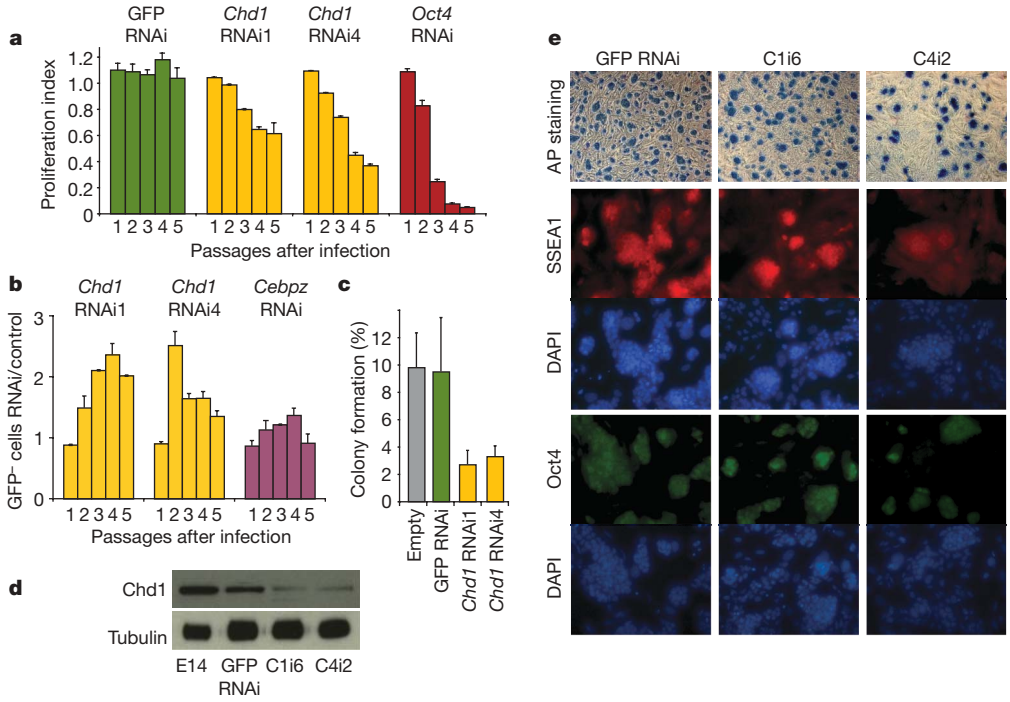


Figure 1 | *Chd1* RNAi ES cells have decreased self-renewal but maintain expression of markers of the undifferentiated state. a, A competition assay shows that downregulation of *Chd1* in Oct4-GiP ES cells using two independent shRNAs (*Chd1* RNAi1 and *Chd1* RNAi4) leads to a decreased proliferative capacity. The proliferation index represents the population of mCherry^{+} cells relative to cells infected with empty vector, at each time point. Experiments were done in triplicate and are represented as mean and s.d. ($n = 3$). **b**, Downregulation of *Chd1* leads to reduced activity of the Oct4–GFP reporter, as measured by the ratio between the percentage of GFP-negative cells in RNAi and in empty vector. *Chd1* RNAi cells have a twofold increase of GFP-negative cells relative to controls for at least one passage. Downregulation of a different gene that affects only proliferation (*Cebpz*) has little effect on Oct4–GFP expression. Data are mean and s.d. ($n = 3$). **c**, The efficiency of colony formation in *Chd1* RNAi ES cells was decreased relative to controls (empty and GFP RNAi). Data are mean and s.d. ($n = 3$), and are representative of two independent experiments. **d**, Immunoblot with whole cell extracts from parental E14 cells, GFP RNAi and two *Chd1* RNAi ES cell clones (C1i6 and C4i2) using antibodies against Chd1 or α-tubulin. **e**, *Chd1* RNAi cells still express markers of undifferentiated ES cells, such as alkaline phosphatase (AP; shown in bright field; original magnification, ×100), and SSEA1 and Oct4 (shown with immunofluorescence; original magnification, ×400). Nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI).

still able to form ES-like colonies (Fig. 1c), unlike *Oct4* RNAi ES cells. ES cell clones constitutively expressing either of the two shRNAs against *Chd1* were established, and sustained *Chd1* downregulation was verified by qRT–PCR (see below Supplementary Fig. 5) and western blot (Fig. 1d). Control lines were established using empty and GFP RNAi viruses. As described later, the two shRNAs targeting *Chd1* led to identical phenotypes in marker gene expression, transcriptional profile, differentiation potential and chromatin state, relative to controls. Results were validated in the two independent ES cell lines Oct4-GiP and E14. The data are from analyses in standard E14 ES cells not expressing GFP. *Chd1* RNAi ES cells, even though they have a self-renewal defect, form compact colonies and express markers of ES cells, such as SSEA1, alkaline phosphatase and Oct4 (Fig. 1e), indicating that they maintain at least some aspects of the undifferentiated state.

Chd1 is required for ES cell pluripotency

To gain insight into the state of *Chd1* RNAi ES cells, we determined their global gene expression profiles using Affymetrix mouse Gene 1.0 ST microarrays (Supplementary Fig. 4). We anticipated that we would find a pattern of downregulated genes in *Chd1* RNAi cells, because Chd1 is known to be associated with active transcription[13]. As expected, both *Chd1* and *Oct4* were found to be downregulated in *Chd1* RNAi ES cells. Surprisingly, however, very few other genes were significantly downregulated (only 25 genes were downregulated more than twofold and none more than threefold at 90% confidence, Supplementary Fig. 4b and Supplementary Data 1). These data indicate that, at least with the low levels of *Chd1* still present in *Chd1* RNAi ES cells, there is a global maintenance of the ES cell transcriptome. On the other hand, a larger group of genes was upregulated in *Chd1* RNAi ES cells (Supplementary Fig. 4b and Supplementary Data 1). A Gene Ontology (GO) analysis of the list of upregulated transcripts showed a significant enrichment for genes involved in neurogenesis (Supplementary Fig. 4c), such as nestin and *Blbp* (also known as *Fabp7*), which was confirmed by qRT–PCR (Supplementary Fig. 5d).

The maintenance of the ES cell transcriptome and the unexpected expression of neural markers were further analysed by immunofluorescence for nestin, Blbp and Oct4. Oct4 was detected in *Chd1* RNAi colonies, but cells between the colonies stained strongly for Blbp (Fig. 2a) and nestin (Supplementary Fig. 4e). No staining for nestin or Blbp was detected in control ES cells. These results were confirmed with fluorescence-activated cell sorting (FACS) using the ES cell marker SSEA1 (Supplementary Fig. 5). In summary, *Chd1* RNAi ES cells can be propagated with many of the hallmarks of the undifferentiated state, but have a high propensity for neuronal differentiation.

We next tested the differentiation potential of *Chd1* RNAi ES cells *in vitro* by the formation of embryoid bodies. *Chd1* RNAi embryoid bodies did not form the typical outer layer of primitive endoderm, as marked by immunofluorescence of embryoid body sections with Afp and Gata4 (Fig. 2b). Similarly, yolk sac endoderm cysts were reduced or not observed in *Chd1* RNAi embryoid bodies (Supplementary Fig. 6a), which showed downregulation of primitive endoderm markers (*Gata4*, *Afp*, *Hnf4a* and *Lamb*) by qRT–PCR (Supplementary Fig. 6b). The loss of primitive endoderm in *Chd1* RNAi embryoid bodies was comparable to that in embryoid bodies lacking an essential regulator of primitive endoderm, *Gata6* (ref. 16) (Fig. 2b). Beating foci, indicative of cardiac mesoderm differentiation, were not detected in *Chd1* RNAi embryoid bodies, whereas they could be readily

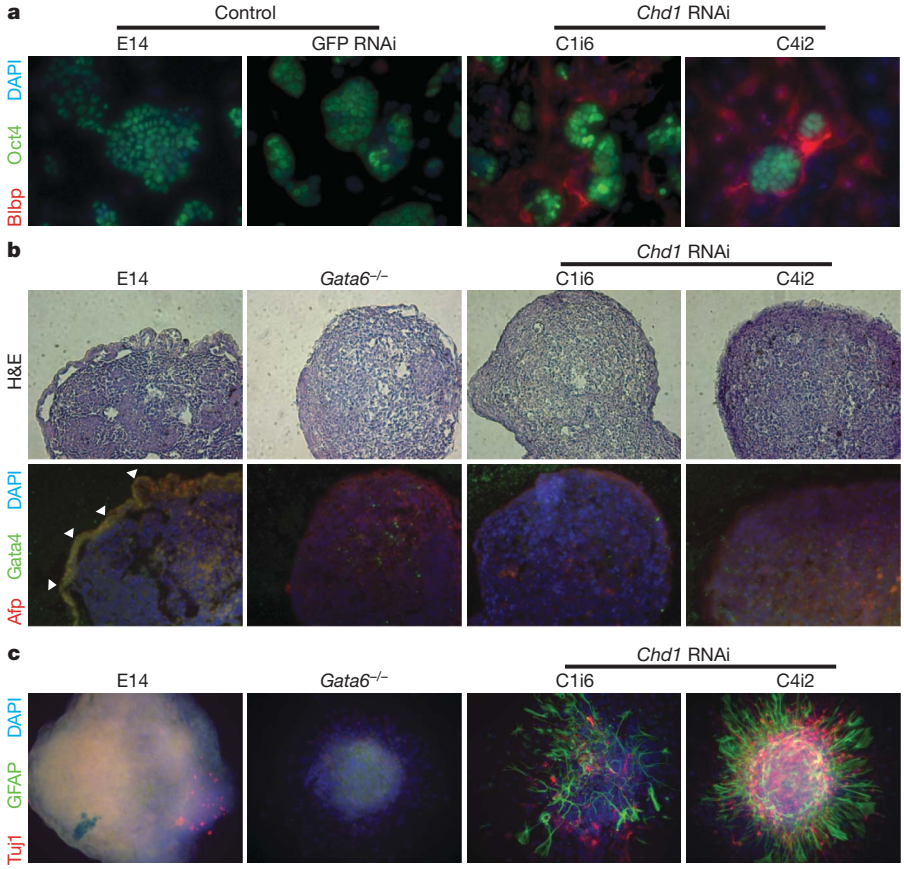


Figure 2 | Chd1 is required for ES cell pluripotency. a, Immunofluorescence analysis of *Chd1* RNAi cells shows expression of Blbp (red), but in a population not expressing the ES cell marker Oct4 (green). Original magnification, ×400. **b**, Paraffin sections of 6 day embryoid bodies were stained for Afp (red) and Gata4 (green), or haematoxylin and eosin (H&E). The loss of the primitive endoderm layer (highlighted in control GFP RNAi by white arrowheads) in *Chd1* RNAi embryoid bodies is similar to that observed in embryoid bodies mutant for *Gata6*. Original magnification, ×100. **c**, A significant increase in neural differentiation is observed, as detected by staining embryoid bodies (plated on matrigel) for astrocytes (GFAP in green) and neurons (Tuj1 in red), in 12 day *Chd1* RNAi embryoid bodies, relative to controls. Nuclei were stained with DAPI. Original magnification, ×200.

quantified in control embryoid bodies (Supplementary Fig. 6c). However, beating foci were also not observed in *Gata6*$^{-/-}$ embryoid bodies, indicating that the loss of cardiac mesoderm differentiation in *Chd1* RNAi embryoid bodies may be secondary to the loss of primitive endoderm. Immunostaining of *Chd1* RNAi embryoid bodies plated on matrigel showed a marked increase of neurons (stained with Tuj1) and astrocytes (stained with Gfap) relative to controls (Fig. 2c). This increase in neural differentiation is not secondary to the loss of primitive endoderm, because *Gata6*$^{-/-}$ embryoid bodies did not show such phenotype (Fig. 2c). Furthermore, *Chd1* RNAi ES cells gave rise to teratomas with abundant neuronal differentiation when compared to wild-type ES cells (Supplementary Fig. 7). These results indicate that downregulation of Chd1 leads to loss of primitive endoderm, with consequential loss of cardiac mesoderm differentiation, and abnormally high levels of neural differentiation that derives from a propensity already detected in the undifferentiated state.

Chd1 is a euchromatin protein in ES cells

We then sought to understand the potential changes in the chromatin state of *Chd1* RNAi ES cells that may underlie their differentiation defects. Previous studies[11–13] indicated that Chd1 associates with euchromatin by binding to H3K4me3, although genome-wide location studies had not been performed. We carried out chromatin immunoprecipitation (ChIP)-chip for Chd1 in wild-type ES cells, and compared the genome-wide location of Chd1 to that of H3K4me3, RNA polymerase II (Pol II) and H3K27me3. These data showed that Chd1 binding strongly correlates with that of Pol II and H3K4me3 (Fig. 3 and Supplementary Fig. 8). Bivalent domains, simultaneously enriched for both the activating H3K4me3 mark and repressive H3K27me3 mark[17], are largely devoid of Chd1 (Fig. 3a and Supplementary Fig. 8). Interestingly, GO analysis indicated that the strongest Chd1 and Pol II targets are enriched for roles in DNA binding, translation and chromatin assembly genes, and that this enrichment is not strictly correlated with expression levels (Fig. 3b and Supplementary Data 2). Chd1 binding also correlates with H3K4me3 enrichment after differentiation: during embryoid body formation, the levels of both H3K4me3 and Chd1 are decreased at the *Oct4* promoter and increased at the endodermal regulator *Gata4* promoter (Supplementary Fig. 9). These data indicate that Chd1 associates globally with euchromatin in ES cells, and may preferentially target genes with roles in chromatin organization and transcription.

Chd1 is required for maintenance of open chromatin

To investigate the effects of Chd1 downregulation on ES cell chromatin, we performed immunofluorescence for histone marks of euchromatin and heterochromatin. Surprisingly, foci of heterochromatin marks such as H3K9me3 and HP1γ (also known as Cbx3), which normally appear as dispersed foci in ES cells[5], were markedly increased in *Chd1* RNAi ES cells (Fig. 4a and Supplementary Fig. 10). No obvious differences were observed in staining for H3K4me3 or H3K27me3 between *Chd1* RNAi ES cells and controls (data not shown). As described earlier, *Chd1* RNAi ES cells are prone to spontaneous neural

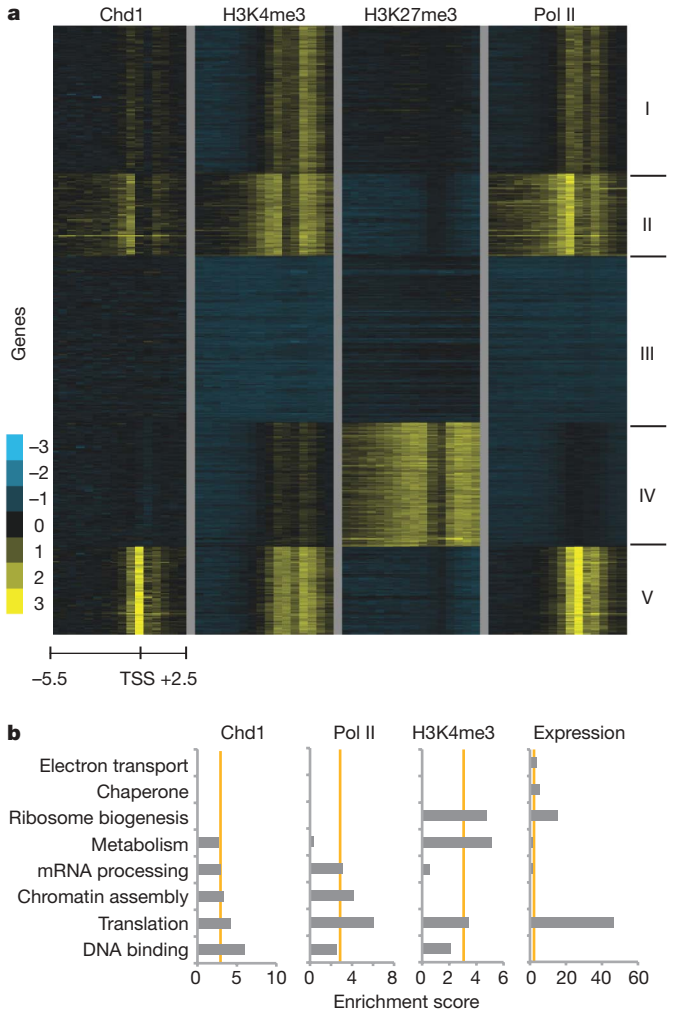


Figure 3 | Chd1 associates with euchromatic promoter regions in ES cells. **a**, K-means clustering of Chd1, H3K4me3, H3K27me3 and RNA Pol II binding in ES cells. Chd1 binding correlates with binding of H3K4me3 and RNA Pol II but is excluded from bivalent domains (cluster IV) in ES cells. Each row represents the binding pattern along the −5.5 kilobase (kb) to +2.5 kb promoter region relative to the transcription start site (TSS). **b**, Functional categorization of Chd1 targets. Gene Ontology (GO) terms associated with the 200 genes most strongly bound by Chd1 or RNA Pol II, or enriched for H3K4me3, as well as the top 200 genes in expression level in ES cells. Categories above an enrichment score of three (*x* axis) are considered significantly enriched. Sixty-nine genes overlap between the Chd1 and the RNA Pol II top 200 gene lists, versus 27 for Chd1 and expression, and 13 for Chd1 and H3K4me3.

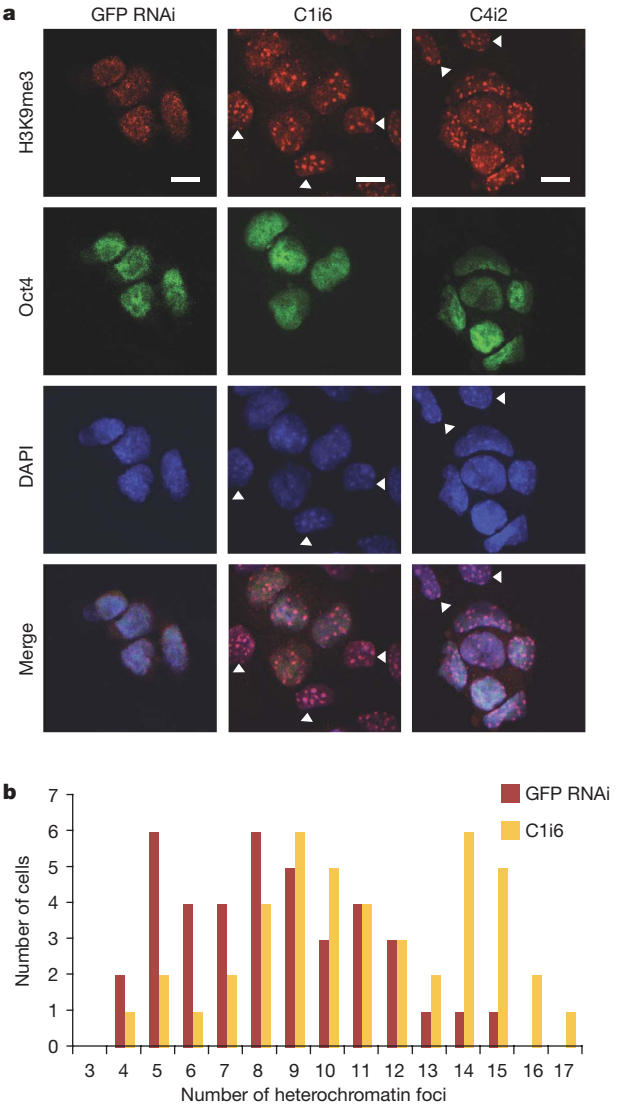


Figure 4 | Chd1 is required to maintain open chromatin in ES cells. **a**, Analysis of H3K9me3 staining by immunofluorescence. Co-staining for H3K9me3 and Oct4 distinguishes between ES-like cells (Oct4-positive) and differentiating cells (Oct4-negative, white arrowheads). Oct4-positive *Chd1* RNAi ES-like cells have increased heterochromatin foci. Scale bar, 10 μm **b**, Quantification of the increase of heterochromatin foci per nucleus in *Chd1* RNAi ES-like cells, as seen by H3K9me3 staining in Oct4-positive cells; $P < 0.0005$.

differentiation, and it has been shown that ES cell-derived neural precursors accumulate heterochromatin[5]. It was therefore important to evaluate whether the accumulation of heterochromatin is a consequence of commitment to the neural lineage, or whether it is present in ES-like cells before differentiation. Co-staining with H3K9me3 and Oct4 revealed that Oct4-positive ES-like cells, located in the centre of compact colonies that stain for other markers of the undifferentiated state (Fig. 1e), have accumulated high levels of heterochromatin in *Chd1* RNAi cells (Fig. 4a, quantified in 4b). Moreover, we analysed the global chromatin dynamics of *Chd1* RNAi cells by fluorescence recovery after photobleaching (FRAP) using a GFP-tagged histone H1. H1 is a linker protein involved in condensing nucleosomes that has been shown to rapidly exchange in the hyperdynamic chromatin of undifferentiated ES cells[5]. H1 showed a significant decrease in recovery in heterochromatin of *Chd1* RNAi ES cells, indicating that the rapid exchange of H1 is compromised (Supplementary Fig. 11). These results indicate that, despite a global maintenance of the transcriptome, morphology, and marker gene expression of ES cells, *Chd1* RNAi ES-like cells are not fully undifferentiated: their chromatin is condensed.

Heterochromatin formation is induced by methylation of H3K9 by the enzymes ESET (also known as Setdb1), Suv39H1/2, G9a (Ehmt2) or Glp (Ehmt1)[18], and reversed by the action of H3K9 demethylases such as Jmjd1a (Kdm3a) and Jmjd2c (Kdm4c). Jmjd1a and Jmjd2c have been shown to regulate genes expressed in ES cells and to repress differentiation[19]. All of these H3K9 methyltransferases and demethylases are expressed in *Chd1* RNAi ES cells at similar levels to control ES cells (Supplementary Data 1). Therefore, the accumulation of heterochromatin in *Chd1* RNAi ES cells is not likely to be due to the differential expression of known H3K9 methyltransferases or demethylases. These results indicate that the capacity to induce heterochromatin formation exists in undifferentiated ES cells, despite the presence of H3K9 demethylases, and that heterochromatinization is countered, directly or indirectly, by Chd1. Our data suggest that ES cells exist in a dynamic state of opposing epigenetic influences of euchromatin and heterochromatin, and that the

euchromatin protein Chd1 is required to maintain the heterochromatin-poor pluripotent stem cell state.

Chd1 is required for efficient induction of pluripotency

Given the role of Chd1 in maintaining pluripotency of ES cells, we hypothesized that it may also be involved in the re-acquisition of pluripotency during somatic cell reprogramming. We analysed the effect of *Chd1* downregulation (Fig. 5a, b) in the reprogramming of Oct4–GFP mouse embryonic fibroblasts (MEFs) to induced pluripotent stem (iPS) cells[20–24]. Downregulation of *Chd1*, using two independent shRNAs in three separate experiments, led to a significant reduction in the number of iPS cell colonies, scored both by morphology and GFP expression. This was not due to a delay in colony formation, as colony counts at later time points showed the same relative reduction in reprogramming efficiency after *Chd1* RNAi treatment (data not shown). The iPS cell colonies that did form in *Chd1* RNAi wells either had not been infected by the RNAi virus or had silenced it, as assessed both by *Chd1* qRT–PCR and mCherry fluorescence (Supplementary Figs 12 and 13). Downregulation of *Chd1* could potentially affect proliferation of MEFs, which would confound the calculation of reprogramming efficiency. However, no significant changes in MEF growth rates were found between control and *Chd1* RNAi (Fig. 5c). Moreover, *Chd1* RNAi does not alter the expression level of the exogenous reprogramming factors (data not shown). In summary, our data show that downregulation of Chd1 does not affect the expansion of fibroblasts but inhibits their reprogramming by induction of pluripotency.

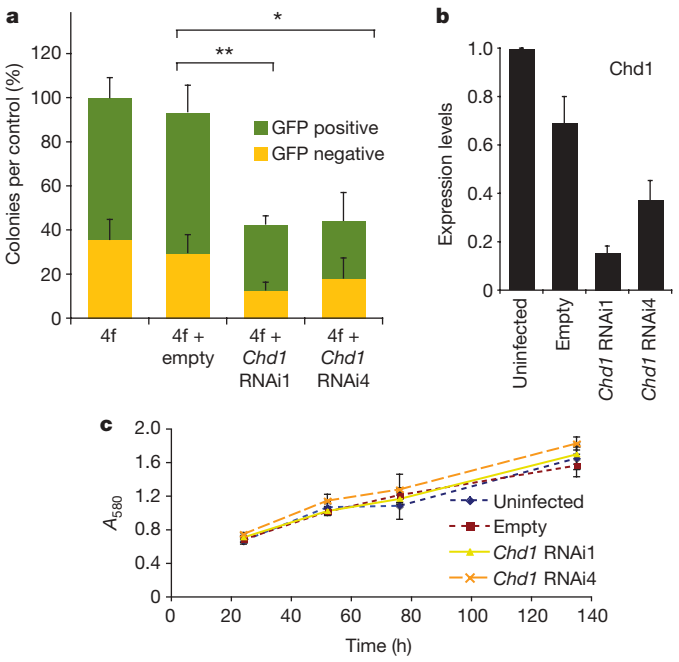


Figure 5 | Chd1 is required for efficient induction of pluripotency. a, The percentage of reprogrammed colonies was scored both by morphology and GFP expression, and normalized to the total number of colonies obtained in the control (four factors only, '4f'). The values are represented as mean and s.d. of the averages of three independent experiments, each one done in duplicate or triplicate. The total number of colonies in control wells (4f or 4f + empty) in the three separate experiments varied between 200 and 500 per 6-well. The efficiency of induction of pluripotency is significantly reduced after *Chd1* RNAi treatment. Unpaired *t*-test was performed using the total number of colonies obtained, comparing the control (empty) with RNAi against *Chd1* (*Chd1* RNAi1 and *Chd1* RNAi4). $*P < 0.0001$, $**P < 0.00001$. **b**, *Chd1* downregulation after RNAi was confirmed by qRT–PCR. The values are represented as mean of absolute expression and s.d. ($n = 3$). **c**, Mean growth rate of MEFs measured by the MTT assay and s.d. ($n = 6$). No differences in MEF growth rate were observed after *Chd1* RNAi treatment.

Discussion

We show here that Chd1 is required for open chromatin and pluripotency of ES cells. ES cells have been reported to be 'poised' for differentiation by the presence of bivalent domains (marked by H3K4me3 and H3K27me3) in developmental regulatory genes[17]. We speculate that the opposing influences of euchromatin and heterochromatin (marked by H3K9me3) may be a further mechanism for maintaining ES cells in a state poised for differentiation. Chd1 is also highly expressed in human ES cells relative to differentiated cells (Supplementary Fig. 14)[25], suggesting that its role in pluripotent stem cells may be conserved. Interestingly, it is possible that other stem cells maintain their differentiation potential using a similar mechanism, because *Chd1* has also been identified as a gene upregulated in adult haematopoietic and neural stem/progenitor cells[26]. Furthermore, our data show that Chd1 is required for efficient generation of iPS cells. Fibroblasts have much higher levels of heterochromatin than pluripotent stem cells[5] (and data not shown), and therefore a global opening of the chromatin is expected to be a component of reprogramming. Our data suggest that Chd1 may contribute to opening the chromatin and enabling transcription-factor-mediated reprogramming to occur, although the precise mechanisms remain to be determined.

It is unclear how Chd1, a protein associated with euchromatin, acts to counter heterochromatinization. Recent genetic studies in yeast indicate that euchromatin-associated factors prevent spreading of heterochromatin to euchromatic regions[27,28]. Further studies will be required to determine whether Chd1 has a similar role in ES cells. An intriguing possibility is that Chd1 may mediate incorporation of the histone variant H3.3, which is generally associated with active genes and is less prone to H3K9 methylation[29]. In support of this model, Chd1 has recently been shown to be required in the *Drosophila* oocyte for incorporation of H3.3 into sperm chromatin, a step necessary for development[30]. In addition, a genome-wide analysis of the chromatin state of *Chd1* RNAi ES cells may provide insight into the differential sensitivity of endoderm and neural lineages to chromatin condensation.

METHODS SUMMARY

The RNAi screen was performed in Oct4-GiP ES cells using the lentiviral vector pSicoR-mCherry. A proliferation index was derived from a competition assay between infected and non-infected cells analysed by flow cytometry over five passages. Oct4-GiP and E14 *Chd1* RNAi and control ES cell clones were expanded from mCherry$^+$ cells isolated by FACS, and validated by qRT–PCR for *Chd1* downregulation. ES clones and embryoid bodies were grown in standard culture conditions. For expression microarray experiments, RNA from E14 *Chd1* RNAi and control clones was amplified and hybridized to Affymetrix Mouse Gene 1.0 ST arrays. Normalized data was analysed in dChip. Gene Ontology analyses were done with MAPPFinder or DAVID. For ChIP-chip, immunoprecipitated DNA was amplified and hybridized to Agilent G4490 promoter arrays. Data were extracted as previously described[31] and visualized using Cluster 3.0 and Treeview. ChIP, qRT–PCR and western blotting were performed using standard protocols. Immunofluorescence was performed on matrigel-coated chamber slides, with ES cells or embryoid bodies plated 2 days before fixation, or on paraffin-sectioned embryoid bodies. Teratomas were generated by subcutaneous injection of ES cells into SCID mice. FRAP analysis[5] and induction of pluripotency[24] were performed as described previously. Cell proliferation was measured using the MTT assay.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 April; accepted 18 June 2009.
Published online 8 July 2009.

1. Mikkelsen, T. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
2. Reddien, P. W. & Sanchez-Alvarado, A. Fundamentals of planarian regeneration. *Annu. Rev. Cell Dev. Biol.* **20**, 725–757 (2004).
3. Terstappen, L. W. *et al.* Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34$^+$CD38$^-$ progenitor cells. *Blood* **77**, 1218–1227 (1991).
4. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).

5. Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10,** 105–116 (2006).
6. Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2,** 437–447 (2008).
7. Grskovic, M. *et al.* Systematic identification of *cis*-regulatory sequences active in mouse and human embryonic stem cells. *PLoS Genet.* **3,** e145 (2007).
8. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101,** 10380–10385 (2004).
9. Woodage, T. *et al.* Characterization of the CHD family of proteins. *Proc. Natl Acad. Sci. USA* **94,** 11472–11477 (1997).
10. Sims, R. J. *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280,** 41789–41792 (2005).
11. Simic, R. *et al.* Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J.* **22,** 1846–1856 (2003).
12. Stokes, D. G., Tartof, K. D. & Perry, R. P. CHD1 is concentrated in interbands and puffed regions of *Drosophila* polytene chromosomes. *Proc. Natl Acad. Sci. USA* **93,** 7137–7142 (1996).
13. Sims, R. J. *et al.* Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell* **28,** 665–676 (2007).
14. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
15. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24,** 372–376 (2000).
16. Morrisey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* **12,** 3579–3590 (1998).
17. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
18. Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14,** 286–298 (2002).
19. Loh, Y.-H. *et al.* Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21,** 2545–2557 (2007).
20. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
21. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
22. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1,** 55–70 (2007).
23. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
24. Blelloch, R., Venere, M., Yen, J. & Ramalho-Santos, M. Generation of induced pluripotent stem cells in the absence of drug selection. *Cell Stem Cell* **1,** 245–247 (2007).
25. Skottman, H. *et al.* Gene expression signatures of seven individual human embryonic stem cell lines. *Stem Cells* **23,** 1343–1456 (2005).
26. Ramalho-Santos, M. *et al.* ''Stemness'': transcriptional profiling of embryonic and adult stem cells. *Science* **298,** 597–600 (2002).
27. Kimura, A., Umehara, T. & Horikoshi, M. Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. *Nature Genet.* **32,** 370–377 (2002).
28. Venkatasubrahmanyam, S. *et al.* Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. *Proc. Natl Acad. Sci. USA* **104,** 16609–16614 (2007).
29. McKittrick, E., Gafken, P. R., Ahmad, K. & Henikoff, S. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc. Natl Acad. Sci. USA* **6,** 1525–1530 (2004).
30. Konev, A. Y. *et al.* CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin *in vivo*. *Science* **317,** 1087–1090 (2007).
31. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors wish to thank A. Smith for Oct4-GiP ES cells, M. S. Parmacek for *Gata6*$^{-/-}$ ES cells, R. P. Perry and D. G. Stokes for the Chd1 antibody, M. Bigos and V. Stepps at the Flow Cytometry Core Facility and L. Ta and C. Barker at the Genomics Core Facility of the Gladstone Institutes for expert assistance, C. Chiu for technical assistance, members of the Santos laboratory, in particular M. Grskovic for advice, and A. Kriegstein, D. Melton, A. Alvarez-Buylla, D. Stainier, R. Blelloch, B. Panning, J. Reiter and M. Grskovic for discussions and critical reading of the manuscript. A.G.-M. was the recipient of a predoctoral fellowship from the Foundation for Science and Technology (POCI2010/FSE), Portugal. R.S. was the recipient of a CIRM postdoctoral training grant. This work was supported by grants from the Sandler Family to M.T.M., CIRM Young Investigator Award and NIH Director's New Innovator Award to K.P., Israel Science Foundation (ISF 215/07), European Union (IRG-206872) and Alon Fellowship to E.M., and NIH Director's New Innovator Award, California Institute for Regenerative Medicine and Juvenile Diabetes Research Foundation to M.R.-S.

Author Contributions A.G.-M., J.R.-S., M.T.M., K.P., E.M. and M.R.-S. planned the project; A.G.-M., A.A., F.P., R.S., M.J.M. and A.H. performed the experiments; A.G.-M., A.A., F.P., R.S., M.J.M., K.P., E.M. and M.R.-S. analysed the data; and A.G.-M., K.P., E.M. and M.R.-S. wrote the manuscript.

Author Information Expression and ChIP-chip microarray data are deposited in the Gene Expression Omnibus (GEO) under accession number GSE16462. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.R.-S. (mrsantos@diabetes.ucsf.edu).

METHODS

ES cell culture and differentiation. Mouse E14 and Oct4-GiP ES cells[32] were plated on 0.1% gelatin-coated plates or on a feeder layer of irradiated MEFs, and maintained in DMEM (Invitrogen) supplemented with 15% knockout serum replacement (Invitrogen), 1 mM L-glutamine, 0.1 mM nonessential amino acids, 100 μg ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol and recombinant LIF. Mouse *Gata6*$^{-/-}$ ES cells[16] were grown in identical conditions except that fetal bovine serum was used instead of knockout serum replacement. Embryoid bodies were formed by plating ES cells in non-attachment conditions (suspension culture) with ES cell medium with fetal bovine serum and in the absence of LIF, for 16 days. Contractile foci were counted under an inverted microscope using triplicates of 10-cm dishes per ES cell clone.

RNAi and competition assay. The genes tested in the RNAi screen were the following: *Ap2gamma* (also known as *Tcfap2c*), *Brca1*, *Cebpz*, *Chd1*, *Ddx18*, *Dmrt1*, *Dppa2*, *Dppa3* (*Stella*), *Dppa4*, *Eed*, *Foxd3*, *Hells*, *Mybl2*, *c-Myc*, *Mycbp*, *Nmyc1* (*Mycn*), *Nanog*, *Nfya*, *Nfyb*, *Nr0b1*, *Nr5a2*, *Oct4*, *Pramel4*, *Pramel5*, *Rex1* (*Zfp42*), *Rex2*, *Rbm35a* (*Esrp1*), *Sall4*, *Six4*, *Sox2*, *Suz12*, *Tcfcp2l1*, *Terf1*, *Tex292* (*Cirh1a*), *Utf1*, *Zic3*, A030007L17Rik (*Ggc2*), and Affymetrix MG_U74Av2 probe sets 160906_i_at, 135189_f_at, 97154_f_at and 98524_f_at. shRNA sequences were selected according to published criteria[33]: GFP RNAi, ACAGCCACAACGTCTATAT; *Oct4* RNAi, GAACCTGGCTAAGCTTCCA; *Chd1* RNAi1, ACATTATGATGGAGCTAAA; *Chd1* RNAi4, GTGCTACT ACAACCATTTA. All other sequences are in Supplementary Table 1. Oligonucleotides coding for the shRNAs were designed and cloned into the lentiviral vector pSicoR-mCherry as described[7]. Lentiviruses were produced as described[8]. For transduction, 10^6 ES cells were incubated with virus in 1 ml of ES cell medium (multiplicity of infection 5–10). After 1 h rotating at 37 °C, 2.5–3×10^5 cells were plated per gelatinized well of a 12-well plate. A competition assay[34] was performed by analysing cells that were passaged every two or three days. Flow cytometry was performed on a LSRII and analysed using the Flojo software. Proliferation index was measured, for every passage, by dividing the percentage of mCherry$^+$ (shRNA) with the percentage of mCherry$^+$ (empty virus). Loss of Oct4–GFP activity was measured by dividing the percentage of GFP$^-$ cells (shRNA) with the percentage of GFP$^-$ (empty virus). The calculation of the loss of Oct4/GFP expression was done with total GFP$^-$ cells rather than just GFP$^-$/mCherry$^+$ to account for potential silencing of the mCherry construct after differentiation or non-cell autonomous effects. Proliferation index data are averages of triplicates ($n = 3$) and standard error bars. mCherry$^+$ ES cells were isolated using a FACSDiVa (BD Biosciences) cell sorter.

Colony formation assay and clonal derivation. E14 and Oct4-GiP ES cells were infected with lentiviruses containing shRNAs or empty virus alone, as described above. mCherry$^+$ cells were sorted on day 5 after infection using a FACSDiVa (BD Biosciences) cell sorter. Five thousand cells were plated per 10-cm dish in triplicates. After 10 days in culture, individual clones were picked per each condition (empty virus, GFP RNAi, *Chd1* RNAi1 or RNAi4) and propagated in standard ES cell growth conditions. Plates were stained for alkaline phosphatase using a Vector kit and colonies were counted. Results are averages of triplicates and standard error bars.

Expression microarrays. Uninfected parental E14 ES cells, one clone infected with empty pSicoR-mCherry, one GFP RNAi clone and four *Chd1* RNAi clones, three from *Chd1* RNAi1 (C1i5, C1i6 and C1i9) and one from *Chd1* RNAi4 (C4i2) were grown on gelatin in ES cell culture medium. Total RNA was isolated using the RNeasy kit (Qiagen) with in-column DNase digestion. Three-hundred nanograms of total RNA per sample were amplified and hybridized to Affymetrix Mouse Gene 1.0 ST arrays according to the manufacturer's instructions at the Genomics Core Facility of the Gladstone Institutes. These arrays assay for the expression of about 35,500 transcripts. Data were normalized using robust multi-array normalization. Hierarchical clustering and calculations of differential gene expression were done using dChip (http://www.dchip.org)[35]. The full normalized data are in Supplementary Data 1. The lower bound of the 90% confidence interval of the fold change (LCB) was used as a conservative estimate of the fold change. Five-hundred-and-thirty-one transcripts with LCB > 2 in *Chd1* RNAi relative to controls were analysed in MAPPFinder[36] for enrichment of gene ontology terms. Terms with *P* values adjusted for multiple testing ≤ 0.01 were considered enriched.

Immunohistochemistry. ES cells or embryoid bodies were plated on chamber glass slides pre-coated with matrigel. ES cells were plated on a layer of irradiated MEFs. After 2 days, cells were fixed with 4% paraformaldehyde, permeabilized with PBT (PBS plus 0.1% Triton X-100) and blocked with 2% BSA plus 1% goat or donkey serum in PBT. Slides were immunostained with primary antibody in blocking solution. Alternatively, embryoid bodies in suspension (at day 6) were fixed, paraffin-embedded, sectioned (50 μm), and stained for haematoxylin and eosin or immunostained.

Primary antibodies used: SSEA1 (MC-480, Developmental Studies Hybridoma Bank (DSHB), 1:200), Oct4 (sc5279, Santa Cruz, 1:100; sc9081; Santa Cruz; 1:50), nestin (MAB353, Chemicon, 1:200), BLBP (ab32423, Abcam, 1:200), Afp (sc8977, Santa Cruz. 1:200), Gata4 (sc1237, Santa Cruz, 1:50), Tuj1 (MMS-435P, Covance, 1:250), GFAP (Z0334, Dako, 1:500), H3K4me3 (ab8580, Abcam, 1:200), H3K9me3 (07-449, Upstate, 1:100; ab8898, Abcam, 1:100), H3K27me3 (07-449, Upstate, 1:100), HP1gamma (MAB3450, Chemicon, 1:500). Secondary antibodies: Alexa Fluor 488/594 conjugated secondary antibodies (anti-mouse, anti-rabbit, or anti-goat, 1:500, Molecular Probes). Nuclei were counterstained with DAPI. The MC-480 antibody developed by D. Solter was obtained from the DSHB developed under the auspices of the NICHD and maintained by University of Iowa.

qRT–PCR. RNA was isolated according to the RNeasy kit (Qiagen), and reverse-transcribed using the iScript first strand cDNAsynthesis kit (BioRad). The cDNA reaction was diluted 1:5 in TE buffer and used in Sybr Green real-time PCR reactions (BioRad). Housekeeping genes used were ubiquitin-b and ribosomal protein L7. PCR primer sequences are available on request. Reactions were run in replicates on a MyiQ qPCR machine (BioRad) according to the manufacturer's instructions. Cycle threshold values were imported into the REST software[37] for fold-change calculations, using the housekeeping genes as controls. Values are presented in $\log_2$ scale or in absolute expression levels compared with parental E14 RNA unless otherwise indicated.

SSEA1 cell sorting. ES cells (GFP RNAi control and *Chd1* RNAi clones) were collected by trypsinization, washed in ice-cold PBS, first resuspended in staining medium (HBSS, Ca/Mg-free, no phenol red, 2% FBS) with primary mouse antibody anti-SSEA1 (MC-480, DSHB, 1:50) for 30 min on ice, and then in secondary anti-mouse IgM-PE (406507, BioLegends, 1:100) for 30 min on ice. Propidium iodide (P3566; Invitrogen) was added and live SSEA1$^+$ and SSEA1$^-$ cells were isolated using a FACSDiVa (BD Biosciences) cell sorter.

Western blotting. Whole cell extracts were prepared and measured using the Bradford assay (BioRad) for protein content. From whole ES cell extracts 30 μg of protein were resolved on SDS–PAGE gel (10%) using a rabbit antibody against Chd1 (1:2,000, from R. Perry[12]) and a goat anti-rabbit HRP (1:10,000). The loading control used was α-tubulin, detected with a mouse antibody (1:1,000, Sigma T9026). From embryoid body extracts (at day 12), 10 μg of protein were resolved on a 10% SDS–PAGE gel using an antibody against Tuj1 (1:1,000) and anti-goat HRP (1:10,000). Detection was performed using the ECL kit according to manufacturer's instructions (Amersham).

FRAP analysis. Transfection of H1–GFP into ES cells and FRAP analysis were performed as described[5].

Generation of teratomas. Teratomas were produced by injecting 3×10^6 cells subcutaneously in the flanks of SCID mice. Tumour tissue samples developed in 12 weeks and were fixed overnight in 4% paraformaldehyde before paraffin embedding. Sections were stained with haematoxylin and eosin with a standard protocol.

ChIP. ChIP was performed essentially as described by Upstate Biotechnology, with some minor changes described below: chromatin was cross-linked by incubating cells on plates with PBS containing 20 mM dimethylpimelimidate (DMP; Sigma) and 0.25% dimethylsulphoxide (DMSO) for 1 h at room temperature. Cells were re-fixed with 2% paraformaldehyde for another hour at room temperature, scraped and centrifuged at 1,350*g* for 5 min. Pellets were resuspended in SDS lysis buffer and sonicated to obtain fragments of ~200–1,000 base pairs (bp) as verified on a gel. Reactions were centrifuged at 13,000*g* for 10 min and the supernatants were used. Antibodies used (3 μg each): Chd1 (PAB-10569, Orbigen), H3K4me3 (ab8580, Abcam) IgG (ab46540, Abcam). DNA was purified by phenol–chloroform extraction, followed by ethanol precipitation. DNA concentration was determined using a Nanodrop spectrophotometer (NanoDrop Technologies) and 5–10 ng were used in Sybr Green real-time PCR reactions (see earlier) ran in duplicates or triplicates. Primer sequences are available on request. Fold enrichment over input was calculated using the $2^{\Delta Ct}$ method corrected with IgG C_t values. The HoxA3 primer set was used as a control gene because it corresponds to a region known to lack H3K4me3 (ref. 17).

ChIP-chip. ChIP and hybridization onto Agilent promoter microarrays was performed as described[22]. In brief 500 μg of crosslinked ES chromatin was immunoprecipitated with 10 μg of Chd1 antibody (Allele Biotech PAB-10568) or hypophosphorylated RNA polymerase II (8WG16). The eluate was reverse crosslinked, RNase- and proteinase-K-treated and purified. Equal amounts of input and immunoprecipitated samples were amplified using the WGA2 kit (Sigma), labelled with the Bioprime kit (Invitrogen) and hybridized onto Agilent mouse promoter arrays (G4490) according to manufacturer's instructions. Data were extracted as previously described[31]. Data were visualized using the Cluster 3.0 and Treeview programs. Bound genes were determined using the Young laboratory algorithm[38]. H3K4me3 and H3K27me3 data, as well as the algorithm to generate the 500-bp window presentation were previously published[22]. In K-means clustering, each row

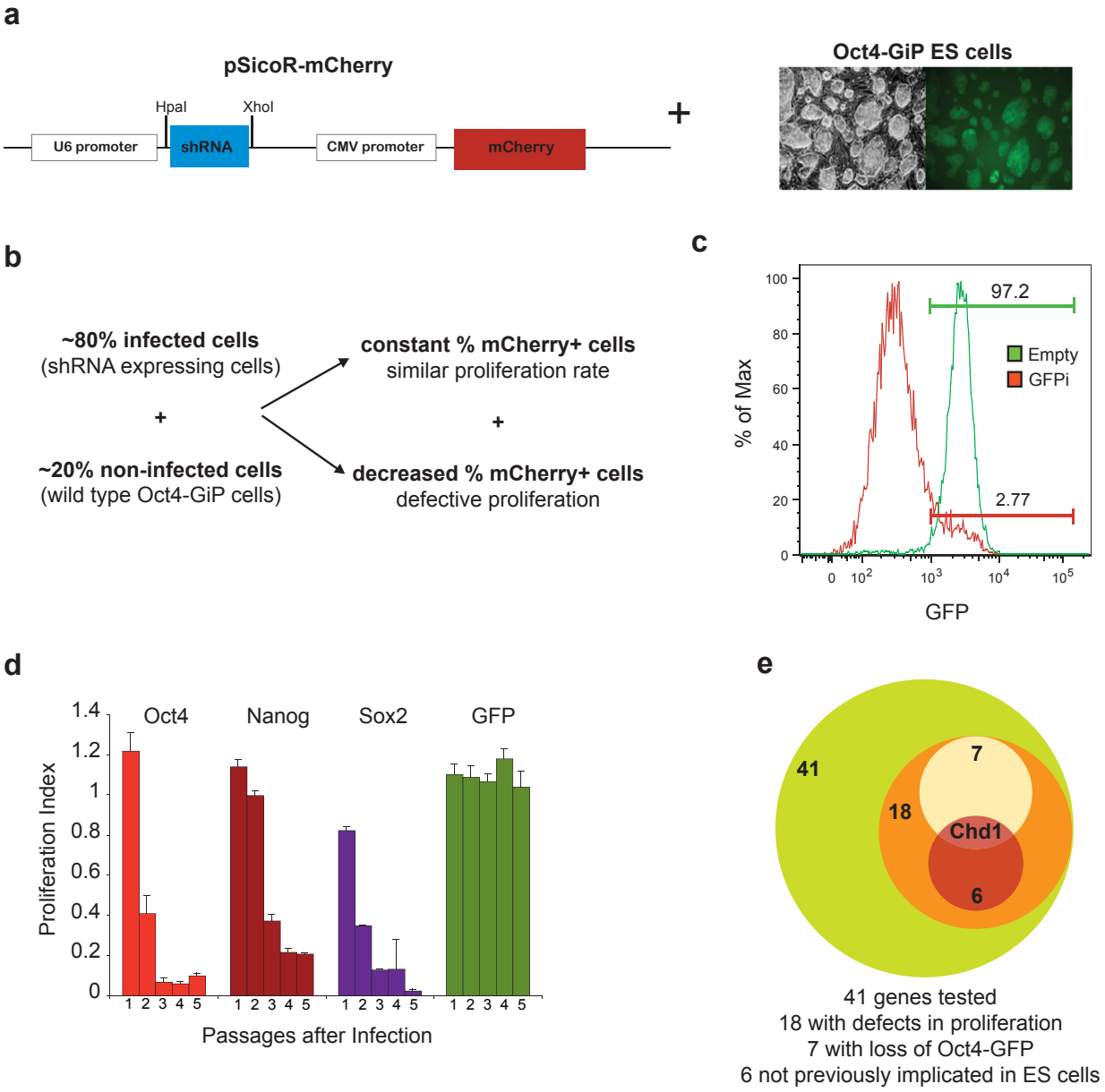
represents the binding pattern along the −5.5 to +2.5 kb promoter region relative to the TSS, reiterated four times to present the data for each immunoprecipitation. The 8 kb promoter region is divided into sixteen 500 bp fragments that display the average log ratio of probe signal intensity with blue, yellow and grey representing lower-than-average, higher-than-average and missing values for enrichment due to lack of probes in those regions, respectively. The odds ratio for binary correlation of Chd1 binding strength was calculated as the ratio of the probability of a gene being bound by Chd1 divided by the probability of it being bound by H3K4me3 (or RNA Pol II or H3K27me3) to the probability of a gene being bound by Chd1 divided by the probability of the gene being unbound by H3K4me3 (or RNA Pol II or H3K27me3).

Reprogramming. Reprogramming was performed as previously described[24], with minor changes: Oct4–GFP MEFs were reprogrammed using lentiviral infection (day 0) of four transcription factors (*Oct4*, *Sox2*, *nMyc* and *Klf4*). RNAi lentiviral vectors (empty, *Chd1* RNAi1, *Chd1* RNAi4) were used to infect MEFs 4 days before the addition of the four factors. At day 1, MEFs were also plated for a MTT assay to quantify growth rates and for RNA collection for qRT–PCR for *Chd1*. An optimization of the protocol was also used and described in Supplementary Fig. 12 (as reported previously[39]). iPS colonies were scored (13 to 16 days after addition of the four factors) by GFP fluorescence, using a scale according to the number of cells in a colony that were GFP positive, as described in Supplementary Fig. 12b (GFP-positive refers to all the colonies with any GFP-positive cells), or by their morphology under bright field (GFP-negative).

MTT assay. The growth rate of MEFs was measured using an indirect method. Yellow MTT is reduced by mitochondrial enzymes into a purple formazan and the absorbance measured as a result of the number of viable cells. For the MTT assay, MEFs were plated at 5,000 cells per well in a 96-well plate and analysed 24, 52, 76 and 135 h after plating. At the indicated time points, ES cell medium (without LIF) was replaced with 100 ml 1 mg ml^{-1} 3-(4,5-dimethylthiazol-2-yl)-2.5-diphenyltetrazolium bromide (MTT) (Molecular Probes) in DMEM. After incubation at 37 °C for 3 h, the MTT solution was removed. One-hundred millilitres of DMSO was added to dissolve precipitate for 10 min at 37 °C and 5 min at room temperature. Absorbance was recorded at 540 nm using a Spectramax M2 microplate reader (Molecular Devices).

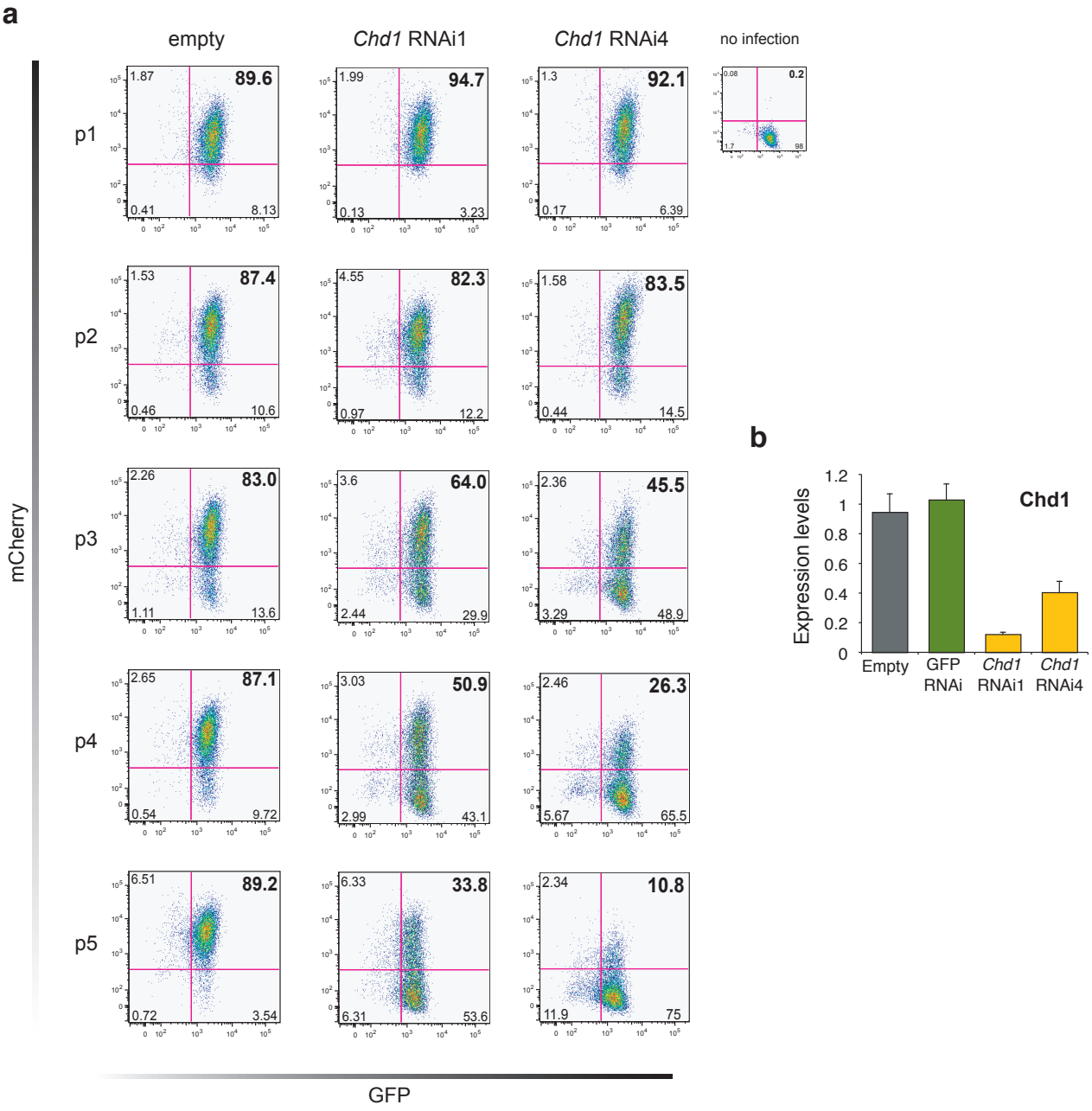
32. Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416,** 545–548 (2002).
33. Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nature Biotechnol.* **22,** 326–330 (2004).
34. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442,** 533–538 (2006).
35. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98,** 31–36 (2001).
36. Doniger, S. W. *et al.* MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.* **4,** R7 (2003).
37. Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30,** e36 (2002).
38. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
39. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnol.* **26,** 916–924 (2008).

SUPPLEMENTARY INFORMATION



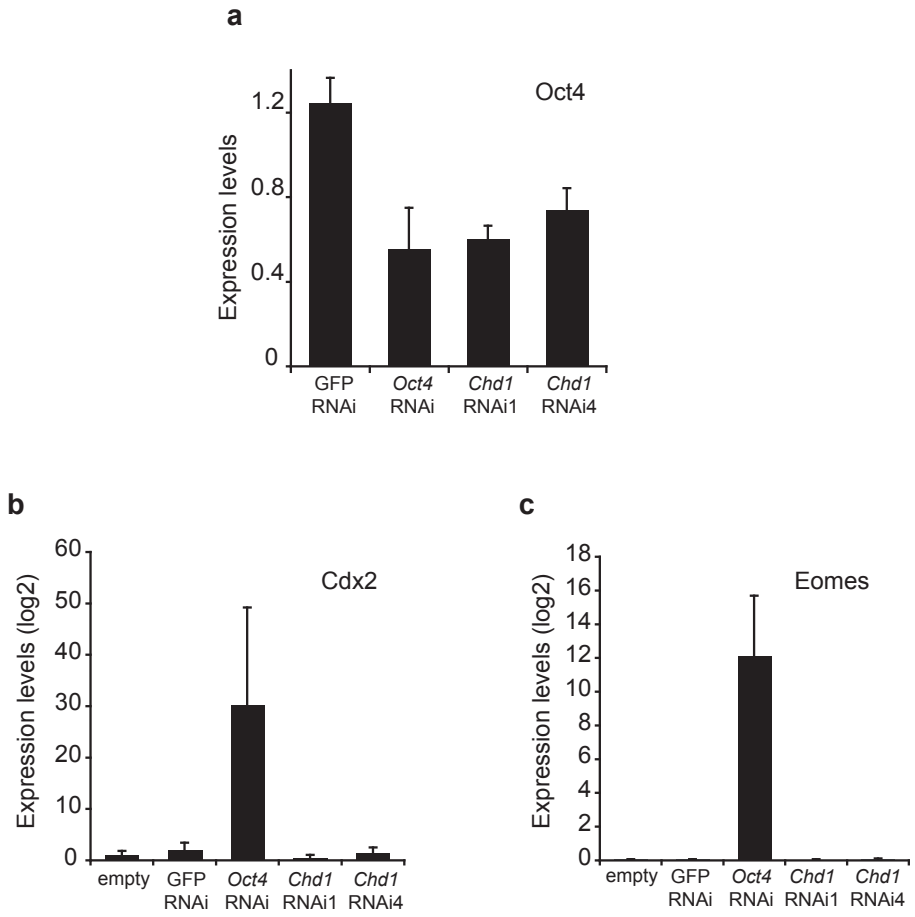
Supplementary Figure 1

Supplementary Figure 1. RNAi screen in mouse ES cells. a, A loss-of-function screen for 41 candidate genes in ES cells. The lentiviral vector pSicoR-mCherry used transcribes a shRNA (under the control of human *U6* promoter) and a reporter red fluorescent protein, mCherry; the ES cell line used expresses GFP under the control of the *Oct4* promoter (*Oct4*-GiP). **b,** A competition assay was performed between infected cells (mCherry+, undergoing RNAi for the gene of interest) and non-infected cells. The proportion of mCherry+ cells was measured by FACS analysis for 5 passages. **c,** FACS plot showing the percentage of GFP+ cells in Oct4-GiP ES cells infected with either the empty vector or with shRNA against GFP. **d,** Knockdown of the three major transcription factors known to regulate ES cells (*Oct4, Sox2* and *Nanog*) leads to decreased proliferation. The proliferation index was measured by comparing the percentage of infected cells in each passage with cells infected with a control virus. Down-regulation of GFP did not affect proliferation of ES cells. The values are represented as mean ± s.d. (n=3) **e,** Summary of the RNAi screen highlighting *Chd1* as a novel regulator of ES cells. 7 genes showed both loss of *Oct4*-GFP activity and proliferation defects (*Chd1, FoxD3, Nanog, Oct4, Sall4, Sox2, Utf1*). 6 of the genes that showed proliferation defects had not previously been implicated in ES cells (*Chd1, Cebpz, Ddx18, NFYa, NFYb, Terf1*), and were confirmed by two independent shRNAs. A detailed description of the RNAi phenotype of *NFYa* and *NFYb* have recently been described.[7]



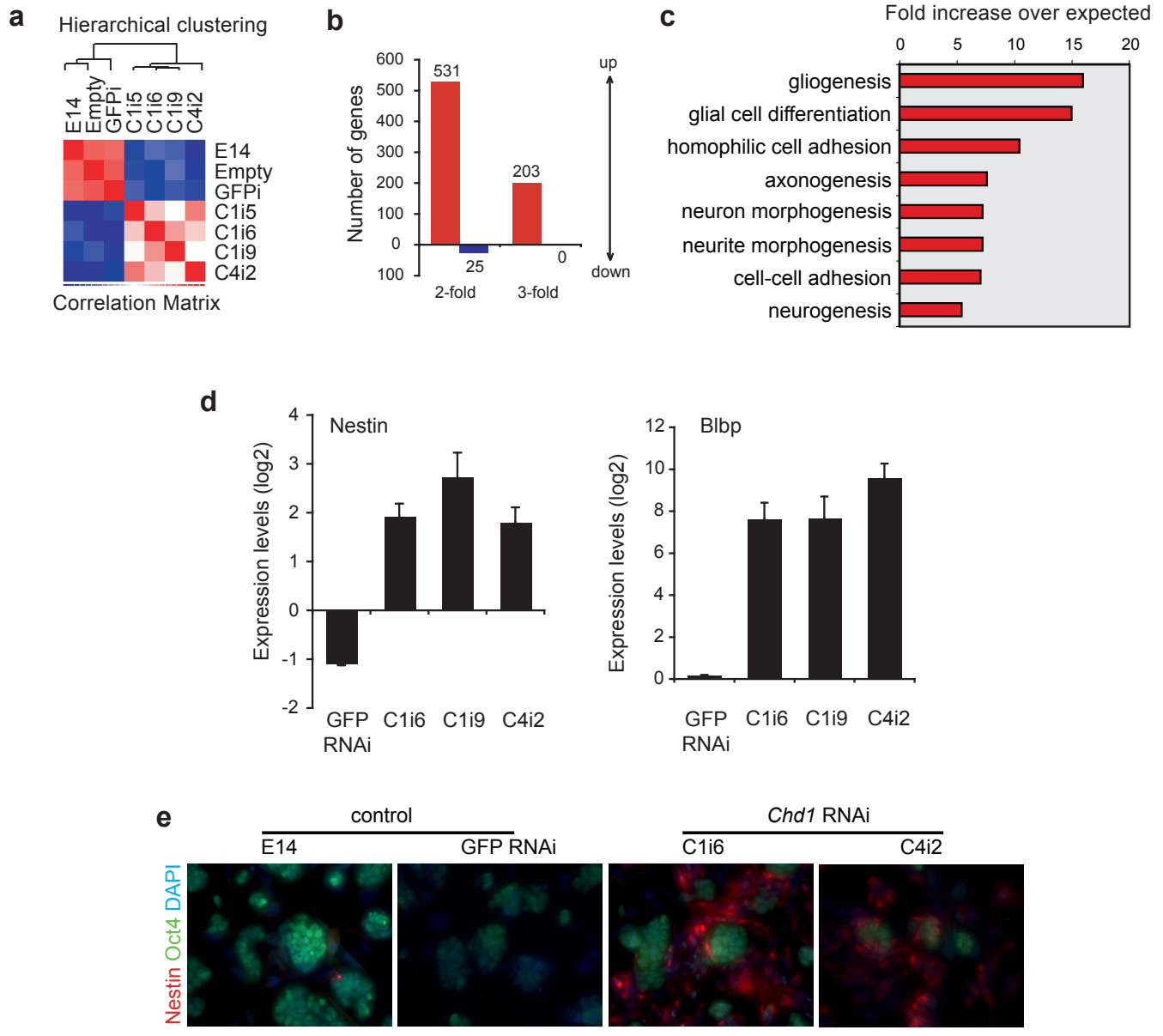
Supplementary Figure 2

Supplementary Figure 2. *Chd1* RNAi in mouse ES cells. a, FACS plots showing mCherry and GFP fluorescence of Oct4-GiP ES cells infected with pSicoR-mCherry with shRNAs against *Chd1* or empty vector, over 5 passages. Highlighted on the top right side of each plot is the percentage of mCherry+/GFP+ cells. **b,** *Chd1* down-regulation upon RNAi was confirmed by qRT-PCR on cells isolated by FACS for mCherry+. The values are represented as mean of absolute expression ± s.d. (n=3).



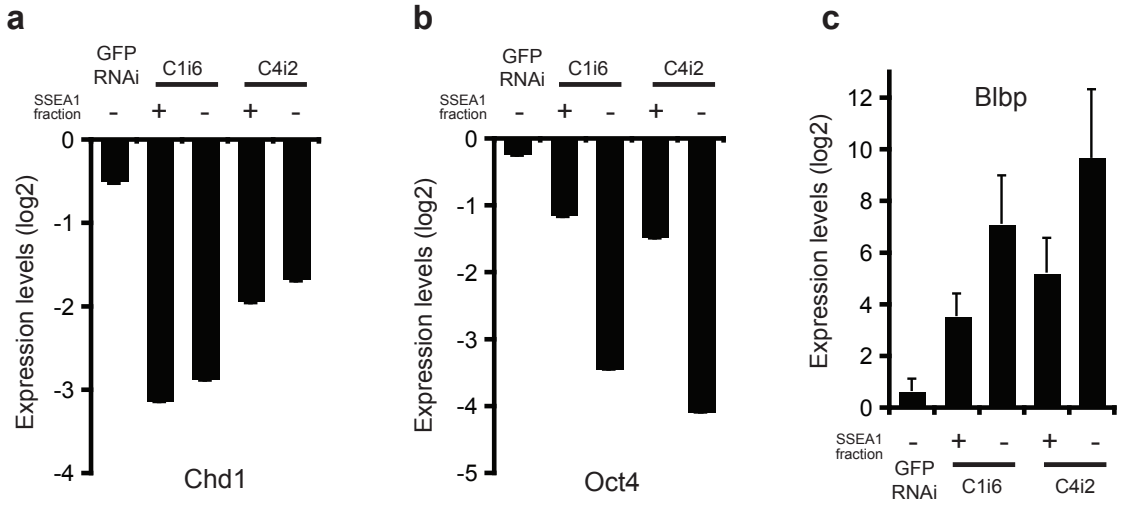
Supplementary Figure 3

Supplementary Figure 3. Knockdown of *Chd1* leads to a reduction in the level of *Oct4* but does not induce expression of trophectoderm markers. a, Expression levels of *Oct4* were analyzed by qRT-PCR in infected E14 cells (isolated by FACS for mCherry) on passage 2. *Chd1* RNAi, using either shRNA, leads to a reduction in the levels of *Oct4* mRNA to about half, similar to that observed in *Oct4* RNAi. The values are represented as mean ± s.d. (n=3). **b, c,** *Oct4* RNAi, but not *Chd1* RNAi, leads to expression of the trophectoderm markers *Cdx2* **(b)** and *Eomes* **(c)**, measured by qRT-PCR. The values are represented as mean ± s.d. (n=3).



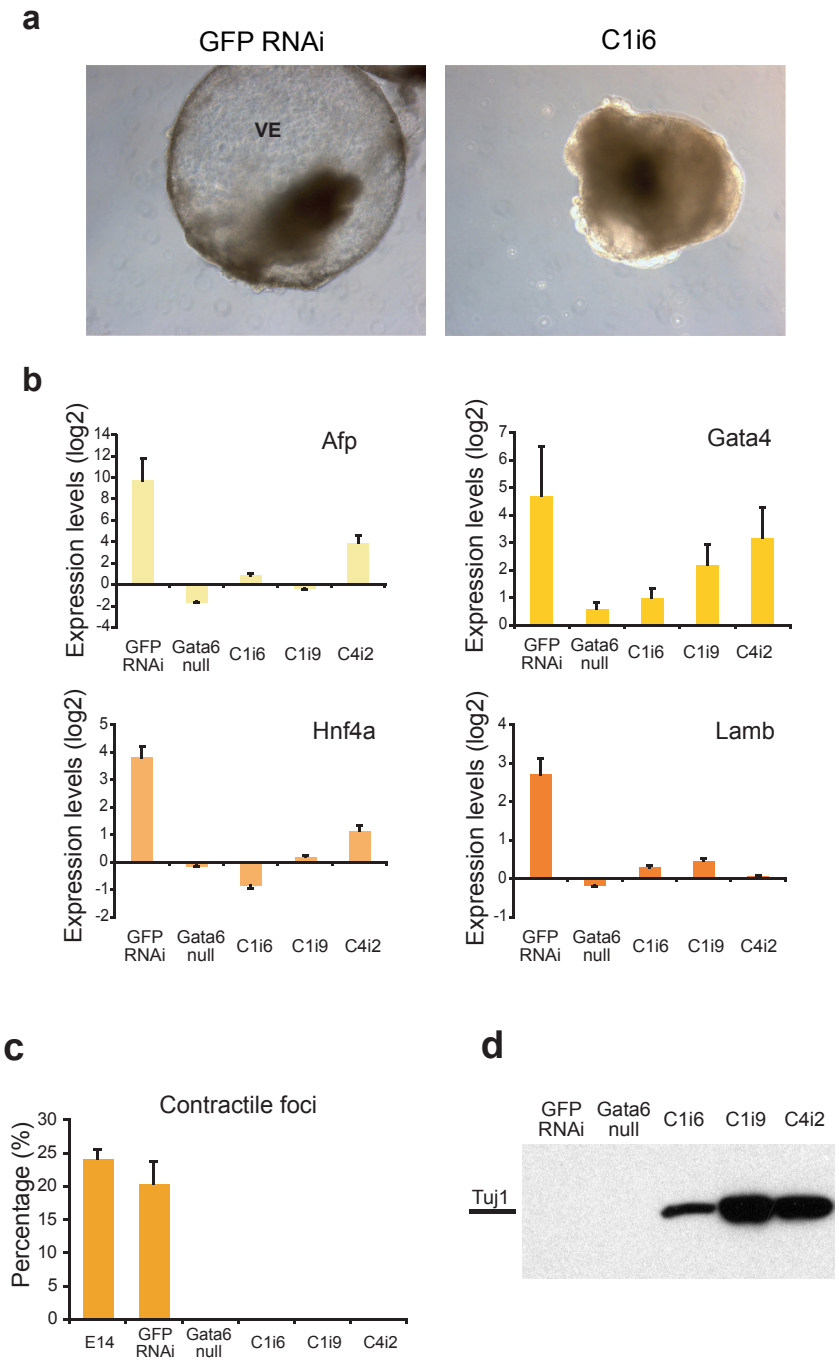
Supplementary Figure 4

Supplementary Figure 4. Propensity for neural differentiation in Chd1i cells . a, Microarray analysis of 4 Chd1i clones, three from Chd1 shRNAi1 (C1i5, C1i6 and C1i9) and one from Chd1 shRNAi4 (C4i2), and 3 control cell lines, E14, empty virus-infected, and GFP RNAi. Hierarchical clustering revealed that the transcriptional profiles of the four Chd1i ES cell lines cluster together and separately from the controls. **b,** Few genes are down-regulated in Chd1i ES cells relative to controls, but a significantly larger subset of genes are up-regulated. **c,** The subset of up-regulated transcripts is enriched for genes with roles in neurogenesis, as determined by Gene Ontology term analysis. All terms shown have p-values adjusted for multiple testing ≤0.01. **d,** Up-regulation of neural markers in *Chd1* RNAi cells, first detected with expression profiling using microarrays, was confirmed with qRT-PCR for *Nestin* and *Blbp*. The values are represented as mean ± s.d. (n=3). **e,** Immunofluorescence analysis of *Chd1* RNAi cells shows expression of Nestin (red), but in a population not expressing the ES cell marker Oct4 (green). Nuclei were stained with DAPI (blue). Original magnification, x400.



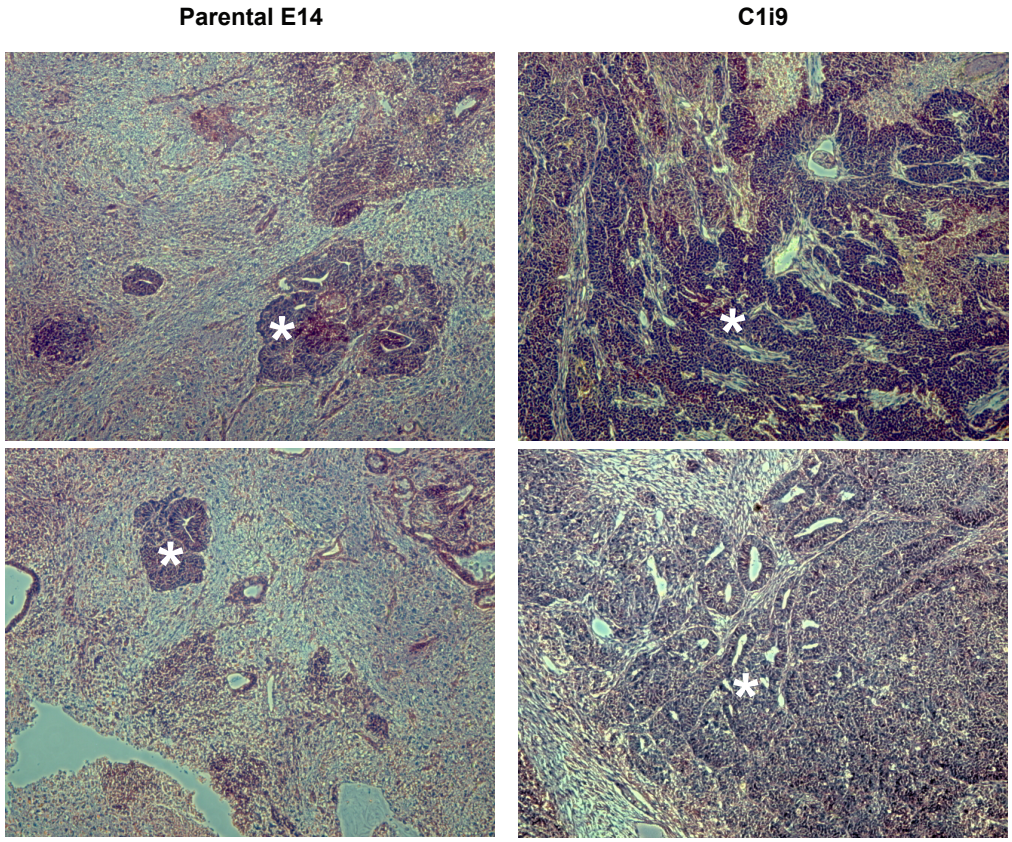
Supplementary Figure 5

Supplementary Figure 5. Analysis of *Chd1* RNAi cells differentiation using the ES cell surface marker SSEA1. Live cells (PI negative) of *Chd1* RNAi clones and control GFP RNAi were sorted using the surface marker SSEA1. RNA from both SSEA1+ and SSEA1- fractions was collected and analyzed by qRT-PCR. **a,** Down-regulation of *Chd1* was confirmed in SSEA1+ and SSEA1- cells. **b,** SSEA1- *Chd1* RNAi cells have consistently lower levels of *Oct4* when compared to SSEA1+ *Chd1* RNAi cells. **c,** *Blbp* is already detected in *Chd1* RNAi SSEA1+ cells and is highly induced in SSEA1- *Chd1* RNAi cells, but not in GFP RNAi cells. The values are represented as mean ± s.d. (n=3), compared with control GFP RNAi SSEA1+ fraction, and are representative of 2 independent FACS experiments.



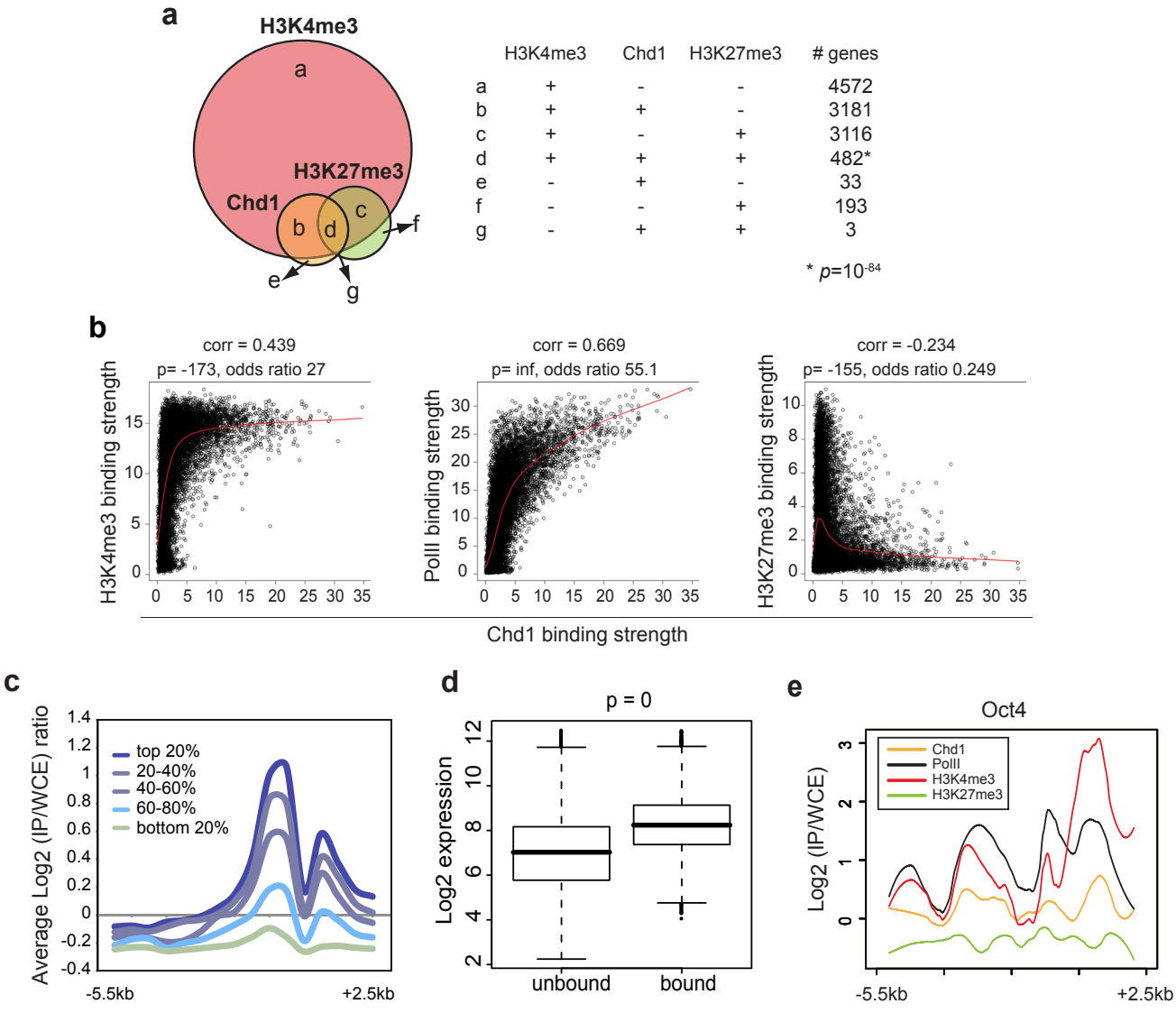
Supplementary Figure 6

Supplementary Figure 6. Chd1i cells have impaired differentiation. a, ES cells were cultured in non-attachment conditions and without LIF to form embryoid bodies (EBs). Chd1 RNAi EBs lack visceral endoderm cysts (V.E.), as seen in bright field images (10 day EBs). Original magnification, x100. **b,** Expression of primitive endoderm markers was analyzed in EBs at day 6. All three *Chd1* RNAi clones tested showed reduced expression of *Afp, Gata4, Hnf4a* and *Lamb*, similar to *Gata6*-/- cells. The values are represented as mean ± s.d. (n=3). **c,** EBs plated for 12 days were analyzed for foci of contractile cardiac muscle. *Gata6*-/- cells and Chd1-deficient cells did not have any EBs with contractile foci. The values are represented as mean ± s.d. (n=3). **d,** Whole cell extracts of day 6 EBs were collected and equal amounts of total protein were loaded for western blot analysis. The neuronal marker Tuj1 is strongly detected in Chd1 RNAi EBs.



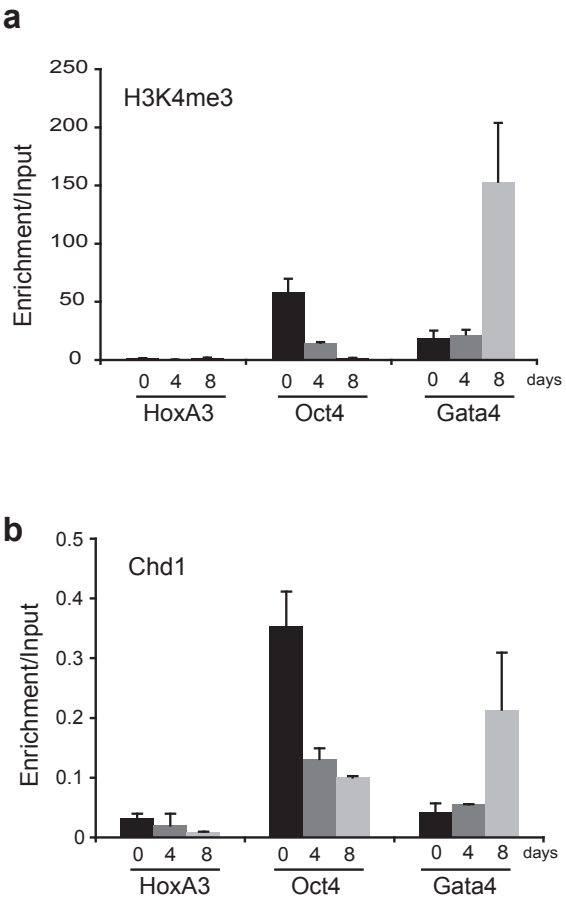
Supplementary Figure 7

Supplementary Figure 7. Teratomas from *Chd1* RNAi ES cells have abundant neural differentiation. Parental E14 cells and Chd1 RNAi ES cells were introduced into immunocompromised SCID mice to form teratomas. Analysis of the teratomas after 12 weeks shows abundant neuroectoderm tissue (white asterisk) as compared to parental ES cells. Original magnification, x100.



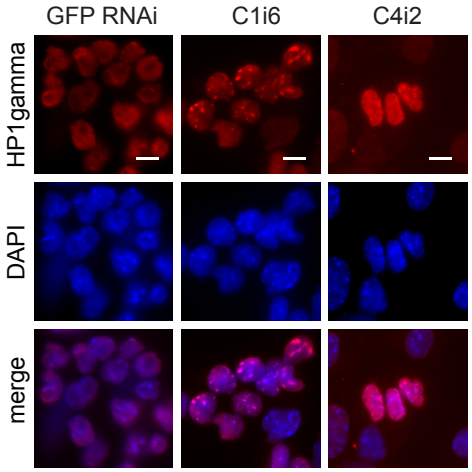
Supplementary Figure 8

Supplementary Figure 8. Genome-wide analysis of Chd1 binding in ES cells using chromatin immunoprecipitation and microarrays. a, Chd1 targets are trimethylated at histone H3 K4. Venn diagram demonstrating the overlap of Chd1, H3K4me3 and H3K27me3 target genes. Numbers of genes for each section of the Venn diagram are given in the table. Target genes were determined using the Young lab algorithm as described in the Materials and Methods section. Only 482 genes with bivalent domains (positive for both H3K4me3 and H3K27me3) are bound by Chd1. The data indicate that Chd1 is significantly under-represented in bivalent domains as determined by the hypergeometric test for the ratio of the probability of Chd1 targets found in bivalent genes to that found in non-bivalent genes. **b,** Genes strongly bound by Chd1 are characterized by high enrichment of H3K4me3 and RNA PolII and lack H3K27me3. Binary correlation of Chd1 binding strength with that of H3K4me3 (left panel), RNA PolII (middle panel) and H3K27me3 (right panel) for each gene (black dot) present on the promoter array. The Pearson correlation value, the –log 10 of the p value of the correlation as determined by Fisher's exact test and the odds ratio are presented above each plot. Lowess normalization was used to generate the smoother indicated by the red line, revealing the anti-correlation of Chd1 binding with H3K27me3 and positive correlation with RNA PolII and H3K4me3. **c,** Chd1 bound targets are highly expressed. All genes were classified into 5 groups depending on their expression as indicated. Average log2(IP/input) ratios for 16 500bp windows along 8kb promoter region for each expression group were plotted. Note enrichment of Chd1 is highest in the 20% most highly expressed genes in ES cells. **d,** Genes that are bound by Chd1 in ES cells have a higher median of expression than those that are not bound. The *p*-value was calculated using the t-test. **e,** Chd1 is enriched immediately adjacent to transcription start site of *Oct4*. The *Oct4* gene is bound by Chd1 (following the Young lab algorithm). The graph plots log2(IP/input) ratios of probes for Chd1, H3K4me3, H3K27me3 and PolII binding or enrichment over the 8kb promoter region of the *Oct4* gene.



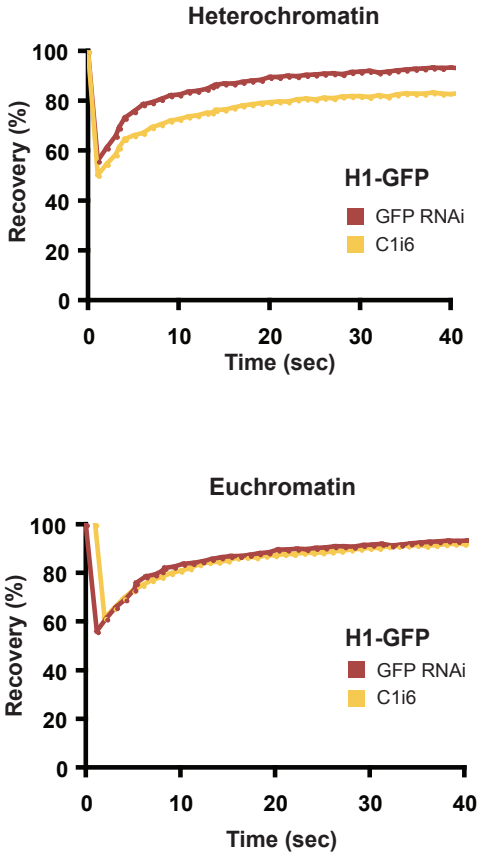
Supplementary Figure 9

Supplementary Figure 9. Chd1 binding coincides with H3K4me3 upon embryoid differentiation. a, b, During embryoid body differentiation of mouse E14 ES cells (days 0, 4, 8) both H3K4me3 enrichment **(a)** and Chd1 binding **(b)** are lost in the promoter region of *Oct4* , and increase at the promoter of *Gata4*. The values are represented as mean ± s.d. (n=3). The promoter of HoxA3 is a control gene known not to be marked with H3K4me3 in ES cells[17].



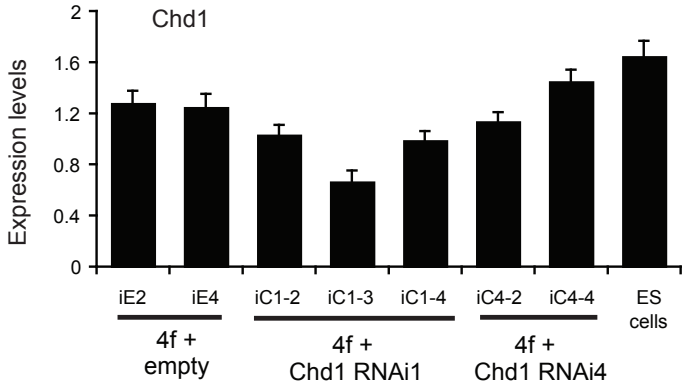
Supplementary Figure 10

Supplementary Figure 10. Heterochromatin foci in *Chd1* RNAi cells. Immunofluorescence of *Chd1* RNAi cells for the heterochromatin mark HP1gamma shows accumulation in heterochromatin foci, whereas localization in the GFP RNAi cells is diffuse throughout the nucleoplasm. Scale bar, 10μm.



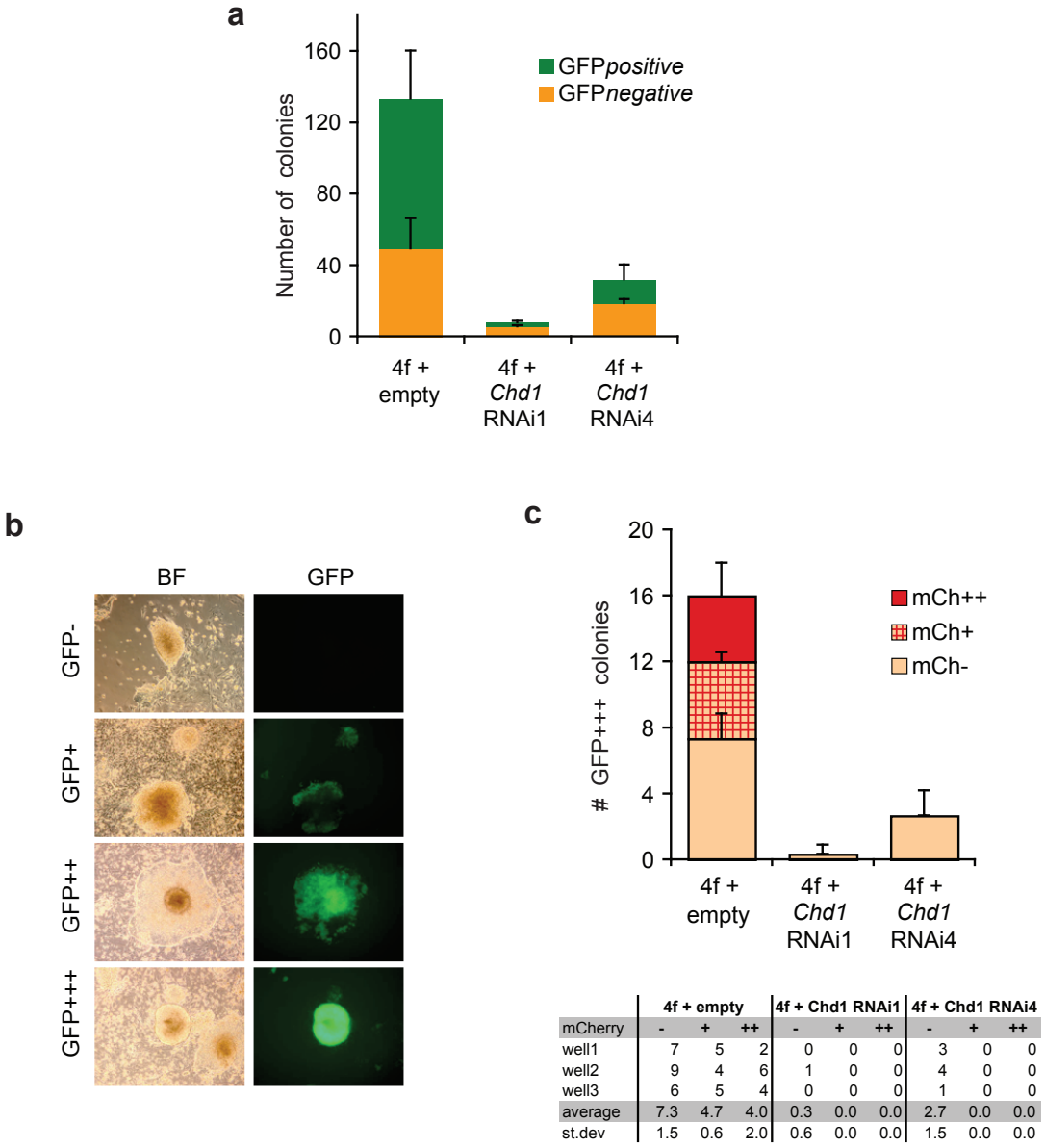
Supplementary Figure 11

Supplementary Figure 11. FRAP analysis of H1-GFP in control GFP RNAi and *Chd1* RNAi ES-like cells. *Chd1* RNAi ES-like cells have a significant delay in H1-GFP protein exchange in heterochromatin but not euchromatin. Values are representative of two independent experiments and the error bars are between 3 and 8%.



Supplementary Figure 12

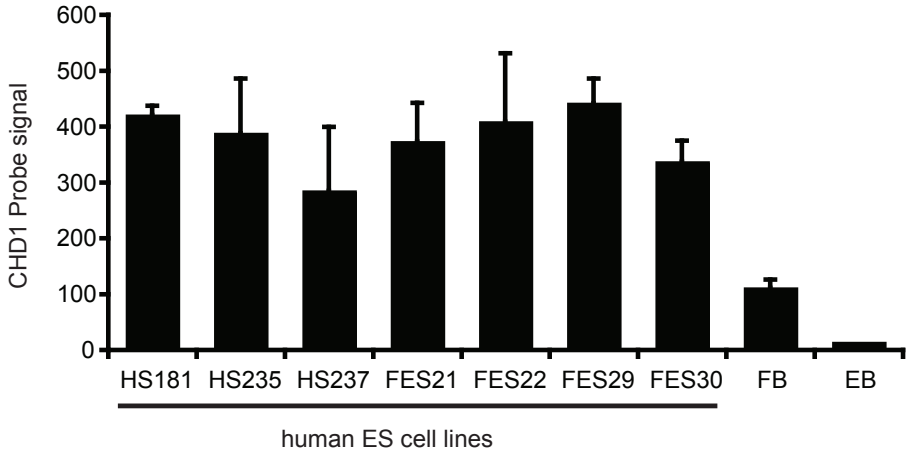
Supplementary Figure 12. *Chd1* levels in reprogrammed iPS cells. Expression of *Chd1* was analyzed in colonies expanded from empty vector- or Chd1RNAi vector-infected cells (along with the four reprogramming factors). Data are shown relative to a 4F+Empty iPS clone (iE1). No signifcant differences in *Chd1* expression were detected between the control and the *Chd1* RNAi iPS lines. The values are represented as mean ± s.d. (n=3).



	4f + empty			4f + Chd1 RNAi1			4f + Chd1 RNAi4		
mCherry	-	+	++	-	+	++	-	+	++
well1	7	5	2	0	0	0	3	0	0
well2	9	4	6	1	0	0	4	0	0
well3	6	5	4	0	0	0	1	0	0
average	7.3	4.7	4.0	0.3	0.0	0.0	2.7	0.0	0.0
st.dev	1.5	0.6	2.0	0.6	0.0	0.0	1.5	0.0	0.0

Supplementary Figure 13

Supplementary Figure 13. Analysis of the expression of the lentiviral vector for RNAi upon reprogramming using the reporter gene mCherry. This assay was performed with a higher infection efficiency[38], in which the RNAi virus is added at lower cell density (40,000 cells per 6well) and at the same time as the four reprogramming factors. **a,** The number of colonies was scored both by morphology and GFP expression. The values are represented as mean ± s.d. (n=3). **b,** Oct4-GFP colonies were scored using 4 categories (GFP-, GFP+, GFP++, GFP+++) depending on the number of GFP positive cells. Original magnification, x200. **c,** Only GFP+++ colonies were scored according to their mCherry expression. These represent colonies that are considered fully reprogrammed. No GFP+++/mCherry+ colonies were found when *Chd1* RNAi vector was added. The values are represented as mean ± s.d. (n=3). Number of colonies are shown in the table below.



Supplementary Figure 14

Supplementary Figure 14. Microarray analysis of *CHD1* expression in several human ES cell lines. Transcriptional profile of seven different independently-derived human ES cell lines show up-regulation of *CHD1* when compared to foreskin fibroblasts (FB) and Embryoid Body-derived differentiated cells (EB)[25].

Chapter 3.

Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells

Marica Grskovic, Christina Chaivorapol, **Alexandre Gaspar-Maia**, Hao Li, Miguel Ramalho-Santos. (2007) Plos Genetics 3(8) pp. e145

Systematic Identification of *cis*-Regulatory Sequences Active in Mouse and Human Embryonic Stem Cells

Marica Grskovic[1,2☯], Christina Chaivorapol[3,4,5☯], Alexandre Gaspar-Maia[1,2,6], Hao Li[3,4,5*], Miguel Ramalho-Santos[1,2*]

1 Institute for Regeneration Medicine, University of California San Francisco, San Francisco, California, United States of America, **2** Diabetes Center, University of California San Francisco, San Francisco, California, United States of America, **3** Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California, United States of America, **4** California Institute for Quantitative Biomedical Research, University of California San Francisco, San Francisco, California, United States of America, **5** Graduate Program in Biological and Medical Informatics; University of California San Francisco, San Francisco, California, United States of America, **6** Doctoral Program in Biomedicine and Experimental Biology, Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal

Understanding the transcriptional regulation of pluripotent cells is of fundamental interest and will greatly inform efforts aimed at directing differentiation of embryonic stem (ES) cells or reprogramming somatic cells. We first analyzed the transcriptional profiles of mouse ES cells and primordial germ cells and identified genes upregulated in pluripotent cells both in vitro and in vivo. These genes are enriched for roles in transcription, chromatin remodeling, cell cycle, and DNA repair. We developed a novel computational algorithm, CompMoby, which combines analyses of sequences both aligned and non-aligned between different genomes with a probabilistic segmentation model to systematically predict short DNA motifs that regulate gene expression. CompMoby was used to identify conserved overrepresented motifs in genes upregulated in pluripotent cells. We show that the motifs are preferentially active in undifferentiated mouse ES and embryonic germ cells in a sequence-specific manner, and that they can act as enhancers in the context of an endogenous promoter. Importantly, the activity of the motifs is conserved in human ES cells. We further show that the transcription factor NF-Y specifically binds to one of the motifs, is differentially expressed during ES cell differentiation, and is required for ES cell proliferation. This study provides novel insights into the transcriptional regulatory networks of pluripotent cells. Our results suggest that this systematic approach can be broadly applied to understanding transcriptional networks in mammalian species.

Citation: Grskovic M, Chaivorapol C, Gaspar-Maia A, Li H, Ramalho-Santos M (2007) Systematic identification of *cis*-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet 3(8): e145. doi:10.1371/journal.pgen.0030145

Introduction

Pluripotent stem cells can give rise to all fetal and adult cell lineages, including the germline. The prototypical pluripotent stem cells are embryonic stem (ES) cells [1,2]. ES cells are a remarkable model for the study of early development and hold promise as a source for cell replacement therapies [3]. To successfully manipulate ES cells in culture, it is important to understand the mechanisms by which ES cells maintain their self renewal and pluripotency.

ES cells are derived from the inner cell mass of the blastocyst, a group of cells that gives rise to all cells of the fetus. After the blastocyst implants in the uterus and gastrulation ensues, most cells of the embryo lose the ability to give rise to pluripotent stem cells, except for primordial germ cells (PGCs) [4,5]. PGCs are the germline precursors that give rise to sperm or eggs. When cultured in vitro, PGCs give rise to embryonic germ (EG) cells, pluripotent stem cells very similar to ES cells [6,7].

Several regulatory pathways that control ES cell pluripotency and self renewal have recently been identified (reviewed in [8]). Factors involved include the leukemia inhibitory factor (LIF) and BMP signaling pathways [9–12], and transcription factors Nanog [13,14] and Oct4 [15,16]. Interestingly, the signaling pathways do not appear to be conserved between mouse and human ES cells [17–20], but the transcriptional regulators Oct4 and Nanog are required in ES cells of both species [21–23]. Recent studies indicate that transcription factors other than Oct4 and Nanog are also important for maintenance of the ES cell state [24,25]. A major goal will be to obtain a complete description of the transcriptional regulatory networks of ES cells.

The increasing availability of whole genome sequences and high-throughput experimental methods, such as microarrays, have led to the development of systematic approaches for deciphering transcriptional regulation. Such analyses generally lead to the identification of sets of genes whose expression is coregulated. It has been shown that genes within a coregulated set often share common *cis*-regulatory motifs, corresponding to transcription factor binding sites, in

Editor: Anne C. Ferguson-Smith, University of Cambridge, afsmith@mole.bio.cam.ac.uk

Received March 26, 2007; **Accepted** July 10, 2007; **Published** August 31, 2007

Abbreviations: ChIP, chromatin immunoprecipitation; dpc, days post coitum; EB, embryoid body; EG, embryonic germ; ES, embryonic stem; LIF, leukemia inhibitory factor; PGC, primordial germ cell; RA, retinoic acid; RNAi, RNA interference; RT-PCR, reverse transcriptase PCR; SGM, somatic cells of the genital ridge/mesonephros area; shRNA, short hairpin RNA; TK, thymidine kinase

* To whom correspondence should be addressed. E-mail: haoli@genome.ucsf.edu (HL); mrsantos@diabetes.ucsf.edu (MRS)

☯ These authors contributed equally to this work.

Author Summary

Embryonic stem cells have two remarkable properties: they can proliferate very rapidly, and they can give rise to all of the body's cell types. Understanding how gene activity is regulated in embryonic stem cells will be an important step towards therapeutic applications. The activity of genes is regulated by proteins called transcription factors, which bind to stretches of DNA sequences that act as on or off switches. We identified genes that are active in mouse embryonic stem cells but not in differentiated cells. We reasoned that if these genes have similar patterns of activity, they may be regulated by the same transcription factors. We therefore developed a computational approach that takes information on gene activity and predicts DNA sequences that may act as switches. Using this approach, we discovered new DNA switches that regulate gene activity in mouse and human embryonic stem cells. Furthermore, we identified a transcription factor that binds to one of these DNA switches and is important for the rapid proliferation of embryonic stem cells. Our approach sheds light on the genetic regulation of embryonic stem cells and will be broadly applicable to questions of how gene activity is regulated in other cell types of interest.

their upstream genomic sequences (for reviews see [26,27]). A number of computational algorithms have been developed to identify such regulatory motifs. These algorithms include enumeration of overrepresented substrings or regular expression patterns, local multiple sequence alignment, or sequence segmentation to decompose the DNA sequence into the most plausible set of motifs [28–32]. The strategy of identifying clusters of coregulated genes by expression profiling followed by a computational search for regulatory motifs has been successfully applied to a number of questions, mostly in lower eukaryotes such as yeast. For mammalian species, the problem is much more challenging [33], as the genomes are more complex and regulation often involves combinatorial action of transcription factors [34]. Examples of computational de novo motif discovery followed by experimental validation in mammalian species are scarce. One experimentally validated case recently reported using motif discovery to identify mouse transcription factors that regulate oxidative phosphorylation [35].

Recent algorithms targeted at higher eukaryotes use interspecies comparisons to identify functional motifs in orthologous promoters [36–38], because functional elements are subject to selective pressure and tend to evolve more slowly than nonfunctional sequences [34]. These algorithms typically use conserved blocks of DNA sequence that can be aligned to reduce the background noise. However, alignment-based approaches can miss important sequence motifs as many regulatory sequences do not fall into conserved regions [39,40].

In this paper, we used a combination of gene expression profiling with computational genomic analyses and biochemical assays to systematically identify novel *cis*-regulatory sequences that control gene expression in pluripotent stem cells. To gain insight into the transcriptional regulatory networks of pluripotent cells, we compared the gene expression profiles of ES cells and PGCs to embryonic and adult somatic cell types. We identified clusters of genes upregulated in ES cells and PGCs, which include several known markers of pluripotency. To identify regulatory motifs that control gene expression within these clusters, we developed a novel algorithm, CompMoby. This algorithm combines the strategies of comparative genomics with DNA sequence segmentation to identify sets of motifs in the upstream regions of coregulated genes. Using CompMoby, we identified motifs that are statistically overrepresented in genes upregulated in pluripotent cells and highly conserved across multiple mammalian species. We demonstrate that several of the predicted motifs are novel regulatory elements of gene expression in mouse and human ES cells. Finally, we show that the transcription factor NF-Y binds to one of the motifs, is differentially expressed during ES cell differentiation, and is required for ES cell proliferation.

Results

Genes Upregulated in Pluripotent Cells Are Involved in Transcription, Chromatin Remodeling, Cell Cycle, and DNA Repair

The identification of the gene expression profiles of PGCs and neighboring somatic cells of the genital ridge/mesonephros area (SGM) is described elsewhere (Wei et al., submitted). Briefly, PGCs and SGM cells were isolated by fluorescence-activated cell sorting from 11.5-d post coitum (dpc) mouse embryos carrying the Oct4/EGFP transgene. This construct has been shown to drive expression of EGFP specifically in PGCs [41]. We then identified the gene expression profiles of PGCs and SGM cells using Affymetrix U74Av2 microarrays. The raw data can be obtained from ArrayExpress. The complete normalized expression data can be found in Dataset S1. We compared the gene expression profiles of PGCs and SGM cells to those of embryonic and adult stem cells, and adult differentiated tissues [42]. Hierarchical clustering revealed similarities at the gene expression level between PGCs and ES cells (Figure 1A). Furthermore, the transcription profile of ES cells is more similar to PGCs than to that of adult stem cells.

This result suggests that aspects of the transcriptional regulation of pluripotency of ES cells are maintained in PGCs during embryogenesis. We therefore sought to identify clusters of genes upregulated in ES cells and PGCs, but not in other cell types. Figure 1A depicts a composite cluster of 230 probe sets upregulated (in red) in pluripotent cells, and downregulated or not expressed (in blue) in adult stem cells and differentiated cells (Figure 1A and Dataset S2). These genes are also largely downregulated upon differentiation of ES cells, a further indication that their expression correlates with the pluripotent state (H. Chipperfield, S. Zhong, D. Melton, and W. Wong, personal communication). This cluster includes several known markers of pluripotency (see below).

We used Onto-Express [43] to search the Gene Ontology database for functional categories overrepresented in the cluster of genes upregulated in pluripotent cells (Figure 1B). The full list of Gene Ontology categories can be found in Datasets S3, S4, S5, and S6. Overall, our data indicate that pluripotent cells are highly enriched for nuclear activities related to cell cycle, DNA repair, transcription, and chromatin remodeling.

Computational Identification of Putative Regulatory Motifs

Genes coexpressed in pluripotent cells may be (at least in part) coregulated by the same transcription factors. It follows

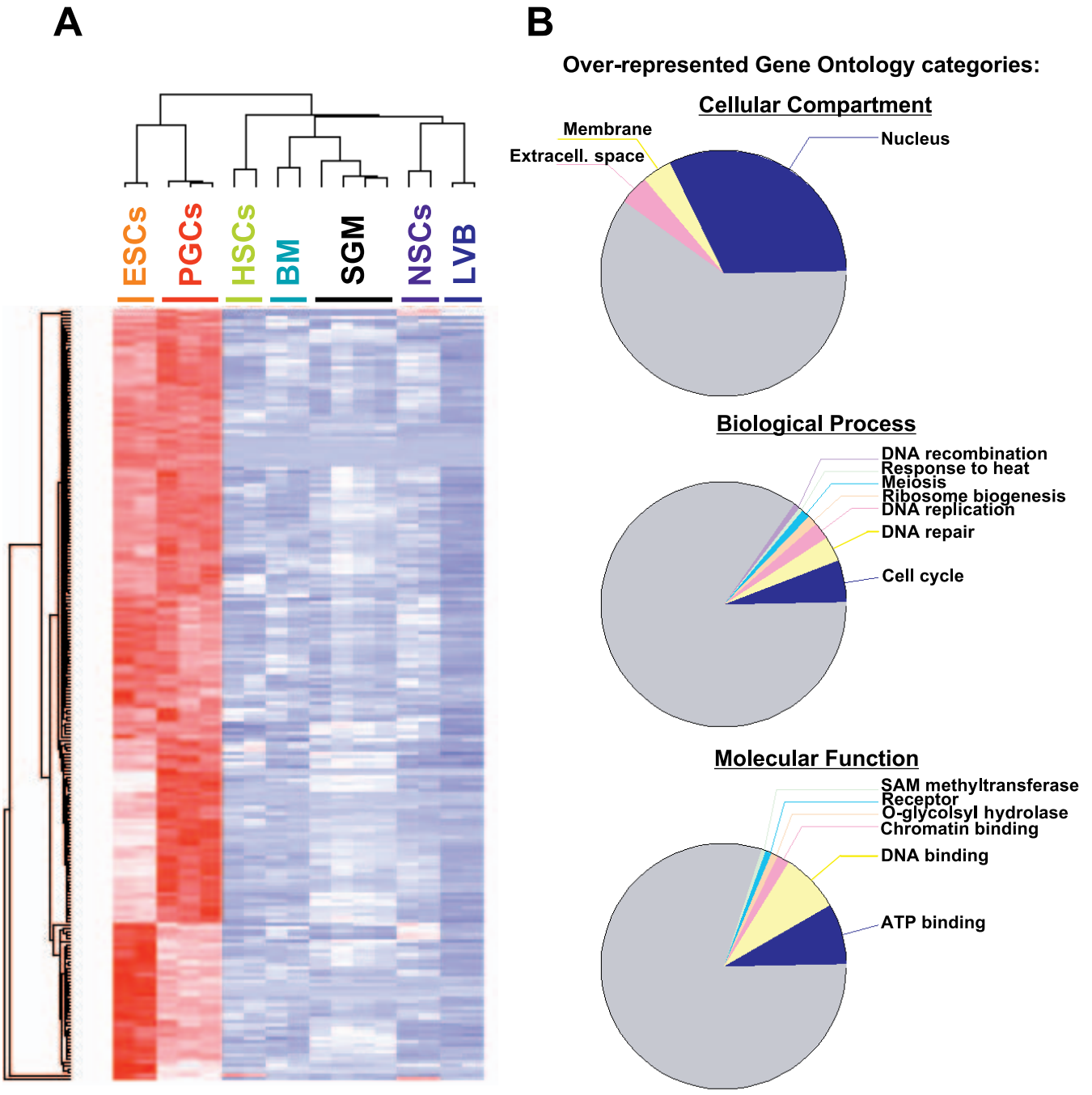


Figure 1. Identification of Genes Upregulated in Pluripotent Cells

(A) Gene expression profiling of ES cells and PGCs. Hierarchical clustering was used to identify genes upregulated in ES cells and PGCs relative to SGMs, bone marrow (BM), hematopoietic stem cells (HSCs), lateral ventricles of the brain (LVB), and neural stem/progenitor cells (NSCs) [42]. The cluster shown depicts gene expression of 230 probe sets upregulated in ESCs and PGCs. Red means the gene is upregulated, blue means it is downregulated.

(B) Functional annotation of genes upregulated in ES cells and PGCs. Gene Ontology analysis was performed to identify functional categories overrepresented in the cluster of genes upregulated in pluripotent cells (A). The top pie chart represents the ontology "cellular component," the middle pie chart represents the ontology "biological process," and the bottom pie chart represents the ontology "molecular function." Categories shown are significantly overrepresented at $p < 0.005$. Grey slices represent categories with $p > 0.005$.

doi:10.1371/journal.pgen.0030145.g001

that transcription factor binding sites responsible for driving gene expression in pluripotent cells are likely to be over-represented in the *cis*-acting regions of those genes. We therefore took a computational approach to identify DNA motifs that are statistically overrepresented in the putative promoter and enhancer regions of genes upregulated in pluripotent cells. To reduce noise in our computational analysis, we derived a smaller subset of genes, those with highly significant changes in expression (standard deviation/mean > 0.6). A smaller cluster of 55 probe sets (Dataset S7) was obtained that includes several known markers of pluripotency, such as Oct4, Nanog, Gdf3, Dppa2, Esg1, Utf1, and Tera.

To identify *cis*-regulatory motifs involved in transcriptional regulation of pluripotency-associated genes, we developed CompMoby (Comparative MobyDick), which improves upon MobyDick [28] by incorporating a flexible analysis of evolutionary conservation (Figure 2A). From a set of coregulated genes, CompMoby builds multiple dictionaries (lists of motifs) from the upstream noncoding sequences of individual genomes as well as sequences conserved across species. The motifs of these dictionaries are clustered to obtain a final dictionary of motif clusters. CompMoby then screens for motif clusters that are overrepresented in the set of coregulated genes compared to the entire genome.

Functional elements may not reside within conserved regions [39,40], and an advantage of CompMoby is that it does not solely rely on sequence alignments, but also uses information from individual genomes. By combining these two sets of information, CompMoby can identify conserved sites that are aligned, sites that are conserved but not aligned, and nonconserved sites. CompMoby is flexible; motifs do not have to be exactly conserved across species, since clustering the multiple dictionaries derived from different sets of sequences will group motifs related to each other by a few mutations.

We employed CompMoby to identify putative *cis*-acting motifs in the upstream sequences that may be shared among upregulated genes in mouse pluripotent cells and their human orthologs (Datasets S8–S13). From our final dictionary (Figure 2B and Datasets S14 and S15), we selected ten motif clusters and used promoter alignment data between human, mouse, rat, and dog [44] to systematically identify highly conserved motifs and their flanking regions (Figure 2C and Dataset S16) within the promoters of genes upregulated in pluripotent cells. We chose 25 different motifs and their flanking regions from our top ten motif clusters for further experimental characterization (underlined sequences in Figure 2B).

It is important to note that two of the predicted motifs correspond to putative binding sites of known transcriptional regulators of ES cells. The motif 7 cluster contains the sequence ATTACAAT, which has been implicated in Sox2 binding [45]. This sequence and its flanking regions are conserved in the upstream sequences of the Nanog gene in human, mouse, rat, and dog (Figure 2C). Interestingly, the conserved sequence corresponds to the recently described binding site for Oct4 and Sox2 that is required for *Nanog* expression [45,46], indicating that we have indeed identified a functional motif that regulates a pluripotency-associated gene. Cluster 8 contains a palindromic motif that matches the known canonical binding motif for Myc [47,48]. Although several other members of the basic helix-loop-helix family can bind this motif [49], it is interesting to note that c-Myc has recently been implicated in the regulation of self-renewal and pluripotency in mouse ES cells [50], and that it is part of a cocktail of factors capable of inducing pluripotency [51]. These results demonstrated the power of CombMoby and suggested that the other novel identified motifs may also be functional.

Identification of Novel Regulatory Motifs in Mouse ES Cells

We next sought to assess the transcriptional regulatory activity of the predicted motifs. We transfected mouse ES cells with Firefly luciferase reporter constructs containing the motifs upstream of a heterologous thymidine kinase (TK) minimal promoter (Figure 3). Each construct contained a motif and its flanking sequences (median length 30 bp) present in at least two repeats (table in Figure 3A; Dataset S18). As a positive control, we used a 242-bp fragment of the *Oct4* distal enhancer (DE) (Oct4, Figure 3A). Since one of our predicted motifs together with its flanking sequence has already been shown to regulate Nanog and be sufficient for gene expression in ES cells [45,46], we used it as an additional positive control (Nng, Figure 3A). Both controls contain an Oct4/Sox2 binding site, the only known enhancer element shown to specifically regulate expression of several genes preferentially expressed in ES cells [45,52–56].

When compared to the construct containing only the TK promoter (Figure 3A), 14 out of 25 motif-containing constructs showed a change in luciferase expression in transfected ES cells (Figure 3B, red bars), suggesting that the predicted motifs are functional enhancers of transcription. Notably, the enhancer activity of several motifs was higher than the activity of the previously identified Oct4/Sox2 enhancer (Figure 3B, compare motifs 1a, 1b, 2a, and 2c to Nng).

For some motifs, we tested different numbers of repeats (Figure 3A and Figure S1) and found that the increase in luciferase expression was directly proportional to the number of repeats, further suggesting that the predicted motifs act as transcriptional activators. To determine if the observed transcriptional activation is specific for pluripotent ES cells, we transfected several differentiated cell types: HEK293, NIH-3T3, and ES cells differentiated either by formation of embryoid bodies (EBs) or addition of retinoic acid (RA) (Figure 3B, blue bars). Several motifs showed decreased activity in differentiated cells compared to ES cells, indicating that they are preferentially active as transcriptional enhancers in pluripotent ES cells.

Two of the tested motifs appear to have repressing activities. While showing little activity in ES cells, motif 4b seems to downregulate expression in NIH 3T3s (which are transformed mouse embryonic fibroblasts) and EB cells

Figure 2. Computational Analysis and Identification of Regulatory Motifs

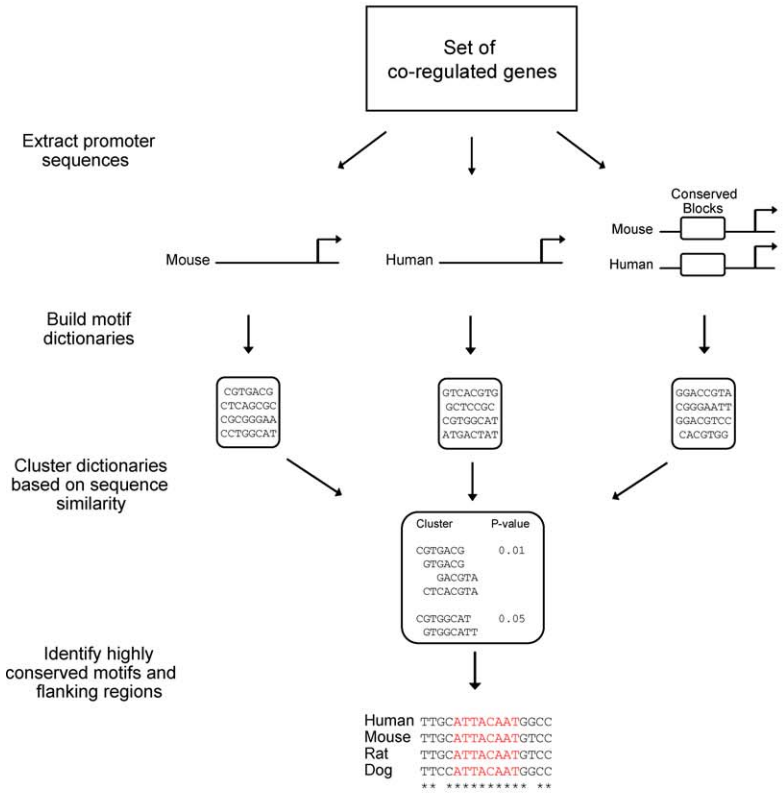
(A) Schematic diagram of the CompMoby algorithm.

(B) Top ten predicted motif clusters from CompMoby analysis of 2,000-bp sequences upstream from the transcriptional start site of 55 probe sets upregulated in pluripotent cells. Upstream sequence sets are given as Mouse (M), Human (H), and Conserved blocks between mouse and human (C). The fourth column lists $-\log_{10}$ *p*-values, which are calculated based on overrepresentation of the motif cluster in sequences upstream of the 55 probe sets relative to all other probe sets. *p*-Values are Bonferroni corrected for multiple testing. Underlined motifs were chosen for experimental characterization.

(C) Examples of motifs (red) found by CompMoby to be highly conserved across four mammalian species. Asterisks denote bases conserved across all four species.

doi:10.1371/journal.pgen.0030145.g002

A

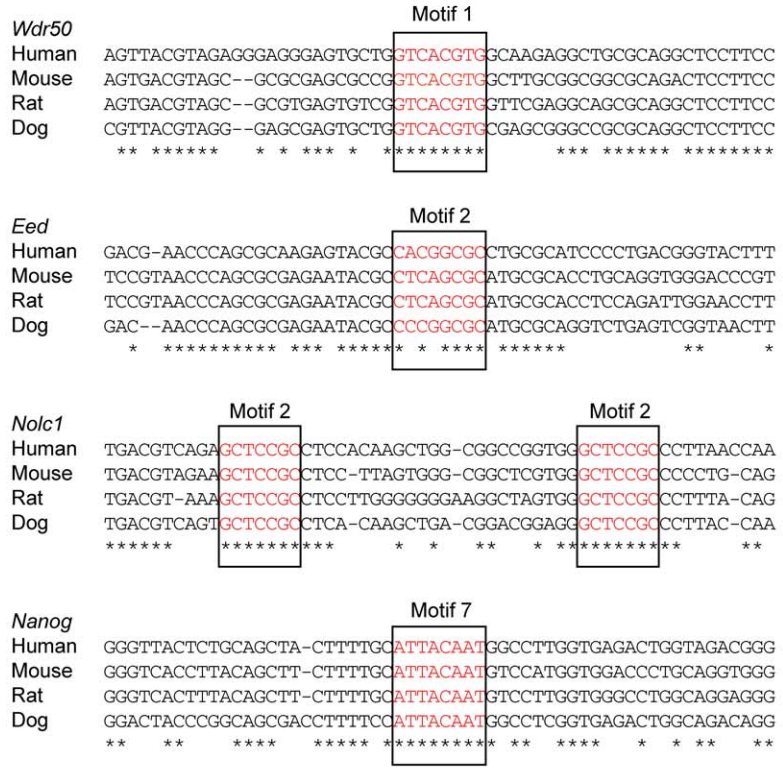


B

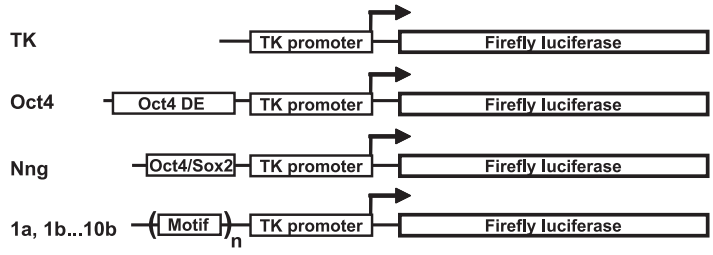
Motif Number	Sequence Set	Motif Cluster	-log$_{10}$P
1	M	CGGTCACG^{r}	0.45
	M	CGTCACG^{r}	
	H	CGGTCAC^{r}	
	H	CGGTCACC^{r}	
	H	GTCACGTG	
	C	GGTCACGT	
2	M	CTCAGCGT^{r}	7.08
	M	CTCAGCGC	
	M	CCTCTGCG^{r}	
	M	CCGCTGCGC	
	M	CGCTGCGCT	
	M	CGCGGCGC^{r}	
	M	CGCGGCG	
	H	GCTCCGCG	
	C	GCTCGGCG^{r}	
	M & H	GCTCCGC	
3	M	TACGGTAT	4.95
	M	GTACGGTCC^{r}	
	M & C	GGTACCGT	
	M & C	GGTACGGT^{r}	
4	M	ACGGAATTC	3.73
	M	GACGGAATT	
	M	CGGACGGA	
	C	CGGAATTC	
	C	AACGGAAT^{r}	
5	M	GACGTTCG	2.63
	M	GGACGTTT^{r}	
	M	TGGACGTT^{r}	
	H	GAGACGTC	
	H	GTGACGTC	
	C	GGACGTCC	
	M & C	GACGTCCG	
6	M	GCCAATCAG^{r}	0.87
	M	CCAATCAG	
	M	GCCAATCA^{r}	
	H	GGCCAATCA	
	C	CAATCACT^{r}	
	M & H	GCCAATCAG	
7	M	TGACAATA	1.42
	H	ATGACTAT	
	C	AACAATAC^{r}	
	C	ATTACAAT	
	C	GTAACAAT	
	C	TAACAATA^{r}	
	M & C	ATGACAAT	
8	M	CACGTGGC	0.82
	M	CCTGGCAT	
	H	CGTGGCAT	
	C	ACGTGG	
	M & C	CACGTGG	
9	M	CACGACAT	2.25
	M	CACGACG	
	H	ACCACGAC^{r}	
	H	CACGACAC^{r}	
	H	CACGACTC^{r}	
	C	GCACGACG	
	C	CACGACGA^{r}	
	C	ACGACAT	
10	M	ATTGAAGT	2.78
	C	TGACTGAA	
	C	GACTGAAA	
	C	CTACTGAA	
	C	GACTGAAG	

^{r}denotes reverse complement.

C



A



Motif	n	Motif	n
1a	2	6a	2
1b	3	6b	2
1c	3	6c	2
1d	2	7a	2
2a	2	7b	2
2b	2	8a	2x2
2c	2x2	8b	4
3a	5	8c	2
3b	2	9a	2
4a	2	9b	5
4b	2	10a	2
5a	4	10b	2

B

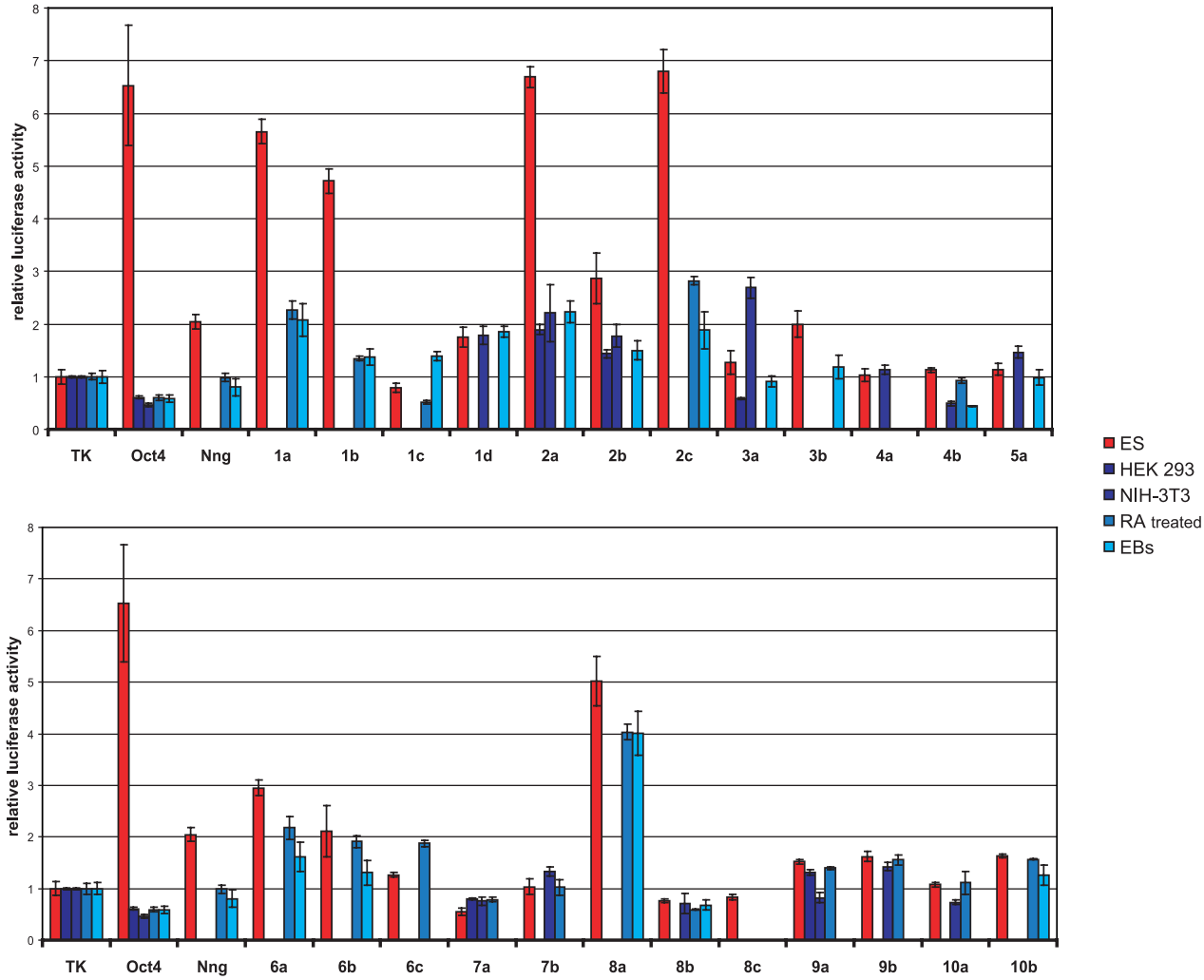


Figure 3. Experimental Validation of Predicted Regulatory Motifs

(A) Schematic representation of constructs used in this study. A fragment of the *Oct4* distal enhancer (DE, bp −2,181 to −1,939) (Oct4), and the Oct4/Sox2 regulatory motif identified in cluster 7 and comprising a part of the *Nanog* promoter (bp −190 to −156) (Nng) were used as positive controls. Predicted motif sequences (1a, 1b, . . ., 10b) were fused to a construct containing Firefly luciferase driven by a minimal TK promoter. The number of repeats for a given motif is indicated in the right panel. Constructs 2c and 8a bear two repeats of genomic sequence that each contains two predicted

regulatory motifs. (B) Regulatory activity of predicted motifs in mouse ES cells. Constructs described in (A) were transfected in undifferentiated mouse ES cells (red bars); or HEK293, NIH-3T3, and ES cells differentiated by formation of EBs or addition of RA (blue bars). A plasmid containing Renilla luciferase driven by the TK promoter was cotransfected. After 24 h, cells were lysed and assayed for luciferase activities. Firefly values were corrected for Renilla values, and the activities of the tested constructs were calculated relative to the activity of the TK construct, which was set to 1. For undifferentiated mouse ES cells, each construct was tested in several independent experiments (five on average), and representative results are shown. For HEK293, NIH-3T3, EB-derived, and RA-differentiated ES cells, one representative experiment is shown for each (blue bars). Bars represent averages of triplicates performed in each single experiment. Error bars depict standard deviation.
doi:10.1371/journal.pgen.0030145.g003

(Figure 3B). Therefore, motif 4b might bind a repressor necessary for downregulation of genes expressed in ES cells upon differentiation. In contrast, motif 7a appears to confer repression preferentially in ES cells, suggesting that it may control the levels of ES cell-expressed genes (Figure 3B). In summary, we have identified several novel *cis*-acting motifs that are sufficient to regulate gene expression preferentially in undifferentiated mouse ES cells. These results demonstrate that CompMoby can successfully predict functional motifs in mammals and even compares favorably to studies in organisms with simpler genomes such as yeast [57].

We decided to focus on eight motifs that showed the most interesting levels of activity in undifferentiated versus differentiated cells (Figure 4A). To confirm that the predicted motifs are indeed responsible for this activity, we performed mutational analyses by introducing point mutations at every second position of a motif sequence, with the exception of motif 1a and 8a. Motifs 1a and 8a, while clearly belonging to distinct motif clusters, have the sequence CACGTG in common (Figures 2B and 4A). CACGTG has been previously identified as a binding site for c-Myc and several other transcription factors of the basic helix-loop-helix family (see above), and a point mutation in CACGTG inhibits binding of Myc proteins [58]. Therefore, we decided to introduce a single point mutation in motifs 1a and 8a (1aM1, and 8aM, respectively). When transfected into mouse ES cells (Figure 4B), both mutated motifs 1a and 8a showed a drastic reduction in activity. This result indicates that their enhancer activity is regulated by the CACGTG sequence, likely through the binding of a basic helix-loop-helix transcription factor. The construct tested for motif 1a contains GC-rich sequences in the regions flanking the predicted motif. Particular GC-rich sequences may be bound by the Sp1 transcription factor [59]. To test the contribution of flanking sequences to the activity of motif 1a, we introduced four point mutations in each of the flanking regions (1aM3), four point mutations in the motif 1a sequence (1aM2), and both combined (1aM4). Our results indicate that the enhancing activity of motif 1a is due to the predicted motif sequence, and not to the flanking regions (Figure 4B). Similarly, the activity of all other motifs was significantly reduced or abolished when mutated, indicating that the predicted motif sequences are responsible for their activity.

To determine whether the identified motifs are specifically active in ES cells only, we tested their activity in another pluripotent cell type, EG cells. We found that all of the tested motifs have comparable levels of activity in both pluripotent cell types (Figure 4C). Next, we investigated whether any of the motifs are required for regulation of gene expression in the context of an endogenous promoter. One of the genes we identified as upregulated in pluripotent cells is *Eed* (Datasets S2 and S7), a component of chromatin remodeling complexes that regulate transcriptional silencing in ES cells [60]. The regulatory sequences of the *Eed* gene have not been described, and the genomic sequence upstream of *Eed* contains several of our predicted motifs. Point mutations in the sequences representing predicted motifs 2 and 6 significantly reduced the activity of the *Eed* promoter (Figure 4D), indicating that motifs 2 and 6 are necessary for maximal expression driven by the endogenous *Eed* promoter. Future studies will be necessary for complete dissection of the *Eed* promoter. Nevertheless, our proof-of principle experiments demonstrate that at least some of the motifs identified as sufficient to enhance transcription of a heterologous promoter in pluripotent stem cells are also functional enhancers of an endogenous promoter. In summary, we have identified novel, bona fide regulatory motifs present in genes preferentially expressed in pluripotent cells. We anticipate that our approach will greatly accelerate the dissection of enhancer/promoter elements of pluripotency-associated genes.

The Identified Regulatory Motifs Are Active in Human ES Cells

Comparative DNA sequence analysis of pluripotency-associated genes revealed a high degree of conservation between mouse and human for several of the identified motifs (Figure 2C). To address whether the motifs active in mouse ES cells are also sufficient to activate transcription in human ES cells, we compared the expression levels of constructs bearing the identified motifs (Figure 5, red bars) with their mutated counterparts (blue bars). Interestingly, all of the motifs showed significant regulatory activity in human ES cells with levels similar to those in mouse ES cells. Likewise, the activity was diminished or abolished upon point mutations. These results underscore the power of our approach to predict and identify regulatory elements and suggest a strong degree of conservation in the transcriptional regulatory networks in mouse and human ES cells.

Proteins Present in ES cells, Including the Transcription Factor NF-Y, Bind Sequence-Specifically to the Motifs

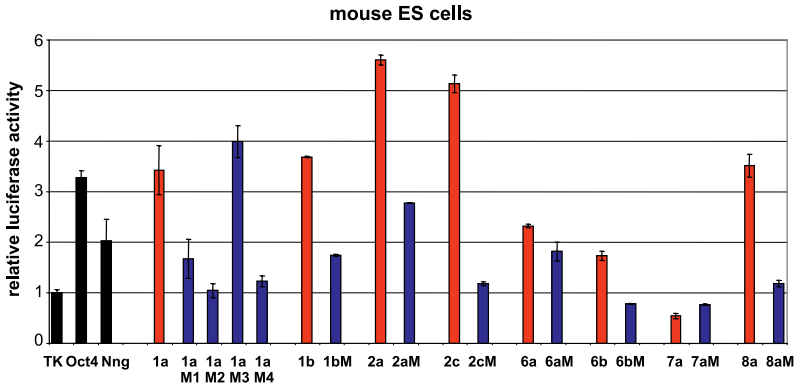
For several of the motifs, we performed electrophoretic mobility shift assays (EMSA) with biotin-labeled motif sequences. In the presence of ES cell nuclear protein extracts, shifted bands were observed for motifs 1a, 1b, 2a and 8a (Figure 6A). Several of these bands represent specific protein-motif complexes, as they were efficiently competed in the presence of excess unlabeled wild-type (1a, 1b, 2a and 8a, respectively) but not mutated probe (1aM, 1bM, 2aM and 8aM, respectively) (Figure 6A). These results show that proteins present in ES cells bind sequence-specifically to the motifs.

To identify putative transcription factors that can bind to the motifs, we systematically searched databases of known transcription factors. We did not find a match for most of the motifs, e.g., 1 and 2. We are particularly interested in identifying transcription factors that bind to these motifs, as they showed very high enhancer activity specifically in

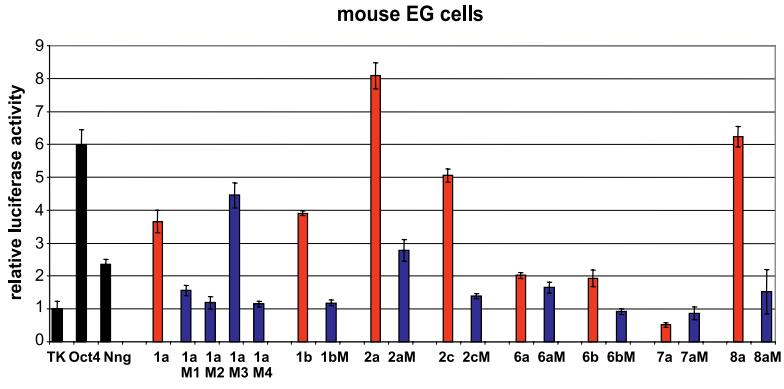
A

	WT motif sequence		Mutated motif sequence
1a	GCGCGAGCGCCGGTCACGTGGCTTGCGGCGGCG	1aM1	GCGCGAGCGCCGGTCTCGTGGCTTGCGGCGGCG
		1aM2	GCGCGAGCGCCGTTAAAGGGCCTTGCGGCGGCG
		1aM3	GCGAGATCTCAGGTCACGTGGCTTTCTGAGTCG
		1aM4	GCGAGATCTCAGTTAAAGGGCTTTCTGAGTCG
1b	GGGCGTTTGTTTACGTGACGTTGCGGAGTAGGCA	1bM	GGGCGTTTGTTTACTTTAAGTTGCGGAGTAGGCA
2a	AACCCAGCGCGAGAATACGCCTCAGCGCATGCGCAC	2aM	AACCCAGCGCGAGAATACGCATAATCTCATGCGCAC
2c	ATCTGACGTAGAAGCTCCGCCTCCTTAGTGGGCGGCTCGTGGGCTCCGCCCCTGCAGGATGATGACGCGTGT	2cM	ATCTGACGTAGAATATAAGCCTCCTTAGTGGGCGGCTCGTGGTATAAGCCCCTGCAGGATGATGACGCGTG
4b	CAAAGGTAATGCGCGGGAATTTGCAAGTTTTCCCGCCT	4bM	CAAAGGTAATGCACTGTACTTTGCAAGTTTTCCCGCCT
6a	CGTTAGACTCGCCAATCAGCGGCGCCTCA	6aM	CGTTAGACTCTCAACTAATCGGCGCCTCA
6b	TCCGCAGCTACCCAATCAGCTGGGGTCGCC	6bM	TCCGCAGCTACACCAGCATCTGGGGTCGCC
7a	CAAATGACAAATGACAATATAGAAACAATT	7aM	CAAATGACAAAGGTCCAGATAGAAACAATT
8a	GACTATGCATCCACGTGGTACGTTCCCATCACGTGGCAGGATTCCTC	8aM	GACTATGCATCCTCGTGGTACGTTCCCATCTCGTGGCAGGATTCCTC

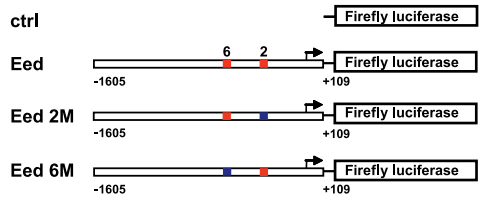
B



C



D



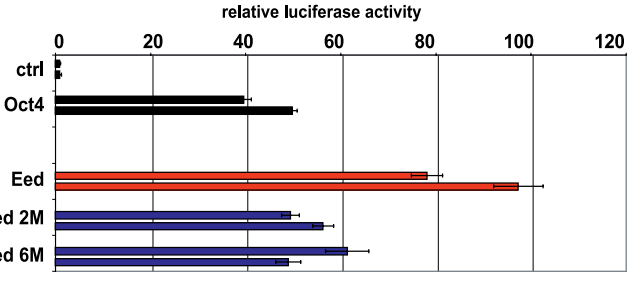


Figure 4. The Predicted Regulatory Motifs Are Sequence-Specific, Active in Both Mouse ES and EG Cells, and Act as Functional Enhancers of the Endogenous Eed Promoter

(A) Mutations in the predicted regulatory motifs. Sequences containing the predicted regulatory motifs (depicted in red) flanked by endogenous sequences are shown on the left. Sequences containing point mutations (depicted in blue) are shown on the right.
(B) Regulatory activity of the predicted motifs and their mutated counterparts in mouse ES cells. Data were collected and analyzed as described in Figure 3B. Representative results from five independent experiments are shown. Bars represent averages of triplicates performed in each single experiment. Error bars depict standard deviation. Wild-type sequences, red bars; mutated sequences, blue bars.
(C) Regulatory activity of predicted motifs and their mutated counterparts in mouse EG cells. Data were collected and analyzed as described in Figure 3B. Representative results from two independent experiments are shown. Bars represent averages of triplicates performed in each single experiment, error bars depict standard deviation. Wild-type sequences, red bars; mutated sequences, blue bars.
(D) Activity of regulatory motifs 2 and 6 present in Eed upstream genomic sequence. A 1.7-kb fragment of Eed upstream genomic sequence (bp −1,605 to +109 relative to the transcription start site) was cloned and fused to the Firefly luciferase reporter gene (Eed). Mouse ES cells were transfected, and the activity of the Eed construct was compared to the activities of the construct containing luciferase reporter gene alone (ctrl), a TK-bearing construct containing the Oct4 DE (Oct4), an Eed construct containing four point mutations in motif 2 (Eed 2M), and an Eed construct containing four point mutations in motif 6 (Eed 6M). Data were collected and analyzed as described in Figure 3B. Results from two independent experiments are shown. Bars represent averages of triplicates performed in each single experiment; error bars depict standard deviation. Wild-type sequences, red bars; mutated sequences, blue bars. The same mutations also significantly reduce activity of the Eed promoter in EG cells (unpublished data).
doi:10.1371/journal.pgen.0030145.g004

undifferentiated ES cells (Figure 3B; compare motifs 1a, 1b, 2a, and 2c to Oct4), but these transcription factors will have to be identified with unbiased biochemical or genetic approaches.

For motif 6 we were able to take a candidate factor approach. Two variants of the motif 6 (motifs 6a and 6b) contain a CCAAT box that when mutated caused a reduction in enhancer activity (Figures 4 and 5). CCAAT boxes have been shown to act as enhancers of transcription [61]. To identify the motif 6 binding factor(s), we performed EMSA. When biotin-labeled motif 6b was incubated with ES cell nuclear protein extracts, several shifted bands were observed (lane 2, Figure 6B). Excess unlabeled motif 6a and 6b, but not motif 6bM where the CCAAT box was mutated (Figure 4A), eliminated the binding of the major band, indicating that the band represents factor(s) specifically bound to the CCAAT box (lanes 3–5, Figure 6B). Several proteins able to bind CCAAT boxes have been described [62]. Among these is a heterotrimeric factor NF-Y (composed of NF-YA, NF-YB, and NF-YC subunits), which requires a high degree of conservation of the CCAAT sequence [63,64]. To determine whether NF-Y binds to the CCAAT box of motifs 6a and 6b, we performed additional EMSAs in which ES cell extracts were preincubated with anti-NF-Y antibodies (lanes 6–9, Figure 6B). The major motif 6b–protein complex was found to be specifically supershifted by antibodies against both NF-YA and NF-YB (lanes 6–9, Figure 6A), but not by an unrelated antibody (lanes 10 and 11, Figure 6B). These data show that NF-Y binds to motif 6, indicating that the NF-Y binding site is conserved and overrepresented in *cis*-acting regions of genes preferentially expressed in pluripotent cells.

To confirm that NF-Y binds directly to the promoters of genes upregulated in pluripotent cells, we performed chromatin immunoprecipitation (ChIP) real-time PCR. *Cdc25c* is a known target of NF-Y [65] and *Sall4* and *Zic3* are two of the genes with the highest levels of upregulation in pluripotent cells, present in the cluster in Figure 1A, that contain consensus NF-Y sites. Our ChIP data show that NF-Y binds to the CCAAT-containing regions of these genes in ES cells, but not to a control gene *(Rpl15)* that is not upregulated in ES cells (Figure 6C).

NF-Y Is Differentially Expressed during ES Cell Differentiation

Even though NF-Y is expressed in several cell types and tissues [61], there is strong evidence for its differential expression: NF-YA and NF-YC are highly upregulated in mouse oocytes (40-fold and 12-fold, respectively, relative to the median expression in 60 other tissues) [66]. NF-YB was recently identified in a screen for genes upregulated in the

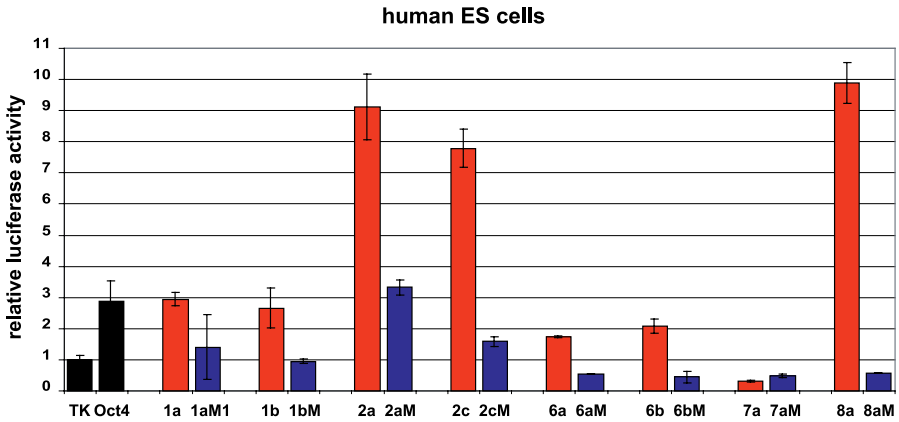


Figure 5. The Activity of the Identified Motifs Is Conserved in Human ES Cells

Cells were transfected and data were analyzed as described in Figure 3B. Results from one of two independent experiments are shown. Bars represent average of duplicates performed in each single experiment, error bars depict standard deviation. Wild-type sequences, red bars; mutated sequences, blue bars.
doi:10.1371/journal.pgen.0030145.g005

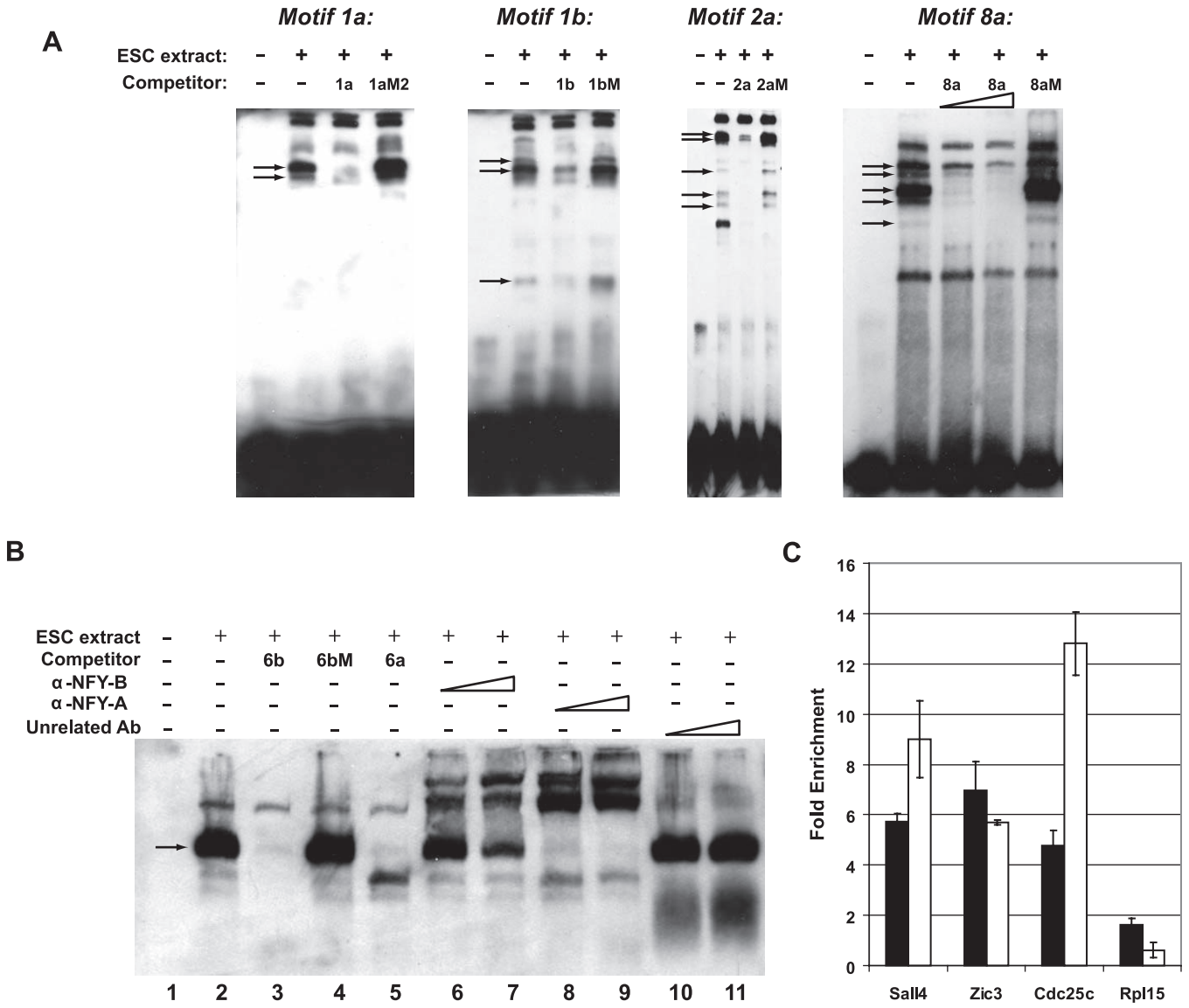


Figure 6. Proteins Present in ES Cells, Including NF-Y, Bind Sequence-Specifically to the Motifs

(A) EMSA using motifs 1a, 1b, 2a and 8a. EMSA was performed with a double-stranded, biotin-labeled oligonucleotide containing the corresponding motif in the absence or presence of ES cell nuclear extracts. Where indicated, ES cell nuclear extracts were preincubated with a 200-fold molar excess of unlabeled competitor double-stranded oligonucleotides. For motif 8a, ES cell nuclear extracts were also preincubated with a 100-fold molar excess of unlabeled competitor double-stranded oligonucleotides. Arrows denote specific bands.

(B) EMSA using motif 6. EMSA was performed with a double-stranded, biotin-labeled oligonucleotide containing motif 6b in the absence (lane 1) or presence (lanes 2–11) of ES cell nuclear extracts. ES cell nuclear extracts were preincubated with a 200-fold molar excess of unlabeled competitor double-stranded oligonucleotides (lanes 3–5), increasing amounts of α-NF-YB (lanes 6 and 7), α-NF-YA (lanes 8 and 9), or unrelated antibody (lanes 10 and 11). Arrow denotes the major specific band.

(C) ChIP-real time PCR. Chromatin was precipitated from ES cell nuclear extracts using α-NF-YB or unrelated antibody. Data represent fold enrichment in the α-NF-YB precipitation relative to the unrelated antibody precipitation. *Cdc25c* is a known target of NF-Y [65] and *Sall4* and *Zic3* are two of the genes with the highest levels of upregulation in pluripotent cells (Figure 1A) that contain consensus NF-Y sites. *Rpl15* is a control gene that is not upregulated in ES cells. Black and white bars represent independent experiments performed with different ES cell nuclear extracts.

doi:10.1371/journal.pgen.0030145.g006

inner cell mass of the mouse blastocyst [67]. In addition, alternative splicing produces two different NF-YA isoforms: NF-YA(long) and NF-YA(short) [68].

We analyzed expression of NF-YA (including long and short isoforms), NF-YB, and NF-YC during mouse and human ES cell differentiation (Figure 7A). The levels of NF-Y mRNAs were analyzed by real-time reverse transcriptase PCR (RT-PCR) in mouse ES cells and differentiated ES cells either treated with RA or induced to form EBs. Interestingly, the expression of the two isoforms of NF-YA (long and short) changed in opposite directions; while the levels of NF-YA(long) increased with ES cell differentiation, NF-YA(short) was significantly downregulated with differentiation. At day 6 of RA-induced differentiation, NF-YA(short) was not detected, and at day 7 it was detected at low levels (8-fold reduction) (Figure 7A). The expression of NF-YB subunit was modestly reduced (up to 2.5-fold), while that of NF-YC did not change considerably during the course of differentiation.

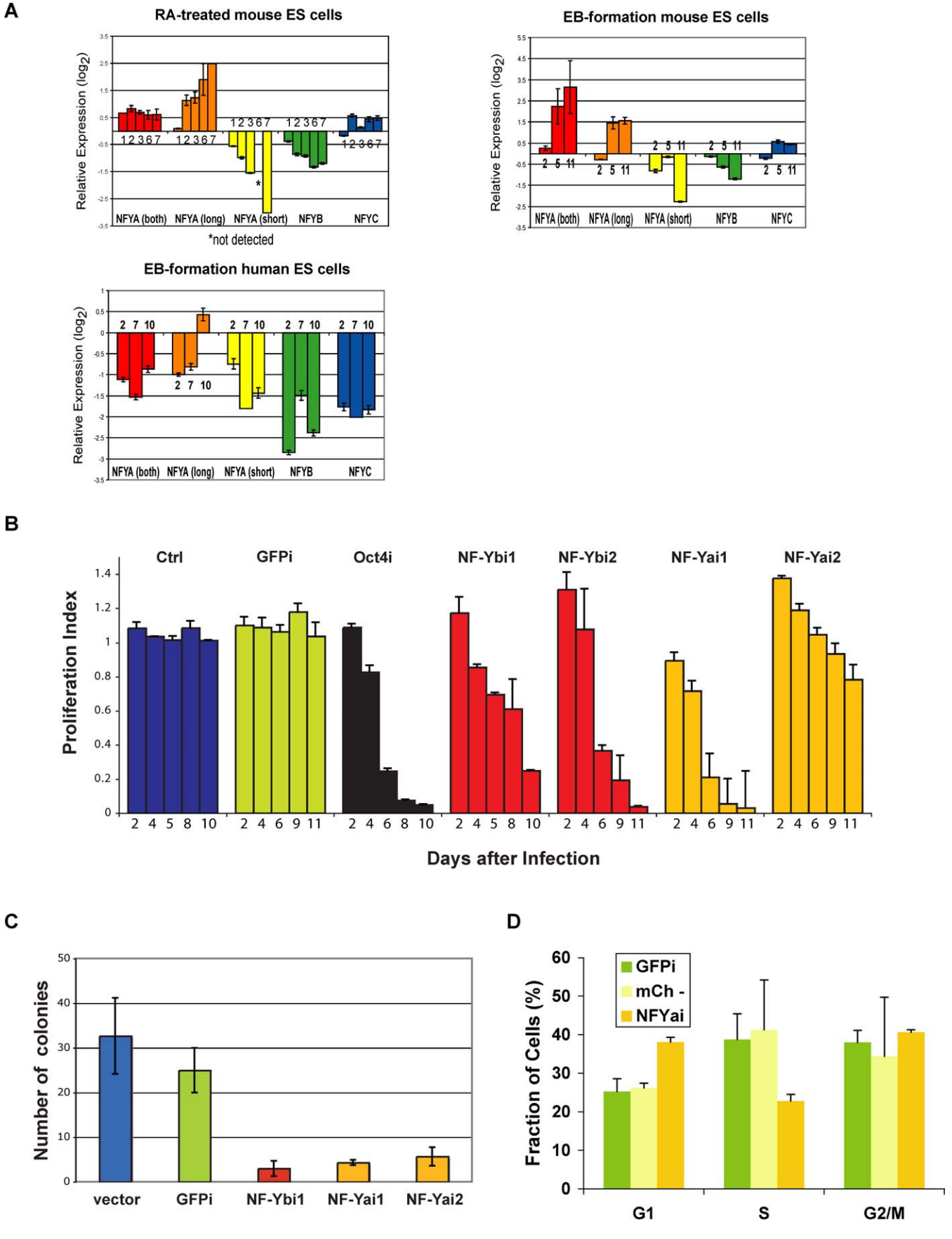

A
RA-treated mouse ES cells
Relative Expression (log2)
NFYA (both)
NFYA (long)
NFYA (short)
NFYB
NFYC
*not detected
EB-formation mouse ES cells
EB-formation human ES cells
B
Ctrl
GFPi
Oct4i
NF-Ybi1
NF-Ybi2
NF-Yai1
NF-Yai2
Proliferation Index
Days after Infection
C
Number of colonies
vector
GFPi
NF-Ybi1
NF-Yai1
NF-Yai2
D
Fraction of Cells (%)
GFPi
mCh -
NFYai
G1
S
G2/M

Figure 7. NF-Y Is Differentially Expressed during ES Cell Differentiation and Is Required for ES Cell Proliferation

(A) Expression levels of NF-YA, NF-YA(long), NF-YA(short), NF-YB, and NF-YC during differentiation of ES cells. Real-time RT-PCR of RA-treated mouse ES cells, upper left panel; EB formation by mouse ES cells, upper right panel; EB formation by human ES cells, lower left panel. Fold-changes were calculated relative to undifferentiated ES cells using the REST software [94] and housekeeping genes as controls. A representative of at least three experiments (each performed in duplicate) is shown. Days of differentiation are indicated next to the bars.

(B) RNAi-based competition assay. Mouse ES cells were infected with a lentiviral vector that induces RNAi and labels the cells with a red fluorescent marker, mCherry. The percentage of cells undergoing RNAi (mCherry+) was measured in a competition assay with noninfected wild-type cells over time. The ratio [mCherry+ cells in RNAi against target gene/mCherry+ cells in RNAi against GFP] gives a proliferation index. In the case of cells undergoing RNAi against GFP (GFPi), the ratio was calculated using cells infected with empty lentiviral vector as control. This index is expected to remain at 1 over time if there are no effects of RNAi against the target gene on proliferation and to be less than 1 if there are defects in cell proliferation. We validated our approach with RNAi against Oct4 (black bars). Downregulation of NF-YA or NF-YB leads to defective proliferation of ES cells (orange and red bars, respectively), while the unrelated control sequence or downregulation of GFP has no effect (blue and green bars, respectively). Results from one of 3–5 independent experiments are shown. Bars represent averages of duplicates (NF-YB, control) or triplicates (NF-YA, GFP, Oct4) performed in a single experiment.

(C) Colony-formation assay. Control (vector and GFPi) cells and cells undergoing RNAi against NF-YA or B were sorted and plated at low density (300 cells per well of a 6-well plate) in triplicates. Colonies were counted after 10 d of culture.

(D) Cell-cycle analysis. Cells infected with a lentivirus leading to NF-YA knockdown were sorted for mCherry+ cells (NF-YAi), which are undergoing RNAi against NF-YA, or mCherry- cells (mCh-), which correspond to in-plate control noninfected cells with which the NF-YA knockdown ES cells are competing. An additional control used was ES GFPi cells. Cells were stained using propidium iodide and analyzed for DNA content using flow cytometry. Samples were analyzed in triplicates. Error bars depict standard deviation.

doi:10.1371/journal.pgen.0030145.g007

During differentiation of human ES cells, the expression of all NF-Y subunits was significantly reduced (Figure 7A). There is a lack of concordance of expression patterns for NF-YA(both), NF-YA(long), and NF-YC between the mouse and human EBs. It is possible that not all of the NF-Y subunits are regulated in identical manner in mouse and human ES cells, or that the cells forming in differentiated EBs, which are very heterogeneous cell populations, differ in nature or proportion in both species and have different levels of some NF-Y subunits (particularly NF-YA[long] and NF-YC). Nevertheless, our results show that NF-Y subunits, in particular NF-YA(short) and NF-YB, are downregulated during differentiation of mouse and human ES cells, suggesting that a specific subunit composition of NF-Y may be critical for ES cells.

NF-Y Regulates ES Cell Proliferation

NF-Y has been implicated in promoting proliferation [69,70] and inhibiting differentiation [71] and senescence [72]. NF-YA mutant mice have been reported to display early embryonic lethality, as no mutant embryos were observed at the earliest stage analyzed (8.5 dpc) [70]. However, the function of NF-Y in ES cells had not been examined. We sought to investigate the role of NF-Y in proliferation of ES cells, using RNA interference (RNAi) in combination with a recently described competition assay [25]. This assay measures the ability of cells undergoing RNAi and grown in the presence of wild-type cells to maintain the rapid cell proliferation that characterizes wild-type ES cells (Figure 7B). The results are therefore a measure of the growth rate of ES cells undergoing RNAi relative to that of wild-type cells. We tested the effect of lentivirus-mediated NF-Y knockdown in mouse ES cells. As all three NF-Y subunits are required for sequence-specific DNA binding and downregulation of any subunit is expected to impair NF-Y binding to DNA [63,64], we infected GFP-expressing ES cells [73] with either NF-YA or NF-YB short hairpin RNAs (shRNAs). Similar to the control Oct4 shRNA [25], cells infected with NF-YA or NF-YB shRNAs were selectively out-competed by wild-type cells over time (NF-Yai1 and NF-Ybi1, Figure 7B). To confirm the RNAi specificity and exclude possible off-target effects, we tested NF-YA and NF-YB shRNAs that target a different region of the mRNAs (NF-Yai2 and NF-Ybi2, respectively, Figure 7B). In addition, an unrelated shRNA and an shRNA targeting GFP transcript were used to exclude the possibility that the effects observed were due to sequestration of the RNAi machinery, rather than depletion of specific gene products (ctrl and GFPi, respectively, Figure 7B). While the percentage of cells infected with the unrelated or GFP shRNAs did not change significantly, cells infected with the NF-Yai2 and NF-Ybi2 shRNAs were out-competed by noninfected, wild-type cells. The specificity of the NF-YA and NF-YB knockdown was confirmed by real-time RT-PCR (unpublished data). Our preliminary results using real-time RT-PCR for several differentiation markers do not reveal induction of differentiation upon RNAi. This suggests that the primary role of NF-Y may be to maintain the high proliferative capacity of ES cells.

We characterized the role of NF-Y in ES cell proliferation in more detail. Plating cells at low density revealed a strong decrease in the clonogenic potential of ES cells undergoing RNAi against NF-YA or B, relative to control cells (Figure 7C). Using staining for alkaline phosphatase, a marker of undifferentiated ES cell colonies, we did not observe partially stained or unstained colonies upon NF-Y knockdown. The very few colonies that formed were still alkaline phosphatase positive, and may be due to less than 100% pure FACS isolation of cells undergoing RNAi prior to plating, or to incomplete knockdown of NF-Y to levels that still allow colony formation. These data indicate that knockdown of NF-YA or NF-YB compromises the clonogenic potential of mouse ES cells, reducing it by 5–10-fold. Cell cycle analysis using NF-YA knockdown ES cells revealed an increased proportion of cells in G1 and a decreased proportion of cells in S phase (Figure 7D). Taken together, these results indicate that inhibition of NF-Y function leads to defects in ES cell proliferation that correlates with an accumulation of cells at the G1/S transition of the cell cycle.

Discussion

In this study we report a systematic approach that combines comprehensive expression analysis of coregulated genes, computational de novo motif prediction, biochemical validation of *cis*-regulatory elements, and identification of transcription factors that bind to those elements in pluri-

potent stem cells. Our methodology can be used with any set of coregulated genes, and, as such, is broadly applicable to the characterization of transcriptional regulatory networks. The approach we describe compares favorably to the standard experimental method to identify regulatory sequences, which relies on time-consuming dissection of large noncoding regions of a single gene. When compared to other methods to identify *cis*-regulatory elements, like ChIP in combination with microarrays (ChIP-chip) or paired-end ditag sequencing (ChIP-PET), our approach has two principal advantages: it does not require prior knowledge of the critical transcription factors whose targets are to be investigated, and it is not limited by the number of cells available for analysis. In particular, we have been able to generate reliable expression data from as low as 500–1,000 cells (unpublished data), whereas current ChIP-chip and ChIP-PET methods require several million cells [23,74,75]. Thus, we envision that the approach described here will be particularly useful for the characterization of transcriptional networks that regulate cell fate decisions during embryonic development and stem cell differentiation.

We identified short DNA sequence motifs that are highly active in undifferentiated ES cells but not in differentiated cells (Figure 3B, motifs 1 and 2). Importantly, the level of activity of these motifs is significantly higher than that of the Oct4/Sox2 element in the Nanog promoter (Figure 3B, compare motifs 1a, 1b, 2a, and 2c to Nng). These results indicate that we identified enhancer elements that are bound by transcriptional factors preferentially active in undifferentiated mouse and human ES cells. The availability of EMSAs for motifs 1 and 2 and of mutated versions that highly reduce or abolish motif activity (Figures 4, 5, and 6A) should facilitate the unbiased identification of the transcription factors that bind to these motifs.

An important validation of our systematic analysis of *cis*-regulatory elements active in ES cells is the identification of NF-Y as a transcription factor that binds specifically to one of those elements and regulates ES cell proliferation. In support of our findings, the NF-Y binding site was detected as overrepresented in genomic regions bound by Oct4 and Sox2 in human ES cells (Qing Zhou and Wing Wong, personal communication). It is possible that NF-Y contributes to the regulation of the peculiar cell cycle pattern of ES cells, with a short G1 phase and insensitivity to the Rb pathway (reviewed in [76]). NF-Y had previously been shown to regulate cell proliferation in other experimental paradigms [69,70], but its role in early embryonic development remains poorly understood. The strong upregulation of subunits of NF-Y in oocytes [66] and the ICM [67], and the early arrest of NF-YA mutant embryos [70], indicates that NF-Y plays important roles during early embryogenesis. It is also worth noting the dramatic difference in expression of NF-YA isoforms during ES cell differentiation (Figure 7A). Both NF-YA isoforms contain a glutamine-rich region that is reduced in the short isoform of NF-YA [68]. The glutamine-rich region of NF-YA has been shown to activate transcription [68,77,78] and it is also a protein–protein interaction domain [79]. The functional significance of the two NF-YA isoforms remains to be elucidated, although recent data indicate that NF-YA(short) promotes self-renewal of hematopoietic stem cells [80]. Future studies will address the specific contribution of NF-Y and its different subunits, in particular NF-YA(short), in ES cells.

ES cells may be governed at the molecular level by the action of cell-specific transcription factors, such as Oct4 and Nanog, and factors that are also expressed in other cell types, such as NF-Y, c-Myc [50], and Stat3 [11]. Interestingly, NF-Y binds to the promoter of Sall4 (Figure 6C), an essential ES cell regulator [24]. It will be important to identify the target genes that are regulated by NF-Y in ES cells. We expect that the combination of ChIP-chip and expression profiling will reveal the contribution of NF-Y to the transcriptional program of ES cells.

In summary, we report here the identification of clusters of genes upregulated in pluripotent cells, the development of a novel algorithm for discovery of short *cis*-acting regulatory motifs, the validation of the activity of several novel motifs in mouse and human pluripotent stem cells, and the identification of transcription factor NF-Y as a regulator of gene expression in ES cells that is required for their proliferation. Genetic and biochemical approaches should allow the identification of other transcription factors that bind to the motifs. Our results provide a basis for understanding the transcriptional regulatory networks that underlie early mammalian embryogenesis and ES cell self-renewal and pluripotency.

Materials and Methods

Microarray data analysis. The isolation of PGCs and SGM cells from 11.5 dpc Oct4/GFP transgenic mouse embryos and the identification of their transcription profiles is described elsewhere (Wei et al, submitted). Briefly, we used 20,000–30,000 PGCs or SGM cells per replicate sample, and analyzed 3–4 replicates per tissue using Affymetrix U74Av2 arrays (http://www.affymetrix.com), which assay for the expression of about 12,000 genes. We normalized, modeled, and clustered gene expression profiles (Dataset S1) using the dChip software (http://biosun1.harvard.edu/complab/dchip/) [81]. We compared the gene expression profiles of PGCs and SGM cells with those of embryonic and adult stem cells that we had previously described [42]. Hierarchical clustering was used to identify clusters of genes associated with pluripotency. A total of 230 probe sets were selected and used for Gene Ontology term analyses with the Onto-Express software (http://vortex.cs.wayne.edu/) [43]. *p*-Values for significance of overrepresentation of functional annotations were calculated in Onto-Express using a hypergeometric distribution and corrected for multiple testing using false discovery rate. For motif discovery, a cluster of 55 probe sets (included in the 230 probe sets used for Gene Ontology term analyses) was selected by the additional criteria: downregulation in differentiation of ES cells towards EBs (H. Chipperfield, S. Zhong, D. Melton, and W. Wong, personal communication); standard deviation/mean > 0.6. These 55 probe sets are listed in Dataset S7. Detailed protocols are available upon request.

Computational methods. To identify putative motifs shared among the pluripotency-associated gene cluster, Affymetrix probe sets were mapped to Ensembl (http://www.ensembl.org) gene annotation v.27 (Dataset S17) [82,83]. Both 1,000 bp and 2,000 bp of the intergenic sequences upstream from the transcriptional start site were extracted for each gene. For each of the different lengths, three sets of sequences were extracted from Ensembl; the first set contained the sequences of all annotated genes in the cluster in mouse (build 33), the second set contained the sequences from the orthologous human (build 35) genes in the cluster obtained from Ensembl mapping [83,84], and the last set consisted only of concatenated blocks of mouse promoter sequences that were conserved between mouse and human within the specified upstream sequence length. Pairwise alignments between mouse (mm5) and human (hg17) were obtained from the University of California Santa Cruz Genome Browser database (http://genome.ucsc.edu) [85–87].

Next, the three sets of upstream sequences were repeat masked (http://www.repeatmasker.org) and then used as input for the MobyDick algorithm [28,88] to build three dictionaries of putative motifs. The adjustable parameters used were as described [57], except

MaxP, which was set to 0.1. Similar motifs from all dictionaries were grouped into one final dictionary of motif clusters using the CAST clustering algorithm [89]. All pairs of motifs in the dictionaries were scored based on a gapless pairwise alignment using a simple mutation model [57], after which CAST was applied with the threshold parameter set at 0.55 (the lower bound of the normalized score averaged over all pairs in a cluster).

Following the clustering step, we calculated a *p*-value to identify motif clusters that were significantly overrepresented in the pluripotency-associated gene cluster compared to a background contrast set. As our contrast set, we used about 8,500 mouse promoter regions from the genes on the Affymetrix U74Av2 platform not in the pluripotency-associated gene cluster. To calculate the *p*-values, we counted the number of occurrences of each motif cluster in the contrast set and calculated the expected number of occurrences based on a random distribution throughout the genes, N_{exp}. We then counted the number of occurrences of each motif cluster within the mouse promoter regions of the pluripotency-associated gene cluster, N_{obs}. Poisson distribution was then used to calculate the probability p of observing the number of occurrences equal to or greater than N_{obs} by chance given the expected number of occurrences N_{exp}. These *p*-values were Bonferroni corrected by the number of clusters.

Ten motif clusters (Table S1) were selected for further experimental characterization based on the following criteria: $-\log_{10}$ *p*-values greater than zero after correction for multiple testing, evolutionary conservation across different promoter dictionaries, copy number less than 50, and nonrepetitive elements. From our list of ten motif clusters, we identified highly conserved motifs by searching for all occurrences of each motif within the mouse sequence of the alignment between human, mouse, rat, and dog [44] and extended the motif to the flanking regions if the flanking regions were also highly conserved (Dataset S16). A highly conserved position was defined to be a nucleotide base that was conserved across all four species.

Construction of reporter vectors. The 242-bp fragment of *Oct4* DE (−2,181 to −1,939) was PCR amplified from GOF18ΔPE/EGFP plasmid [90] using primers Oct4__2 and Oct4__3 (Dataset S18) containing BglII and BamHI restriction sites. The PCR product was digested and cloned into the BglII/BamHI digested plasmid pFoXLucTK [91]. Nng and motif-containing plasmids were cloned by hybridizing complementary oligos (Dataset S18) that yielded BglII and BamHI restriction site overhangs and ligating them to BglII/BamHI-digested plasmid pFoXTKLuc. Upstream genomic sequences of *Eed* were PCR amplified from the mouse genome with primers Eed2 and Eed4 (Dataset S18) and cloned into the pCRII-TOPO vector (Invitrogen, http://www.invitrogen.com), from which it was subsequently excised by SpeI/XbaI restriction digest and ligated to SpeI/XbaI digested pFoxLuc vector [91]. All plasmids were verified by sequencing.

Cell culture and differentiation. Mouse E14 ES and EG cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum, 1mM L-glutamine, 0.1 mM nonessential amino acids, 100u/mL penicillin, 100 μg/ml streptomycin, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, and recombinant LIF. Mouse Oct4/GFP ES cells [73] were grown in identical conditions except that knockout serum replacement (Invitrogen) was used instead of fetal bovine serum. Human ES H9 cells were cultured in Knockout DMEM supplemented with 20% knockout serum replacement (Invitrogen), 1mM L-glutamine, 0.1 mM nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 10 ng/ml recombinant human FGF-2 on X-ray inactivated mouse embryonic fibroblasts. Embryoid bodies were formed by suspension culturing, and chemical differentiation induction was performed with 0.5 μM all-trans-RA (Sigma, http://www.sigmaaldrich.com/), both in the absence of LIF. HEK293 and NIH-3T3 cells were cultured in DMEM containing 10% fetal bovine serum, 1mM L-glutamine, 100u/ml penicillin, 100 μg/ml streptomycin, and 0.1 mM nonessential amino acids.

Transfection and luciferase reporter assays. 1.5×10^5 cells were plated in 12-well tissue culture plates 24 h prior to transfection. Human ES cells were plated on Matrigel (BD Bioscience, http://www.bdbiosciences.com/), in the absence of mouse feeder cells. Each reporter construct (500 ng) was cotransfected with the pRL-TK vector (200 ng) (Promega http://www.promega.com/) as an internal control using 2 μl of Lipofectamine2000 (Invitrogen), according to manufacturer's instructions. Cells were lysed 24 h after transfection, and luciferase activities were measured using a dual-luciferase assay system (Promega).

Electrophoretic mobility-shift assay. Biotin-labeled double-stranded oligos containing motif 6a and 6b sequences (Figure 4A) were incubated with 10 μg mouse ES cell nuclear extracts using the LightShift Chemiluminescent EMSA Kit (Pierce, http://www.piercenet.com/). The formation of DNA–protein complexes was analyzed by 5% nondenaturing polyacrylamide gel electrophoresis, followed by semi-dry transfer to the GeneScreen membrane (PerkinElmer, http://www.perkinelmer.com/) and biotin detection using the LightShift Kit according to manufacturer's instructions. For supershift assays, 2 μg or 6 μg of α-NF-YA (ab6558; Abcam, http://www.abcam.com/) or α-NF-YB (ab6559, Abcam) were added.

ChIP. ChIP was performed essentially as described in [23] and by Upstate Biotechnology (http://www.upstate.com). Briefly, chromatin was cross-linked using 1% formaldehyde for 10 min, the reaction was quenched with 1/20 volume of 2.5 M glycine and centrifuged at 1,350 ×g for 5 min, and the pellet was washed with PBS and sonicated to obtain fragments of ~100–600 bp, as verified on a gel. Reactions were centrifuged at 20,000 ×g for 10 min and the supernatants were used for incubations with α-NF-YB (FL207; Santa Cruz Biotechnology, http://www.scbt.com/) or α-V5 (ab9137, Abcam) overnight at 4 °C. Dynal Protein G beads (Invitrogen) were used for magnetic recovery of antibody-bound chromatin following the manufacturer's instructions. Crosslinking was reversed by incubation at 65 °C overnight. Reactions were digested with RNAse A and Proteinase K and DNA was purified by phenol–chloroform extraction and ethanol precipitation. DNA concentration was determined using a Nanodrop (NanoDrop Technologies, http://www.nanodrop.com/) and 8 ng were used in Sybr Green real-time PCRs (see below) ran in duplicates or triplicates. Primer sequences are available upon request. Fold enrichment was calculated using the $2^{\Delta Ct}$ method. The gene *Rpl15* was used as control. All PCRs were verified on a gel for the presence of a single band of the correct size.

RNAi and competition assay. shRNA sequences were selected according to published criteria [92]: GFPi-ACAGCCACAACGTCTATAT, Oct4i-GAACCTGGCTAAGCTTCCA, NF-YBi1-GTAGTTCTAGCTCTATCAA, NF-YBi2-GACTAATTGAGGTGTTAAT, NF-YAi1-GAGACAGTTTAGAGAGTAA, NF-YAi2- GAAGTGTTGAGGACATTCA, and control-ACAGCCACAACGTCTATAT. Oligos coding for the shRNAs were designed and cloned into the lentiviral vector pSicoR-mCherry as described [93]. pSicoR-mCherry was generated by replacing mCherry for GFP in pSicoR.

Lentiviruses were produced as described [93]. For transduction, 10^6 ES cells were incubated with virus in 1 ml of ES cell medium (multiplicity of infection 5–10). After 1 h rotating at 37 °C, 2.5×10^5–3×10^5 cells were plated per gelatinized well of a 12-well plate. Cells were passaged and a sample collected for analysis every 2 d. The percentage of mCherry+ cells was determined and mCherry+ and mCherry− cells were isolated using a FACSDiVa (BD Biosciences) cell sorter.

Real-time RT-PCR. RNA was isolated and reverse transcribed using the iScript first strand cDNA synthesis kit (Bio-Rad Laboratories, http://www.bio-rad.com/). The cDNA reaction was diluted 1:5 in TE and used in Sybr Green real-time PCRs (Bio-Rad Laboratories). PCR primers were designed to amplify 100–200-bp fragments spanning two exons at the 3′ end of the gene. Housekeeping genes used were *Ppia* (for mouse), *Ubb,* and *ribosomal protein L7* (for mouse and human), which were determined from the microarray data to not be differentially expressed in the samples analyzed. PCR primer sequences are available upon request. Reactions were run in replicates on a MyiQ qPCR machine (Bio-Rad Laboratories) according to the manufacturer's instructions. Only samples with single and matching end-point melting curve peaks were used for subsequent analysis. Cycle threshold values were imported into the REST software [94] for fold-change calculations, using the housekeeping genes as controls.

Colony formation assay. Cells were infected with lentiviruses containing shRNAs and mCherry, as described above. mCherry+ and mCherry− cells were isolated using a FACSDiVa (BD Biosciences) cell sorter. Three hundred cells were plated per well of a 6-well plate in triplicates. After 10 d in culture, cells were stained for alkaline phosphatase using a Vector kit (http://www.vectorlabs.com/) and colonies were counted.

Cell cycle analysis. Cells were infected with lentiviruses leading to the expression of shRNAs and mCherry, as described above. mCherry+ and mCherry− cells were isolated using a FACSDiVa (BD Biosciences) cell sorter. Cells were washed twice with cold PBS, resuspended at concentration of 2×10^6 cells/ml in PBS and fixed with cold ethanol. After overnight incubation at 4 °C, cells were washed twice and resuspended in 160 μl PBS containing 1% BSA. Twenty microliters of propidium iodide (0.5 mg/ml) and 20 μl of RNase A (10 mg/ml) were added, cells were incubated at 37 °C for 30 min, and analysis was performed using a FACScalibur flow cytometer and FloJo.

Supporting Information

Dataset S1. Normalized Microarray Expression Data
Found at doi:10.1371/journal.pgen.0030145.sd001 (6.3 MB XLS).

Dataset S2. Genes in Microarray Expression Cluster from Figure 1A
Found at doi:10.1371/journal.pgen.0030145.sd002 (96 KB XLS).

Dataset S3. Gene Ontology Analysis of All Categories
Found at doi:10.1371/journal.pgen.0030145.sd003 (785 KB XLS).

Dataset S4. Gene Ontology Analysis of Cellular Compartment
Found at doi:10.1371/journal.pgen.0030145.sd004 (61 KB XLS).

Dataset S5. Gene Ontology Analysis of Biological Process
Found at doi:10.1371/journal.pgen.0030145.sd005 (100 KB XLS).

Dataset S6. Gene Ontology Analysis of Molecular Function
Found at doi:10.1371/journal.pgen.0030145.sd006 (88 KB XLS).

Dataset S7. Affymetrix Probe Sets Used in CompMoby Analysis
Found at doi:10.1371/journal.pgen.0030145.sd007 (30 KB XLS).

Dataset S8. CompMoby Results for 1-kb Mouse Dictionary
Found at doi:10.1371/journal.pgen.0030145.sd008 (41 KB XLS).

Dataset S9. CompMoby Results for 1-kb Human Dictionary
Found at doi:10.1371/journal.pgen.0030145.sd009 (48 KB XLS).

Dataset S10. CompMoby Results for 1-kb Conserved Blocks Dictionary
Found at doi:10.1371/journal.pgen.0030145.sd010 (42 KB XLS).

Dataset S11. CompMoby Results for 2-kb Mouse Dictionary
Found at doi:10.1371/journal.pgen.0030145.sd011 (47 KB XLS).

Dataset S12. CompMoby Results for 2-kb Human Dictionary
Found at doi:10.1371/journal.pgen.0030145.sd012 (42 KB XLS).

Dataset S13. CompMoby Results for 2-kb Conserved Blocks Dictionary
Found at doi:10.1371/journal.pgen.0030145.sd013 (46 KB XLS).

Dataset S14. CompMoby Results for 1-kb Final Dictionary Where All 1-kb Dictionaries Are Clustered by Sequence Similarity
Found at doi:10.1371/journal.pgen.0030145.sd014 (59 KB XLS).

Dataset S15. CompMoby Results for 2-kb Final Dictionary Where All 2-kb Dictionaries Are Clustered by Sequence Similarity
Found at doi:10.1371/journal.pgen.0030145.sd015 (40 KB XLS).

Dataset S16. Conservation of Motif Clusters across Four Species Alignment of Human, Mouse, Rat, and Dog
Found at doi:10.1371/journal.pgen.0030145.sd016 (27 KB XLS).

Dataset S17. Ensembl Probes Used for CompMoby Analysis
Found at doi:10.1371/journal.pgen.0030145.sd017 (14 KB XLS).

Dataset S18. Oligos Used for Motifs
Found at doi:10.1371/journal.pgen.0030145.sd018 (22 KB XLS).

Figure S1. Increase in the Number of Motif Repeats Leads to a Proportional Increase in Motif Activity

Predicted motif sequences 1b, 2c and 8a were fused to a construct containing Firefly luciferase driven by a minimal TK promoter. The number of repeats for a given motif is indicated below corresponding bar graph. Constructs 2c and 8a bear one or two repeats of genomic sequence that each contain two identical predicted regulatory motifs. The constructs were transfected in undifferentiated mouse ES cells. A plasmid containing Renilla luciferase driven by the TK promoter was cotransfected. After 24 h, cells were lysed and assayed for luciferase activities. Firefly values were corrected for Renilla values, and the activities of the tested constructs were calculated relative to the activity of the TK construct, which was set to 1. Representative results from two to five independent experiments are shown. Bars represent averages of triplicates performed in each single experiment. Error bars depict standard deviation.

Found at doi:10.1371/journal.pgen.0030145.sg001 (220 KB PDF).

Table S1. Position of Motifs from Transcriptional Start Sites
Found at doi:10.1371/journal.pgen.0030145.st001 (40 KB XLS).

Accession Numbers

The gene expression profiles of ES cells, PGCs, and somatic cells can be obtained from ArrayExpress (http://www.ebi.ac.uk/arrayexpress/), accession number E-MEXP-1158.

Acknowledgments

We thank Stuart Smith and Michael German for pFOXLuc and pFOXLucTK plasmids, Fanny Polesso and Clara Chiu for excellent technical support, and Amy Heidersbach and Michael McManus for producing lentiviruses. We are grateful to Qing Zhou, Hiram Chipperfield, Sheng Zhong, Doug Melton, and Wing Wong for sharing data prior to publication. We are indebted to Kehkooi Kee, Connie Wong, Kelly Haston, and Renee Reijo Pera for advice on culturing human ES cells and sharing reagents, and Cory Nicholas for human EB cDNA. We thank Renee Reijo Pera, Robert Blelloch, Susan Fisher, Katerina Kechris, Brian Tuch, Jeremy Reiter, Kevin Corbit, and Ryan Baugh for helpful discussions and critical reading of the manuscript.

Author contributions. MG, CC, AGM, HL, and MRS conceived and designed the experiments and analyzed the data. MG, CC, and AGM performed the experiments. MG, CC, HL, and MRS wrote the paper.

Funding. This work was supported by European Molecular Biology Organization and California Institute for Regenerative Medicine (CIRM) postdoctoral fellowships to MG; predoctoral fellowships from the National Science Foundation and the University of California San Francisco (UCSF) Mentorship and Research Assistantship Program to CC; a predoctoral fellowship from the Faculty of Science and Technology in Portugal to AGM; and grants from the UCSF Institute for Regeneration Medicine, UCSF School of Medicine Research Evaluation and Allocation Committee, the National Institute of Diabetes and Digestive and Kidney Diseases Diabetes Endocrinology Research Center, and CIRM to MRS. CC and HL acknowledge partial support from National Institutes of Health grant GM070808 and the Packard Fellowship in Science and Engineering to HL.

Competing interests. The authors have declared that no competing interests exist.

References

1. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. Nature 292: 154–156.
2. Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78: 7634–7638.
3. Pera MF, Trounson AO (2004) Human embryonic stem cells: Prospects for development. Development 131: 5515–5525.
4. Donovan PJ, de Miguel MP (2003) Turning germ cells into stem cells. Curr Opin Genet Dev 13: 463–471.
5. Matsui Y, Okamura D (2005) Mechanisms of germ-cell specification in mouse embryos. Bioessays 27: 136–143.
6. Matsui Y, Zsebo K, Hogan BL (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 70: 841–847.
7. Resnick JL, Bixler LS, Cheng L, Donovan PJ (1992) Long-term proliferation of mouse primordial germ cells in culture. Nature 359: 550–551.
8. Boiani M, Scholer HR (2005) Regulatory networks in embryo-derived pluripotent stem cells. Nat Rev Mol Cell Biol 6: 872–884.
9. Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, et al. (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336: 688–690.
10. Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, et al. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336: 684–687.
11. Niwa H, Burdon T, Chambers I, Smith A (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev 12: 2048–2060.
12. Ying QL, Nichols J, Chambers I, Smith A (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115: 281–292.
13. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, et al. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113: 631–642.

14. Chambers I, Colby D, Robertson M, Nichols J, Lee S, et al. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113: 643–655.
15. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, et al. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95: 379–391.
16. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24: 372–376.
17. Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A (2000) Embryonic stem cell lines from human blastocysts: Somatic differentiation in vitro. Nat Biotechnol 18: 399–404.
18. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, et al. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282: 1145–1147.
19. Pera MF, Andrade J, Houssami S, Reubinoff B, Trounson A, et al. (2004) Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. J Cell Sci 117: 1269–1280.
20. Xu RH, Chen X, Li DS, Li R, Addicks GC, et al. (2002) BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nat Biotechnol 20: 1261–1264.
21. Hay DC, Sutherland L, Clark J, Burdon T (2004) Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells 22: 225–235.
22. Zaehres H, Lensch MW, Daheron L, Stewart SA, Itskovitz-Eldor J, et al. (2005) High-efficiency RNA interference in human embryonic stem cells. Stem Cells 23: 299–305.
23. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122: 947–956.
24. Zhang J, Tam WL, Tong GQ, Wu Q, Chan HY, 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.
25. Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, et al. (2006) Dissecting self-renewal in stem cells with RNA interference. Nature 442: 533–538.
26. Pennacchio LA, Rubin EM (2001) Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet 2: 100–109.
27. MacIsaac KD, Fraenkel E (2006) Practical strategies for discovering regulatory DNA sequence motifs. PLoS Comput Biol 2: e36. doi:10.1371/journal.pcbi.0020036
28. Bussemaker HJ, Li H, Siggia ED (2000) Building a dictionary for genomes: Identification of presumptive regulatory sites by statistical analysis. Proc Natl Acad Sci U S A 97: 10096–10100.
29. Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics 15: 563–577.
30. Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, et al. (1993) Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment. Science 262: 208–214.
31. Liu X, Brutlag DL, Liu JS (2001) BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput: 127–138.
32. Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3: 21–29.
33. Tompa M, Li N, Bailey TL, Church GM, De Moor B, et al. (2005) Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 23: 137–144.
34. GuhaThakurta D (2006) Computational identification of transcriptional regulatory elements in DNA sequence. Nucleic Acids Res 34: 3585–3598.
35. Mootha VK, Handschin C, Arlow D, Xie X, St Pierre J, et al. (2004) Erralpha and Gabpa/b specify PGC-1alpha-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. Proc Natl Acad Sci U S A 101: 6570–6575.
36. Wang T, Stormo GD (2003) Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics 19: 2369–2380.
37. Liu Y, Liu XS, Wei L, Altman RB, Batzoglou S (2004) Eukaryotic regulatory element conservation analysis and identification using comparative genomics. Genome Res 14: 451–458.
38. Siddharthan R, Siggia ED, van Nimwegen E (2005) PhyloGibbs: A Gibbs sampling motif finder that incorporates phylogeny. PLoS Comput Biol 1: e67. doi:10.1371/journal.pcbi.0010067
39. Emberly E, Rajewsky N, Siggia ED (2003) Conservation of regulatory elements between two species of Drosophila. BMC Bioinformatics 4: 57.
40. Dermitzakis ET, Clark AG (2002) Evolution of transcription factor binding sites in Mammalian gene regulatory regions: Conservation and turnover. Mol Biol Evol 19: 1114–1121.
41. Yoshimizu T, Sugiyama N, De Felice M, Yeom YI, Ohbo K, et al. (1999) Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Dev Growth Differ 41: 675–684.
42. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells. Science 298: 597–600.
43. Khatri P, Draghici S, Ostermeier GC, Krawetz SA (2002) Profiling gene expression using onto-express. Genomics 79: 266–270.
44. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434: 338–345.
45. Kuroda T, Tada M, Kubota H, Kimura H, Hatano SY, et al. (2005) Octamer and Sox elements are required for transcriptional *cis* regulation of Nanog gene expression. Mol Cell Biol 25: 2475–2485.
46. Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, et al. (2005) Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem 280: 24731–24737.
47. Blackwell TK, Huang J, Ma A, Kretzner L, Alt FW, et al. (1993) Binding of myc proteins to canonical and noncanonical DNA sequences. Mol Cell Biol 13: 5216–5224.
48. Blackwell TK, Kretzner L, Blackwood EM, Eisenman RN, Weintraub H (1990) Sequence-specific DNA binding by the c-Myc protein. Science 250: 1149–1151.
49. Massari ME, Murre C (2000) Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. Mol Cell Biol 20: 429–440.
50. Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, et al. (2005) LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development 132: 885–896.
51. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663–676.
52. Yuan H, Corbi N, Basilico C, 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.
53. Tomioka M, Nishimoto M, Miyagi S, Katayanagi T, Fukui N, et al. (2002) Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res 30: 3202–3213.
54. Tokuzawa Y, Kaiho E, Maruyama M, Takahashi K, Mitsui K, et al. (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.
55. Nishimoto M, Fukushima A, Okuda A, Muramatsu M (1999) The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol Cell Biol 19: 5453–5465.
56. Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F (2005) Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem 280: 5307–5317.
57. Patil CK, Li H, Walter P (2004) Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. PLoS Biol 2: e246. doi:10.1371/journal.pbio.0020246
58. Perini G, Diolaiti D, Porro A, Della Valle G (2005) In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. Proc Natl Acad Sci U S A 102: 12117–12122.
59. Safe S, Abdelrahim M (2005) Sp transcription factor family and its role in cancer. Eur J Cancer 41: 2438–2448.
60. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349–353.
61. Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239: 15–27.
62. Matuoka K, Chen KY (2002) Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. Ageing Res Rev 1: 639–651.
63. Maity SN, Sinha S, Ruteshouser EC, de Crombrugghe B (1992) Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. J Biol Chem 267: 16574–16580.
64. Sinha S, Maity SN, Lu J, de Crombrugghe B (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc Natl Acad Sci U S A 92: 1624–1628.
65. Zwicker J, Gross C, Lucibello FC, Truss M, Ehlert F, et al. (1995) Cell cycle regulation of cdc25C transcription is mediated by the periodic repression of the glutamine-rich activators NF-Y and Sp1. Nucleic Acids Res 23: 3822–3830.
66. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101: 6062–6067.
67. Yoshikawa T, Piao Y, Zhong J, Matoba R, Carter MG, et al. (2006) High-throughput screen for genes predominantly expressed in the ICM of mouse blastocysts by whole mount in situ hybridization. Gene Expr Patterns 6: 213–224.
68. Li XY, Hooft van Huijsduijnen R, Mantovani R, Benoist C, Mathis D (1992) Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. J Biol Chem 267: 8984–8990.
69. Hu Q, Maity SN (2000) Stable expression of a dominant negative mutant of CCAAT binding factor/NF-Y in mouse fibroblast cells resulting in retardation of cell growth and inhibition of transcription of various cellular genes. J Biol Chem 275: 4435–4444.
70. Bhattacharya A, Deng JM, Zhang Z, Behringer R, de Crombrugghe B, et al. (2003) The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. Cancer Res 63: 8167–8172.
71. Farina A, Manni I, Fontemaggi G, Tiainen M, Cenciarelli C, et al. (1999) Down-regulation of cyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex. Oncogene 18: 2818–2827.
72. Matuoka K, Chen KY (2000) Possible role of subunit A of nuclear factor Y

(NF-YA) in normal human diploid fibroblasts during senescence. Biogerontology 1: 261–271.
73. Ying QL, Nichols J, Evans EP, Smith AG (2002) Changing potency by spontaneous fusion. Nature 416: 545–548.
74. Weinmann AS, Farnham PJ (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods 26: 37–47.
75. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, et al. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38: 431–440.
76. Burdon T, Smith A, Savatier P (2002) Signalling, cell cycle and pluripotency in embryonic stem cells. Trends Cell Biol 12: 432–438.
77. Coustry F, Maity SN, Sinha S, de Crombrugghe B (1996) The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit. J Biol Chem 271: 14485–14491.
78. Serra E, Zemzoumi K, di Silvio A, Mantovani R, Lardans V, et al. (1998) Conservation and divergence of NF-Y transcriptional activation function. Nucleic Acids Res 26: 3800–3805.
79. Coustry F, Sinha S, Maity SN, Crombrugghe B (1998) The two activation domains of the CCAAT-binding factor CBF interact with the dTAFII110 component of the Drosophila TFIID complex. Biochem J 331: 291–297.
80. Zhu J, Zhang Y, Joe GJ, Pompetti R, Emerson SG (2005) NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. Proc Natl Acad Sci U S A 102: 11728–11733.
81. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 98: 31–36.
82. Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, et al. (2004) The Ensembl automatic gene annotation system. Genome Res 14: 942–950.
83. Birney E, Andrews D, Bevan P, Caccamo M, Cameron G, et al. (2004) Ensembl 2004. Nucleic Acids Res 32: D468–D470.
84. Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, et al. (2004) An overview of Ensembl. Genome Res 14: 925–928.
85. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, et al. (2003) The UCSC Genome Browser Database. Nucleic Acids Res 31: 51–54.
86. Chiaromonte F, Yap VB, Miller W (2002) Scoring pairwise genomic sequence alignments. Pac Symp Biocomput: 115–126.
87. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, et al. (2003) Human-mouse alignments with BLASTZ. Genome Res 13: 103–107.
88. Bussemaker HJ, Li H, Siggia ED (2000) Regulatory element detection using a probabilistic segmentation model. Proc Int Conf Intell Syst Mol Biol 8: 67–74.
89. Ben-Dor A, Shamir R, Yakhini Z (1999) Clustering gene expression patterns. J Comput Biol 6: 281–297.
90. Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, et al. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122: 881–894.
91. Smith SB, Ee HC, Conners JR, German MS (1999) Paired-homeodomain transcription factor PAX4 acts as a transcriptional repressor in early pancreatic development. Mol Cell Biol 19: 8272–8280.
92. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, et al. (2004) Rational siRNA design for RNA interference. Nat Biotechnol 22: 326–330.
93. Ventura A, Meissner A, Dillon CP, McManus M, Sharp PA, et al. (2004) Cre-lox-regulated conditional RNA interference from transgenes. Proc Natl Acad Sci U S A 101: 10380–10385.
94. Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30: e36.

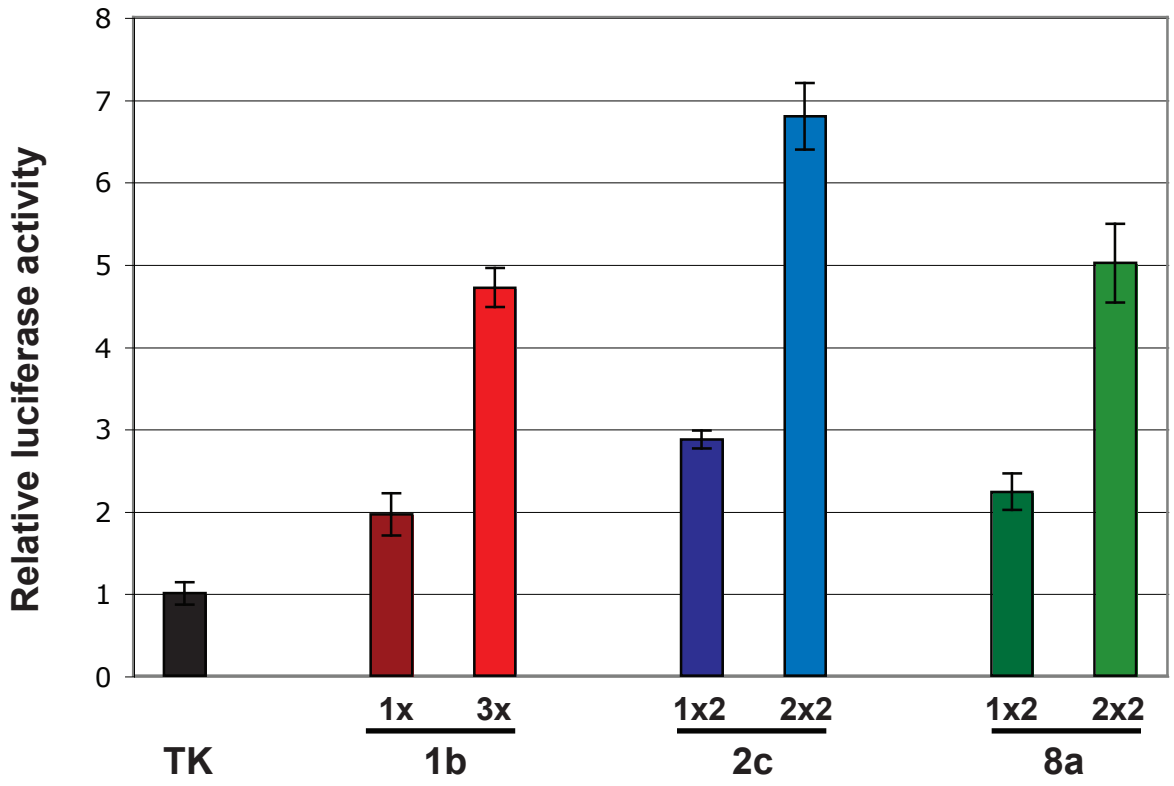


Figure S1. Increase in the Number of Motif Repeats Leads to a Proportional Increase in Motif Activity. Predicted motif sequences 1b, 2c and 8a were fused to a construct containing Firefly luciferase driven by a minimal TK promoter. The number of repeats for a given motif is indicated below corresponding bar graph. Constructs 2c and 8a bear one or two repeats of genomic sequence that each contain two identical predicted regulatory motifs. The constructs were transfected in undifferentiated mouse ES cells. A plasmid containing Renilla luciferase driven by the TK promoter was cotransfected. After 24 h, cells were lysed and assayed for luciferase activities. Firefly values were corrected for Renilla values, and the activities of the tested constructs were calculated relative to the activity of the TK construct, which was set to 1. Representative results from two to five independent experiments are shown. Bars represent averages of triplicates performed in each single experiment. Error bars depict standard deviation.

Chapter 4.

High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming

Connie C Wong, **Alexandre Gaspar-Maia**, Miguel Ramalho-Santos, Renee A Reijo Pera (2008) Plos One 3(4) pp. e1955

OPEN ACCESS Freely available online

PLoS one

High-Efficiency Stem Cell Fusion-Mediated Assay Reveals Sall4 as an Enhancer of Reprogramming

Connie C. Wong[1,2☯¤], Alexandre Gaspar-Maia[2,3,4], Miguel Ramalho-Santos[2,3☯*], Renee A. Reijo Pera[1,2¤*]

1 Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, California, United States of America, **2** Institute for Regeneration Medicine, University of California San Francisco, San Francisco, California, United States of America, **3** Diabetes Center, University of California San Francisco, San Francisco, California, United States of America, **4** Doctoral Program in Biomedicine and Experimental Biology, Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal

Abstract

Several methods allow reprogramming of differentiated somatic cells to embryonic stem cell-like cells. However, the process of reprogramming remains inefficient and the underlying molecular mechanisms are poorly understood. Here, we report the optimization of somatic cell fusion with embryonic stem cells in order to provide an efficient, quantitative assay to screen for factors that facilitate reprogramming. Following optimization, we achieved a reprogramming efficiency 15–590 fold higher than previous protocols. This allowed observation of cellular events during the reprogramming process. Moreover, we demonstrate that overexpression of the Spalt transcription factor, Sall4, which was previously identified as a regulator of embryonic stem cell pluripotency and early mouse development, can enhance reprogramming. The reprogramming activity of Sall4 is independent of an N-terminal domain implicated in recruiting the nucleosome remodeling and deacetylase corepressor complex, a global transcriptional repressor. These results indicate that improvements in reprogramming assays, including fusion assays, may allow the systematic identification and molecular characterization of enhancers of somatic cell reprogramming.

Citation: Wong CC, Gaspar-Maia A, Ramalho-Santos M, Reijo Pera RA (2008) High-Efficiency Stem Cell Fusion-Mediated Assay Reveals Sall4 as an Enhancer of Reprogramming. PLoS ONE 3(4): e1955. doi:10.1371/journal.pone.0001955

Editor: Thomas Zwaka, Baylor College of Medicine, United States of America

Received December 21, 2007; **Accepted** March 2, 2008; **Published** April 16, 2008

Funding: This research was supported by funds from the UCSF Institute for Regeneration Medicine and Juvenile Diabetes Research Foundation (to MRS) and from the Stanford University Institute for Stem Cell Biology and Regenerative Medicine, and the California Institute of Regenerative Medicine (to RARP). The funders were not directly involved in the study or the preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mrsantos@diabetes.ucsf.edu (MR); reneer@stanford.edu (RR)

¤ Current address: Institute for Stem Cell Biology and Regenerative Medicine, Department of Obstetrics and Gynecology, Stanford University, Palo Alto, California, United States of America

☯ These authors contributed equally to the work.

Introduction

The developmental programs of somatic cells are characterized by remarkably stable patterns of gene expression and repression. Nonetheless, through nuclear reprogramming, the developmental programs of somatic cells may be erased and redirected [1–6]. In recent years, much attention has been given to nuclear reprogramming of somatic cells in hopes of generating patient-specific embryonic stem cells (ESCs) that might provide valuable tools for basic science studies and potential novel therapeutics [7,8].

Nuclear reprogramming was first demonstrated as an integral part of mammalian development; following fusion of the egg and sperm, the fused gametic nucleus must be reprogrammed, through a series of changes that include DNA demethylation and chromatin remodeling, to that of an embryonic cell if development is to be successful [5,6,9]. In methods such as somatic cell nuclear transfer (SCNT), the nucleus of a somatic cell is transferred to an enucleated oocyte for reprogramming to an embryonic cell state, through the use of the endogenous machinery [3,10,11]. Methods other than SCNT have also been used to reprogram somatic cells including fusion with ESCs and genetic reprogramming via co-expression of pluripotency-associated genes [12–16]. Each of these methods has advantages and limitations. For example, although SCNT takes advantage of endogenous programs, it requires the use of oocytes that may be in short supply [17]. In the case of cell fusion, although the cells are in great supply, the procedure results in the formation of tetraploid cells that are genetically unstable [12,18–20]. Finally, although genetic reprogramming by co-expression of the stem cell factors Oct4, Sox2, c-myc and Klf4 is remarkable in that it yields ESCs capable of contributing to both the somatic and germ cell lineages, use of the reprogrammed cells to generate offspring results in increased tumorigenesis in progeny [13–16]. Moreover, in all methods, the efficiency of reprogramming is very low, suggesting that additional components of the reprogramming pathways remain to be identified.

In this study, we sought to optimize cell fusion reprogramming protocols, based on fusion of somatic cells and ESCs, in order to screen for enhancers of somatic cell reprogramming. We reasoned that if a factor functions in reprogramming, overexpression of that factor in somatic cells might increase the efficiency with which the cells can be reprogrammed. Thus, we tested whether overexpression of the following factors, individually, increased reprogramming efficiency of MEFs: Oct4, Nanog, Sox2, and Sall4.

Results

Optimization of an Efficient Reprogramming Assay

Several different protocols have been developed to reprogram somatic cells via cell fusion with ESCs, with protocol efficiencies typically less than 0.001% (i.e. ranging from approximately 1 reprogramming event per 1×10^5 to 4×10^6 total somatic cells) [12,20]. Such low efficiencies lead to technical difficulties in screening for positive regulators of somatic cell reprogramming. Thus, we sought to establish an efficient and quantitative reprogramming assay via cell fusion between mouse ESCs and G418-resistant (Rosa26) mouse embryonic fibroblasts (MEFs) that carry the *Oct4-gfp* transgene [20,21]. We began by exploring conditions required for efficient fusion. Traditionally, cells are fused in suspension in 50% polyethylene glycol [12,18–20]. However, we found that the fusion efficiency was substantially increased by both fusing the ESCs and MEFs in adherent cultures and increasing polyethlyene glycol from 50 to 56%. FACS (fluorescent-activated cell sorting) analysis of MEFs and ESCs, which were fluorescently labeled with Vybrant DiD and Vybrant DiO respectively, indicated that the fusion efficiency was 4.6 +/− 0.1% at 5 h post-fusion (Figure 1A).

The first visible, qualitative evidence of reprogramming (within 24 to 48 h post-fusion) was the expression of the *Oct4-gfp* transgene, which was normally silent in MEFs [22]. As time progressed, the reprogrammed MEFs gradually obtained the morphology of ESCs, as reflected by comparisons of forward and side scatter profiles of the GFP-positive MEFs at 24, 48 and 72 h post-fusion (Figure 1B).

Reprogramming efficiency was quantified by determining: 1) the percentage of cells that expressed *Oct4-gfp* and 2) the number of G418-resistant, stem cell-like colonies formed. The percentage of GFP positive cells was measured by FACS at 24 h and 48 h post-fusion, using wildtype MEFs (without the *Oct4-gfp* transgene) as a negative control (Figure 1C). The number of GFP positive particles from the wt MEFs was subtracted from that of *Oct4-GFP* MEFs in order to eliminate any background fluorescence from our calculations. Typical results indicated that the percentage of GFP positive cells, among the total MEF population at 24 h post-fusion, was 0.029 +/− 0.008% and that the number of reprogrammed colonies obtained by this method was found to be as many as 1 in 6.8×10^3 total MEF cells, a value 15 to 590 fold higher than previously reported with other reprogramming assays available [12,18–20]. These reprogrammed cells can be expanded

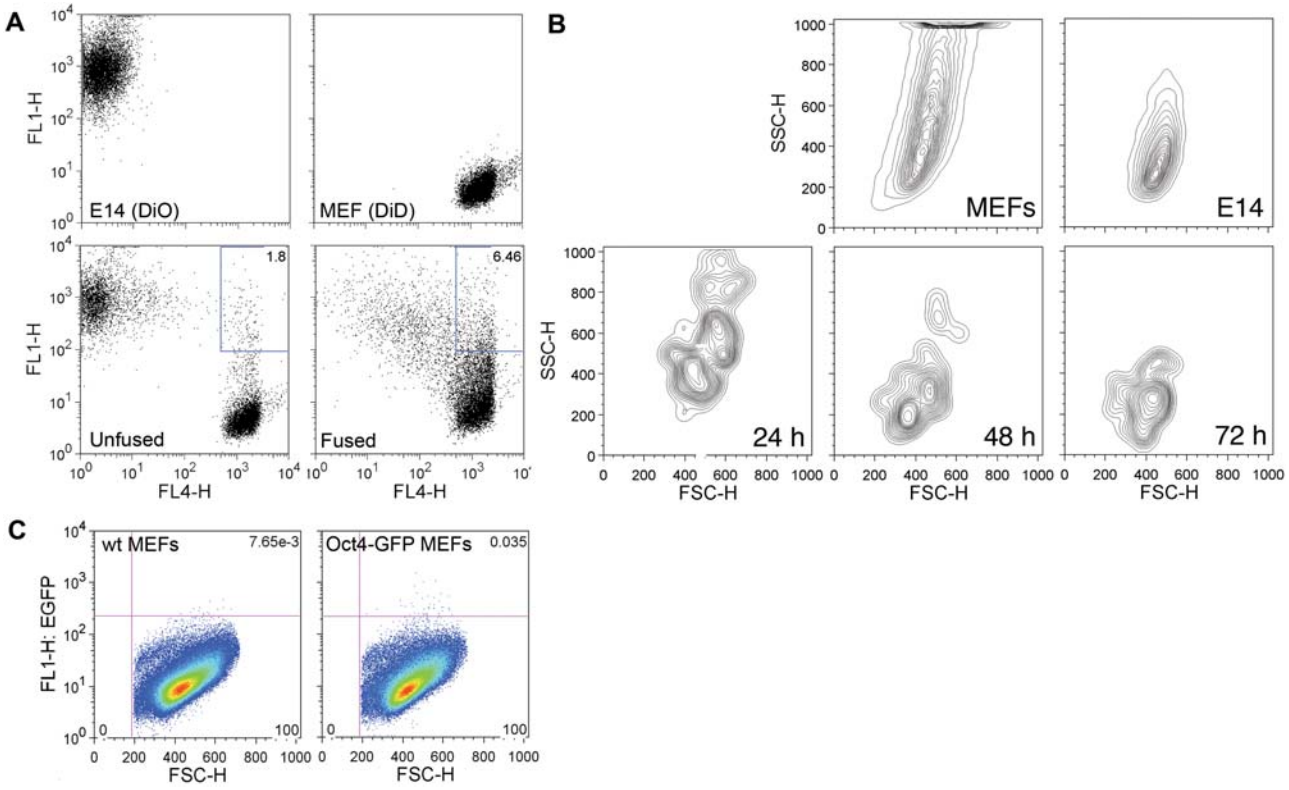


Figure 1. Establishment of an efficient fusion assay. A) Fusion efficiency of the assay. The MEFs and ESCs were stained with the fluorescent dyes Vybrant DiD and Vybrant DiO respectively before fusion. Fusion efficiency was determined by FACS analysis at 5 h post-fusion, using an unfused mixture of cells as a negative control. Note that previous studies have shown that cell surface dyes rarely diffuse across the cell membranes of stained cells [41]. The dual-labeled cells in the unfused population was most likely due to non-specific binding between the ESCs and MEFs. The green fluorescent dye Vybrant DiO was only used in the determination of fusion efficiency but not in a typical reprogramming experiment (due to interference with the observation of GFP signal). B) Morphology change of MEFs during reprogramming. The forward- and side-scatter profiles of the GFP positive cells were FACS analyzed at 24, 48 and 72 h post-fusion. The morphology of the reprogrammed MEFs changed with time to resemble that of the ESCs. C) Quantification of GFP expression. The number of GFP positive cells was FACS analyzed at 24 and 48 h post-fusion (right panel), using wildtype MEFs that did not carry the *Oct4-gfp* transgene but had undergone identical fusion treatment with ESCs as a negative control (left panel). The number of GFP positive cells from wt MEFs was subtracted from that of *Oct4-gfp* MEFs in all calculations.
doi:10.1371/journal.pone.0001955.g001

in ESC culture conditions, are tetraploid and express markers of pluripotency (data not shown).

Potential Enhancers of Somatic Reprogramming

The significant improvements in reprogramming efficiency, brought about by the modifications described above, allow for scaling of the assay to multiple-well formats. We next tested whether our protocol was suitable for quantitative analysis of potential reprogramming factors, including those implicated in maintaining pluripotency and in early embryo development: Oct4, Sox2, Nanog and Sall4 [23–25]. Previously, Oct4, Sox2, and Nanog have been shown to function in somatic cell reprogramming, whereas the role of Sall4 in this process has not been explored [14,18]. A schematic of the protocol is shown (Figure 2A).

Aliquots of G418-resistant MEFs, carrying the *Oct4-gfp* transgene were infected with lentivirus constructs that expressed one of the candidate factors 72 h prior to fusion. The overexpression of candidate proteins was confirmed by Western blotting (Figure 2B). 24 h prior to fusion, infected MEFs were harvested and labeled with the fluorescent dye Vybrant DiD (Figure 2C). The fluorescently-labeled MEFs and unstained ESCs were then plated together in triplicate wells (Figure 2D); the visible overexpression of the red fluorescent protein mCherry indicated proper production and infection of the lentiviruses. Cells were harvested at 24 h and 48 h post-fusion, and the percentage of GFP-positive cells among the DiD-positive MEF population was determined. G418 was then added to the remaining well of fused cells 48 h post-fusion and subsequently at 10 days post-fusion, the number of G418-resistant, GFP-positive colonies was determined.

The onset of *Oct4-gfp* expression provides an initial measure of reprogramming. The percentage of GFP-positive cells, in the population of MEFs that overexpressed each candidate gene, was compared to that of uninfected MEFs and MEFs that overexpressed the negative control proteins, firefly luciferase and the red fluorescent protein, mCherry. At 24 h post-fusion, GFP expression in MEFs that overexpressed negative control proteins was similar to that of the uninfected control, indicating that lentiviral-mediated protein overexpression did not affect GFP expression in MEFs carrying the *Oct4-gfp* transgene (Figure 3A). Unexpectedly, however, MEFs that overexpressed the known reprogramming facilitators, Oct4, Nanog and Sox2, also did not show a significant increase in *Oct4-gfp* expression relative to controls.

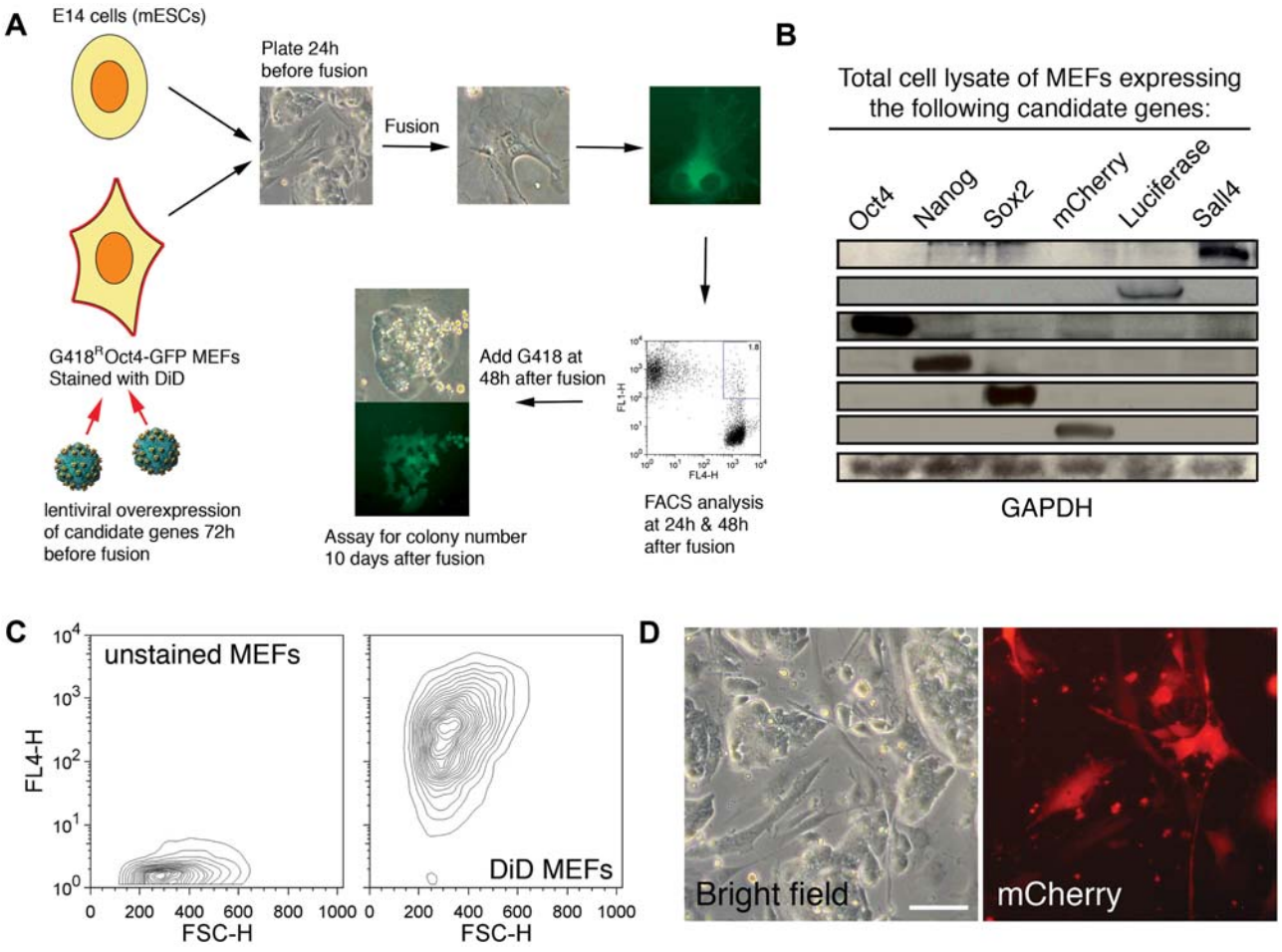


Figure 2. Screen of positive regulators of somatic cell reprogramming. A) Schematic of the screen. Candidate genes were overexpressed in *Oct4-gfp*, G418-resistant MEFs via lentivirus infection 72 h before fusion. B) Successful lentiviral overexpression was verified by Western blotting, as well as expression of the positive control mCherry. C) Infected MEFs were harvested at 24 h before fusion and stained with the fluorescent dye Vybrant DiD. Prepared MEFs were plated with unstained ESCs. GFP expression was FACS analyzed at 24 and 48 h post-fusion. G418 was added to the fused cells at 48 h post-fusion, and the formation of G418-resistant, GFP positive colonies was assayed 10 days post-fusion. D) The visible overexpression of mCherry in infected MEFs indicated the effectiveness of our lentiviral overexpression system. Scale bar represents 50 μm.
doi:10.1371/journal.pone.0001955.g002

Although this was unexpected, the lack of enhanced GFP expression when Oct4, Nanog and Sox2 were overexpressed in MEFs might be attributed to several possibilities (this is further described in the discussion section below). In contrast, we observed that the percentage of GFP-positive cells in MEFs that overexpressed the Spalt transcription factor, Sall4, increased 7-fold relative to controls. The comparison of *Oct4-gfp* expression at 48 h post-fusion was similar to that at 24 h (Figure 3B). MEFs that overexpressed Sall4 consistently demonstrated the highest percentage of GFP-positive cells compared to the other candidate genes.

A second measure of reprogramming is colony formation. With G418 drug selection, GFP-positive colonies began to appear within 5 days post-fusion. The total number of colonies was recorded 10 days post-fusion (Figure 3C). The number of reprogrammed colonies formed by MEFs overexpressing negative control proteins was again similar to the uninfected control, confirming that the lentivirus itself did not affect reprogramming. Nanog, Sox2 and Sall4 all showed a significant increase in the number of reprogrammed colonies relative to controls ($p<0.05$). However, the overexpression of Oct4 in MEFs did not promote formation of reprogrammed colonies in these assays (Figure 3C).

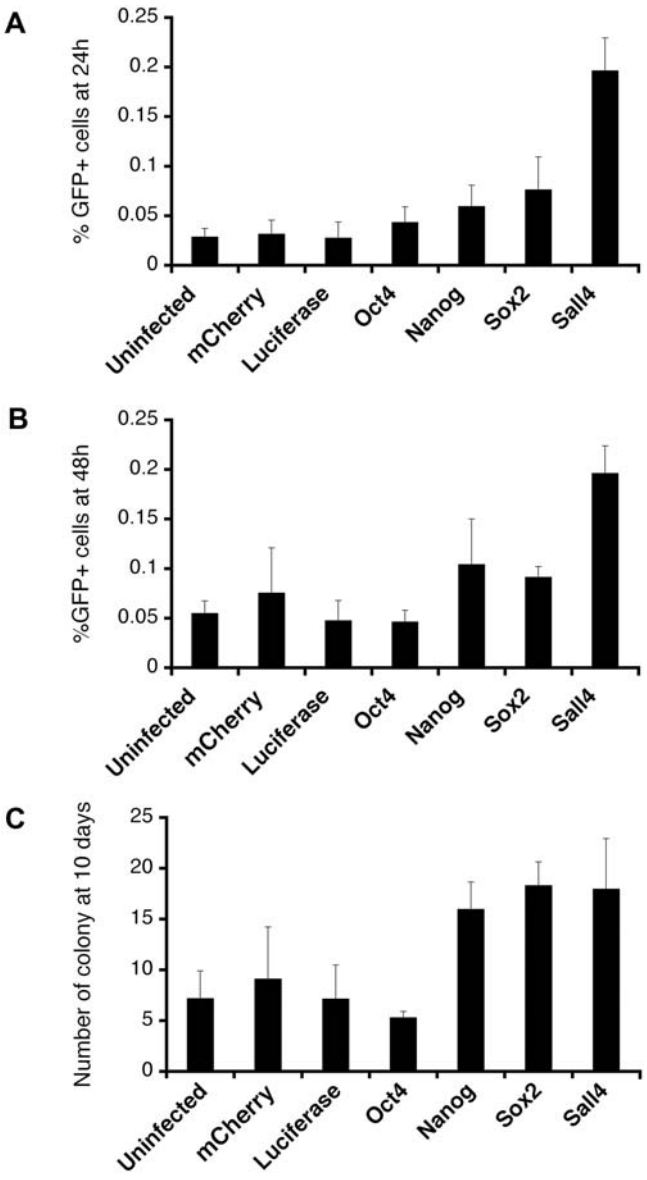


Figure 3. Overexpression of Sall4 enhanced *Oct4-gfp* expression and ES cell-like colony formation in MEFs during reprogramming. The percentage of GFP positive MEFs at: A) 24 h and B) 48 h post-fusion. C) The number of GFP positive, G418-resistant colonies in 1 well of a 6-well plate, 10 days post-fusion. The overexpression of Sall4 positively enhanced both *Oct4-gfp* expression and colony formation of MEFs upon reprogramming.
doi:10.1371/journal.pone.0001955.g003

Confirmation of Sall4 as an Enhancer of Reprogramming by Cell Fusion

As described above, results indicated that Sall4 was likely a positive regulator of somatic cell reprogramming, contributing to both early activation of *Oct4* in the somatic cells and formation of reprogrammed colonies. However, given this data, we also considered whether Sall4 might directly activate the *Oct4-gfp* transgene in the absence of overall reprogramming. To examine this alternative possibility, we tested whether increased GFP expression at 24 h and 48 h post-fusion was due to transcription activity of Sall4 alone. For this purpose, we overexpressed Sall4, as well as the negative controls, in MEFs carrying the *Oct4-gfp* transgene. Half of the infected MEFs were then cultured alone, and the remainder was fused with ESCs. Then, when the 24 h time point would typically be analyzed in a fusion experiment, cells were harvested and the percentage of GFP-positive cells was determined (Figure 4A). We found that the observed increase in the number of GFP-positive cells was dependent on fusion with ESCs; MEFs that overexpressed Sall4 but were not fused with ESCs did not demonstrate increased *Oct4-gfp* expression. This indicated that increased GFP expression in cells overexpressing Sall4 is not a direct effect of Sall4 interacting with the *Oct4* promoter of the *Oct4-gfp* transgene, but rather is a result of the enhancement of reprogramming.

Next, we tested whether overexpression of Sall4 altered the growth rate of MEFs, thus leading to an increased number of colonies unrelated to reprogramming. For this purpose, we overexpressed the negative control, mCherry, and Sall4 in MEFs, plated the cells and determined cell number every 24 h as shown (Figure 4B). An independent clone of Sall4 of identical sequence was used in this experiment as a duplicate; clone 1 was the construct used in all other experiments described in this study. Results indicated that overexpression of Sall4 did not increase, but instead slightly decreased, the growth rate of MEFs relative to the control.

We also addressed whether expression of Sall4 enhanced plating efficiency of ESCs. For this purpose, we used both wildtype ESCs and subcloned lines of tetraploid (4N, Figure 4C) reprogrammed cells. In order to test the effect of Sall4 overexpression on ESC colony formation efficiency, we infected ESCs and 4N cells with constructs that expressed the negative control proteins and Sall4. We observed under a microscope that the fluorescent intensity of the negative control, mCherry, was significantly lower in infected ESCs than in infected MEFs from previous experiments; Western blotting also suggested that the expression level of Sall4 in infected ESCs was not significantly increased relative to endogenous levels (data not shown). This may reflect different activity, or susceptibility to silencing, of the CMV promoter in MEFs relative to ESCs [26]. Nonetheless, we reasoned that the lower expression levels in ESCs parallel observations during reprogramming: when reprogrammed mCherry-infected MEFs gain ESC-like characteristics after fusion, there appears to be a sharp decline of mCherry expression during colony formation (data not shown). After

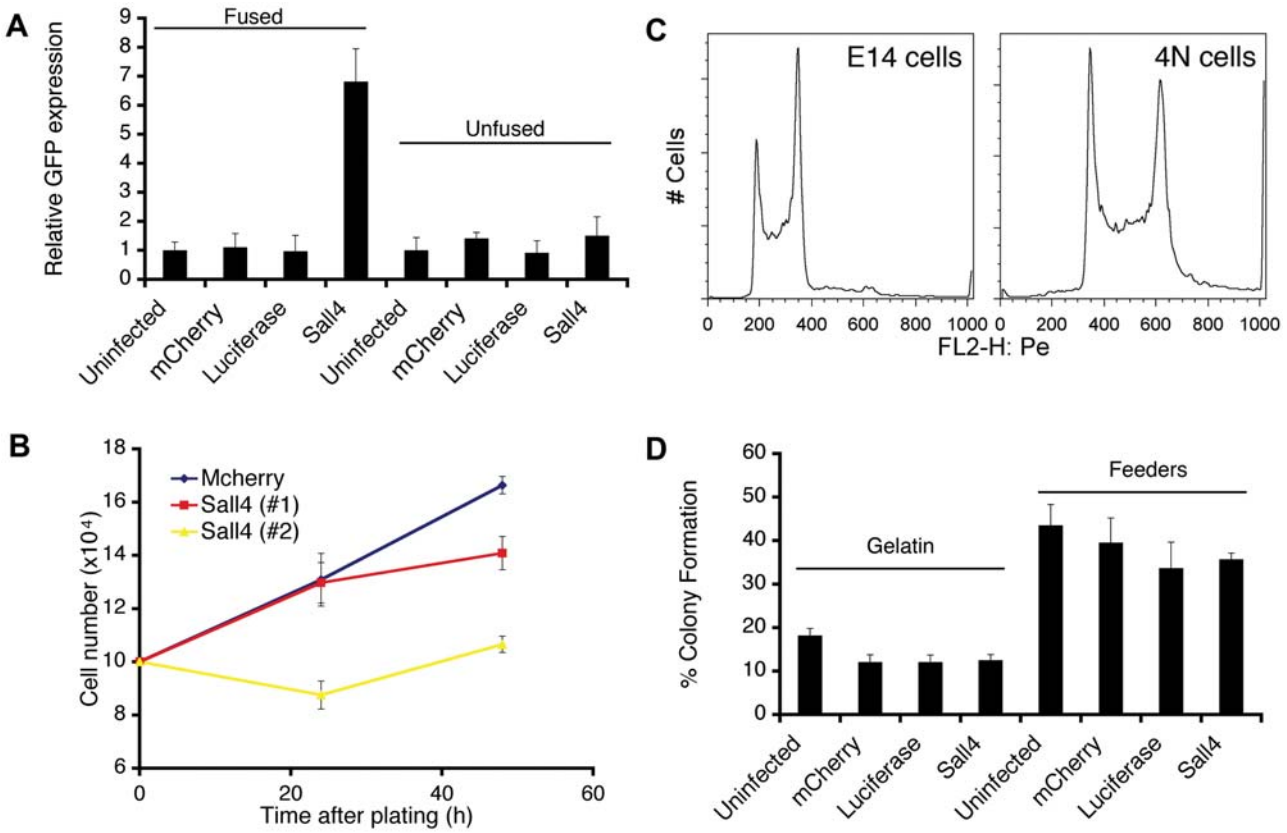


Figure 4. Sall4 is a *bona fide* positive regulator of reprogramming. A) Overexpression of Sall4 in *Oct4-gfp* MEFs did not induce GFP expression. Sall4, mCherry and luciferase were overexpressed in *Oct4-gfp* MEFs via lentivirus infection. Half of the infected MEFs was fused to ESCs as described, while the other half was not. Only MEFs overexpressing Sall4 and fused to ESCs showed an increased number of GFP positive cells when compared to the negative controls, indicating that overexpression of Sall4 alone did not induce GFP expression. The numbers of GFP positive cells in the infected cells relative to that of the uninfected cells were shown. B) Overexpression of Sall4 did not increase cell doubling time in MEFs. mCherry and two different constructs of Sall4 were overexpressed in MEFs, which were plated onto 6 well plates and assayed for cell number every 24 h. Note that another clone of Sall4 of identical sequence was used as a duplicate. C, D) Infection of Sall4-expressing lentiviruses did not increase the colony formation efficiency in ESCs. Both ESCs and previously reprogrammed MEFs that were tetraploid were infected with lentiviruses expressing Sall4, mCherry or luciferase. D) The infected cells were plated either on gelatin or on feeder cells. The number of colonies formed was assayed after 7 days.
doi:10.1371/journal.pone.0001955.g004

infection, we plated the infected ESCs and 4N cells with and without feeders, and assayed colony formation after 7 days (Figure 4D). The efficiency of forming colonies in uninfected ESCs, or ESCs infected with constructs expressing the negative controls mCherry and luciferase, was approximately 20% on gelatin and 45% on feeder cells, similar to previously reported values [27]. We observed that cells infected with the Sall4 construct also did not demonstrate enhanced ability to form colonies in either the presence or absence of feeders. Similar results where obtained with expression of Sall4 in the 4N reprogrammed ESCs (data not shown).

Structure-Function Studies of Reprogramming by Cell Fusion

Taken together, the data described above suggested that Sall4 is a positive regulator of somatic reprogramming. Sall4 is a zinc finger transcription factor expressed in cells of the early embryo and the germ line, and is required for maintenance of pluripotency [28–30]. Sall4 may act as both a positive transcriptional regulator of genes such as *Oct4* [29] and as a transcriptional repressor [31]. The Sall family of proteins contains an N-terminal 12-amino acid motif that recruits the nucleosome remodeling and deacetylase corepressor (NuRD) complex, which is involved in global transcriptional repression and regulation of specific developmental processes [31,32]. The C-terminal region of Sall4 has also been shown recently to contain weak transcription repression activity as well [33].

We sought to determine if our quantitative protocol for reprogramming could be used to dissect the structure-function relationships of factors implicated in reprogramming, such as Sall4. Thus, we tested whether the N-terminal 12-amino acid motif of Sall4 is required for somatic cell reprogramming. For this purpose, we generated a truncated Sall4 mutant (Sall4 d12) that lacked the N-terminal 12-amino acid motif and repeated the fusion assays (Figure S1). We found that overexpression of Sall4 d12 resulted in both early activation of *Oct4-gfp* (Figure 5A) and in increased numbers of ESC-like colonies (Figure 5B), similar to results with wildtype Sall4. We also noted that overexpression of Sall4 d12 did not alter the GFP expression pattern or growth rate of MEFs carrying the *Oct4-gfp* transgene, nor did overexpression increase colony formation efficiency (Figure S2, S3, S4). These data show that the enhancement of somatic cell reprogramming by Sall4 does not require the N-terminal domain of the protein that has been implicated in recruiting the NuRD complex.

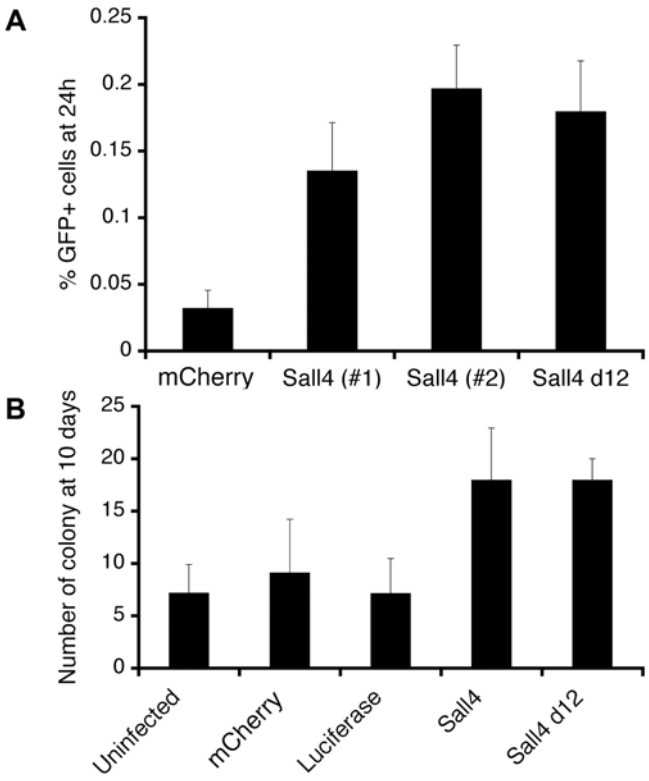


Figure 5. The N-terminal domain is not required for Sall4 function in reprogramming. A) Sall4 d12 mutant behaved similarly to wt Sall4 in reprogramming. Sall4 d12, as well as two clones of wt Sall4 of identical sequences, were overexpressed in *Oct4-gfp* MEFs, which were then fused to ESCs and assayed for GFP expression as described. B) The overexpression of Sall4 d12 resulted in a similar increase in the number of *Oct4-gfp* cells as wt Sall4, as well as a similar increase in the number of reprogrammed colonies.
doi:10.1371/journal.pone.0001955.g005

Furthermore, these data suggest that improvements in the fusion assay may provide a useful platform for future structure-function studies of regulators of reprogramming.

Discussion

In this study, we optimized the cell fusion reprogramming assay. The assay makes use of G418-resistant, *Oct4-gfp* MEFs and mouse ESCs. Due to improved fusion and reprogramming efficiencies, the assay is now potentially amenable to screening formats as demonstrated here with the analysis of overexpression of the pluripotency factors Oct4, Nanog, Sox2 and Sall4. Moreover, by taking advantage of the fact that the *Oct4-gfp* transgene is activated within the first 24–48 h of reprogramming, the assay allows for further physical and molecular characterization of the reprogramming process by microscopy and FACS.

Sall4: An Enhancer of Reprogramming

In this study, we demonstrate that the transcription factor Sall4 can enhance somatic cell reprogramming as evidenced by both enhanced *Oct4-gfp* expression and colony formation. Previously, Sall4 had not been shown to function in somatic cell reprogramming. Sall4 is a member of the Spalt family of transcription factors which was originally identified in *Drosophila* as a homeotic gene required for head and tail development [28,34,35]. In mammals, Sall4 is essential for early embryo development including establishment and maintenance of the early cell lineages of the inner cell mass [30]. Sall4 is also essential for the maintenance of pluripotency and self-renewal of ESCs and for their derivation from blastocysts [30]. Although Sall4 may act as a transcription factor that regulates numerous genes, one of the few known target genes is *Oct4* [29]. Recent studies show that Sall4 interacts with Nanog to control the expression of Oct4 [36]. Together, Oct4, Nanog, Sox2 and Sall4 form a regulatory circuit to maintain pluripotency of ESCs, prompting our exploration of these factors [36–38]. Our results suggesting that Sall4 enhances reprogramming in cell fusion prompts further analyses regarding whether it may enhance reprogramming in other reprogramming strategies; in addition, it is very likely that additional enhancers remain to be identified.

Comparisons to Other Reprogramming Assays

A previous report by Silva and colleagues demonstrated that overexpression of Nanog in mouse ESCs enhances reprogramming of neural stem cells nearly 200-fold and reprogramming of MEFs 10-fold as measured by colony formation [18]. In the current study, when we overexpressed Nanog in MEFs rather than in ESCs, surprisingly, we achieved only a 3-fold increase in reprogramming efficiency as judged by colony formation. Further consideration and comparison of these studies is merited: First, it is apparent from several studies, including that of Silva and colleagues, that it is more difficult to reprogram somatic cells such as MEFs than neural stem cells [18], perhaps due to the state of differentiation of MEFs and/or epigenetic status of key pluripotency genes. Second, we note that overexpression of Nanog in ESCs resulted in greater enhancement of reprogramming efficiency compared to overexpression in MEFs. This observation might reflect fundamental differences between the two studies. Since Nanog is an important pluripotency factor, it is highly likely that the overexpression of Nanog in ESCs may reinforce the pluripotency regulatory circuit, or stem cell properties, of ESCs. In contrast, in our study, the overexpression of Nanog and other positive regulators of reprogramming in MEFs most likely enhances reprogramming by priming and preparing the somatic cell genome for reprogramming. Thus, we suspect, from comparisons of this data to that from other publications, that the latter is a far less efficient process than reinforcing the pluripotency regulatory circuit of ESCs.

Recently, several reports have demonstrated that MEFs can be reprogrammed by co-overexpressing the pluripotency factors Oct4, Sox2, C-myc and Klf4 [13–15]. In our study, neither Oct4 nor Sox2 overexpression in the somatic compartment led to early activation of *Oct4* during reprogramming of MEFs, and only Sox2 led to increased numbers of reprogrammed colonies. It is possible that activation of *Oct4* is not one of the earliest events to occur during reprogramming, and clearly that not all factors that facilitate reprogramming will lead to early activation of *Oct4*. Thus, the lack of early Oct4 activation in our assay does not preclude a factor from being an enhancer of reprogramming. Furthermore, the expression level of Oct4 is regulated in a precise manner in ESCs such that an increase in Oct4 expression level leads to differentiation into primitive endoderm and mesoderm, whereas a decrease results in trophectoderm formation [39]. Thus, we speculate that overexpression of Oct4 alone without other reprogramming factors may actually inhibit reprogramming.

Conclusions

Finally we note that it is possible that the role of reprogramming factors may differ depending on the method of reprogramming, be it SCNT, cell fusion or over-expression of a subset of genetic

factors. The fusion reprogramming assay as optimized here is useful for identification and characterization of new regulators or enhancers of somatic reprogramming and may bypass some of the difficulties with other methods. Together, the array of methods for reprogramming holds great promise for the generation of patient-specific stem cells for use in diverse basic and clinical studies in the future.

Materials and Methods

Cell Culture

Mouse ESCs (E14) were cultured on plates coated with 0.1% gelatin (Sigma-Aldrich, Steinheim, Germany) in ESC medium [(Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Carlsbad, CA) supplemented with 15% knockout serum replacement (Invitrogen), 2 mM L-glutamine (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.57 μM beta-mercaptoethanol (Sigma-Aldrich), 1% penicillin/streptomycin (Invitrogen), and 0.3% leukemia inhibitory factor]. MEFs were harvested from (*Rosa26 X Oct4-gfp*) transgenic mice as described [20,21] and cultured on gelatinized plates in MEF medium [DMEM supplemented with 10% fetal bovine serum (Hyclone Labs, Logan, UT), 2 mM L-glutamine, 1 mM sodium pyruvate, 1% non-essential amino acids and 1% penicillin/streptomycin]. 293 cells for lentivirus production were cultured on gelatinized plates in MEF medium.

Lentiviral Vectors and Overexpression

The lentivirus overexpression vector pLove has been described [40]. The candidate genes were cloned individually into pEntr-1A (Invitrogen), then subcloned into pLove using the Gateway® technology (Invitrogen) according to the manufacturer's protocol. The companion vectors for lentivirus production, pMDL, pRSV and pVSV-G, were gifts from Dr. Michael McManus (University of California, San Francisco, CA).

The 293 cells were plated on 15-cm plates at 80000 cells/cm^2 12–24 h before transfection. 4 μg of pLove and 1.3 μg each of pMDL, pRSV and pVSV-G were transfected into 293 cells with FuGENE 6 transfection reagent (Roche Applied Science, Indianapolis, IN) according to the manufacturer's protocol. Supernatant containing mature lentivirus was harvested at 48 h to 72 h after transfection and filtered with 0.45 μm PVDF syringe filters (Millipore, Billerica, MA). For infection, 10 ml of the filtered supernatant and 5 ml of fresh MEF medium was added to MEFs cultured in 10-cm plates for 24 h. The cells were then rinsed thoroughly 3× with DMEM and continued to culture in fresh MEF medium for another 24 h. Overexpression of candidate genes was verified by Western blotting. MEFs overexpressing the candidate genes were harvested and homogenized in RIPA buffer [50 mM Tris, 150 mM NaCl, 0.5% sodium deoxycholate, 1% NP-40, 0.1% SDS, pH 8.0] at 100000 cells/μl. After a clarifying centrifugation step at 12000 rpm for 20 min at 4°C, 30 μl of 6× Lammeli buffer [0.3 M Tris pH 6.8, 36% glycerol, 10% SDS, 120 μg/ml bromophenol blue] and 2 μl of betamercaptoethanol were added to 60 μl of cell lysate, of which 20 μl was loaded per lane on a 10% SDS-polyacrylamid gel. Western blotting was performed using a goat anti-V5 antibody (Abcam, Cambridge, MA) to detect the expression of all V5-tagged candidate genes, and a goad anti-GAPDH antibody (Abcam) to detect the expression of GAPDH as a loading control.

Cell Fusion Assay

At 24 h before fusion, G418-resistant, *Oct4-gfp* MEFs overexpressing the candidate genes were stained with 0.5 μl/ml Vybrant DiD (Invitrogen) in DMEM for 20 min at 37°C. The cells were thoroughly rinsed 3× with phosphate buffered saline (PBS) before trypsinized and replated on 6-well plates with unstained ESCs; both MEFs and ESCs were seeded at 3×10^5 cells per well in ESC medium. During fusion, the cells were first rinsed 1× with 2 ml PBS (pH 7.4) per well, then primed with 1 ml 50 μM sodium dodecyl sulphate (Sigma-Aldrich) in PBS for 3 min at 37°C before incubating with 1 ml 56% PEG-3350 (Sigma-Aldrich) resuspended in PBS for 1 min at 37°C [41]. DMEM was then added to the wells at 1 ml/min to dilute the PEG solution for up to 5 ml. The cells were rinsed 1× with 2 ml DMEM, 1× with 2 ml ESC medium before returning to 3 ml ESC medium. The medium was fully replaced daily post-fusion. At 24 h and 48 h post-fusion, the fused cells were harvested and resuspended in PBS-1% bovine serum albumin (Sigma-Aldrich) before assaying for GFP expression with a FACSCalibur (BD Biosciences, San Jose, CA). 200 μg/ml G418 solution (Invitrogen) was added at 48 h post-fusion to begin the selection for reprogrammed colonies. In order to control for background fluorescence in our FACS analysis, we fused both wt MEFs that did not contain any *gfp* transgene, and MEFs that carried the *Oct4-gfp* transgene to ESCs independently in our fusion experiments. We measured the number of GFP positive cells in both populations, and we subtracted the number of GFP positive particles of the wt MEFs from that of *Oct4-gfp* MEFs in order to eliminate background fluorescence from our calculations. All fusion experiments were repeated between 3–6 times. The data were then pooled and the average and standard deviation were calculated. Post-hoc tests following a univariate analysis of variance (ANOVA) show that average number of colony for Nanog, Sall4, and Sox2 are significantly different from those of the uninfected, luciferase and mCherry controls ($p<0.05$).

For the analysis of fusion efficiency described in Figure 1A, MEFs were stained with 0.5 μl/ml Vybrant DiD (Invitrogen) and ESCs were stained with 0.5 μl/ml Vybrant DiO (Invitrogen) in DMEM for 20 min at 37°C before cell fusion. Cells were allowed to recover in ESC medium for 5 h before FACS analysis.

Supporting Information

Figure S1 Sall4 d12 overexpression. Overexpression of Sall4 d12 in Oct4-GFP MEFs was verified via Western blotting using antibodies against Sall4 (gifts from Dr. Huck-Hui Ng from Nanyang Technological University, Singapore). Sall4 was expressed in wildtype mouse ESCs but not uninfected MEFs.
Found at: doi:10.1371/journal.pone.0001955.s001 (0.11 MB TIF)

Figure S2 Overexpression of Sall4 d12 in Oct4-gfp MEFs did not induce GFP expression. Sall4 d12 was overexpressed in Oct4-gfp MEFs and the activation of Oct4-GFP was measured as described in the main text and Figure 4A.
Found at: doi:10.1371/journal.pone.0001955.s002 (0.21 MB TIF)

Figure S3 Overexpression of Sall4 d12 did not increase cell doubling time in MEFs. Sall4 d12 was overexpressed in Oct4-gfp MEFs and the doubling time of MEFs was measured as described in the main text and Figure 4B.
Found at: doi:10.1371/journal.pone.0001955.s003 (0.26 MB TIF)

Figure S4 Overexpression of Sall4 d12 did not increase the colony formation efficiency in MEFs. Sall4 d12 was overexpressed in E14 and the tetraploid reprogrammed MEFs, and the colony forming efficiency of the infected ESCs was measured as described in the main text and Figure 4D.
Found at: doi:10.1371/journal.pone.0001955.s004 (0.21 MB TIF)

Acknowledgments

We thank Dr Michael McManus and the UCSF Lentiviral Core Facility for valuable assistance and reagents (293 cells, pSicoR-GFP pMDL, pRSV and pVSV-G). We also thank members of the Ramalho-Santos and Reijo Pera laboratories for valuable discussions and comments on the manuscripts.

Author Contributions

Conceived and designed the experiments: CW MR. Performed the experiments: CW MR. Analyzed the data: CW MR RR. Contributed reagents/materials/analysis tools: CW AG. Wrote the paper: CW RR.

References

1. Yamanaka S (2007) Strategies and New Developments in the Generation of Patient-Specific Pluripotent Stem Cells. Cell Stem Cells 1: 39–49.
2. Eilertsen KJ, Power RA, Harkins LL, Misica P (2007) Targeting cellular memory to reprogram the epigenome, restore potential, and improve somatic cell nuclear transfer. Anim Reprod Sci 98: 129–146.
3. Hochedlinger K, Jaenisch R (2006) Nuclear reprogramming and pluripotency. Nature 441: 1061–1067.
4. Alberio R, Campbell KH, Johnson AD (2006) Reprogramming somatic cells into stem cells. Reproduction 132: 709–720.
5. Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nature Reviews, Genetics 3: 662–673.
6. Surani MA (2001) Reprogramming of genome function through epigenetic inheritance. Nature 414: 122–128.
7. Yang X, Smith SL, Tian XC, Lewin HA, Renard JP, et al. (2007) Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. Nat Genet 39: 295–302.
8. Cervera RP, Stojkovic M (2007) Human Embryonic Stem Cell Derivation and Nuclear Transfer: Impact on Regenerative Therapeutics and Drug Discovery. Clin Pharmacol Ther.
9. Deans B, Griffin CS, O'Regan P, Jasin M, Thacker J (2003) Homologous recombination deficiency leads to profound genetic instability in cells derived from Xrcc2-knockout mice. Cancer Res 63: 8181–8187.
10. Hochedlinger K, Jaenisch R (2002) Nuclear transplantation: lessons from frogs and mice. Curr Opin Cell Biol 14: 741–748.
11. Wilmut I, Schnieke A, McWhir J, Kind A, Campbell K (1997) Viable offspring derived from fetal and adult mammalian cells. Nature 385: 810–813.
12. Cowan CA, Atienza J, Melton DA, Eggan K (2005) Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science 309: 1369–1373.
13. Okita K, Ichisaka T, Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. Nature.
14. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663–676.
15. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, et al. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature.
16. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, et al. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cells 1: 55–70.
17. Hochedlinger K, Jaenisch R (2003) Nuclear transplantation, embryonic stem cells, and the potential for cell therapy. N Engl J Med 349: 275–286.
18. Silva J, Chambers I, Pollard S, Smith A (2006) Nanog promotes transfer of pluripotency after cell fusion. Nature 441: 997–1001.
19. Yu J, Vodyanik MA, He P, Slukvin II, Thomson JA (2006) Human embryonic stem cells reprogram myeloid precursors following cell-cell fusion. Stem Cells 24: 168–176.
20. Tada M, Takahama Y, Abe K, Nakatsuji N, Tada T (2001) Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol 11: 1553–1558.
21. Yoshimizu T, Sugiyama N, De Felice M, Yeom YI, Ohbo K, et al. (1999) Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Dev Growth Differ 41: 675–684.
22. Han DW, Do JT, Gentile L, Stehling M, Lee HT, et al. (2008) Pluripotential reprogramming of the somatic genome in hybrid cells occurs with the first cell cycle. Stem Cells 26: 445–454.
23. Niwa H (2007) How is pluripotency determined and maintained? Development 134: 635–646.
24. Surani MA, Hayashi K, Hajkova P (2007) Genetic and epigenetic regulators of pluripotency. Cell 128: 747–762.
25. Pan G, Thomson JA (2007) Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res 17: 42–49.
26. Xia X, Zhang Y, Zieth CR, Zhang SC (2007) Transgenes delivered by lentiviral vector are suppressed in human embryonic stem cells in a promoter-dependent manner. Stem Cells Dev 16: 167–176.
27. Ying QL, Nichols J, Chambers I, Smith A (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115: 281–292.
28. Sweetman D, Munsterberg A (2006) The vertebrate spalt genes in development and disease. Dev Biol 293: 285–293.
29. Zhang J, Tam WL, Tong GQ, Wu Q, Chan HY, 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.
30. Elling U, Klasen C, Eisenberger T, Anlag K, Treier M (2006) Murine inner cell mass-derived lineages depend on Sall4 function. Proc Natl Acad Sci U S A 103: 16319–16324.
31. Bowen NJ, Fujita N, Kajita M, Wade PA (2004) Mi-2/NuRD: multiple complexes for many purposes. Biochim Biophys Acta 1677: 52–57.
32. Lauberth SM, 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.
33. Bohm J, Kaiser FJ, Borozdin W, Depping R, Kohlhase J (2007) Synergistic cooperation of Sall4 and Cyclin D1 in transcriptional repression. Biochem Biophys Res Commun 356: 773–779.
34. Frei E, Schuh R, Baumgartner S, Burri M, Noll M, et al. (1988) Molecular characterization of spalt, a homeotic gene required for head and tail development in the Drosophila embryo. Embo J 7: 197–204.
35. Jurgens G (1988) Head and tail development of the Drosophila embryo involves spalt, a novel homeotic gene. Embo J 7: 189–196.
36. Wu Q, Chen X, Zhang J, Loh YH, Low TY, et al. (2006) Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem 281: 24090–24094.
37. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, et al. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38: 431–440.
38. Wang J, Rao S, Chu J, Shen X, Levasseur DN, et al. (2006) A protein interaction network for pluripotency of embryonic stem cells. Nature 444: 364–368.
39. Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24: 372–376.
40. Blelloch R, Venere M, Yen J, Ramalho-Santos M (2007) Generation of induced pluripotent stem cells in the absence of drug selection. Cell Stem Cells 1: 245–247.
41. Wong C, Stearns T (2003) Centrosome number is controlled by a centrosome-intrinsic block to reduplication. Nature Cell Biol 5: 539–544.

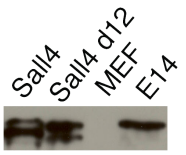


Figure S1 Sall4 d12 overexpression. Overexpression of Sall4 d12 in Oct4-GFP MEFs was verified via Western blotting using antibodies against Sall4 (gifts from Dr. Huck-Hui Ng from Nanyang Technological University, Singapore). Sall4 was expressed in wild type mouse ESCs but not uninfected MEFs.

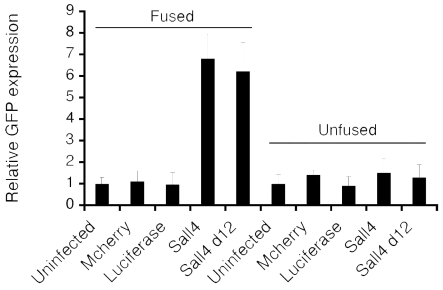


Figure S2 Overexpression of Sall4 d12 in Oct4-gfp MEFs did not induce GFP expression. Sall4 d12 was overexpressed in Oct4-gfp MEFs and the activation of Oct4-GFP was measured as described in the main text and Figure 4A.

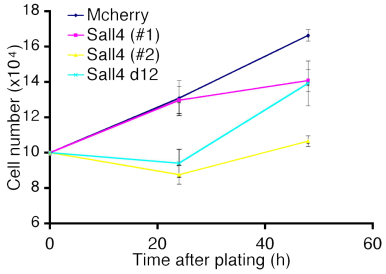


Figure S3 Overexpression of Sall4 d12 did not increase cell doubling time in MEFs. Sall4 d12 was overexpressed in Oct4-gfp MEFs and the doubling time of MEFs was measured as described in the main text and Figure 4B.

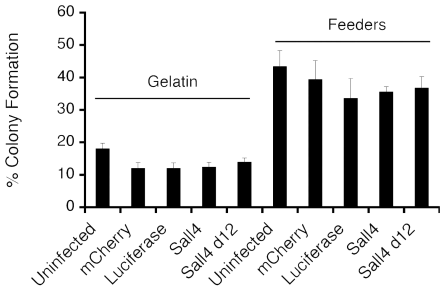


Figure S4 Overexpression of Sall4 d12 did not increase the colony formation efficiency in MEFs. Sall4 d12 was overexpressed in E14 and the tetraploid reprogrammed MEFs, and the colony forming efficiency of the infected ESCs was measured as described in the main text and Figure 4D.

Chapter 5.

Discussion and future directions

Discussion and future directions

RNAi screen

The first aim of my thesis project was to develop an RNAi screen to analyze the effects of the loss-of-function of pluripotent-associated genes in ES cells. A couple of RNAi-based screens in ES cells have been published during this period, using different approaches (Ivanova et al., 2006; Fazzio et al., 2008). In the first RNAi screen, a competition assay was performed between infected cells (undergoing RNAi) and wild type cells (uninfected). The percentage of both infected and uninfected cells was analyzed by flow cytometry, making use of a fluorescent reporter present in the lentiviral vector. In this screen 70 genes were tested, from a list of down-regulated genes upon retinoic acid induced differentiation, through microarray analysis, of which they chose transcription factors or genes containing DNA-binding domains. In this study, aiming specifically for the transcriptional network, 3 novel regulators of ES cell self-renewal were identified (Esrrb, Tbx3, Tcl1) and they were shown to be effectors of the ES cell core network, as its down-regulation could be rescued by over-expression of Nanog (Ivanova et al., 2006).

The second study was performed to analyze the effects in cell growth of the loss of function of more than a thousand loci encoding chromatin proteins. The strategy was based on a RNAi library comprised of RNA digested with an endoribonuclease, EsiRNAs (Kittler et al., 2007). Sixty-eight proteins were identified with knockdown phenotype, including 7 sub-units of the Tip60-p400 acetyltransferase and remodeling complex. This activating complex recognizes independently the histone mark H3K4me3 and Nanog binding sites, integrating the transcriptional network and a chromatin mark to regulate gene expression (Fazzio et al., 2008).

With some important similarities with our approach, especially when compared with the first study, there were still major differences, which allowed for different discoveries. Firstly, the candidate list used was different. The genes tested in our screen were chosen out of a list of genes that were up-regulated (when compared to differentiated cells) in two populations of pluripotent stem cells: one representing an *in vitro* cell line, ES cells, and the other, an *in vivo* representative, PGCs. The rationale being that using *in vitro* and *in vivo* cells, the genes represented would not include some genes possibly being expressed due to the culturing conditions. Our study included 41 genes (see Supplementary Fig.1 in Chapter 2), a list that included mostly transcription factors, chromatin enzymes, DNA/RNA binding genes, oncogenes and some unknown genes. Secondly, the competition assay analysis performed included two reporters to analyze through flow cytometry. The first one enabled us to distinguish infected and uninfected cells, and second one to report for the effect of the down-regulation of a particular gene on the expression of the known regulator and marker of ES cells, Oct4. In fact, by making use of an Oct4-GFP ES cell line, we were able to compare the RNAi effect on the expression of Oct4, in every passage. In effect, this allowed for a more complex analysis of the RNAi phenotype and confirmed through down-regulation of known ES cell regulators, such as Oct4, Nanog, Sox2 and Sall4. This work led to the discovery of 18 genes

with self-renewal defects, 7 of which showed down-regulation of Oct4, and 6 of which were novel regulators of ES cells (Chd1, Cbf, Ddx18, NF-Ya, NF-Yb, Terf1). All the novel regulators were screened using two short hairpin sequences, to verify specificity and to avoid off-targeting effects. The only novel regulator that had both the self-renewal and loss of Oct4-GFP expression phenotypes was the chromatin remodeler Chd1. The lentiviral-based RNAi system (Ventura et al., 2004) was then used to understand the biological function of Chd1 in the undifferentiated state and in the context of reprogramming.

Further studies with the other novel regulators may also shed light on other aspects of ES cell biology. In a joint effort in our lab, the role of NF-Y in ES cells was also dissected (Chapter 3). This transcription factor was identified as a novel regulator of ES cells, mostly through its role in cell proliferation. In this study a novel computational approach was used to predict sequence motifs that can act as enhancers, which were subsequently validated by functional assays. Candidate DNA binding proteins were then tested for their ability to bind specifically to those motifs. This allowed for the identification of 8 conserved motifs, of which one is recognized and bound by NF-Y. The analysis of the down-regulation of two NF-Y sub-units (NF-Ya and NF-Yb), with the use of the RNAi system developed, revealed its role as a regulator of ES cell proliferation. The analysis was performed using both the competition assay and a colony-forming assay, which assesses the ability of ES cells to form single colonies. In addition, the loss-of-function of NF-Y leads to an accumulation of ES cells at G1 phase of their cell cycle, as analyzed by PI staining with flow cytometry.

Furthermore, as a follow-up to the RNAi screen, I started to develop tools to better understand the role of Sall4 in pluripotency. Sall4 is a zinc finger transcription factor highly expressed in the early embryo and germ line, and is required for the maintenance of pluripotency and for the proper development of the embryo (Elling et al., 2006; Zhang et al., 2006). It was also shown to cooperate with Nanog in the transcription network that underlies the ES cell state (Wu et al., 2006) and to be a core element in the ES cell protein interaction network (Wang et al., 2006). Interestingly, the Sall family contains an N-terminal 12-amino acid sequence known to recruit the repressing remodeling complex NuRD (Lauberth and Rauchman, 2006). I cloned a Sall4 over-expressing vector with the wild type sequence, and with a deletion lacking the 12 amino acid sequence in the N-terminal to test if the Sall4 function is required for the recruitment of the NuRD complex. In the context of reprogramming using a cell-fusion assay, the full and the deletion sequences were tested to enhance their efficiency (Chapter 4). Sall4 was demonstrated to enhance the efficiency of reprogramming with both constructs, suggesting that its role in reprogramming is independent of the recruitment of this repressor complex. Further analysis is necessary to understand how Sall4 acts in cooperation with the NuRD complex, knowing that Sall4 binds to many of the genes highly expressed in ES cells in cooperation with Nanog. It is possible that the recruitment of such repressive complexes is context dependent, and it is still unclear if this interaction indeed happens in ES cells. Further biochemical studies would be necessary to understand the interplay between Sall4 and NuRD in the context of ES cells.

In summary, the RNAi screen developed allowed for the discovery of 6 novel

ES cells regulators. Loss of function of NF-Y analysis was integrated in another study involving computational prediction and functional validation of ES cell regulating motifs. The transcription factor Sall4 was also studied in the context of enhancement of reprogramming efficiency using a cell fusion assay. Finally, Chd1 RNAi showed proliferation defects in ES cells and loss of Oct4-GFP expression. I then used the RNAi system to investigate the role of Chd1 in ES cells and in iPS reprogramming.

Role of Chd1 in ES cells

In the study here presented (in Chapter 2), it is shown that the euchromatin protein Chd1, a known chromatin remodeler that recognizes the active mark H3K4me3 (Flanagan et al., 2005; Sims et al., 2005), is an ES cell regulator. Using the RNAi technology, with constitutive expression of the shRNA via a lentiviral vector (Ventura et al., 2004), we were able to dissect the role of Chd1 in ES cells. Depletion of Chd1 leads to an impaired proliferation and loss of pluripotency (Gaspar-Maia et al., 2009). It is possible that the phenotypes observed correspond to a situation closer to a hypomorph, rather then a complete absence of Chd1. Because of the intrinsic caveats of using an RNAi system, two cell lines (E14 and Oct4-GiP) were analyzed holding similar phenotypes and at least 3 colonies expressing the lentiviral vector were expanded per cell line, representing the two shRNAs tested.

Surprisingly, down-regulation of Chd1 in ES cells did not affect the expression levels of the majority of the genes, as analyzed by microarrays, and the cells still presented ES cell markers. The surprise comes from the fact that Chd1 is a factor widely associated with active gene transcription, and its depletion would be expected to have a stronger effect in gene levels. One simple explanation could be the presence of minimal amounts of Chd1, due to the efficiency of RNAi. Another explanation could be the presence of alternative molecules, such as Chd2, that could substitute Chd1 in these complexes for proper gene transcription. Both explanations can be complementary, and go along with other observations that human CHD1 depletion (again using RNAi) does not affect the pre-mRNA levels, but affects the dynamics of the processing, as part of the splicing machinery (Sims et al., 2007).

Nevertheless, Chd1-deficient cells lost their pluripotency capacity as these cells were no longer able to differentiate *in vitro* into primitive endodermal cells, and had a higher propensity to differentiate into the neural lineage, as seen both with EB formation and with teratomas. Further analysis shows that, in ES cells, Chd1 binds to euchromatic and active regions all throughout the genome, with a significant overlap with regions marked with H3K4me3 and bound by RNA polymerase II. Down-regulation of Chd1 leads to heterochromatinization of the nucleus, with increased foci of compacted chromatin marked both by H3K9me3 and Heterochromatin protein 1 (HP1). The histone protein exchange was also impaired in these foci, when evaluated by FRAP analysis. It is known that upon differentiation, specifically into the neuronal lineage, chromatin becomes more compacted with increased foci of heterochromatin and with the concomitant reduction in histone mobility within the nucleus (Meshorer, 2007). We, therefore, analyzed carefully whether this heterochromatinization was a consequence of the propensity of these cells to turn the neuronal program. Nuclear immunostaining shows that the foci of heterochromatin precede the loss of expression of the ES cell marker Oct4. This suggests that Chd1 is acting to prevent heterochromatin from forming in ES cells.

The mechanism by which Chd1 prevents heterochromatinization, tilting the balance between euchromatin and heterochromatin towards the former, is still not known. Global anti-silencing mechanisms have been studied in other species, such as yeast, and despite the differences, it may help us understand the principles that govern this battle between heterochromatin and euchromatin. The Sir proteins (Silent

information regulator) form the ordered compact structure of heterochromatin in yeast, and bind preferentially to the telomeric regions. Two redundant mechanisms have been shown to prevent the spreading of the Sir proteins: the incorporation of a histone variant H2AZ and the methylation of H3K4 mediated by the methyltransferase Set1. In this study it is clear that incorporation of specific histone variants or altering the chemical surface of histone prevents binding of these repressive proteins (Venkatasubrahmanyam et al., 2007). Another important anti-silencing mechanism works through histone hyperacetylation, which prevents Sir proteins from binding (Kimura et al., 2002). More recently, with a single cell resolution, the establishment of silencing mediated by the Sir family protein Sir3, was traced. Introducing the Sir3 protein in a Sir3 mutant cell line, and tracing the onset of the repressive complexes, it was possible to uncover the complexity of this silencing mechanism, involving acetyltransferases (Sas2), histone methyltransferases (Dot1 and Set1) and demethylases (Jhd2). The dynamic process of silencing and activation is played by the competing roles of the addition and removal of the methyl group in H3K4 and H3K79 (Osborne et al., 2009).

To explain the role of Chd1, it is important to take in consideration its role in transcription as an elongation and splicing factor (Sims et al., 2007) and its remodeling activity as an ATPase SWI/SNF like protein (Lusser et al., 2005). One possible mechanism by which Chd1 maintains an open chromatin in ES cells is through deposition of the euchromatic histone variant H3.3, which is generally associated with active genes and is less prone to H3K9 methylation (McKittrick et al., 2004). H3.3 is incorporated by the chaperone HIRA specifically at promoter regions of active genes, and, together with methylation of H3K4 and H3 acetylation, marks these genomic regions through mitosis (Chow et al., 2005). This is thought to be a mechanism by which cells maintain a specific transcriptional memory, as reported in frog reprogramming assays (Ng and Gurdon, 2008). H3.3 incorporation was also shown to mark specific promoter regions during T-cell activation (Sutcliffe et al., 2009). Chd1 has recently been shown to be required in the Drosophila oocyte for incorporation of H3.3 into sperm chromatin (Konev et al., 2007). Chd1 depleted sperm cells are unable to incorporate H3.3, necessary for decondensation, and fail to fuse the male and the female pronucleus after fertilization. It also shows the broad impact for male chromatin that H3.3 incorporation has, and makes us hypothesize that a similar mechanism may maintain open chromatin in ES cells. Unfortunately, mammalian H3.3 variant differs from Histone H3.1 by only four amino acids, and no specific antibody (to date) has been able to distinguish them. This would allow an easy analysis of the regions where H3.3 is enriched, and to study the effect of Chd1 depletion on the deposition of H3.3 incorporation.

Chd1 may also directly protect H3K4me3 from demethylation. Not just by remodeling chromatin variants with enriched H3K4me3, but through direct interaction with histone methyltransferases, such as Ash2 (Sims et al., 2007), necessary for the maintenance of the active mark. This histone mark is known to prevent the binding of repressive complexes such as the NuRD deacetylation complex (Nishioka et al., 2002; Zegerman et al., 2002) and the DNA methyltransferase subunit DNMT3L (Ooi et al., 2007). It can also act in the balance between addition and removal of methyl groups in H3K9. This involves the histone demethylases Jmjd1a and Jmjd2c (Loh et al., 2007), and histone methyltransferases, like SETDB1 (Schultz et al., 2002),

G9a/GLP (Tachibana et al., 2002; Tachibana et al., 2005), Suv39H1 (Rea et al., 2000) and Suv39H2 (Peters et al., 2001), all known to be present in ES cells. Analyzing the expression levels of such players upon down-regulation of Chd1 does not show any differences that would explain the heterochromatinization phenotype. It is important to stress that the foci formed in Chd1 deficient cells, may not be consequence of a dramatic increase of H3K9 methylation, but it could be a result of a higher order compaction of nucleosomes. Preliminary data analyzing H3K9me3 protein levels in these cells do not seem to reveal significant methylation increase.

The opening of the chromatin can also be complemented by histone hyperacetylation, as it was shown in yeast (Kimura et al., 2002). In fact, Chd1 could indirectly mediate the binding of the histone acetylase and remodeling complex Tip60/p400, that recognizes H3K4me3. When this mark is reduced in ES cells, the binding of this complex to its targets is impaired (Fazzio et al., 2008). Here again, the acetylation status of Chd1 deficient cells is still unknown. All of these different mechanisms may orchestrate a complex dynamic regulation of open versus compact chromatin, and I believe genetic studies, similar to the ones performed in yeast, will be important to dissect them.

It remains unclear how this heterochromatinization may act to prevent expression of the endodermal markers, and if the propensity for neuronal differentiation is connected to it. It would be interesting to further analyze H3K9me3 accumulation in Chd1-deficient cells. Through a genome-wide approach (ChIP-chip or ChIP-seq), it would be possible to find regions where this repressive mark is enriched. One problem with such approach would be to circumvent the heterogeneity of the population growing in culture, as it has been shown to have cells expressing either the ES cell marker Oct4, or the neural progenitor markers Blbp or Nestin. Appropriate surface markers for both ES-like cells and neural progenitors should be used to separate these two populations using FACS. This analysis would give insights on the mechanism by which Chd1-deficient cells lose their pluripotency and shed light on the apparent contradiction between the formation of heterochromatin and the unaltered expression levels of most genes in Chd1-deficient cells as seen by microarrays. Heterochromatin marks would be expected mostly in gene-poor regions but could also determine specific silencing of endodermal regulators, such as Gata4 and Gata6 (Morrisey et al., 1996; Morrisey et al., 1998). These genes could potentially be silenced in Chd1-deficient ES cells, but as their expression levels are already low in the undifferentiated state, the microarray data would not be able to reveal differences between wild-type and Chd1-deficient cells. Heterochromatin localization could also help us understand the propensity for neural differentiation. One possibility is that Chd1 may be required for the expression of a repressor of neural differentiation, either as a direct transcriptional activator, or through the opening of chromatin. Candidate genes such as REST and Co-REST (Andrés et al., 1999; Chong et al., 1995) are known to prevent the neuronal fate through recruitment of repressive HDAC complexes (Huang et al., 1999). In fact, REST is highly expressed in ES cells, but recent studies have shown that it may have a broader repressive effect, since its depletion leads to a general up regulation of markers from all lineages (Singh et al., 2008). Another possibility is a more general imbalance of the signaling pathways that are necessary to maintain the undifferentiated state (Silva and Smith, 2008). This metastable state could be generally impaired, and for

instance, the Fgf/Erk signaling could promote a bias lineage commitment towards neural differentiation. It is important to note that the spontaneous differentiation was noted in ES cells grown either in gelatin or with a feeder layer of mouse fibroblasts, which means that the possible signaling fluctuations will be consistent in both culturing conditions.

It would also be interesting to analyze the effect of Chd1 over-expression in ES cells. As Chd1 is important for pluripotency, what effects on the cell would we see with higher levels of Chd1? Would the cells be able to differentiate at all if a major open chromatin remodeler prevented silencing of ES cell specific genes and microRNAs? Would this anti-silencing machinery be prevalent towards the transcriptional cues and other chromatin changes that allow for differentiation? It is hard to believe that Chd1 could dictate such strong phenotype, and eventually render ES cells unable to differentiate. However, some examples of deletion of repressive complexes show "confused" ES cells. Deletion of the NuRD repressive complex protein Mbd3 in ES cells shows various aberrant phenotypes, like the atypical expression of trophectodermal markers, and the maintenance of the ES cell state in the absence of LIF (Kaji et al., 2006); G9a mutants (a H3K9 methyltransferase) also show inefficient silencing of the ES cell regulator Oct4, even after induction of differentiation (Feldman et al., 2006). Even though it is hard to compare the absence of a repressor with the excess of an activator, it surely gives us interesting insights on how the regulation of the undifferentiated state is governed. The evidence so far is that the transcription network (and extracellular signals) may be prevalent as a determinant for the ES cell state and for lineage specification, where chromatin modifiers would act mostly to set the stage for the execution of transcription. In fact, most deletions in chromatin remodelers have mild effects in ES cell proliferation and marker expression, showing only differentiation defects, a critical role in the developing embryo. Addition of the exogenous protein in these mutants is able to rescue the differentiation defects (Niwa, 2007a). This may also be the case with the maintenance of open chromatin by Chd1. Over-expression of Chd1 could instead, allow for the dissection of the critical domains that are involved in opening the chromatin, through the expression of different domains of the protein, as it has been shown in yeast (Simic et al., 2003). Unfortunately, it has been quite challenging to build a stable construct to express Chd1, mostly due to its size, and to the low efficiency of expression of these vectors (with such long transcripts) in ES cells.

Thus, the maintenance of an open chromatin in ES cells through Chd1, and its specific roles in the execution of pluripotency, allowing for lineage commitment into all germ layers, open several avenues. Most of the possibilities raised here are still technically challenging to be tested, and will require novel approaches, which I will discuss below.

Role of Chd1 in reprogramming

As most cells of the body share the same genetic information as the original zygote, reprogramming of a differentiated cell to support the development of an entire organism can teach us a lot about the meaning of the undifferentiated state and issues related to lineage specification such as epigenetic memory and transcriptional and epigenetic regulation. Clearly, reprogramming is a very complex sequence of events that involves overcoming epigenetic barriers, the silencing of the somatic cell program, and resetting of the self-renewing and pluripotency programs. Since we learned about how Chd1 is necessary for the maintenance of the undifferentiated state, it was pertinent to evaluate whether it would be important for the re-acquisition of pluripotency during somatic cell reprogramming, using over-expression of the four factors, Oct4, Sox2, N-Myc and Klf4 (Blelloch et al., 2007). By playing the differentiation process backwards we sought to better understand the role of Chd1 in the maintenance and induction of pluripotency. Making use of the same RNAi technology, Chd1 was depleted in mouse embryonic fibroblasts right before the infection with the four lentiviral vectors. Efficiency of the induction of reprogramming by these 4 factors decreased significantly with Chd1 down-regulation. Since we were analyzing efficiency of reprogramming by the number of iPS colonies, the role of Chd1 in inducing pluripotency could be masked by its possible effect in proliferation of MEFs. We ruled out that possibility by using an indirect measure of proliferation, the MTT assay, allowing MEFs to grow in the same conditions as in the reprogramming assay, with and without Chd1 RNAi. Depletion of Chd1 had no significant effect in proliferation of MEFs. Interestingly, the iPS colonies formed after Chd1 RNAi did not express the RNAi vector, as seen by both expression of the reporter mCherry and the levels of Chd1 in these colonies. This means the RNAi vector had either been silenced in the process of reprogramming or the iPS colonies formed had not been infected with the virus in the first place. More analysis of these colonies would be able to distinguish these two. An easy way to detect the lentiviral integration could be done by probing the DNA using a southern blot. With the limitations of only analyzing the effect of depletion of Chd1 on the end point of reprogramming, these data suggest that Chd1 does not affect expansion of fibroblasts, but inhibits their induction of pluripotency.

When this experiment was first designed, the expectation was that one could learn about the role of Chd1 in pluripotency, specially in maintaining an open chromatin, while the process of opening the chromatin was happening, i.e., during the process of reprogramming. However, it became apparent that, since our knowledge about the process is still limited, it would be very difficult to “see” it as it happens, and investigate the effect of Chd1 depletion. At this point, we are not able to know which cells are going to be reprogrammed, therefore, tracing individual cells through the reprogramming process is still very difficult. In addition, to date we rely on a few markers that indicate changes in somatic cells towards reprogramming. In this study we made use of a MEF cell line that expresses GFP under the control of Oct4, allowing for the analysis of the state of the colonies, with the expression of a known ES cell regulator. But other markers, such as Alkaline phosphatase (AP) and SSEA1, are expressed earlier in the process, which can be used to address the effect of Chd1 at those earlier stages. In fact, preliminary unpublished data indicates that indeed AP positive colonies are also reduced upon Chd1 depletion. This

suggests that Chd1 may be required early in the process, which goes along with the idea of a factor that works mostly on setting the stage for the activation of the transcriptional network.

A critical experiment to functionally address the role of Chd1 in iPS reprogramming would be the analysis of the effect of over-expression in the efficiency of the process. MEFs already express Chd1 but its levels are increased in ES cells and iPS cells (Mikkelsen et al., 2008 and unpublished data). If ectopic expression of Chd1 increases the efficiency of the process, it would be interesting to test if Chd1 could substitute for any of the four factors, or eventually two factors. It has been suggested that Myc is a major contributor in the early events of reprogramming, and its role is clearly related to the multiple targets that may act with specific pathways but also through a more global activation mechanism (Sridharan et al., 2009). Even though Klf4 shares most of its targets with Oct4 and Sox2 (Sridharan et al., 2009), it may also be dispensable in certain conditions, such as an earlier opening of the chromatin. As in ES cells, over-expression of different Chd1 variants could also clarify if there is an interdependency of the structural domains of Chd1, using the reprogramming assay to test it (Woodage et al., 1997).

As it has already been mentioned, the use of oncogenes (Myc) and retro or lentivirus in the current iPS cell establishment protocol raises safety concerns. To generate clinical quality iPS cells, the development of novel reprogramming methods that avoid permanent genetic modification is highly desired. Understanding how the opening of a permissive chromatin can affect the efficiency could encourage different strategies to generate safe iPS cells. This is probably the biggest impact in a short term, for this kind of research. A better understanding of reprogramming may bring technical advantages in terms of safer protocols that allow for its broader use in clinical and therapeutic approaches. In the long term, learning how the cell reverses its somatic program, goes back to an undifferentiated state and then is able to differentiate and establish silencing mechanisms to preserve its new identity, may lead to new approaches and protocols for differentiation. Some cell lineages are still hard to obtain with current protocols, but it may be possible to overcome specific barriers, by using certain chromatin factors. So far, the majority of the effectors of these differentiation protocols have been signaling pathways and transcription factors, but chromatin remodelers may be critical in directing cell fates. As an example, the chromatin remodeling subunit Baf60c has been used to direct ectopic differentiation of mouse mesoderm into beating cardiomyocytes (Takeuchi and Bruneau, 2009).

Open chromatin and pluripotency

An open chromatin largely devoid of heterochromatin is a hallmark of stem cells, from Planarians (Reddien and Sanchez-Alvarado, 2004) to Mammals (Spangrude et al., 1988; Terstappen et al., 1991). Pluripotent Embryonic Stem (ES) cells have an open, hyper-dynamic chromatin, and accumulate regions of heterochromatin upon differentiation (Meshorer et al., 2006). While the open chromatin state in ES cells has been described at different levels, it is still not clear whether this open chromatin is indeed essential for ES cell pluripotency. In this thesis, it is suggested that indeed open chromatin is necessary for ES cells to differentiate into all three germ layers, and that the chromatin remodeler Chd1 plays a role in maintaining euchromatin from being heterochromatinized. Moreover, since the induction of pluripotency requires overcoming epigenetic barriers and a global opening of chromatin, it is possible that Chd1 is involved in that process, together with other players. Treatment with inhibitors of chromatin silencing agents increases efficiency of the reprogramming process: inhibition of DNA methylation (Huangfu et al., 2008a; Huangfu et al., 2008b), histone deacetylation (Mikkelsen et al., 2008) and histone K9 methylation (Shi et al., 2008). In order to understand the link between open chromatin and pluripotency it is now necessary to dissect how Chd1 establishes the boundaries between open and closed chromatin states, and whether the role in maintaining pluripotency results from a local or global effect in transcription (Figure 5.1).

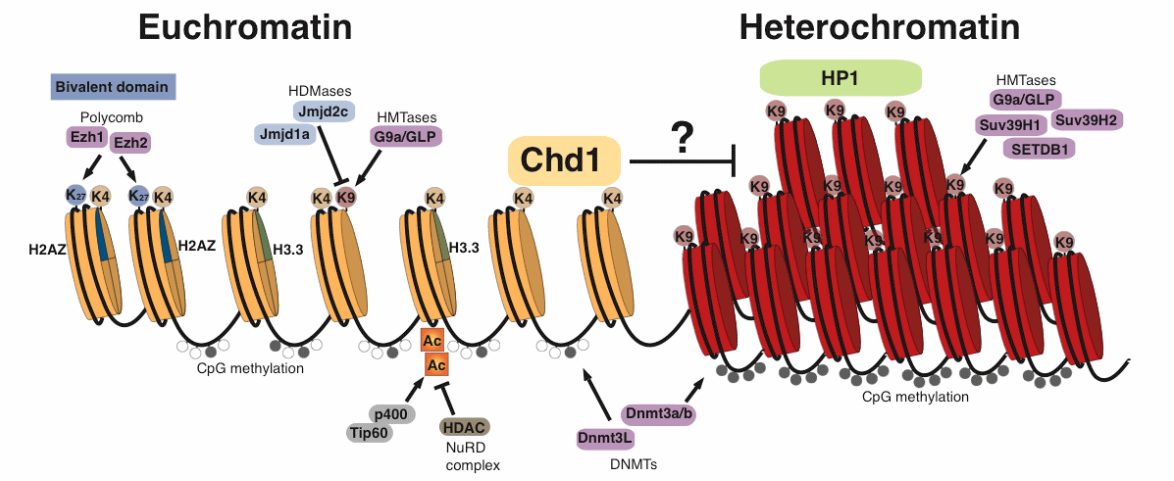


Figure 5.1 The balance between euchromatin and heterochromatin in ES cells is established through Chd1. Several epigenetic regulators orchestrate the transcriptional regulation through modification of marks and incorporation of different core nucleosome (yellow and red cylinders) compositions that alter access and efficiency of the transcriptional machinery. The main histone marks, the active H3K4me3 and the repressive H3K9me3 and H3K27me3 (here represented by the circles K4, K9 and K27) are positively regulated by histone methyltransferases (HTMases) and negatively regulated by histone demethylases (HDMases). Active (K4) and repressive (K27) marks can be present in the promoter regions of developmental genes, to prevent their expression, but allowing rapid activation (bivalent domains). Histone acetylation also marks active chromatin, and the acetyl group (the orange square Ac) can be added through complexes like Tip60/p400 and removed by histone deacetylases (HDACs) that can be part of repressive complexes like NuRD. Different histone variants (H2AZ and H3.3) can be incorporated into the nucleosome by chromatin remodelers, such as Chd1. DNA (black line) methylation is typically present in CpG islands in promoter regions and heterochromatin. It can be hypermethylated (dark grey circles), as the result of DNA methyltransferases (DNMTs), but in euchromatic regions DNA is unmethylated (white circles). Chd1 is essential for the maintenance of an open chromatin state, possibly defining the boundaries between euchromatin and heterochromatin (marked by H3K9me3 and HP1), through an unknown mechanism.

These sorts of studies require nuclear resolution and *in vivo* chromatin structure analysis that are technically challenging, but one could speculate some strategies that may become available. Live imaging of the nucleus coupled with an inducible RNAi and over-expression systems, would allow us to study in detail the onset of the heterochromatinization. This requires fluorescent markers compatible with an *in vivo* setting. Maybe we are far from developing such markers to look at chromatin modifications, but for this particular case, it would be possible to follow an HP1-GFP protein and to make use of DNA tagged regions (genetically engineered, as in Belmont, 2001) to look at specific loci and their relationship with HP1 (for example). These loci would come from a candidate list of genes possibly marked by H3K9me3 in Chd1-deficient cells. Single cell microarrays could also help us identify the consequences of heterochromatin foci formation with the cell state through its expression. The use of genetically modified cell lines and over-expression of the chromatin regulators possibly involved, would then enable us to establish the role of each of the players mentioned. Such approaches could be used both in ES cells as in iPS reprogramming, where the complex induction of an undifferentiated state involves dramatic changes and stochastic events, which would gain much from single cell analysis.

Chd1 in other contexts

The fact that Chd1 has been shown to be up-regulated in mouse (Efroni et al., 2008), human ES cells (Skottman et al., 2005), primordial germ cells (Grskovic et al., 2007) and adult stem cells (Ramalho-Santos et al., 2002) may indicate a common mechanism for its action in maintaining an open chromatin in pluri/multipotency of these cells. Clearly, the role of Chd1 in human ES cells and in induction of pluripotency in human somatic cells is the next step. It will also be interesting to see how the loss of Chd1 in hematopoetic stem cells and neural stem cells affects their ability to proliferate and differentiate. Chd1 could also be involved in setting the stage for lineage specification in these adult stem cells, even if it is easier to speculate that Chd1 could have a more local range of influence in these cells.

It will be also very interesting to understand the role this chromatin remodeler *in vivo*. Because of critical role that most chromatin remodelers have in several functions of the cell and in development, deletion mutants are typically embryonic lethal (de la Serna et al., 2006). On the other hand, mutant ES cells in most cases can be derived. Unfortunately there is no genetic deletion available in mice. A *knockout* mouse model would allow for a better understanding of the role of Chd1 during development. And genetic manipulation strategies that can target Chd1 in specific cell types, would also enable us to study particular adult stem cell populations.

In summary, my work has provided a link between pluripotency and open chromatin in ES cells, indicating that the balance between euchromatin and heterochromatin may be established by Chd1 and that Chd1 is essential for the cells to maintain all the lineage commitment programs available prior to differentiation. This, I hope, paves the way for new approaches in iPS reprogramming and setting a new defining paradigm in understanding the undifferentiated state.

Appendix

RNA interference in embryonic stem cells and the prospects for future therapies

Amy Heidersbach, **Alexandre Gaspar-Maia**, Michael T McManus, Miguel Ramalho-Santos (2006). Gene Therapy 3(6): 478-86

Gene Therapy (2006) 13, 478–486

www.nature.com/gt

RNA interference in embryonic stem cells and the prospects for future therapies

A Heidersbach[1], A Gaspar-Maia[1,2], MT McManus[1] and M Ramalho-Santos[1,2]
[1]Diabetes Center, University of California, San Francisco, CA, USA and [2]Program in Developmental and Stem Cell Biology, University of California, San Francisco, CA, USA

In 1998, two distinct and exciting scientific fields emerged which have profoundly shaped the current direction of biomedical research. The discovery of RNA interference (RNAi) and the derivation of human embryonic stem (ES) cells have yielded exciting new possibilities for researchers and clinicians alike. While fundamentally different, aspects from these two fields may be combined to yield extraordinary scientific and medical benefits. Here, we review the prospects of combining RNAi and ES cell manipulation for both basic research and future therapies, as well as current limitations and obstacles that need to be overcome.
Gene Therapy (2006) **13,** 478–486. doi:10.1038/sj.gt.3302723; published online 16 February 2006

Keywords: *miRNA; ES cells; RNAi; siRNA; differentiation; cell-replacement therapy*

Introduction

The path from discovery to therapy is a long and difficult one. The development of any potential therapy requires: (1) a good model and exhaustive *in vivo* characterization of the disease to be treated, (2) disease target identification and drug discovery and (3) therapeutic development and clinical testing. The combinatorial use of RNA interference (RNAi) and embryonic stem (ES) cells may provide new tools for all of the stages of therapeutic development and yield extraordinary benefits. However, the use of biological tools in potential therapies requires an extensive knowledge of their biological functions. While much work remains to be done in both of these fields before the biology of these systems is clearly defined, significant insights have been gained in the last few years.

Biological function

RNAi

RNAi is a mechanism of post-transcriptional silencing which acts through degradation of mRNA transcripts by the action of homologous short RNA species. Since its characterization in 1998 by Fire *et al*.,[1] RNAi has been the subject of intense investigation, the driving force behind which is twofold. First, RNAi is an ancient evolutionarily conserved mechanism of gene regulation, which is thought to be present in many, if not all, eukaryotic model systems. It has been shown to play an essential role in processes ranging from developmental regulation of gene expression to viral immunity. The second reason RNAi has intrigued the scientific and biomedical communities relates to its practical applications, both in the lab as well as in potential therapies.

Early studies investigated the ability of long dsRNA (generally ranging from 500 to 1000 nucleotides) to initiate an RNAi response in *Caenorhabditis elegans* and plants. These studies showed that dsRNA was able to silence homologous mRNA transcripts,[1,2] resulting in a measurable decrease in gene-specific expression. Additional studies, however, have shown that, in many types of mammalian cells, exposure to long dsRNA generates a non-specific immune response directed by dsRNA-dependent protein kinase (PKR) (for a review, see Kumar and Carmichael[3]). As a result, instead of the sequence-specific mRNA degradation seen in *C. elegans*, the PKR-directed interferon pathway can trigger a global shutdown of translation and apoptosis. Gene-specific silencing by RNAi was successfully achieved in mammalian tissue culture cells in 2001 through the introduction of shorter dsRNA species (less than 21 bp) into cells.[4] These short interfering RNAs (siRNAs) can specifically inactivate genes, minimizing the interferon response. This discovery enabled the use of RNAi-based tools for the large-scale manipulation of gene expression in mammalian systems. It is, however, worth noting that dsRNA does not induce the interferon pathway in all mammalian cell types. Notably, specific silencing has been reported in mouse oocytes/zygotes, ES cells and embryonal carcinoma (EC) cells using long dsRNA.[5–7]

While the general RNAi mechanism is conserved throughout the plant and animal kingdom, some variations in the pathway do exist. In addition to its function as a post-transcriptional gene silencer, there have been several studies which suggest that RNAi may play an important role in the nucleus as a transcriptional gene regulator (for recent reviews, see Matzke and Birchler,[8] Verdel and Moazed,[9] and Bernstein and Allus[10]). In

Correspondence: Dr M Ramalho-Santos, Program in Developmental and Stem Cell Biology, UCSF, 513 Parnassus Avenue, HSW1201G, San Francisco, CA 94143-0525, USA.
E-mail: mrsantos@diabetes.ucsf.edu
Received 17 August 2005; revised 21 December 2005; accepted 3 January 2006; published online 16 February 2006

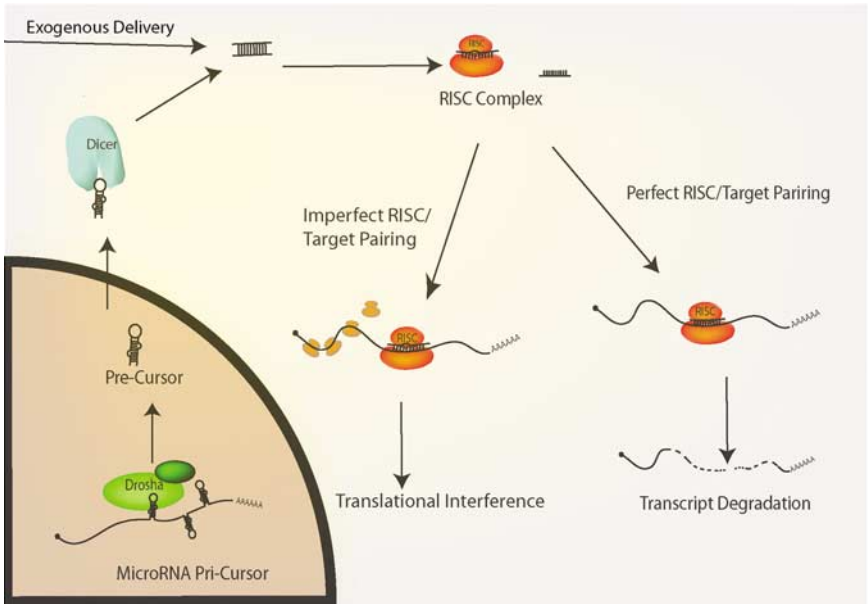


Figure 1 miRNA processing and RNAi in mammals. miRNA processing begins in the nucleus where long primary miRNAs are recognized by the associated proteins Drosha and Pasha. These enzymes cleave the long pre-cursor into short ~75 nucleotide (pre-cursor) hairpins which are then transported out of the nucleus. In the cytoplasm, the enzyme Dicer recognizes the pre-cursor and cleaves it into a 19–21 nucleotides RNA duplex with characteristic 2 nucleotides 3′ overhangs. A single strand of the duplex originating from a Dicer-processed pre-cursor or an exogenous siRNA is incorporated into the RNAi silencing complex (RISC). Depending on the degree of complementarity between the siRNA and its target mRNA, RISC may either block the translation machinery or cleave the target.

particular, the mechanism by which long dsRNA is tolerated and processed in various organisms differs. In plants and invertebrates, where mRNA transcript degradation is far more common than in mammalian systems, siRNAs are processed from either long dsRNA or very long hairpin species. In these systems, siRNAs can be generated from both strands of the RNA duplex and multiple siRNAs can be generated from a single long RNA. In fact, an important function of the plant 'immune' system is dependent upon its ability to take exogenous dsRNA, like that introduced by an invading virus, and process the dsRNA into siRNA. After an amplification step catalyzed by the RNA-dependent RNA polymerase (RdRP), these 'preprogrammed' siRNA species seek out and target other homologous viral particles (for recent reviews, see McManus,[11] Herr[12] and Wang and Metzlaff[13]). Instead of inducing viral immunity as seen in plants, the introduction of long dsRNA in a mammalian system induces the interferon response generating universal gene silencing and apoptosis. Such differences profoundly affect the interspecies application of RNAi. An effective long dsRNA-based gene-silencing technique in *C. elegans*, for example, could likely be ineffective or dangerous in a mammalian system. This is why a thorough understanding of such differences is essential before any potential gene therapy can make the transition from the lab to clinical trials.

Several years after the initial description of RNAi, data emerged that suggested a wide variety of organisms might use small RNAs to regulate gene expression.[14–16] This discovery shed light on the earlier finding that the important developmental gene in *C. elegans lin-4* does not encode a protein but instead a non-coding RNA species.[17] Further research has revealed that the *lin-4* gene is not a unique anomaly but only the first of thousands of microRNAs (miRNAs) to be identified.

miRNAs are transcribed as long primary transcripts (can be more than 2000 nucleotides). These primary transcripts are processed by the Drosha enzyme to yield a short hairpin miRNA precursor of approximately 75 nucleotides. This precursor is exported from the nucleus, where it is processed by the Dicer enzyme to yield a transiently existing ~21 nucleotide RNA duplex. Depending on the thermodynamic asymmetry of the duplex, a single strand is preferentially loaded into the silencing complex. This short RNA strand is considered to be the mature miRNA. The overall pathway is displayed in Figure 1, although there are many additional factors and details that are omitted (for recent reviews, see Hammond,[18] Tomari and Zamore[19] and Hutvagner[20]).

The RNAi pathway, induced through the introduction of synthetic siRNA or short hairpin RNA (shRNA), is very similar to the pathway by which the endogenous miRNAs are processed. In general, a high degree of complementarity between the mRNA transcript and the loaded silencing complex usually leads to RNAi-like degradation of the target transcript. If, however, there is a lower degree of sequence complementarity, the loaded complex may merely interfere with translational machinery, inhibiting protein production, through a largely unknown mechanism and usually leaving the mRNA intact. In mammals, miRNAs tend to mediate their effects through translational repression, although exceptions to this general rule can be found. The mouse miRNA miR-196, for example, pairs exactly with the Hoxb8 mRNA transcript and directs an RNAi-induced mRNA cleavage.[21,22]

Although the mechanism of RNAi has only very recently been elucidated, RNAi has quickly become one of the most popular methods of gene silencing in the lab. There are several benefits to utilizing RNAi over other gene silencing methods. siRNA silencing strategies, for

example, have been shown to be 100 times more effective than other antisense oligonucleotides (ODNs), at silencing the same target.[23] Although chemical modification of synthetic ODNs have led to more efficient delivery, they also tend to result in a decreased sequence specificity as well as an increase in toxic side effects. Additionally, RNAi-induced silencing tends to be more stable and have fewer toxic side effects when compared with silencing induced by other synthetic nucleotides such as DNA oligos or ribozymes, perhaps because RNAi harnesses an endogenous cellular pathway.

Studies in mouse ES cells have shown that embryos derived from RNAi-treated ES cells can recapitulate the phenotype of the conventionally derived null animal.[24] Use of RNAi to induce gene-silencing offers several advantages when compared with the practice of generating targeted genetic deletions in mouse 'knockouts'. In addition to the substantial time and cost required to produce a 'knockout', complex models involving the altered function of multiple genes may be very difficult to produce with current approaches. Techniques such as blastocyst injection of RNAi-encoding viruses or the implantation of RNAi-modified ES cells can produce transgenic animals in months rather than years. These techniques can be used to study normal tissue function and disease by varying the level of gene expression instead of completely abolishing it. In effect, this may provide researchers with a molecular 'tuning dial' instead of simply an on/off switch. It should be noted that RNAi-based knockdown strategies are not likely to replace conventional gene knockout techniques, but instead provide a complementary tool that may have particular advantages in gene therapies.[25]

Depending on the model system being studied, or the disease to be treated, a wide variety of methods may be employed to induce RNAi-mediated gene silencing, each of which has its own distinct advantages and disadvantages. In *C. elegans*, inducing stable RNAi is as simple as soaking the animals in a solution of dsRNA or feeding them transformed bacteria which produce long dsRNA.[26,27] Unfortunately, for those researchers not studying nematodes, these delivery methods cannot be applied to most model systems. Probably, the simplest and most versatile method used in the lab to silence gene expression *in vitro* is to design siRNA duplexes which target a gene of interest and insert them into cells using a variety of transfection techniques. Although this strategy is both rapid and inexpensive, there are limitations. Unlike *C. elegans*, mammalian cells do not contain RdRP for the amplification of siRNA. As a result, the effects of transfected or injected siRNA in a mammalian cell decrease as the moiety is diluted with cell division. Thus, a simple injection delivery method of siRNA does not provide stable long-lasting RNAi silencing in mammals. Long-term stability, however, may not be necessary in some gene therapies, such as two recently approved by the FDA for clinical trials, for the treatment of AIDS-induced age-related macular degeneration. These therapies involve a local injection of the 'naked' unpackaged siRNA directly into the eye.[28,29] Local delivery may reduce the likelihood of potential off-target effects elsewhere in the body and the transient nature of the treatment may actually be beneficial because it limits unknown, potentially negative side effects that may occur from long-term expression.

Local delivery of naked RNAs can be effective when targeting accessible organs such as the eye, the skin or the lungs. For less accessible organs, however, viral vectors may be a useful alternative. These systems often encode shRNAs, which are processed much like miRNAs into siRNAs. Unlike naked RNA, some of these viral vectors, such as adeno-associated viruses (AAV) and lentiviruses, can integrate into the host genome leading to a more permanent expression of a shRNA or siRNA. These viruses can infect non-dividing cells, such as primary neurons, making them important tools in therapies for diseases that target the CNS. shRNA-encoding AAVs have, for example, been used to effectively silence a deleterious gene in the brains of mice with spinocerebellar ataxia, which is similar to the human neurodegenerative disorder Huntington's disease.[30]

Lentiviral-based systems have recently become a very popular way to deliver small RNAs. A variety of lentiviral plasmids are available, containing various selectable markers driven by many different promoters (including inducible systems).[31–36] When pseudotyped with VSV-G, lentiviruses are highly tropic for stem cells and can easily be produced in high-titer if concentrated by high-speed ultra-centrifugation. These characteristics have made lentiviruses very useful in the lab both in ES cell culture as well as in the creation of transgenic animals from modified ES cells (for a review, see Pfeifer[37]). Although still a relatively new strategy, shRNA-encoding lentiviruses hold promise for therapies. Such vectors, ironically derived from the human lentivirus HIV, have shown promise in silencing various components necessary for HIV infection and replication (for a review, see Cullen[38]).

Large-scale screens of RNAi libraries are very useful tools for identifying novel gene function and dissecting the biology of cellular pathways. Various academic and commercial groups have created multiple types of RNAi libraries. These libraries range from collections of shRNAs or siRNAs designed to target a specific gene or group of genes, to constructs derived from the enzymatic digestion of cDNAs.[39–42] RNAi screens are very versatile because they can be applied to most systems. Thus far, RNAi screens have helped to identify novel genes involved in everything from cell division to apoptosis to fat metabolism. These tools can be especially useful in dissecting complex regulatory pathways. RNAi screens are usually carried out by one of two methods. One method involves the transfection (or infection) of a pooled RNAi library into cells followed by selection and analysis of cells expressing a phenotype of interest. Alternately, large-scale RNAi libraries can be arrayed and analyzed in a high-throughput manner. RNAi screens are simple and cost-effective tools for elucidating gene function and dissecting biological pathways, and they are also rapidly becoming essential in the process of identification and validation of potential gene therapy targets. The use of RNAi libraries could allow the rapid identification of effective targets minimizing investment on the development of drugs against ineffective targets.

ES cells

Stem cells are characterized by their ability to proliferate in an undifferentiated state and to give rise to differentiated progeny. There are two major kinds of stem

cells: ES cells and adult stem cells. ES cells can be expanded extensively in culture because of their self-renewing capacity. They are also pluripotent, that is, they have the capacity to generate differentiated progeny from all three embryonic germ layers (endoderm, mesoderm and ectoderm)[43,44] as well as the germ line.[45] In contrast to ES cells, adult stem cells such as neural stem cells or hematopoietic stem cells have a more restricted differentiation capacity and usually generate cells of the tissue from which they are derived. Adult stem cells are maintained throughout the life of the organism by their ability to self-renew.

ES cells were first derived in 1981 from the inner cell mass of the mouse blastocyst.[43,44] Before the derivation of mouse ES cells (mES cells), it had been shown that some tumors called teratocarcinomas behaved as a pluripotent and self-renewing population *in vitro*.[46,47] Cell lines derived from these tumors are called EC cells.[48] ES cell lines have most of the molecular, morphological and growth characteristics of EC cell lines. Unlike EC cells, mES cells can contribute to all tissues when injected into blastocysts, including to the germ line.[45] In 1992, another pluripotent cell type was isolated, this time from mouse primordial germ cells (PGCs). These cells are called embryonic germ (EG) cells and resemble both mES and EC cells.[49,50]

In 1998, the same year of the discovery of the RNAi pathway, a major event for ES cell research took place: the derivation of human embryonic stem (hES) cell lines[51] (Figure 2). hES cells are derived from the inner cell mass of blastocysts at about 1 week post-fertilization. Like mES cells, hES cells are a self-renewing and pluripotent population. Injection of hES cells into immunocompromised mice results in the formation of teratomas, containing cells from the three embryonic layers. Owing to obvious ethical reasons, it is unclear whether these cells can contribute to a human embryo when introduced into the blastocyst. Also in 1998, human embryonic germ (hEG) cells were derived from gonadal ridges containing PGCs (5–9 weeks post-fertilization).[52]

Both mouse and human ES cells can be propagated in the presence of serum and co-cultured with a layer of fibroblasts. However, they require different signals to self-renew. mES cells require leukemia inhibitor factor (LIF) and bone morphogenic proteins (BMPs), whereas hES require fibroblast growth factor (FGF) and suppression of BMP signaling.[53–57] Despite differences in the signals for self-renewal, the regulation by transcription factors appears to be conserved between mES and hES cells. In particular, the transcription factors Oct4[58] and Nanog[59,60] are required to maintain both mouse and human ES cells in an undifferentiated state. This requirement was recapitulated with RNAi: downregulation of Oct4 or Nanog in mES cells and hES cells led to the loss of pluripotency and self-renewal capacities.[61–64] This suggests that both transcription factors have similar roles in mouse and human ES cells. Little is known about the mechanisms by which these transcription factors maintain pluripotency, particularly because their targets have for the most part not been identified. In addition, the regulation of the cell cycle and of lineage commitment in ES cells remains poorly understood. Clearly, much work remains to be carried out to dissect the regulation of ES cell self-renewal and pluripotency and RNAi is likely to be a very powerful tool for this purpose.

The derivation of hES cells opened the possibility of using nuclear transfer techniques to produce cell lines that carry the genetic information of a human donor. By taking any somatic cell from an individual, and introducing its nucleus into an enucleated oocyte, it is possible to generate a new diploid oocyte with a nuclear genome identical to the donor. This oocyte can form a blastocyst, from which it is possible to derive nuclear transfer (NT)-ES cells. Proof-of-principle for this approach has been provided in the mouse.[65] These cells can then be used for therapeutic purposes which will be further discussed. For ethical reasons, it is important to distinguish between therapeutic cloning and reproductive cloning. Therapeutic cloning involves

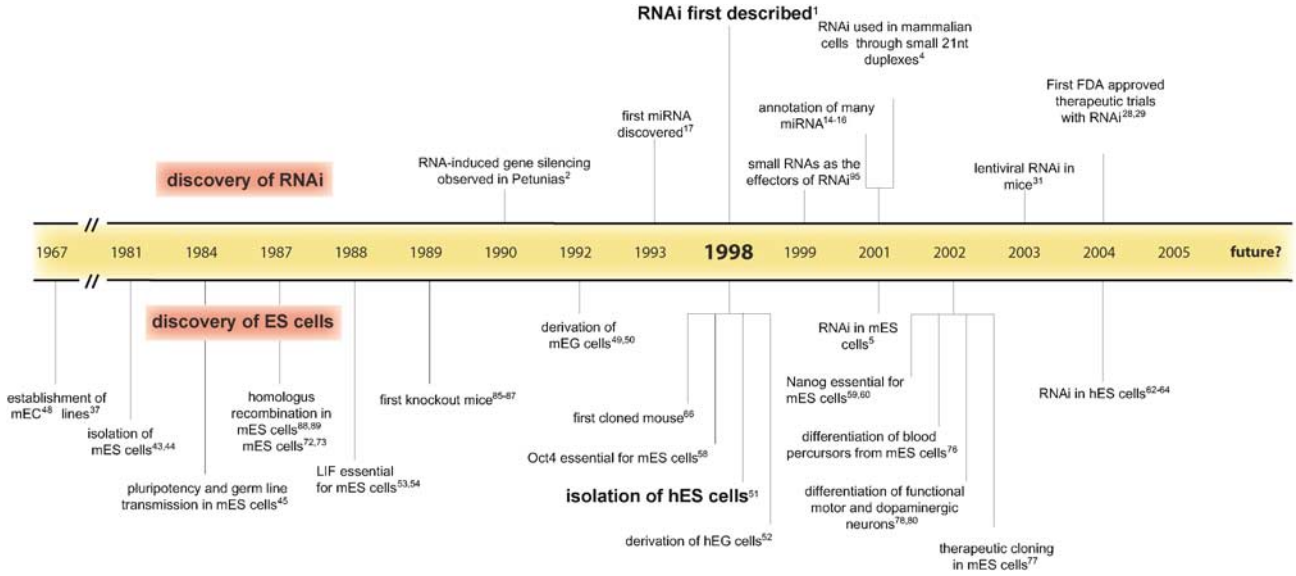


Figure 2 Timeline of RNAi and ES cell discovery. This timeline highlights some of the most important steps in the discovery of both RNAi and ES cells where the year 1998 is an important milestone. Numbers in superscript refer to references in the text.

the use of NT-hES cells for cell-based therapies. Reproductive cloning involves the implantation of a cloned embryo in the uterus to create a entire organism, which has only been reported in some animal models, like the mouse.[66]

The use of RNAi in dissecting ES cell biology and differentiation

So far, the undifferentiated state of ES cells has been studied through gain- and loss-of-function studies that have described the importance of a few genes like Oct4,[58,67] Nanog,[59,60] Foxd3[68] and Sox2.[69] This approach has been limited to the study of mice that are mutant for each of these genes. RNAi allows researchers to test the role of many genes in ES cells without the need to generate mutant mice. The specificity of RNAi can be confirmed by targeting sequences in untranslated regions of the mRNA and then rescuing the phenotype by overexpressing their coding sequence. In addition, using RNAi against various genes simultaneously can help to clarify the pathways that maintain pluripotency. Most of the genes so far shown to regulate the undifferentiated state of ES cells were chosen because of their expression patterns in the early embryo or from functional cDNA overexpression screens. With new technologies, it is possible to identify other candidate regulators of ES cells and study them with loss-of-function screens using RNAi. The availability of new ES cell lines expressing reporter genes under the control of promoters of ES cell-specific genes will allow researchers to monitor the undifferentiated state of ES cells. For example, hES cell lines have been generated that express green fluorescent protein under the control of the Oct4 promoter.[70] Microarray analysis of ES cells suggests that some genes may have an important role in determining the stem cell state, because they are upregulated in these cells when compared to somatic cells.[71] Since conditional RNAi systems are also available, analyzing candidate genes selected from various approaches or performing genome-wide screens by conditional loss-of-function analysis in both mouse and human ES cells can bring essential regulatory pathways to light (Figure 3).

Apart from understanding the undifferentiated state of ES cells, it is of great interest to understand the mechanisms that underlie lineage commitment of ES cells. Owing to the fact that they are pluripotent, ES cells can be differentiated into many if not all cell types. A popular method used to trigger differentiation *in vitro* in ES cells is through the formation of embryoid bodies (EBs), a heterogeneous aggregate of cells that is formed spontaneously in suspension after the removal of LIF. Formation of EBs from mES cells is reported to recapitulate initial steps of cell differentiation in early embryos (for a review, see Keller[72]). Therefore, through detailed study of EB formation it is possible to recapitulate the developmental context and promote differentiation of particular cell types. RNAi can help to dissect these pathways through loss-of-function genetic screens to identify critical genes involved in cell fate decision (Figure 3).

miRNAs are likely to play an important role in ES cell differentiation. It has been shown that several miRNAs are expressed in mouse ES cells. Some miRNAs are immediately suppressed upon ES cell differentiation, whereas others are expressed only after the formation of EBs.[73] Recent studies showed that ES cells lacking the critical RNase for the generation of miRNAs, Dicer, are defective in their proliferation and differentiation. It is still unclear, however, if miRNAs directly regulate the cell cycle or differentiation pathways, or have a more global effect on cell stability.[74,75]

The most successful attempts to differentiate cells from mES cells *in vitro* have shown that ES-derived cells acquire at least some of the molecular, morphological

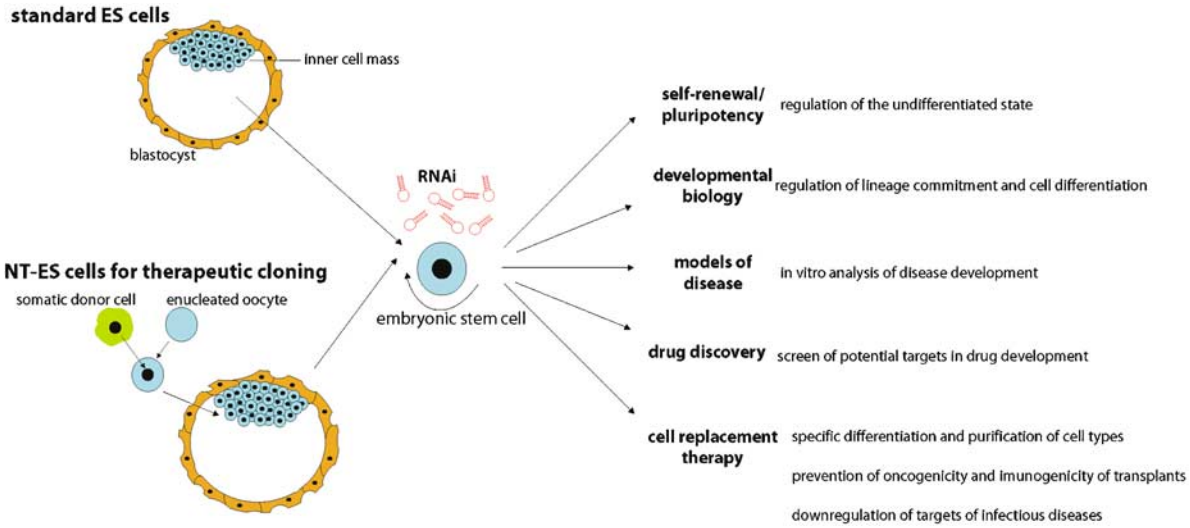


Figure 3 The potential uses of RNAi in ES cells. ES cells can be generated from normal blastocysts (standard) or through blastocysts derived from somatic cell nuclear transfer (NT-ES cells). ES cells can be expanded essentially indefinitely in culture and can give rise to all cell types of the body. RNAi may be useful in a wide variety of studies involving ES cells. Some of these areas include: understanding the basic biology of ES cells and cellular differentiation; modeling disease states *in vitro*; validating new drugs and assessing their toxicity; directing differentiation of cell types of interest from ES cells; controlling the cell cycle and immune repertoire of ES-derived cells to be transplanted and targeting infectious agents.

and functional features of differentiated cells. For example, overexpression of the transcription factor HoxB4 was shown to promote differentiation of mES cells into hematopoietic progenitors.[76] These progenitors were successfully engrafted into irradiated mice resulting in long-term multi-lineage hematopoietic progeny that persisted in secondary recipients. A parallel study showed that the Rag2 (−/−) deficiency could be repaired using homologous recombination on NT-ES cells of these mice.[77] This is an example of therapeutic cloning that combines nuclear transfer with gene correction of ES cells. Neural lineages have also been obtained from mouse ES cells. For example, exposure of differentiating cells to retinoic acid and sonic hedgehog led to differentiation into motor neurons.[78] These neurons formed functional synapses with muscle when transplanted into chick embryo spinal cords.[79] Dopaminergic neurons were also derived from ES cells and used to reverse symptoms of Parkinson's disease in rats.[80] It has been more challenging to differentiate endodermal derivatives, such as liver or pancreatic cells, from ES cells. Nevertheless, the derivation of insulin-producing cells capable of reversing diabetes in mice has been reported.[81,82] Genetic manipulation including gene silencing by RNAi may help to overcome the roadblocks to endoderm differentiation**.** Notwithstanding all the promising advances in differentiating ES cells, one should not expect that it will be possible to obtain all cell types desired or even to engineer whole organs from ES cells, at least not in the near future. For example, in the case of neurons, where there can be distinguished up to 200 different subtypes, it is unlikely that it will be possible to differentiate all of them, and to recapitulate all the neuronal networks. Nevertheless, RNAi has already been used to manipulate ES differentiation. For example, knockdown of the tumor suppressor p53 facilitates differentiation of mouse ES cells into muscle cells,[83] and knockdown of the transcription factor PU.1 favors differentiation of pro-B cells.[84]

The case studies described above lead us to believe that many of the obstacles to the generation of cell types of interest from ES cells can be overcome, and RNAi will be an important tool. Downregulation of critical genes during differentiation may induce either the growth of a specific sub-population of cells or the apoptosis of an undesired one, resulting in both cases in the enrichment for a cell type of interest.

Potential therapeutic applications and limitations of RNAi and ES cells

One may expect that, in a near future, the development of some diseases will be studied *in vitro* using ES cells. ES cells of an animal model for a specific disease, or human ES cells that were derived from a pool of patients, may enable the identification of specific genes involved in the pathophysiology of the disease, as well as the characterization of the impact of mutations or allelic diversity among different groups of patients. It should be possible to obtain hES cells that can be clonally propagated which contain exactly the same genetic information as a human donor, as it has been done in the mouse.[65] Once differentiation is well established for a particular cell type, the etiology of a disease can be studied at the molecular and cellular level using these *in vitro* models, allowing manipulations that would otherwise be impossible. RNAi-based screens will allow the identification of molecular modules essential for disease progression. RNAi screens will permit the use of ES-derived cells to validate therapeutic targets for new drugs that are cell-specific.

Historically, mouse ES cells have been very useful for generating genetically engineered animal[85–87] for research purposes using homologous recombination.[88,89] Recently, however, the focus of ES cell research has been directed towards more clinical applications, such as development of cell replacement and gene therapies. The big challenge is to be able to apply all the knowledge of ES cell biology and to obtain well-defined protocols for differentiation for cell-based therapies, where some damaged tissues may be replaced by ES-derived cells. Standardized hES cells or patient-specific NT-hES cells may be used to enrich for specific cell types using adequate genetic manipulation and culture conditions. The possibility of doing therapeutic cloning is a clear advantage of ES-derived cells therapies over those using adult stem cells. ES cells can also be propagated indefinitely and seem to be more amenable to gene manipulation, providing an inexhaustible cell source for therapy. ES cell pluripotency also enables a broader use of these cells in such therapies. Diseases that involve the loss or damage of a single or very few types of cells are the most attractive candidates for ES cell therapies. Parkinson's disease,[80] lower motor neuron loss and spinal cord injuries and[78,79] type I diabetes mellitus[81,82] are all potentially treatable by these therapies.

In addition to the conventional cell-replacement approaches aimed at repairing damaged tissues, the combination of ES cell and RNAi technologies may result in novel therapies for infectious diseases such as HIV, tuberculosis or malaria. One such strategy for combating the HIV virus has already been reported. It involves isolation of hematopoietic stem cells from an infected individual and treating them with a lentivirus that leads to expression of a shRNA targeted against either viral RNA or against the cellular receptor targeted by HIV (for reviews, see Lee and Rossi[90]). These stem cell populations are then expanded *ex vivo* and reintroduced into the patient. As hematopoietic stem cells give rise to the cells comprising the immune system, it is hoped that such a procedure will confer HIV resistance to the immune system (the main target of HIV). Alternatively, hematopoietic progenitors derived from ES cells carrying RNAi vectors that target HIV infection may be used. This approach would circumvent the need to extract hematopoietic stem cells from the patient, taking advantage of the fact that ES cells can be grown in very large numbers.

Although the potential of RNAi and stem cell-based gene therapies is extremely promising, there are issues of safety and efficiency that must be addressed before any potential therapy can be applied in humans. The combinatorial therapeutic use of RNAi and ES cells, while it may yield great benefit, also compounds the limitations and potential negative side effects which both tools may illicit. Currently, one of the most pertinent limitations involving the use of ES cells is the lack of knowledge regarding the details of ES cell developmental biology. The range of cell types that can currently be derived from ES cells is fairly limited. As a result,

the disease targets of potential stem cell-based therapies are restricted to those affecting the small subset of cell types that can be derived. It is likely that in the near future the number of cell types that can be derived from ES cells will greatly increase, but the goal of complete ES cell-based organ replacement may be far off. Most of the differentiation protocols to obtain a cell type of interest yield a heterogeneous population that contains other cell types as well. It will also be important to achieve cell purity before ES-derived cells can be transplanted into patients. Another major caveat of ES cell-based therapies is the possibility of tumor formation. If a transplant happens to contain contaminating undifferentiated ES cells, these could lead to the formation of teratocarcinomas. Downregulation through RNAi of specific genes involved in regulation of the cell cycle may be a way to avoid these tumors.

Another limitation is the immune response following engraftment of an ES-derived transplant not immunologically matched to the patient. The use of immunosuppressive drugs can prolong the survival of allogenic ES cell progeny. This is not an ideal method due to the fact that the ability of the body to heal is compromised when the immune system is suppressed. Another strategy to enhance the compatibility of the graft is to decrease ES cell expression of cell surface proteins that activate host immune responses (i.e. major histocompatibility complex and costimulatory molecules) or to increase ES cell expression of immune-inhibitory antigens (i.e. Killer cell immunoglobulin-like receptors). These approaches in isolation will not likely permit long-term engraftment of ES cell-derived cells. Another way to prevent immune rejection is through the production of patient-specific ES cells by somatic cell nuclear transfer and thus has become a very popular topic of study (Figure 3).

The use of RNAi itself also presents hurdles that must be overcome. The two main hurdles are effective RNA delivery and specificity of gene silencing. shRNA delivery for some gene therapies, such as those used in two clinical trials recently approved by the FDA (for age-related macular degeneration), could be as simple as the injection of naked RNA.[28,29] Many other techniques for RNAi delivery have been formulated including liposomal carriers, aerosolized vapors and viral vectors, but like any other potential therapeutic treatment these methods must be carefully evaluated for both efficiency as well as any possible off-target effects.

The *in vivo* delivery efficiency of the interfering RNA species to the cell of interest is an issue that deserves great attention. The second issue that needs to be addressed is how effective a construct is at silencing its target sequence and only its target sequence. The fact that miRNAs can effectively silence mRNA transcripts with which they share only partial sequence homology suggests that the problem of off-target silencing is a very real issue which needs to be addressed in any effective gene therapy design.[91] In addition, in some cases, even small 21 nt duplexes appear to be capable of inducing non-specific global silencing directed by the interferon response in mammalian cells.[92,93] On the other hand, especially with respect to viral infection, it has been reported that siRNAs can specifically silence their target RNA and not other transcripts, even when the target and those other transcripts vary in sequence by as little as a single base pair.[94] Only further studies of the relationship between miRNAs and their target sequences can help to answer questions about specificity and off-targeting, as well as define rules by which to design potential therapeutic constructs.

Conclusion

Although the fields of RNAi and ES cell research are in their infancy, it is already possible to envision cell and gene therapies combining both of these strategies. Studies of ES cells differentiation may overcome concerns about the limiting number and purity of cells available for cell-replacement therapies. For this reason it is important to understand the mechanisms that regulate ES cell differentiation. RNAi may allow the discovery of unknown genes involved in pluripotency and lineage commitment, and may be used to direct cell differentiation. In addition, genes that are implicated in the development of specific lineages can be downregulated to enrich cultures of purified cells, eliminating unwanted derivatives. By knocking down genes involved in cell proliferation, the tumorigenic potential of these ES-derived cells may be eliminated. RNAi may also be useful to manipulate the immune repertoire and reduce the probability of rejection of an ES cell-derived transplant. Finally, using RNAi in ES cells may help to model diseases *in vitro* and identify effective drug targets. Given these promising potential applications, we expect that many fascinating discoveries will be made in the years ahead through the combinatorial use of RNAi and ES cells.

Acknowledgements

We thank Jeremy F Reiter, Amander T Clark, David L Boone, and members of the McManus and Ramalho-Santos labs for critical reading of the manuscript.

References

1 Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998; **391**: 806–811.
2 Napoli C, Lemieux C, Jorgensen R. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* 1990; **2**: 279–289.
3 Kumar M, Carmichael GG. Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol Mol Biol Rev* 1998; **62**: 1415–1434.
4 Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 2001; **411**: 494–498.
5 Yang S, Tutton S, Pierce E, Yoon K. Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. *Mol Cell Biol* 2001; **21**: 7807–7816.
6 Billy E, Brondani V, Zhang H, Muller U, Filipowicz W. Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc Natl Acad Sci USA* 2001; **98**: 14428–14433.
7 Wianny F, Zernicka-Goetz M. Specific interference with gene function by double-stranded RNA in early mouse development. *Nat Cell Biol* 2000; **2**: 70–75.
8 Matzke MA, Birchler JA. RNAi-mediated pathways in the nucleus. *Nat Rev Genet* 2005; **6**: 24–35.

9 Verdel A, Moazed D. RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett* 2005; **579**: 5872–5878.
10 Bernstein E, Allis CD. RNA meets chromatin. *Genes Dev* 2005; **19**: 1635–1655.
11 McManus MT. Small RNAs and immunity. *Immunity* 2004; **21**: 747–756.
12 Herr AJ. Pathways through the small RNA world of plants. *FEBS Lett* 2005; **579**: 5879–5888.
13 Wang MB, Metzlaff M. RNA silencing and antiviral defense in plants. *Curr Opin Plant Biol* 2005; **8**: 216–222.
14 Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science* 2001; **294**: 853–858.
15 Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001; **294**: 858–862.
16 Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001; **294**: 862–864.
17 Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993; **75**: 843–854.
18 Hammond SM. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett* 2005; **579**: 5822–5829.
19 Tomari Y, Zamore PD. Perspective: machines for RNAi. *Genes Dev* 2005; **19**: 517–529.
20 Hutvagner G. Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. *FEBS Lett* 2005; **579**: 5850–5857.
21 Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J, Chaudhuri A *et al.* MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet* 2004; **36**: 1079–1083.
22 Yekta S, Shih IH, Bartel DP. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 2004; **304**: 594–596.
23 Miyagishi M, Hayashi M, Taira K. Comparison of the suppressive effects of antisense oligonucleotides and siRNAs directed against the same targets in mammalian cells. *Antisense Nucleic Acid Drug Dev* 2003; **13**: 1–7.
24 Kunath T, Gish G, Lickert H, Jones N, Pawson T, Rossant J. Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nat Biotechnol* 2003; **21**: 559–561.
25 Caplen NJ. Gene therapy progress and prospects. Downregulating gene expression: the impact of RNA interference. *Gene Therapy* 2004; **11**: 1241–1248.
26 Timmons L, Court DL, Fire A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* 2001; **263**: 103–112.
27 Timmons L, Fire A. Specific interference by ingested dsRNA. *Nature* 1998; **395**: 854.
28 http://www.acuitypharma.com.
29 http://www.sirna.com.
30 Xia H, Mao Q, Eliason SL, Harper SQ, Martins IH, Orr HT *et al.* RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nat Med* 2004; **10**: 816–820.
31 Rubinson DA, Dillon CP, Kwiatkowski AV, Sievers C, Yang L, Kopinja J *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat Genet* 2003; **33**: 401–406.
32 Ventura A, Meissner A, Dillon CP, McManus M, Sharp PA, Van Parijs L *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc Natl Acad Sci USA* 2004; **101**: 10380–10385.
33 Wiznerowicz M, Trono D. Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J Virol* 2003; **77**: 8957–8961.
34 Stegmeier F, Hu G, Rickles RJ, Hannon GJ, Elledge SJ. A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc Natl Acad Sci USA* 2005; **102**: 13212–13217.
35 Tiscornia G, Tergaonkar V, Galimi F, Verma IM. CRE recombinase-inducible RNA interference mediated by lentiviral vectors. *Proc Natl Acad Sci USA* 2004; **101**: 7347–7351.
36 Stewart SA, Dykxhoorn DM, Palliser D, Mizuno H, Yu EY, An DS *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* 2003; **9**: 493–501.
37 Pfeifer A. Lentiviral transgenesis. *Transgenic Res* 2004; **13**: 513–522.
38 Cullen BR. Does RNA interference have a future as a treatment for HIV-1 induced disease? *AIDS Rev* 2005; **7**: 22–25.
39 Shirane D, Sugao K, Namiki S, Tanabe M, Iino M, Hirose K. Enzymatic production of RNAi libraries from cDNAs. *Nat Genet* 2004; **36**: 190–196.
40 Sen G, Wehrman TS, Myers JW, Blau HM. Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat Genet* 2004; **36**: 183–189.
41 Dinh A, Mo YY. Alternative approach to generate shRNA from cDNA. *Biotechniques* 2005; **38**: 629–632.
42 Luo B, Heard AD, Lodish HF. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc Natl Acad Sci USA* 2004; **101**: 5494–5499.
43 Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 1981; **292**: 154–156.
44 Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* 1981; **78**: 7634–7638.
45 Bradley A, Evans M, Kaufman MH, Robertson E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 1984; **309**: 255–256.
46 Kleinsmith LJ, Pierce Jr GB. Multipotentiality of single embryonal carcinoma cells. *Cancer Res* 1964; **24**: 1544–1551.
47 Stevens LC. The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev Biol* 1970; **21**: 364–382.
48 Finch BW, Ephrussi B. Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture *in vitro* and their extinction upon hybridization with cells of permanent lines. *Proc Natl Acad Sci USA* 1967; **57**: 615–621.
49 Matsui Y, Zsebo K, Hogan BL. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* 1992; **70**: 841–847.
50 Resnick JL, Bixler LS, Cheng L, Donovan PJ. Long-term proliferation of mouse primordial germ cells in culture. *Nature* 1992; **359**: 550–551.
51 Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* 1998; **282**: 1145–1147.
52 Shamblott MJ, Axelman J, Wang S, Bugg EM, Littlefield JW, Donovan PJ *et al.* Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci USA* 1998; **95**: 13726–13731.
53 Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 1988; **336**: 688–690.
54 Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 1988; **336**: 684–687.
55 Ying QL, Nichols J, Chambers I, Smith A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* 2003; **115**: 281–292.

56 Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C *et al*. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* 2002; **20**: 1261–1264.
57 Xu RH, Peck RM, Li DS, Feng X, Ludwig T, Thomson JA. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Methods* 2005; **2**: 185–190.
58 Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I *et al*. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998; **95**: 379–391.
59 Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K *et al*. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003; **113**: 631–642.
60 Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S *et al*. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003; **113**: 643–655.
61 Velkey JM, O'Shea KS. Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells. *Genesis* 2003; **37**: 18–24.
62 Zaehres H, Lensch MW, Daheron L, Stewart SA, Itskovitz-Eldor J, Daley GQ. High-efficiency RNA interference in human embryonic stem cells. *Stem Cells* 2005; **23**: 299–305.
63 Hay DC, Sutherland L, Clark J, Burdon T. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* 2004; **22**: 225–235.
64 Matin MM, Walsh JR, Gokhale PJ, Draper JS, Bahrami AR, Morton I *et al*. Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. *Stem Cells* 2004; **22**: 659–668.
65 Munsie MJ, Michalska AE, O'Brien CM, Trounson AO, Pera MF, Mountford PS. Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. *Curr Biol* 2000; **10**: 989–992.
66 Wakayama T, Perry AC, Zuccotti M, Johnson KR, Yanagimachi R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* 1998; **394**: 369–374.
67 Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 2000; **24**: 372–376.
68 Hanna LA, Foreman RK, Tarasenko IA, Kessler DS, Labosky PA. Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* 2002; **16**: 2650–2661.
69 Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003; **17**: 126–140.
70 Gerrard L, Zhao D, Clark AJ, Cui W. Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. *Stem Cells* 2005; **23**: 124–133.
71 Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. 'Stemness': transcriptional profiling of embryonic and adult stem cells. *Science* 2002; **298**: 597–600.
72 Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* 2005; **19**: 1129–1155.
73 Houbaviy HB, Murray MF, Sharp PA. Embryonic stem cell-specific MicroRNAs. *Dev Cell* 2003; **5**: 351–358.
74 Murchison EP, Partridge JF, Tam OH, Cheloufi S, Hannon GJ. Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci USA* 2005; **102**: 12135–12140.
75 Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T *et al*. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* 2005; **19**: 489–501.
76 Kyba M, Perlingeiro RC, Daley GQ. HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* 2002; **109**: 29–37.
77 Rideout III WM, Hochedlinger K, Kyba M, Daley GQ, Jaenisch R. Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* 2002; **109**: 17–27.
78 Wichterle H, Lieberam I, Porter JA, Jessell TM. Directed differentiation of embryonic stem cells into motor neurons. *Cell* 2002; **110**: 385–397.
79 Miles GB, Yohn DC, Wichterle H, Jessell TM, Rafuse VF, Brownstone RM. Functional properties of motoneurons derived from mouse embryonic stem cells. *J Neurosci* 2004; **24**: 7848–7858.
80 Kim JH, Auerbach JM, Rodriguez-Gomez JA, Velasco I, Gavin D, Lumelsky N *et al*. Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature* 2002; **418**: 50–56.
81 Soria B, Roche E, Berna G, Leon-Quinto T, Reig JA, Martin F. Insulin-secreting cells derived from embryonic stem cells normalize glycemia in streptozotocin-induced diabetic mice. *Diabetes* 2000; **49**: 157–162.
82 Hori Y, Rulifson IC, Tsai BC, Heit JJ, Cahoy JD, Kim SK. Growth inhibitors promote differentiation of insulin-producing tissue from embryonic stem cells. *Proc Natl Acad Sci USA* 2002; **99**: 16105–16110.
83 He Z, Li J, Zhen C, Feng L, Ding X. Knockdown of p53 by RNAi in ES cells facilitates RA-induced differentiation into muscle cells. *Biochem Biophys Res Commun* 2005; **335**: 676–683.
84 Zou GM, Chen JJ, Yoder MC, Wu W, Rowley JD. Knockdown of Pu.1 by small interfering RNA in CD34+ embryoid body cells derived from mouse ES cells turns cell fate determination to pro-B cells. *Proc Natl Acad Sci USA* 2005; **102**: 13236–13241.
85 Koller BH, Hagemann LJ, Doetschman T, Hagaman JR, Huang S, Williams PJ *et al*. Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. *Proc Natl Acad Sci USA* 1989; **86**: 8927–8931.
86 Thompson S, Clarke AR, Pow AM, Hooper ML, Melton DW. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* 1989; **56**: 313–321.
87 Thomas KR, Capecchi MR. Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* 1990; **346**: 847–850.
88 Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 1987; **51**: 503–512.
89 Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW, Thompson S *et al*. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* 1987; **330**: 576–578.
90 Lee NS, Rossi JJ. Control of HIV-1 replication by RNA interference. *Virus Res* 2004; **102**: 53–58.
91 Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M *et al*. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 2003; **21**: 635–637.
92 Sledz CA, Holko M, de Veer MJ, Silverman RH, Williams BR. Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol* 2003; **5**: 834–839.
93 Bridge AJ, Pebernard S, Ducraux A, Nicoulaz AL, Iggo R. Induction of an interferon response by RNAi vectors in mammalian cells. *Nat Genet* 2003; **34**: 263–264.
94 Gitlin L, Stone JK, Andino R. Poliovirus escape from RNA interference: short interfering RNA-target recognition and implications for therapeutic approaches. *J Virol* 2005; **79**: 1027–1035.

Bibliography

Aasen, T., Raya, A., Barrero, M.J., Garreta, E., Consiglio, A., Gonzalez, F., Vassena, R., Bilić, J., Pekarik, V., Tiscornia, G.*, et al.* (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol *26*, 1276-1284.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Molecular Cell *9*, 1191-1200.

Andrés, M.E., Burger, C., Peral-Rubio, M.J., Battaglioli, E., Anderson, M.E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999). CoREST: a functional corepressor required for regulation of neural-specific gene expression. Proc Natl Acad Sci USA *96*, 9873-9878.

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.

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 & Development *17*, 126-140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Bailey, T.L., and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. Proc Int Conf Intell Syst Mol Biol 3, 21-29.

Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Belmont, A.S. (2001). Visualizing chromosome dynamics with GFP. Trends Cell Biol *11*, 250-257.

Ben-Dor, A., Shamir, R., and Yakhini, Z. (1999). Clustering gene expression patterns. J Comput Biol 6, 281-297.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature 447, 407-412.

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

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes & Development 19, 1635-1655.

Bhattacharya, A., Deng, J.M., Zhang, Z., Behringer, R., de Crombrugghe, B., and Maity, S.N. (2003). The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. Cancer Res 63, 8167-8172.

Billy, E., Brondani, V., Zhang, H., Müller, U., and Filipowicz, W. (2001). Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. Proc Natl Acad Sci USA 98, 14428-14433.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & Development *16*, 6-21.

Biswas, D., Dutta-Biswas, R., and Stillman, D.J. (2007). Chd1 and yFACT act in opposition in regulating transcription. Molecular and Cellular Biology *27*, 6279-6287.

Biswas, D., Takahata, S., Xin, H., Dutta-Biswas, R., Yu, Y., Formosa, T., and Stillman, D.J. (2008). A role for Chd1 and Set2 in negatively regulating DNA replication in Saccharomyces cerevisiae. Genetics *178*, 649-659.

Birney, E., Andrews, D., Bevan, P., Caccamo, M., Cameron, G., Chen, Y., Clarke, L., Coates, G., Cox, T., Cuff, J., et al. (2004a). Ensembl 2004. Nucleic Acids Res 32, D468-470.

Birney, E., Andrews, T.D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., et al. (2004b). An overview of Ensembl. Genome Res 14, 925-928.

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

Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N., and Weintraub, H. (1993). Binding of myc proteins to canonical and noncanonical DNA sequences. Mol Cell Biol 13, 5216-5224.

Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. Science 250, 1149-1151.

Blelloch, R., Venere, M., Yen, J., and Ramalho-Santos, M. (2007). Generation of induced pluripotent stem cells in the absence of drug selection. Cell Stem Cell 1, 245-247.

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

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G.,

Kumar,R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

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.

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.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151-159.

Bridge, A.J., Pebernard, S., Ducraux, A., Nicoulaz, A.L., and Iggo, R. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. Nat Genet 34, 263-264.

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

Brown, S.W., 1966. Heterochromatin. Science 151, pp. 418–425.

Burman, S., Gebuhr, T., Yee, D., la Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G. and Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Molecular Cell *6*, 1287-95.

Bussemaker, H.J., Li, H., and Siggia, E.D. (2000a). Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis. Proc Natl Acad Sci U S A 97, 10096-10100.

Bussemaker, H.J., Li, H., and Siggia, E.D. (2000b). Regulatory element detection using a probabilistic segmentation model. Proc Int Conf Intell Syst Mol Biol 8, 67-74.

Caplen, N.J. (2004). Gene therapy progress and prospects. Downregulating gene expression: the impact of RNA interference. Gene Ther 11, 1241-1248.

Caretti, G., Salsi, V., Vecchi, C., Imbriano, C., and Mantovani, R. (2003). Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. J Biol Chem 278, 30435-30440.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development 132, 885-896.

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.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Chi, A.S., and Bernstein, B.E. (2009). Developmental biology. Pluripotent chromatin state. Science *323*, 220-221.

Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. Pac Symp Biocomput, 115-126.

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J.*, et al.* (2009). Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell *5*, 111-123.

Chong, J.A., Tapia-Ramírez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuller, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. Cell *80*, 949-957.

Chow, C.-M., Georgiou, A., Szutorisz, H., Maia E Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO Rep *6*, 354-360.

Cooper, G.M., and Sidow, A. (2003). Genomic regulatory regions: insights from comparative sequence analysis. Curr Opin Genet Dev 13, 604-610.

Coustry, F., Maity, S.N., Sinha, S., and de Crombrugghe, B. (1996). The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit. J Biol Chem 271, 14485-14491.

Coustry, F., Sinha, S., Maity, S.N., and Crombrugghe, B. (1998). The two activation domains of the CCAAT-binding factor CBF interact with the dTAFII110 component of the Drosophila TFIID complex. Biochem J 331 (Pt 1), 291-297.

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., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K., Young, R.A., Jaenisch, R., and Boyer, L.A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells

and is necessary for lineage commitment. Cell *135*, 649-661.
Cullen, B.R. (2005). Does RNA interference have a future as a treatment for HIV-1 induced disease? AIDS reviews 7, 22-25.
Currie, R.A. (1997). Functional interaction between the DNA binding subunit trimerization domain of NF-Y and the high mobility group protein HMG-I(Y). J Biol Chem 272, 30880-30888.
Currie, R.A. (1998). NF-Y is associated with the histone acetyltransferases GCN5 and P/CAF. J Biol Chem 273, 1430-1434.
Curwen, V., Eyras, E., Andrews, T.D., Clarke, L., Mongin, E., Searle, S.M., and Clamp, M. (2004). The Ensembl automatic gene annotation system. Genome Res 14, 942-950.
Daley, G.Q., Goodell, M.A., and Snyder, E.Y. (2003). Realistic prospects for stem cell therapeutics. Hematology (Am Soc Hematol Educ Program), 398-418.
Davis, S., Aldrich, T.H., Stahl, N., Pan, L., Taga, T., Kishimoto, T., Ip, N.Y., and Yancopoulos, G.D. (1993). LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor. Science *260*, 1805-1808.
de la Serna, I.L., Ohkawa, Y., and Imbalzano, A.N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nat Rev Genet *7*, 461-473.
Delmas, V., Stokes, D.G., and Perry, R.P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proc Natl Acad Sci USA *90*, 2414-2418.
Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol Biol Evol 19, 1114-1121.
Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R.*, et al.* (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science *321*, 1218-1221.
Dinh, A., and Mo, Y.Y. (2005). Alternative approach to generate shRNA from cDNA. BioTechniques 38, 629-632.
Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576-578.
Donovan, P.J., and de Miguel, M.P. (2003). Turning germ cells into stem cells. Curr Opin Genet Dev 13, 463-471.
Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H.*, et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.
Eissenberg, J.C., Shilatifard, A., Dorokhov, N., and Michener, D.E. (2007). Cdk9 is an essential kinase in Drosophila that is required for heat shock gene expression, histone methylation and elongation factor recruitment. Mol Genet Genomics *277*, 101-114.
Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411, 494-498.
Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. Proc Natl Acad Sci USA *103*, 16319-16324.
Emberly, E., Rajewsky, N., and Siggia, E.D. (2003). Conservation of regulatory elements between two species of Drosophila. BMC Bioinformatics 4, 57.
Eminli, S., Utikal, J., Arnold, K., Jaenisch, R., and Hochedlinger, K. (2008). Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. Stem Cells *26*, 2467-2474.
Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.
Farina, A., Manni, I., Fontemaggi, G., Tiainen, M., Cenciarelli, C., Bellorini, M., Mantovani, R., Sacchi, A., and Piaggio, G. (1999). Down-regulation of cyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex. Oncogene 18, 2818-2827.
Fazzio, T.G., Huff, J.T., and Panning, B. (2008). An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. Cell *134*, 162-174.
Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol *8*, 188-194.
Finch, B., and Ephrussi, B. (1967). Retention of multiple developmental potentialities by cells of a mouse testicular teratocarcinoma during prolonged culture *in vitro* and their extinction upon hybridization with cells of permanent lines. Proc Natl Acad Sci USA *57*, 615-621.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806-811.

Flanagan, J.F., Blus, B.J., Kim, D., Clines, K.L., Rastinejad, F., and Khorasanizadeh, S. (2007). Molecular implications of evolutionary differences in CHD double chromodomains. J Mol Biol *369*, 334-342.

Flanagan, J.F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature *438*, 1181-1185.

Ford, A.M., Bennett, C.A., Healy, L.E., Navarro, E., Spooncer, E., and Greaves, M.F. (1992). Immunoglobulin heavy-chained CD3 delta-chain gene enhancers are DNase I-hypersensitive in hematopoietic progenitor cells. Proc Natl Acad Sci USA *89*, 3424-3428.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell *2*, 160-169.

Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W.C., and Wang, Z. (2008). ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. Proc Natl Acad Sci USA *105*, 6656-6661.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., Mcmanus, M.T., Plath, K., Meshorer, E., and Ramalho-Santos, M. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863-868.

Gearing, D.P., Comeau, M.R., Friend, D.J., Gimpel, S.D., Thut, C.J., McGourty, J., Brasher, K.K., King, J.A., Gillis, S., and Mosley, B. (1992). The IL-6 signal transducer, gp130: an oncostatin M receptor and affinity converter for the LIF receptor. Science *255*, 1434-1437.

Gerrard, L., Zhao, D., Clark, A.J., and Cui, W. (2005). Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. Stem Cells 23, 124-133.

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

Gitlin, L., Stone, J.K., and Andino, R. (2005). Poliovirus escape from RNA interference: short interfering RNA-target recognition and implications for therapeutic approaches. J Virol 79, 1027-1035.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Gorrini, C., Squatrito, M., Luise, C., Syed, N., Perna, D., Wark, L., Martinato, F., Sardella, D., Verrecchia, A., Bennett, S.*, et al.* (2007). Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. Nature *448*, 1063-1067.

Grskovic, M., Chaivorapol, C., Gaspar-Maia, A., Li, H., and Ramalho-Santos, M. (2007). Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet *3*, e145.

Guidi, C.J., Sands, A.T., Zambrowicz, B.P., Turner, T.K., Demers, D.A., Webster, W., Smith, T.W., Imbalzano, A.N., and Jones, S.N. (2001). Disruption of Ini1 leads to peri-implantation lethality and tumorigenesis in mice. Molecular and Cellular Biology *21*, 3598-3603.

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

Gurdon, J.B., and Uehlinger, V. (1966). "Fertile" intestine nuclei. Nature *210*, 1240-1241.

Hall, J.A., and Georgel, P.T. (2007). CHD proteins: a diverse family with strong ties. Biochem Cell Biol *85*, 463-476.

Hammond, S.M. (2005). Dicing and slicing: the core machinery of the RNA interference pathway. FEBS Lett 579, 5822-5829.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J.*, et al.* (2009). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell *4*, 513-524.

Hanna, J., 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, L.A., Foreman, R.K., Tarasenko, I.A., Kessler, D.S., and Labosky, P.A. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. Genes & Development *16*, 2650-2661.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.

Hay, D.C., Sutherland, L., Clark, J., and Burdon, T. (2004). Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. Stem Cells 22, 225-235.

He, Z., Li, J., Zhen, C., Feng, L., and Ding, X. (2005). Knockdown of p53 by RNAi in ES cells facilitates RA-induced differentiation into muscle cells. Biochemical and Biophysical Research Communications 335, 676-683.

Heitz, E., 1928. Das Heterochromatin der Moose. *I. Jb. wiss Bot.* 69, pp. 762–818.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol *21*, 133-153.

Herceg, Z., Hulla, W., Gell, D., Cuenin, C., Lleonart, M., Jackson, S., and Wang, Z.Q. (2001). Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression. Nat Genet *29*, 206-211.

Herr, A.J. (2005). Pathways through the small RNA world of plants. FEBS Lett 579, 5879-5888.

Hertz, G.Z., and Stormo, G.D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics 15, 563-577.

Ho, L., Jothi, R., Ronan, J., Cui, K., Zhao, K., and Crabtree, G. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci USA.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. (2008). A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell *3*, 346-353.

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*, 1132-1135.

Hori, Y., Rulifson, I.C., Tsai, B.C., Heit, J.J., Cahoy, J.D., and Kim, S.K. (2002). Growth inhibitors promote differentiation of insulin-producing tissue from embryonic stem cells. Proc Natl Acad Sci USA 99, 16105-16110.

Houbaviy, H., Murray, M., and Sharp, P. (2003). Embryonic Stem Cell-Specific MicroRNAs. Developmental Cell *5*, 351-358.

Hu, Q., and Maity, S.N. (2000). Stable expression of a dominant negative mutant of CCAAT binding factor/NF-Y in mouse fibroblast cells resulting in retardation of cell growth and inhibition of transcription of various cellular genes. J Biol Chem 275, 4435-4444.

Huang, Y., Myers, S.J., and Dingledine, R. (1999). Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. Nat Neurosci *2*, 867-872.

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

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008b). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol *26*, 1269-1275.

Humphrey, R.K., Beattie, G.M., Lopez, A.D., Bucay, N., King, C.C., Firpo, M.T., Rose-John, S., and Hayek, A. (2004). Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. Stem Cells *22*, 522-530.

Hutvagner, G. (2005). Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. FEBS Lett 579, 5850-5857.

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.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21, 635-637.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Molecular and Cellular Biology *24*, 8862-8871.

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

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jiang, W., Shi, Y., Zhao, D., Chen, S., Yong, J., Zhang, J., Qing, T., Sun, X., Zhang, P., Ding, M*., et al.* (2007). *In vitro* derivation of functional insulin-producing cells from human embryonic stem cells. Cell Res *17*, 333-344.

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.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes & Development 19, 489-501.

Kang, L., Wang, J., Zhang, Y., Kou, Z., and Gao, S. (2009). iPS Cells Can Support Full-Term Development of Tetraploid Blastocyst-Complemented Embryos. Stem Cell, 1-4.

Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112-115.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. (2003). The UCSC Genome Browser Database. Nucleic Acids Res 31, 51-54.

Kehle, J., Beuchle, D., Treuheit, S., Christen, B., Kennison, J.A., Bienz, M., and Müller, J. (1998). dMi-2, a hunchback-interacting protein that functions in polycomb repression. Science *282*, 1897-1900.

Keirstead, H.S., Nistor, G., Bernal, G., Totoiu, M., Cloutier, F., Sharp, K., and Steward, O. (2005). Human embryonic stem cell-derived oligodendrocyte progenitor cell transplants remyelinate and restore locomotion after spinal cord injury. J Neurosci *25*, 4694-4705.

Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. Genes & Development *19*, 1129-1155.

Kelley, D.E., Stokes, D.G., and Perry, R.P. (1999). CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin. Chromosoma *108*, 10-25.

Kim, J.B., Zaehres, H., Wu, G., Gentile, L., Ko, K., Sebastiano, V., Araúzo-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.

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

Kim, J.K., Huh, S.O., Choi, H., Lee, K.S., Shin, D., Lee, C., Nam, J.S., Kim, H., Chung, H., Lee, H.W.*, et al.* (2001). Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. Molecular and Cellular Biology *21*, 7787-7795.

Kimura, A., Umehara, T., and Horikoshi, M. (2002). Chromosomal gradient of histone acetylation established by Sas2p and Sir2p functions as a shield against gene silencing. Nat Genet *32*, 370-377.

Kittler, R., Surendranath, V., Heninger, A.-K., Slabicki, M., Theis, M., Putz, G., Franke, K., Caldarelli, A., Grabner, H., Kozak, K.*, et al.* (2007). Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. Nat Meth *4*, 337-344.

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

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

Kohlhase, J., Heinrich, M., Liebers, M., Frohlich Archangelo, L., Reardon, W., and Kispert, A. (2002). Cloning and expression analysis of Sall4, the murine homologue of the gene mutated in Okihiro syndrome. Cytogenet Genome Res 98, 274-277.

Koller, B.H., Hagemann, L.J., Doetschman, T., Hagaman, J.R., Huang, S., Williams, P.J., First, N.L., Maeda, N., and Smithies, O. (1989). Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. Proc Natl Acad Sci USA *86*, 8927-8931.

Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A.V., Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A.*, et al.* (2007). CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin *in vivo*. Science *317*, 1087-1090.

Koonin, E.V., Zhou, S., and Lucchesi, J.C. (1995). The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. Nucleic Acids Research *23*, 4229-4233.

Kumar, M., and Carmichael, G.G. (1998). Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. Microbiol Mol Biol Rev 62, 1415-1434.

Kunath, T., Gish, G., Lickert, H., Jones, N., Pawson, T., and Rossant, J. (2003). Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. Nat Biotechnol 21, 559-561.

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.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S.Y., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol Cell Biol 25, 2475-2485.

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

Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S.*, et al.* (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nat Biotechnol *25*, 1015-1024.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science 294, 853-858.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294, 858-862.

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.

Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., and Wootton, J.C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science 262, 208-214.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-i.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lei, H., Oh, S.P., Okano, M., Jüttermann, 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.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, Q., Herrler, M., Landsberger, N., Kaludov, N., Ogryzko, V.V., Nakatani, Y., and Wolffe, A.P. (1998). Xenopus NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the Xenopus hsp70 promoter in vivo. Embo J 17, 6300-6315.

Li, X.Y., Hooft van Huijsduijnen, R., Mantovani, R., Benoist, C., and Mathis, D. (1992). Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. J Biol Chem 267, 8984-8990.

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., 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.

Liberati, C., di Silvio, A., Ottolenghi, S., and Mantovani, R. (1999). NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices. J Mol Biol 285, 1441-1455.

Liu, X., Brutlag, D.L., and Liu, J.S. (2001). BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput, 127-138.

Liu, Y., Liu, X.S., Wei, L., Altman, R.B., and Batzoglou, S. (2004). Eukaryotic regulatory element conservation analysis and identification using comparative genomics. Genome Res 14, 451-458.

Lodén, M., and van Steensel, B. (2005). Whole-genome views of chromatin structure. Chromosome Res *13*, 289-298.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J.*, et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

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

Lowry, W.E., Richter, L., Yachechko, R., Pyle, A.D., Tchieu, J., Sridharan, R., Clark, A.T., and Plath, K. (2008). Generation of human induced pluripotent stem cells from dermal fibroblasts. Proc Natl Acad Sci USA *105*, 2883-2888.

Luo, B., Heard, A.D., and Lodish, H.F. (2004). Small interfering RNA production by enzymatic engineering of DNA (SPEED). Proc Natl Acad Sci USA 101, 5494-5499.

Lusser, A., Urwin, D.L., and Kadonaga, J.T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat Struct Mol Biol *12*, 160-166.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell *3*, 340-345.

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.

Maity, S.N., Sinha, S., Ruteshouser, E.C., and de Crombrugghe, B. (1992). Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. J Biol

Chem 267, 16574-16580.
Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G., et al. (2004). MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat Genet 36, 1079-1083.
Mantovani, R. (1999). The molecular biology of the CCAAT-binding factor NF-Y. Gene 239, 15-27.
Marfella, C.G.A., Ohkawa, Y., Coles, A.H., Garlick, D.S., Jones, S.N., and Imbalzano, A.N. (2006). Mutation of the SNF2 family member Chd2 affects mouse development and survival. J Cell Physiol *209*, 162-171.
Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J.*, et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.
Martin, G. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. PNAS *78*, 7634-7638.
Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol 20, 429-440.
Matin, M.M., Walsh, J.R., Gokhale, P.J., Draper, J.S., Bahrami, A.R., Morton, I., Moore, H.D., and Andrews, P.W. (2004). Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. Stem Cells 22, 659-668.
Matsui, Y., and Okamura, D. (2005). Mechanisms of germ-cell specification in mouse embryos. Bioessays 27, 136-143.
Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.
Matuoka, K., and Chen, K.Y. (2000). Possible role of subunit A of nuclear factor Y (NF-YA) in normal human diploid fibroblasts during senescence. Biogerontology 1, 261-271.
Matuoka, K., and Chen, K.Y. (2002). Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. Ageing Res Rev 1, 639-651.
Matzke, M.A., and Birchler, J.A. (2005). RNAi-mediated pathways in the nucleus. Nat Rev Genet 6, 24-35.
McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. PNAS, 6.
McManus, M.T. (2004). Small RNAs and immunity. Immunity 21, 747-756.
Meissner, A., Mikkelsen, T., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B., Nusbaum, C., Jaffe, D.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature.
Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nat Biotechnol *25*, 1177-1181.
Meshorer, E. (2007). Chromatin in embryonic stem cell neuronal differentiation. Histol Histopathol *22*, 311-319.
Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol 7, 540-546.
Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Developmental Cell *10*, 105-116.
Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G.*, et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-50.
Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., 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*, 49-55.
Miles, G.B., Yohn, D.C., Wichterle, H., Jessell, T.M., Rafuse, V.F., and Brownstone, R.M. (2004). Functional properties of motoneurons derived from mouse embryonic stem cells. J Neurosci 24, 7848-7858.
Miller, R.A., and Ruddle, F.H. (1976). Pluripotent teratocarcinoma-thymus somatic cell hybrids. Cell *9*, 45-55.
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.
Miyagishi, M., Hayashi, M., and Taira, K. (2003). Comparison of the suppressive effects of antisense oligonucleotides and siRNAs directed against the same targets in mammalian cells. Antisense Nucleic Acid Drug Dev 13, 1-7.
Morrisey, E.E., Ip, H.S., Lu, M.M., and Parmacek, M.S. (1996). GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. Dev Biol *177*, 309-

322.
Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S., and Parmacek, M.S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & Development *12*, 3579-3590.
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.
Murchison, E., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci USA 102, 12135-12140.
Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.
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.
Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell 2, 279-289.
Narayan, A.D., Chase, J.L., Lewis, R.L., Tian, X., Kaufman, D.S., Thomson, J.A., and Zanjani, E.D. (2006). Human embryonic stem cell-derived hematopoietic cells are capable of engrafting primary as well as secondary fetal sheep recipients. Blood *107*, 2180-2183.
Ng, R.K., and Gurdon, J.B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nat Cell Biol *10*, 102-109.
Nichols, J., Chambers, I., Taga, T., and Smith, A. (2001). Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development *128*, 2333-2339.
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.
Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol Cell Biol 19, 5453-5465.
Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes & Development *16*, 479-489.
Niwa, H. (2007a). How is pluripotency determined and maintained? Development *134*, 635-646.
Niwa, H. (2007b). Open conformation chromatin and pluripotency. Genes & Development *21*, 2671-2676.
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., 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.
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 USA *96*, 13839-13844.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.
Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.
Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. Science *322*, 949-953.
Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem *280*, 5307-5317.
Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.
Osborne, E.A., Dudoit, S., and Rine, J. (2009). The establishment of gene silencing at single-cell resolution. Nat Genet *41*, 800-806.
Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nat Biotechnol *25*, 244-248.
Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human

embryonic stem cells. Cell Stem Cell *1*, 299-312.
Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes & Development *10*, 1991-2002.
Park, I., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M., Cowan, C., Hochedlinger, K., and Daley, G. (2008a). Disease-Specific Induced Pluripotent Stem Cells. Cell.
Park, I., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008b). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141-146.
Paro, R., and Hogness, D.S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc Natl Acad Sci USA *88*, 263-267.
Patil, C.K., Li, H., and Walter, P. (2004). Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. PLoS Biol 2, E246.
Pera, M.F., Reubinoff, B., and Trounson, A. (2000). Human embryonic stem cells. J Cell Sci *113 (Pt 1)*, 5-10.
Perini, G., Diolaiti, D., Porro, A., and Della Valle, G. (2005). In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. Proc Natl Acad Sci U S A 102, 12117-12122.
Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A*., et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.
Pfaffl, M.W., Horgan, G.W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30, e36.
Pfeifer, A. (2004). Lentiviral transgenesis. Transgenic Res 13, 513-522.
Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434-438.
Radomska, H.S., Satterthwaite, A.B., Taranenko, N., Narravula, S., Krause, D.S., and Tenen, D.G. (1999). A nuclear factor Y (NFY) site positively regulates the human CD34 stem cell gene. Blood 94, 3772-3780.
Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D.A. (2002). "Stemness": transcriptional profiling of embryonic and adult stem cells. Science *298*, 597-600.
Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D*., et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.
Reddien, P.W., and Sanchez-Alvarado, A. (2004). Fundamentals of Planarian Regenaration Annu Rev Cell Dev Biol *20*, 725-757.
Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature *359*, 550-551.
Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat Biotechnol 18, 399-404.
Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., and Khvorova, A. (2004). Rational siRNA design for RNA interference. Nat Biotechnol 22, 326-330.
Richie, E.R., Schumacher, A., Angel, J.M., Holloway, M., Rinchik, E.M., and Magnuson, T. (2002). The Polycomb-group gene eed regulates thymocyte differentiation and suppresses the development of carcinogen-induced T-cell lymphomas. Oncogene 21, 299-306.
Rideout, W.M., Hochedlinger, K., Kyba, M., Daley, G.Q., and Jaenisch, R. (2002). Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. Cell *109*, 17-27.
Roberts, C.W., Galusha, S.A., McMenamin, M.E., Fletcher, C.D., and Orkin, S.H. (2000). Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. Proc Natl Acad Sci USA *97*, 13796-13800.
Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem 280, 24731-24737.
Rubinson, D.A., Dillon, C.P., Kwiatkowski, A.V., Sievers, C., Yang, L., Kopinja, J., Rooney, D.L., Zhang, M., Ihrig, M.M., McManus, M.T., et al. (2003). A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. Nat Genet 33, 401-406.
Safe, S., and Abdelrahim, M. (2005). Sp transcription factor family and its role in cancer. Eur J Cancer 41, 2438-2448.
Sapojnikova, N., Thorne, A., Myers, F., Staynov, D., and Crane-Robinson, C. (2008). The Chromatin of Active Genes Is Not in a Permanently Open Conformation. J Mol Biol.
Schaniel, C., Ang, Y., Ratnakumar, K., Cormier, C., James, T., Bernstein, E., Lemischka, I. and Paddison, P. (2009) Smarcc1/Baf155 Couples Self-Renewal Gene Repression with Changes in

Chromatin Structure in Mouse Embryonic Stem Cells. Stem Cells.
Schenke-Layland, K., Angelis, E., Rhodes, K.E., Heydarkhan-Hagvall, S., Mikkola, H.K., and Maclellan, W.R. (2007). Collagen IV induces trophoectoderm differentiation of mouse embryonic stem cells. Stem Cells *25*, 1529-1538.
Schnetz, M., Bartels, C., Shastri, K., Balasubramanian, D., Zentner, G., Balaji, R., Zhang, X., Song, L., Wang, Z., Laframboise, T.*, et al.* (2009). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. Genome Research.
Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes & Development *16*, 919-932.
Schwartz, B.E., and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes & Development *19*, 804-814.
Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res 13, 103-107.
Sen, G., Wehrman, T.S., Myers, J.W., and Blau, H.M. (2004). Restriction enzyme-generated siRNA (REGS) vectors and libraries. Nat Genet 36, 183-189.
Serra, E., Zemzoumi, K., di Silvio, A., Mantovani, R., Lardans, V., and Dissous, C. (1998). Conservation and divergence of NF-Y transcriptional activation function. Nucleic Acids Res 26, 3800-3805.
Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., and Gearhart, J.D. (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc Natl Acad Sci USA *95*, 13726-13731.
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. Molecular Cell *32*, 491-502.
Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Schöler, H.R., and Ding, S. (2008). A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell *2*, 525-528.
Shim, J.H., Kim, S.E., Woo, D.H., Kim, S.K., Oh, C.H., McKay, R., and Kim, J.H. (2007). Directed differentiation of human embryonic stem cells towards a pancreatic cell fate. Diabetologia *50*, 1228-1238.
Shirane, D., Sugao, K., Namiki, S., Tanabe, M., Iino, M., and Hirose, K. (2004). Enzymatic production of RNAi libraries from cDNAs. Nat Genet 36, 190-196.
Siddharthan , R., Siggia, E.D., and van Nimwegen, E. (2005). PhyloGibbs: A Gibbs Sampling Motif Finder That Incorporates Phylogeny. PLoS Computational Biology 1, 534-556.
Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532-536.
Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J *22*, 1846-1856.
Sims, R.J., Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem *280*, 41789-41792.
Sims, R.J., Millhouse, S., Chen, C.-F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Molecular Cell *28*, 665-676.
Sims, R.J., and Reinberg, D. (2008). Is there a code embedded in proteins that is based on post-translational modifications? Nat Rev Mol Cell Biol *9*, 815-820.
Singh, A.M., Hamazaki, T., Hankowski, K.E., and Terada, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells *25*, 2534-2542.
Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H., and Majumder, S. (2008). REST maintains self-renewal and pluripotency of embryonic stem cells. Nature, 7.
Sinha, S., Maity, S.N., Lu, J., and de Crombrugghe, B. (1995). Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc Natl Acad Sci U S A 92, 1624-1628.
Skottman, H., Mikkola, M., Lundin, K., Olsson, C., Strömberg, A.-M., Tuuri, T., Otonkoski, T., Hovatta, O., and Lahesmaa, R. (2005). Gene expression signatures of seven individual human embryonic stem cell lines. Stem Cells *23*, 1343-1356.
Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H., and Williams, B.R. (2003). Activation of the interferon system by short-interfering RNAs. Nat Cell Biol 5, 834-839.
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.
Smith, S.B., Ee, H.C., Conners, J.R., and German, M.S. (1999). Paired-homeodomain transcription

factor PAX4 acts as a transcriptional repressor in early pancreatic development. Mol Cell Biol 19, 8272-8280.

Solari, F., and Ahringer, J. (2000). NURD-complex genes antagonise Ras-induced vulval development in Caenorhabditis elegans. Curr Biol *10*, 223-226.

Soria, B., Roche, E., Berná, G., León-Quinto, T., Reig, J.A., and Martín, F. (2000). Insulin-secreting cells derived from embryonic stem cells normalize glycemia in streptozotocin-induced diabetic mice. Diabetes 49, 157-162.

Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58-62.

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.

Srinivasan, S., Armstrong, J.A., Deuring, R., Dahlsveen, I.K., McNeill, H., and Tamkun, J.W. (2005). The Drosophila trithorax group protein Kismet facilitates an early step in transcriptional elongation by RNA Polymerase II. Development *132*, 1623-1635.

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., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. (2008b). Induced pluripotent stem cells generated without viral integration. Science *322*, 945-949.

Stegmeier, F., Hu, G., Rickles, R.J., Hannon, G., and Elledge, S. (2005). A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. Proc Natl Acad Sci USA 102, 13212-13217.

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

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., et al. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA 9, 493-501.

Stokes, D.G., and Perry, R.P. (1995). DNA-binding and chromatin localization properties of CHD1. Molecular and Cellular Biology *15*, 2745-2753.

Stokes, D.G., Tartof, K.D., and Perry, R.P. (1996). CHD1 is concentrated in interbands and puffed regions of Drosophila polytene chromosomes. Proc Natl Acad Sci USA *93*, 7137-7142.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004a). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA 101, 6062-6067.

Suh, M.-R., Lee, Y., Kim, J.Y., Kim, S.-K., Moon, S.-H., Lee, J.Y., Cha, K.-Y., Chung, H.M., Yoon, H.S., Moon, S.Y.*, et al.* (2004). Human embryonic stem cells express a unique set of microRNAs. Dev Biol *270*, 488-498.

Sutcliffe, E., Parish, I., He, Y., Juelich, T., Tierney, M., Rangasamy, D., Milburn, P., Parish, C., Tremethick, D., and Rao, S. (2009). Dynamic histone variant exchange accompanies gene induction in T cells. Mol and Cell Biol 29(7), 1972-86.

Suwińska A., Czołowska R., Ozdzeński W. and Tarkowski A.K. (2008) Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. Dev Biol vol. 322 (1) pp. 133-44.

Szutorisz, H., Georgiou, A., Tora, L., and Dillon, N. (2006). The proteasome restricts permissive transcription at tissue-specific gene loci in embryonic stem cells. Cell *127*, 1375-1388.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H.*, et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes & Development *16*, 1779-1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & Development *19*, 815-826.

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.

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.

Takeuchi, J.K., and Bruneau, B.G. (2009). Directed transdifferentiation of mouse mesoderm to heart

tissue by defined factors. Nature *459*, 708-711.
Tam, W.-L., Lim, C.Y., Han, J., Zhang, J., Ang, Y.-S., Ng, H.-H., Yang, H., and Lim, B. (2008). T-Cell Factor 3 Regulates Embryonic Stem Cell Pluripotency and Self-Renewal by the Transcriptional Control of Multiple Lineage Pathways. Stem Cells *26*, 2019-2031.
Terstappen, L.W., Huang, S., Safford, M., Lansdorp, P.M., and Loken, M.R. (1991). Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells. Blood *77*, 1218-1227.
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.
Testa, A., Donati, G., Yan, P., Romani, F., Huang, T.H., Vigano, M.A., and Mantovani, R. (2005). Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. J Biol Chem 280, 13606-13615.
Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.
Thomas, K.R., and Capecchi, M.R. (1990). Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature *346*, 847-850.
Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., and Melton, D.W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell *56*, 313-321.
Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.
Tian, X., Woll, P.S., Morris, J.K., Linehan, J.L., and Kaufman, D.S. (2006). Hematopoietic engraftment of human embryonic stem cell-derived cells is regulated by recipient innate immunity. Stem Cells *24*, 1370-1380.
Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene 263, 103-112.
Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature 395, 854.
Tiscornia, G., Tergaonkar, V., Galimi, F., and Verma, I.M. (2004). CRE recombinase-inducible RNA interference mediated by lentiviral vectors. Proc Natl Acad Sci USA 101, 7347-7351.
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.
Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes & Development 19, 517-529.
Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res 30, 3202-3213.
Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Molecular Cell *28*, 1-13.
Turner, B.M. (2007). Defining an epigenetic code. Nat Cell Biol *9*, 2-6.
Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci *118*, 4495-4509.
Velkey, J., and O'shea, K. (2003). Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells. Genesis 37, 18-24.
Venkatasubrahmanyam, S., Hwang, W.W., Meneghini, M.D., Tong, A.H.Y., and Madhani, H.D. (2007). Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. Proc Natl Acad Sci USA *104*, 16609-16614.
Ventura, A., Meissner, A., Dillon, C.P., McManus, M., Sharp, P.A., Van Parijs, L., Jaenisch, R., and Jacks, T. (2004). Cre-lox-regulated conditional RNA interference from transgenes. Proc Natl Acad Sci USA *101*, 10380-10385.
Verdel, A., and Moazed, D. (2005). RNAi-directed assembly of heterochromatin in fission yeast. FEBS Lett 579, 5872-5878.
Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97-100.
Waddington, C.H. (1957). The strategy of the genes: a discussion of some aspects of theoretical biology. (London: Allen & Unwin).
Wakayama, T., Perry, A.C., 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.
Walfridsson, J., Khorosjutina, O., Matikainen, P., Gustafsson, C.M., and Ekwall, K. (2007). A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly. EMBO J *26*, 2868-2879.

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., Menendez, P., Shojaei, F., Li, L., Mazurier, F., Dick, J.E., Cerdan, C., Levac, K., and Bhatia, M. (2005a). Generation of hematopoietic repopulating cells from human embryonic stem cells independent of ectopic HOXB4 expression. J Exp Med *201*, 1603-1614.

Wang, M.B., and Metzlaff, M. (2005). RNA silencing and antiviral defense in plants. Curr Opin Plant Biol 8, 216-222.

Wang, T., and Stormo, G.D. (2003). Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics 19, 2369-2380.

Warner, M.H., Roinick, K.L., and Arndt, K.M. (2007). Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. Molecular and Cellular Biology *27*, 6103-6115.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet, 28.

Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., and Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. Nat Biotechnol *26*, 916-924.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.

Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol 2, 70-75.

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

Williams, C.J., Naito, T., Arco, P.G.-D., 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. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Wiznerowicz, M., and Trono, D. (2003). Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. J Virol 77, 8957-8961.

Wong, C.C., Gaspar-Maia, A., Ramalho-Santos, M., and Reijo Pera, R.A. (2008). High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming. PLoS ONE *3*, e1955.

Woodage, T., Basrai, M.A., Baxevanis, A.D., Hieter, P., and Collins, F.S. (1997). Characterization of the CHD family of proteins. Proc Natl Acad Sci USA *94*, 11472-11477.

Wu, Q., Chen, X., Zhang, J., Loh, Y.-H., Low, T.-Y., Zhang, W., Zhang, W., Sze, S.-K., Lim, B., and Ng, H.-H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J Biol Chem *281*, 24090-24094.

Xella, B., Goding, C., Agricola, E., Di Mauro, E., and Caserta, M. (2006). The ISWI and CHD1 chromatin remodelling activities influence ADH2 expression and chromatin organization. Mol Microbiol *59*, 1531-1541.

Xia, H., Mao, Q., Eliason, S.L., Harper, S.Q., Martins, I.H., Orr, H.T., Paulson, H.L., Yang, L., Kotin, R.M., and Davidson, B.L. (2004). RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. Nat Med 10, 816-820.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434, 338-345.

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

Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Meth *2*, 185-190.

Yamanaka, Y., Ralston, A., Stephenson, R.O., and Rossant, J. (2006). Cell and molecular regulation of the mouse blastocyst. Dev Dyn *235*, 2301-2314.

Yan, Y., Yang, D., Zarnowska, E.D., Du, Z., Werbel, B., Valliere, C., Pearce, R.A., Thomson, J.A., and Zhang, S.-C. (2005). Directed differentiation of dopaminergic neuronal subtypes from human embryonic stem cells. Stem Cells *23*, 781-790.

Yang, S., Tutton, S., Pierce, E., and Yoon, K. (2001). Specific double-stranded RNA interference in

undifferentiated mouse embryonic stem cells. Molecular and Cellular Biology 21, 7807-7816.

Yao, S., Chen, S., Clark, J., Hao, E., Beattie, G.M., Hayek, A., and Ding, S. (2006). Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. Proc Natl Acad Sci USA *103*, 6907-6912.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science 304, 594-596.

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.

Yi, F., Pereira, L., and Merrill, B.J. (2008). Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. Stem Cells *26*, 1951-1960.

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*, 519-523.

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

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.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R.*, 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.

Zaehres, H., Lensch, M.W., Daheron, L., Stewart, S.A., Itskovitz-Eldor, J., and Daley, G.Q. (2005). High-efficiency RNA interference in human embryonic stem cells. Stem Cells 23, 299-305.

Zegerman, P., Canas, B., Pappin, D., and Kouzarides, T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J Biol Chem *277*, 11621-11624.

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.

Zhao, R., and Daley, G.Q. (2008). From fibroblasts to iPS cells: induced pluripotency by defined factors. J Cell Biochem *105*, 949-955.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., 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, J., Zhang, Y., Joe, G.J., Pompetti, R., and Emerson, S.G. (2005). NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. Proc Natl Acad Sci U S A 102, 11728-11733.

Zou, G.M., Chen, J.J., Yoder, M.C., Wu, W., and Rowley, J.D. (2005). Knockdown of Pu.1 by small interfering RNA in CD34+ embryoid body cells derived from mouse ES cells turns cell fate determination to pro-B cells. Proc Natl Acad Sci USA 102, 13236-13241.