Genome-wide analysis of ATP-dependent chromatin remodeling functions in embryonic stem cells

Daria Bou Dargham

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à L'Université Paris Sud

NNT : 2015SACLS033

ÉCOLE DOCTORALE 577
Structure et Dynamique des Systèmes Vivants

Spécialité de doctorat : Sciences de la Vie et de la Santé

Par

Mme Daria Bou Dargham

Analyse de la fonction des facteurs de remodelage de chromatine ATP-dépendants dans le contrôle de l'expression du génome des cellules souches embryonnaires

Thèse présentée et soutenue à Saclay, le 13/10/2015 :

Composition du Jury :

Mme Fabienne Malagnac	Professeur, Université Paris Sud	Présidente
M Saadi Khochbin	Directeur de Recherche, CNRS	Rapporteur
M Slimane Ait-Si-Ali	Directeur de Recherche, CNRS	Rapporteur
M Daan Noordermeer	Chargé de Recherche, CNRS	Examinateur
M Eric Soler	Chargé de Recherche, INSERM	Examinateur
M Matthieu Gérard	Chercheur CEA, CEA Saclay	Directeur de thèse

université PARIS-SACLAY | ÉCOLE DOCTORALE Structure et dynamique des systèmes vivants (SDSV)

Titre : Analyse de la fonction des facteurs de remodelage de chromatine ATP-dépendants dans le contrôle de l'expression du génome des cellules souches embryonnaires

Mots clés : Cellules Souches Embryonnaires, remodeleurs de la chromatine, ChIP-seq

Résumé:
Les cellules souches embryonnaires (cellules ES) constituent un excellent système modèle pour étudier les mécanismes épigénétiques contrôlant la transcription du génome mammifère. Un nombre important de membres de la famille des facteurs de remodelage de chromatine ATP-dépendants ont une fonction essentielle pour l'auto-renouvellement des cellules ES, ou au cours de la différentiation. On pense que ces facteurs exercent ces rôles essentiels en régulant l'accessibilité de la chromatine au niveau des éléments régulateurs de la transcription, en modulant la stabilité et le positionnement des nucléosomes. Dans ce projet, nous avons conduit une étude génomique à grande échelle du rôle d'une dizaine des remodeleurs (Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400, Brg1, Smarca3, Smarcad1, Smarca5, ATRX et Chd1l) dans les cellules ES. Une double stratégie expérimentale a été utilisée.

Des expériences d'immunoprécipitation de la chromatine suivi par un séquençage à haute-débit (ChIP-seq) sur des cellules ES étiquetées pour les différents remodeleurs pour étudier leur distribution sur le génome, et un approche transcriptomique sur des cellules déplétées de chaque remodeleur par traitement avec des vecteurs shRNA (knockdown). Nous avons établi les profils de liaison des remodeleurs sur des éléments régulateurs (promoteurs, enhancers et sites CTCF) sur le génome, et montré que ces facteurs occupent toutes les catégories d'éléments régulateurs du génome. La corrélation entre les données ChIP-seq et les données transcriptomiques nous a permis d'analyser le rôle des remodeleurs dans les réseaux de transcription essentiels des cellules ES. Nous avons notamment démontré l'importance particulière de certains remodeleurs comme Brg1, Chd4, Ep400 et Smarcad1 dans la régulation de la transcription chez les cellules ES.

Title : Genome-wide analysis of ATP-dependent chromatin remodeling factors functions in embryonic stem cells

Keywords: Embryonic Stem Cells, chromatin remodeling factors, ChIP-seq

Abstract:
The characteristics of embryonic stem cells (ES cells) make them one of the best models to study the epigenetic regulation exerted by different actors in order to control the transcription of the mammalian genome. Members of the Snf2 family of ATP-dependent chromatin remodeling factors were shown to be of specific importance for ES cell self-renewal and during differentiation. These factors are believed to play essential roles in modifying the chromatin landscape through their capacity to position nucleosomes and determine their occupancy throughout the genome, making the chromatin more or less accessible to DNA binding factors.
In this project, a genome-wide analysis of the function of a number of ATP-dependent chromatin remodelers (Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Brg1, Ep400, ATRX, Smarca3, Smarca5, Smarcad1 and Alc1) in mouse embryonic stem (ES) cells was conducted.

This was done using a double experimental strategy. First, a ChIP-seq (Chromatin Immunoprecipitation followed by deep sequencing) strategy was done on ES cells tagged for each factor in the goal of revealing the genomic binding profiles of the remodeling factors. Second, loss-of-function studies followed by transcriptome analysis in ES cells were performed in order to understand the functional role of remodelers. Data from both studies were correlated to acquire a better understanding of the role of remodelers in the transcriptional network of ES cells.
Specific binding profiles of remodelers on promoters, enhancers and CTCF binding sites were revealed by our study. Transcriptomic data analysis of the deregulated genes upon remodeler factor knockdown, revealed the essential role of Chd4, Ep400, Smarcad1 and Brg1 in the control of transcription of ES cell genes. Altogether, our data highlight how the distinct chromatin remodeling factors cooperate to control the ES cell state.

Acknowledgements

Initially I would like to thank my jury members for accepting to examine my thesis work. Thanks to Dr. Saadi Khochbin and Dr. Slimane Ait-Si-Ali for accepting to be my thesis reporters and carefully reading and commenting my thesis manuscript. Equally, I thank the examiners Dr. Eric Soler, Dr. Daan Noordermeer and Dr. Fabienne Malagnac for accepting to be a part of my thesis defense jury.

I would like to deeply thank my thesis supervisor Dr. Matthieu Gerard for giving me the opportunity to conduct this interesting project in his lab. Thank you for your guidance, advice and great trust during these three years.

I would like to equally thank Dr. Jean-Christophe Andrau for being my thesis tutor and participating in my thesis committees. In addition, I thank Dr. Ute Rogner for being a member of my thesis committees. Thank you for your advices and comments.

I give special thanks to my collegues in our team. I thank Dr. Michel De Chaldée for his numerous advices and help especially at the beginning of my project, thank you as well for all the interesting discussions we had. I would like to thank Hélène Picaud for her everlasting humor and eventual technical help, keep on laughing! I equally thank Sylvie Jounier for her initial guidance and help, thank you for your patience and professionalism.

I give my thanks as well to my collegues in the animal house for their company and help. Thank you Anne-Sophie Chaplault, Sylvain Thessier, Jean-Charles Robillard, Patrick Héry. Thank you for picking me up at occasional work weekends…

I also would like to thank Dr.Sophie Chantalat and Florence Ribierre from CNG for their initial help and implication in my project.

I would like to thank all my collegues in the SBiGeM and the CEA IRTELIS program for giving me the opportunity to conduct this project. Moreover, I thank my doctoral school especially Dr. Pierre Capy for his guidance.

I thank my friends, in particular Dr. Elma El Khouri for her support and huge implication in reading and commenting my thesis manuscript.

Finally, I would like to send my sincere and deep thanks to each of my parents, my two brothers and my husband Ghazi for their continuous encouragement, support and love. Without you I would have never been where I am now.

Only he who attempts the absurd can achieve the impossible (A. Einstein)

Table of Contents

Figures and Tables

Abbreviations

A

AEBP: Adipocyte Enhancer Binding Protein

B

BMP: Bone Morphogenetic Protein

BAF: Brg1-Associated Factor

C

ChIP-seq: Chromatin Immuno-Precipitation followed by deep sequencing

CNOT3: CCR4-NOT transcription complex subunit 3

CHD: Chromo Domain

CFP1: CxxC Finger Protein 1

CBX: Chromo BOX

CTCF: CCCTC binding factor

ChIA-PET: Chromatin Interaction Analysis by Paired-End Tag Sequencing

CpGI: CpG Islands

D

DAX1: Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X

DNMTs: DNA methyltransferases

DHS: DNase I Hypersensitive Site

DBP: DNA binding Protein

E

ES: Embryonic Stem

ESSRB: Estrogen Related Receptor

EP400: E1A-binding Protein p400

EZH2: Enhancer of Zest Homologue 2

EED: Embryonic Ectoderm Development

ERK: Extracelluler Regulated Kinases

EpiS: Epiblast Stem

ENCODE: The Encyclopedia of DNA Elements

F

FOXd2: Forkhead Box d2

FGF2: Fibroblast Growth Factor 2

FAIRE-seq: Formaldehyde-Assisted Isolation of Regulatory Elements followed by deep sequencing

G

GSK3: Glycogen Synthase Kinase 3

GRO-seq: Global-Run-On Sequencing

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

H

HMG: High Mobility Group

HDAC: Histone DeAcetylase

HCP: High CpG Promoters

bHLH: basic Helix Loop Helix

I

iPS: induced Pluripotent Stem

ICM: Inner Cell Mass

Id: Inhibitor of Differentiation

J

JARID2: Jumonji AT Rich Interactive Domain 2

K

KLF: Kruppel-Like Factors

KD: Knock Down

L

LCD1: Lethal Checkpoint-defective DNA damage-sensitive protein 1

LCP: Low CpG Promoters

LIF: Leukemia Inhibitory Factor

LRP: Low density lipoprotein receptor-related protein

M

Myc: Myelocytomatosis

MEK: Mitogen Activated Protein

MBD3: Methyl CpG Binding protein 3

MACS: Model-based Analysis for ChIP-seq

N

NuRD: Nucleosome Remodeling Deacetylase

NFR: Nucleosome Free Region

O

Oct4: Octamer-binding transcription factor 4

P

PRC: Polycomb repressive Complex

PcG: Polycomb Group

PI3K: Phosphatidylinositol-4,5-bisphosphate 3-kinase

R

RT-qPCR: Reverse Transcription-quantitative Polymerase Chain Reaction

RPKM: Reads Per Kilobase per Million mapped reads

S

SetDB1: Set Domain Bifurcated 1

SALL4: Spalt-Like Transcription Factor 4

Sox2: (Sex determining region Y)-box 2

SE: Super Enhancer

STAT3: Signal Transducer and Activator of Transcription 3

SWI/SNF: Switch/sucrose non-fermenting

shRNA: small hairpin RNA

SICER: Spatial Clustering for Identification of ChIP-seq Enriched Regions

T

TBX3: T-box Transcription Factor 3

Tip60: TAT-interactive protein 60

TrxG: Trithorax Group

TE: Typical Enahcer

TAD: Topologically Associated Domain

TGFβ: Transformation Growth Factor beta

TAP: Tandem Affinity Purification

TEV: Tabacco Etch Virus

TF: Transcription Factor

TBP: TATA Binding Protein

TSS: Transcription Start Site

Z

ZFX: Zinc Finger Protein X-linked

INTRODUCTION

Chapter I. Embryonic Stem Cells: Definition and Regulatory Pathways

A. What are Embryonic Stem Cells?

Embryonic stem cells or ES cells are derived from the blastocyst which is formed during embryogenesis and is composed of two parts: an outer layer of cells, the trophectoderm, that will form the placenta and an inner clump of cells, called the inner cell mass (ICM), that is responsible for the formation of the entire body. The isolation and culture of cells derived from the ICM under appropriate conditions, gives rise to Embryonic Stem cells.

1. The Major Characteristics of Embryonic Stem Cells

Embryonic stem cells possess several distinct features that set them apart from other cell types. The two major characteristics that define ES cells are self-renewal and pluripotency or the ability to give rise to all the different types of cells of the body.

a. Self-renewal

ES cells can divide symmetrically and for a long period of time, where mother cells give rise to identical daughter cells in a continuous fashion. This property is defined as self-renewal. ES cells were initially established and maintained by Evans an Kaufman in 1981 (Evans and Kaufman, 1981). The proliferation of ES cells is assured and maintained by the presence of very specific extrinsic factors (like LIF or the Leukemia inhibitory Factor (Smith et al., 1988)) that provoke a series of intrinsic signaling responses necessary for self-renewal preservation.

b. Pluripotency

The first attempt to produce different cell lineages from embryonic stem cells goes back to the experiments done by Evans and Kaufman using mouse embryonic carcinoma cells by forming embryoid bodies (Evans and Kaufman, 1981). Embryoid bodies were given this name due to their marvelously similar composition to actual embryos (Figure 2). The pluripotent nature of mouse ES cells was demonstrated by their ability to contribute to all tissues of adult mice following their injection into host blastocysts (Bradley et al., 1984). In addition to their developmental potential in vivo, ES cells display a remarkable capacity to form differentiated cell types in culture (Keller, 1995). This capacity to give rise to all types of body cells is called pluripotency (Figure 1).

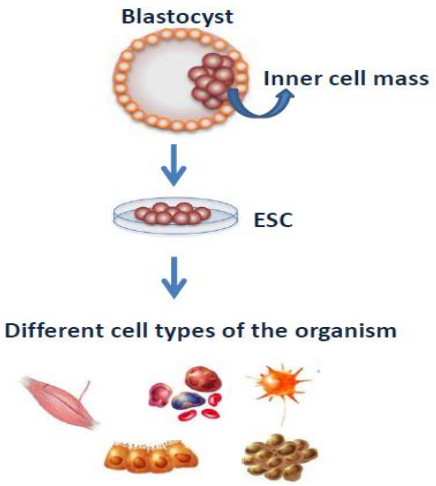


Figure 1 Embryonic Stem Cells and Pluripotency

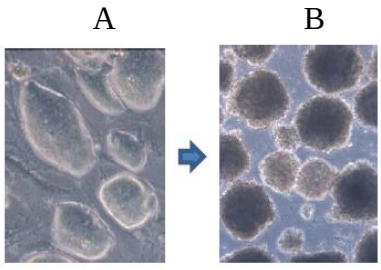


Figure 2 The various aspects of ES cells and Embryoid Bodies. A) Mouse ES cells colonies cultured on feeder cells. B) The differentiation of mouse ES cells. Embryoid body formation (Differentiation day 7)

2. Important signaling pathways that control the mouse ES cells state

The three major signaling pathways that control the ES cell state comprise: The LIF-dependent (Leukemia Inhibiting Factor dependent) signaling pathways, the Wnt pathway and BMP (Bone Morphogenetic Protein) signaling pathway that act with different mechanisms to assure the correct transcriptional profile for ES cell maintenance (Figure 3).

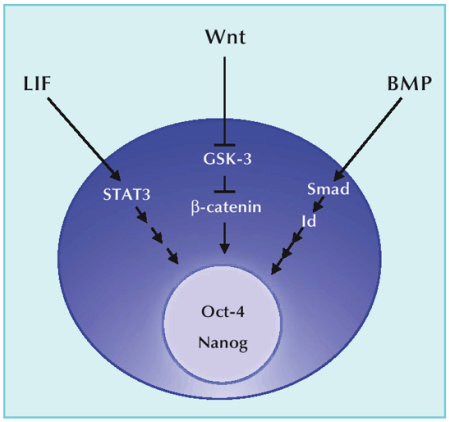


Figure 3 Pathways known to contribute to the maintenance of embryonic stem cell pluripotency (Nakashima et al., 2004)

a. The LIF-dependent signaling pathways

LIF belongs to the interleukin-6 cytokine family. It binds a heterodimeric receptor consisting of the low-affinity LIF receptor and gp130, with downstream signals being transmitted through gp130. The gp130 downstream signaling pathways include: the STAT3, phosphatidylinositol 3-kinase (PI3K) and Ras/Erk pathways (Figure 4). Maintaining the balance among the three pathways allows fine-tuning of the LIF-dependent maintenance of ES cell self-renewal.

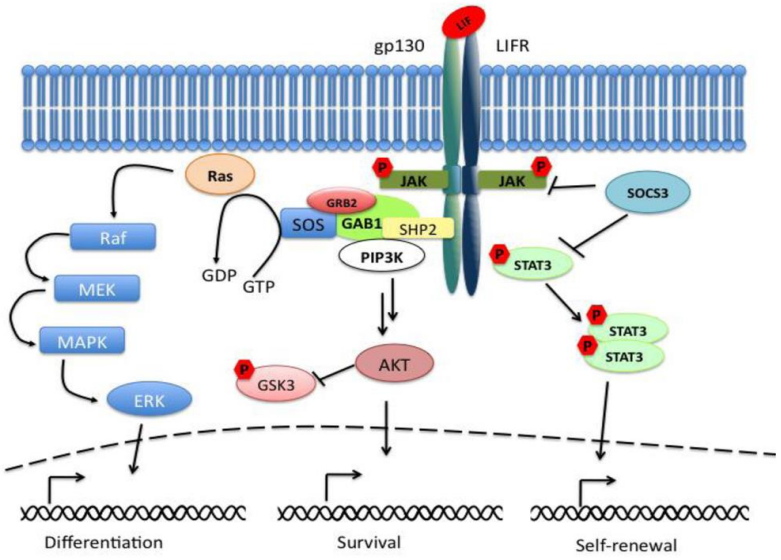


Figure 4 A schematic representation of LIF-dependent pathways (Graf et al., 2011)

The LIF/STAT3 pathway

Signaling through gp130 leads to activation of Janus-associated tyrosine kinases (JAKs), which in turn phosphorylate STAT3. Phosphorylated STAT3 homodimerizes and moves to the nucleus, where it functions as a transcription factor. In mouse ES cells, the STAT3 pathway plays a critical role in the maintenance of self-renewal. When STAT3 is down-regulated, ES cells undergo differentiation (Boeuf et al., 1997; Niwa et al., 1998; Ying et al., 2008). Artificial activation of the STAT3 pathway can maintain ES cell self-renewal even in the absence of LIF (Matsuda et al., 1999).

STAT3 was shown to bind to the regulatory regions of several self-renewal genes in ES cells (Chen et al., 2008a; Kidder et al., 2008). One major role of the LIF/STAT3 pathway is to form transcriptional networks with other key TFs such as Oct3/4, Sox2, Nanog, c-Myc, Klf4 and Esrrb (Transcriptional regulatory networks will be developed in Chapter II). These STAT3-regulated TFs bind to the regulatory regions of other important factors and induce

their expression. In this way, the LIF/STAT3 pathway participates in the formation of self-renewal transcriptional networks.

The LIF/PI3K pathway

PI3K phosphorylates phosphoinositides leading to activation of the serine/threonine protein kinase Akt, a serine/threonine kinase implicated in the regulation of cell cycle progression, cell death, adhesion, migration, metabolism and tumorigenesis (Brazil et al., 2004). Activated Akt then phosphorylates its target molecules, such as glycogen synthase kinase (GSK)-3 and pro-apoptotic BCL2-antagonist of death (BAD) protein, blocking their activity. Similarly to STAT3, the PI3K pathway positively regulates self-renewal as inhibition of PI3K and Akt induces differentiation of mouse and human ES cells, suggesting that PI3K/Akt signaling is necessary for the maintenance of ES cell pluripotency (Paling et al., 2004; Watanabe et al., 2006).

The Ras/Erk (MAPK) pathway

Ras is capable of the sequential activation of the Raf/MEK/Erk kinase cascade, leading to phosphorylation of Erk target molecules, including the transcription factor Elk-1, proapoptotic protein caspase-9, and p90 ribosomal S6 protein kinase (RSK). The activation of the Ras/Erk pathway leads to ES cell differentiation into the endoderm lineage (Yoshida-Koide et al., 2004) while suppressing the Ras/Erk signaling promotes self-renewal (Burdon et al., 1999).

b. Other signaling pathways

The Wnt pathway

Wnt signaling pathways play an important role in the lineage specification of the vertebrate embryo (Amerongen and Nusse, 2009; Clevers, 2006) and regulate pluripotency in ES cells (Nusse et al., 2008; Wend et al., 2010). It was shown that Wnt proteins act through various

Frizzled receptors and LRP5/6 co-receptors to trigger downstream events that eventually cause the inactivation of the β-catenin-degradation complex. By this manner, ES cell pluripotency is facilitated by the inhibition of GSK3 (Aubert et al., 2002; Doble et al., 2007), a protein kinase that phosphorylates β-catenin, marking it for ubiquitin-dependent degradation (Clevers, 2006; MacDonald et al., 2009). Further on, β-catenin interacts with TcF (T-cell Factor) proteins, the terminal nuclear effectors of the Wnt pathway, to regulate the transcription of specific target genes (Figure 5). In the absence of Wnt signaling, TcF proteins act as transcriptional repressors (Behrens et al., 1996; Cadigan, 2002).

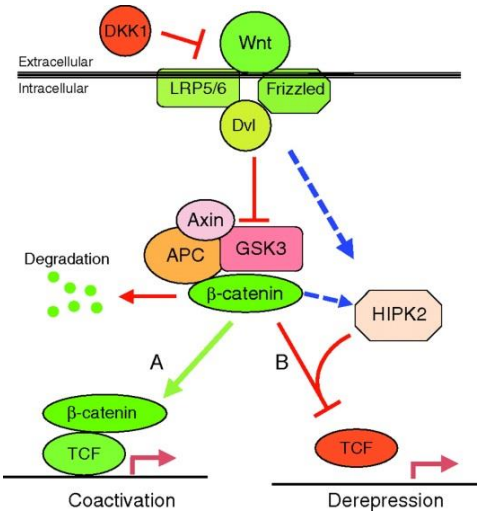


Figure 5 Schematic of the canonical vertebrate Wnt signaling pathway (Sokol, 2011)

TcF3, a member of the TcF family is highly expressed in mouse ES cells, and is critical in early mouse development (Korinek et al., 1998; Merrill et al., 2001; Pereira et al., 2006). It was shown that TcF3 acts to repress Nanog gene in ES cells (Pereira et al., 2006). Interestingly, a study highlights the importance of TcF3 in pluripotency, as it was demonstrated that TcF3 co-occupies the ES cell genome with the pluripotency factors Oct4 and Nanog (Cole et al., 2008). This study suggests that TcF3 affects the balance between pluripotency and differentiation in ES cells (Figure 6). At standard conditions there is a balance between the core TFs, the TcF3 activating complex and the TcF3 repressing complex. Upon TcF3 knockdown, the repressive effect of TcF3 is eliminated and the expression of core

TFs increases, pluripotency is favored and differentiation is inhibited. Upon Wnt stimulation, TcF3 becomes more of an activating complex of pluripotency, assuring the maintenance of the ES cell state. Upon loss of pluripotency gene expression, the Tcf3 repressive complex takes hold of the expression of pluripotency genes and differentiation is favored.

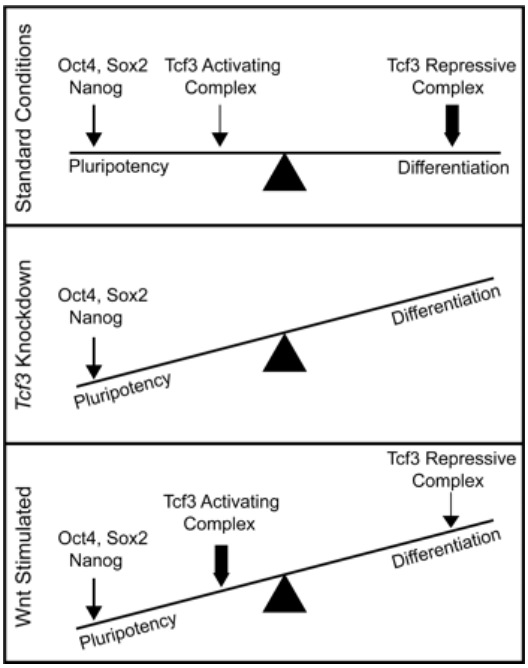


Figure 6 Model depicting the influence of Wnt pathway components on pluripotency and differentiation in ES cells (Cole et al., 2008)

The BMP pathway

BMPs or Bone Morphogenetic Proteins belong to the transformation growth factor beta (TGFβ) superfamily. They were shown to play important roles cell proliferation, differentiation, and apoptosis making them essential actors during embryonic development and pattern formation (Massagué, 1998). It was long known that BMPs protect the pluripotency state of ES cells, inhibit the differentiation into neural lineages (Di-Gregorio et al., 2007; Ying et al., 2003a) and prime ES cells for mesoderm formation (Davis et al., 2004; Lawson et al., 1999; Mishina et al., 1995). BMP activates Smad proteins that in turn act through the transcriptional up regulation of Inhibitor of Differentiation (Id) factors (Ying et

al., 2003a; Zhang et al., 2010), where these factors inhibit bHLH neurogenesis TFs, thus inhibiting neurogenesis (Norton, 2000). Furthermore, the BMP/Id complex was shown to protect pluripotent stem cells from differentiating especially into neural lineages through the maintenance of the expression of E-cadherin (Malaguti et al., 2013). Moreover, the maintenance of the high expression of E-cadherin by BMP/Id was shown to impose a proximal posterior identity on epiblast cells priming them for mesodermal lineages (Malaguti et al., 2013).

The capacity of the BMP signaling pathway to maintain pluripotency is assured by the coordination with LIF signaling pathways. On one hand, BMP/Smad signaling inhibits neural lineages (Tropepe et al., 2001; Ying et al., 2003b) and induces other lineages, on the other hand; LIF/STAT3 signaling inhibits non-neural lineages and possibly regulates Smad function (Ying et al., 2003a) . The balance between LIF and BMP signaling pathways is essential to assure pluripotency of ES cells, as it was shown that the constitutive expression of Smad over rides the effect of LIF/STAT3 and differentiates ES cells into non-neural lineages (Figure 7).

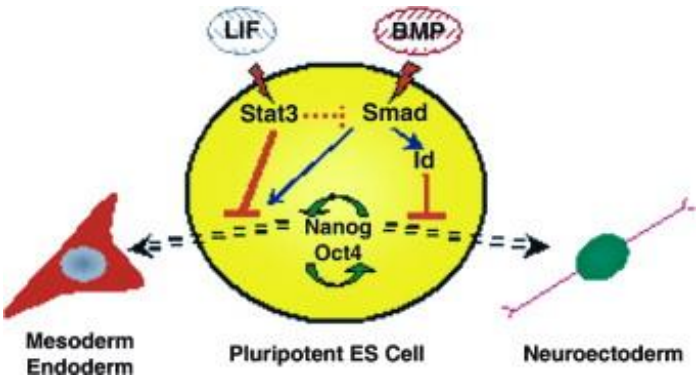


Figure 7 Cooperative lineage restriction by LIF/STAT3 and BMP/Smad (Ying et al., 2003a)

B. Naïve and primed pluripotency

1. Mouse ES cell in serum versus 2i medium

For a long time, the derivation and culture of mouse ES cells was poorly understood. Most of the derived mouse ES cells were from a specific mice strain which is the 129 strain. This strain was thought to have genetic and/or epigenetic profiles permissive for ES cell establishment (Brook and Gardner, 1997). Back then, the mouse ES cell culture conditions used were on feeder cells in serum-supplemented medium, these conditions supplemented the cells with the necessary LIF and BMP required for self-renewal through the activation of STAT pathway and the inhibition of differentiation pathways. However, such culture conditions restrained the derivation of mouse ES cells from most of the other mouse strains. This marked variability between the strains was later shown to be due to the level of ERK signaling (Batlle-Morera et al., 2008), a pathway important to promote differentiation and inhibit self-renewal. For the inhibition of ERK pathway was shown to improve the derivation of mouse ES cells from mouse strains other the 129 strain (Batlle-Morera et al., 2008; Buehr and Smith, 2003).

Later on, this strain specificity was eliminated by the discovery of a new approach where certain kinase inhibitors were used. This novel approach, called the 2i culture approach, makes it possible to culture mouse ES cells without serum by using two small molecules that inhibit kinases in combination with LIF (Ying et al., 2008). The two inhibitors are PD0325901 and CHIR99021 that respectively target the mitogen-activated protein kinase (MEK) and the glycogen synthase kinase-3 (GSK3).

The 2i medium provides a more-tuned environment for mouse ES cells, as a mosaic expression of some of the pluripotency factors (Chambers et al., 2007; Niwa et al., 2009; Toyooka et al., 2008) is observed in serum and eliminated upon culture in 2i conditions (Wray

et al., 2010), establishing a more naïve ground state. However, several studies show that this heterogeneous expression of some of the core TFs is not an only characteristic of ES cells cultured in serum, rather it is also observed in 2i-grown ES cells (Abranches et al., 2014, 2014; Morgani et al., 2013). In these studies the expression of Nanog was observed, were it was shown to be in a continuous fluctuation in individual cells, suggesting a certain role for such fluctuations in the capacity of ES cells to maintain a state where different differentiation opportunities are considered. Another study showed that Oct4 and Sox2 transcriptional heterogeneities are also observed in ES cells under the 2 different culture conditions (Faddah et al., 2013); however, these variations were not as prominent as Nanog fluctuations. This observation supports the hypothesis that the permissive nature of the chromatin and the noisy mRNA transcription observed (Gaspar-Maia et al., 2011) might cause the variable transcription of Oct4 and Sox2, as their expression contrary to that of Nanog seems to be quite homogeneous.

2. EpiS cells versus ES cells

ES cells are derived from the inner cell mass (ICM) of the mature blastocyst (the preimplantation epiblast). After implantation, the epiblast becomes primed for lineage specification and commitment in response to stimuli from the extraembryonic tissues. EpiS cells could be experimentally derived from the epiblast at this stage of development. Mouse ES cells grow in culture as compact domed colonies, while EpiS cells are larger and grow as a monolayer (Figure 8). Several differences demarcate ES cells from EpiS cells, as these two present different developmental stages. Although both cell types retain their pluripotency capacity and express the core TFs, EpiS cells are not able to form chimera when injected to blastocysts (Guo et al., 2009; Rossant, 2008; Tesar et al., 2007).

Moreover, EpiS cells have different signaling mechanisms that control their pluripotency and self-renewal states. Therefore, culture requirements differ between the two cell states, while ES cells need a LIF containing medium to survive, EpiS cells have different requirements including Fgf2 and Activin (Brons et al., 2007; Nichols and Smith, 2009; Tesar et al., 2007). The various differences that characterize ES cells and EpiS cells are listed in Table 1. Compared to ES cells, EpiS cells are defined by a more restricted, primed pluripotency state.

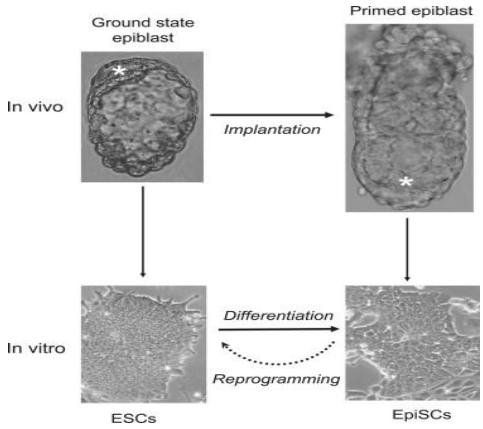


Figure 8 The two phases of pluripotency (Nichols and Smith, 2009)

Property	Ground State	Primed State
Embryonic tissue	Early epiblast	Egg cylinder or embryonic disc
Culture stem cells	Rodent ES cells	Rodent EpiS cells; primate "ES" cells
Blastocyst chimaeras	Yes	No
Teratomas	Yes	Yes
Differentiation bias	None	Variable
Pluripotency factors	Oct4, Nanog, Sox2, Klf2, Klf4	Oct4, Nanog, Sox2
Naïve markers	Rex1, NrOb1, Fgf4	Absent
Specification markers	Absent	Fgf5
Response to LIF/STAT3	Self-renewal	None
Response to Fgf/Erk	Differentiation	Self-renewal
Clonogenicity	High	Low
XX status	XaXa	XaXi
Response to 2i	Self-renewal	Differentiation/death

Table 1 Differences between the ground and primed cell states. Adapted from (Nichols and Smith, 2009)

C. Human ES cells

Human ES cells were first isolated by Thomson and collegues in 1998. These cells were capable to differentiate to the different cell lineages (Thomson et al., 1998). Although the core transcription regulatory network is conserved; however, these cells require different culture conditions. Human "ES" cells do not respond to LIF nor are they sustained by Erk inhibition; on the contrary, they depend on Erk signaling for continued proliferation. This signaling is assured in hESCs by the Activin/Nodal pathway (Figure 9) through Smad2/3 that control the expression of the pluripotency factor, Nanog, which in turn blocks the expression of neuroectoderm genes induced by FGF (Vallier et al., 2009). The most probable explanation for this difference is that human ES cells correspond to a more advanced developmental state, more similar to the mouse EpiS cells (Brons et al., 2007; Tesar et al., 2007).

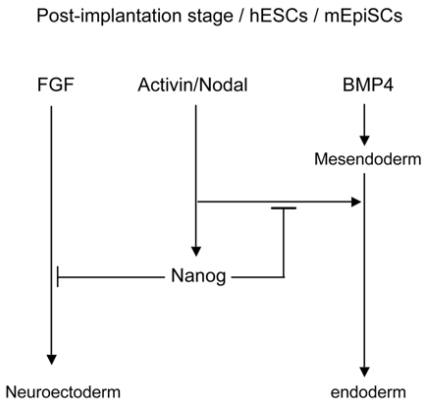


Figure 9 Model explaining the regulation of Nanog in hESCs/mEpiSCs and its function in both cell types (Vallier et al., 2009)

This EpiS cell state might be due to the longer period of culture required for the appearance of human ES cells, that would allow these cells to progress *in vitro* to a state equivalent to post implantation mouse embryos (Figure 10) (Nichols and Smith, 2011).

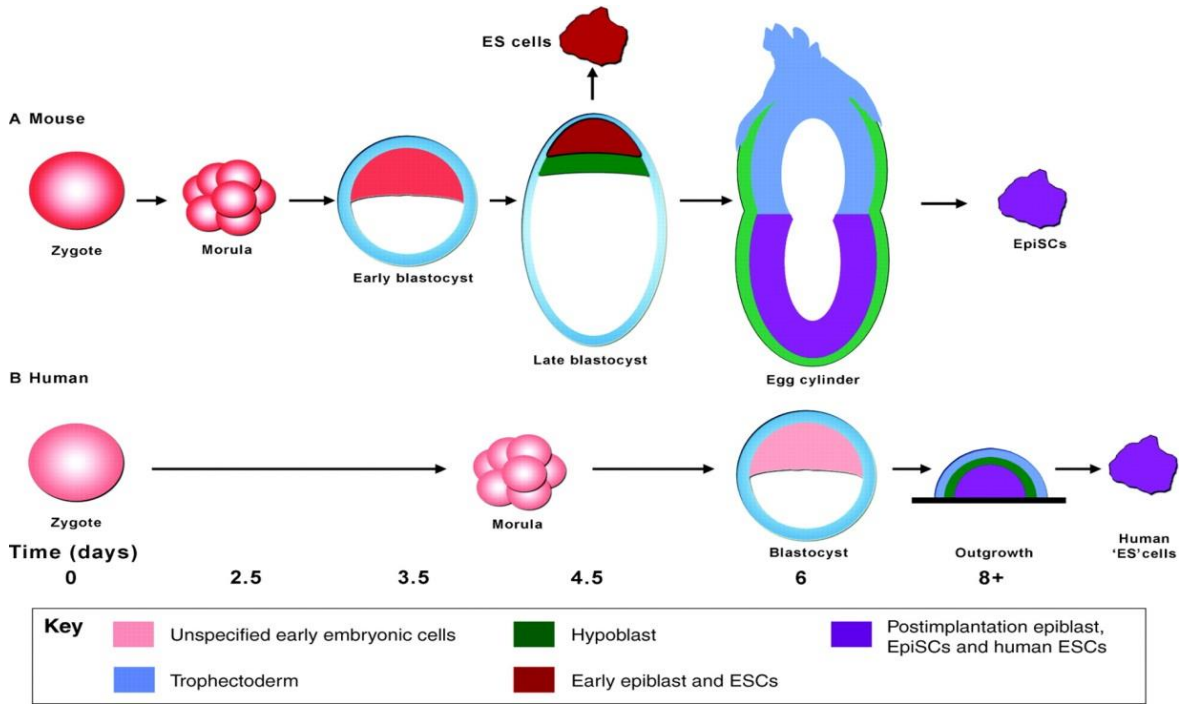


Figure 10 Comparison of pluripotent cell line derivation protocols for (A) mouse and (B) human embryos. This figure highlights the difference in timing of mouse and human development in vivo (in days), and the appearance of pluripotent cell lines *in vitro* (Nichols and Smith, 2011).

Chapter II. The Transcriptional and Epigenetic Control of the Embryonic Stem Cell State

The genome of ES cells is tightly controlled at the transcriptional level. This control is exerted by a large number of actors, including transcription factors (TFs) and elements of the transcriptional machinery, as well as chromatin regulators that determine the epigenetic state of DNA. It has been suggested that ES cells possess a more 'open' chromatin, where chromatin proteins are in a hyper-dynamic state (Meshorer et al., 2006a). This chromatin state would work in parallel with a specific gene expression program that favors the expression of self-renewal genes and poises the differentiation genes. So how do TFs and epigenetic modifications operate in order to maintain pluripotency in ES cells and start differentiation under the appropriate cues?

A. The Transcription Factor Network for ES state maintenance

1. The Core Transcriptional Regulatory Network

Pluripotency maintenance was shown to be mainly controlled and assured by three major TF that form the core regulatory network. These factors are Oct4, Sox2 and Nanog (Chambers and Smith, 2004; Niwa, 2007; Silva and Smith, 2008).

a. Oct4

Oct4 or the octamer-binding transcription factor 4 belongs to the POU family and is encoded by the gene *Pou5f1*. The POU family is characterized by the presence of the POU domain: a 75 amino acid amino-terminal POU specific (POUs) region and a 60 amino-acid carboxyl-

terminal homeodomain (POUh). This domain binds the octamer consensus sequence ATGCAAAT on DNA (Pan et al., 2002).

It has been shown that Oct4 is a pivotal factor and a gatekeeper that prevents ES cells from differentiation (Nichols et al., 1998). Moreover, Oct4 is almost exclusively expressed in ES cells (Nichols et al., 1998; Pesce and Scholer, 2000; Pesce and Schöler, 2001). Studies indicate that during embryonic development, Oct4 is first expressed in all blastomeres. Later on, it becomes only expressed in the ICM; however; at maturity, Oct4 expression becomes restricted to the developing germ cells (Pesce and Scholer, 2000; Pesce and Schöler, 2001). When Oct4 was disrupted in mice, the embryos lacked a pluripotent ICM (Nichols et al., 1998) , further emphasizing its role in pluripotency.

More interestingly, the correct level of expression of Oct4 is essential in the determination of the cell state, as quantitative analysis of Oct4 expression showed that a high level of Oct4 expression drives ES cells towards mesoderm or endoderm lineages, while low levels of Oct4 drive trophectodermal lineages. To maintain the ES cells pluripotency state, a normal level of Oct4 is required (Niwa et al., 2000).

Oct4 is considered to be a pioneer transcription factor, where it was shown not only to bind nucleosome depleted regions and adopt different configurations depending on whether it binds alone or in cooperation with other factors from the core transcriptional regulatory network, but also can bind compacted chromatin structures at the beginning of reprogramming (Soufi et al., 2012). Oct4 was shown to bind the regulatory elements, in particular enhancers, of both pluripotency-related genes and differentiation genes, conferring its role in both activation and repression in order to maintain the ES cell state (Chen et al., 2008b). In addition, Oct4 works mostly in association with the other two core TFs Sox2 and Nanog in order to assure the proper transcriptional regulation (to be more detailed further on).

Proteomic analyses of Oct4 interactome revealed a wide range of actors that interact in a direct or indirect manner with Oct4 (van den Berg et al., 2010a; Pardo et al., 2010). This interactome revealed various actors including TFs such as Esrrb and Klf4, chromatin remodeling complexes such as members of NuRD, esBAF and Tip60 and cofactors most of them known to play an important role in pluripotency maintenance.

b. Sox2

Sox2 or the Sex determining region Y-box 2 is a member of the Sox family of TFs. This family is characterized by the presence of a conserved high-mobility-group (HMG) that binds DNA on an CTTTG(T/A)(T/A) motif (Chambers and Smith, 2004). Sox2 has distinct biological functions and is essential during development. This distinct functionality is due to the interaction of Sox2 with various cofactors during development. Many factors have been shown to influence binding of Sox proteins to their target genes (Wegner, 2010). Sox2 expression is first detected at the morula stage, later on , it becomes located in the ICM of the developing blastocyst and the epiblast (Avilion et al., 2003). The deletion of Sox2 at the zygotic level cause embryonic lethality due the incapacity of the formation of a pluripotent epiblast; however, Sox2 doesn't seem to affect the formation of the trophoectoderm (Wegner, 2010).

Interestingly, during development, Sox2 continues to be expressed, majorly in the central nervous system after gastrulation, conferring its possible role in neural differentiation (Wegner and Stolt, 2005). Sox2 induces neural differentiation by repressing key regulators of other lineage-linked genes (Thomson et al., 2011; Wang et al., 2012; Zhao et al., 2004). This ES cell specification control was not only observed for Sox2 but also for the other two TFs Oct4 and Nanog where they rather promote the differentiation to mesendoderm lineages (Thomson et al., 2011).

c. Oct4/Sox2 complex

Oct4 and Sox2 were found to be co-expressed in several pluripotent cells such as the cells of the morula, ICM, epiblast, and germ cells. It was demonstrated that the Oct4/Sox2 complex works in a synergic way to control transcription, where a physical interaction was detected between the two proteins (Ambrosetti et al., 1997). It was further shown that the octamer elements within the enhancers of Oct4 controlled genes were found in proximity to Sox2-binding sox elements (Avilion et al., 2003) creating a juxtaposed oct-sox binding motif. Moreover, a cooperative Oct4/Sox2 mechanism was observed at Oct4 target genes, where oct-sox binding motifs were present (Loh et al., 2006). Indeed, Sox2 CHIP showed its binding at the majority of Oct4-occupied loci (especially at key regulatory regions of *Pou5f1, Sox2, Nanog, Fgf4* and other pluripotency related genes) emphasizing the cooperation of these two factors to control the gene expression of their targets (Loh et al., 2006), and showing that not only does this Oct4/Sox2 complex control other genes, but also they auto regulate their own enhancer elements. However, Masui et al. demonstrated that Sox2 is not essential for the activation of these oct-sox enhancer motifs, as the deletion effect of Sox2 can be saved by a forced expression of Oct4, conferring its probable role as rather an Oct4 stabilizer at its motifs (Masui et al., 2007).

d. Nanog

Nanog is the last discovered transcription factor of the three core TFs (Chambers et al., 2003; Mitsui et al., 2003). It is a homeobox transcription factor that recognizes the consensus sequences ATTAT (Mitsui et al., 2003). In *Nanog* null mice, the ICM has impaired development. Nanog transcripts were shown to appear at maximal levels between the late morula and the mid-blastocyst and down regulated just prior to implantation (Chambers et al., 2003). It was proposed that Nanog interferes at the blastocyst stage, after the initial action of Oct4 that begins at the morula stage and determines whether cells should remain pluripotent

or become a trophoectoderm. At the later blastocyst stage, Nanog interferes and determines whether cells of the ICM remain pluripotent or differentiate into primitive endoderm (Figure 11) (Mitsui et al., 2003).

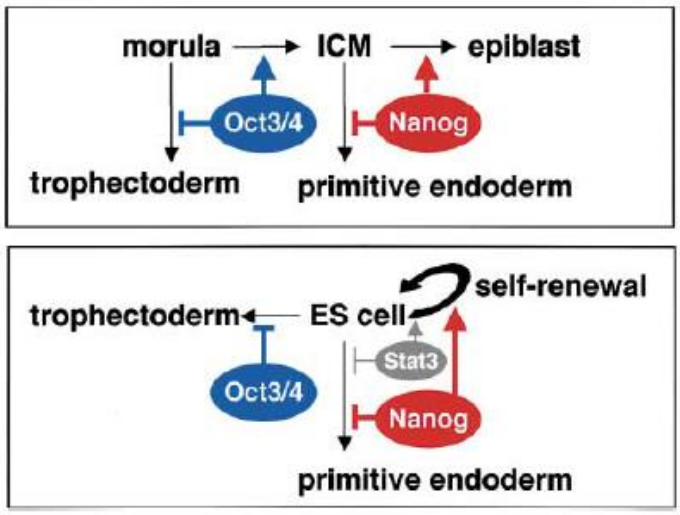


Figure 11 The proposed function of Oct4 and Nanog in preimplantation embryos (upper) and in ES cells (lower) (Mitsui et al., 2003).

Nanog discovery by Mitsui and Chambers have opened the gate to the characterization of LIF/STAT3 independent pathways that control the pluripotency of ES cells. They further demonstrate that the over expression of Nanog is enough to maintain the pluripotency of ES cells; where this was not the case of the over expression of the other core transcription factor Oct4, where ES cells differentiate into endoderm (Niwa et al., 2000). Therefore, it was proposed that Nanog has an essential role in pluripotency maintenance independently from the LIF/STAT3 pathway and the over expression of Nanog is sufficient to maintain the ES cell state in the absence of LIF.

The importance of Nanog in the maintenance of pluripotency was debated in later studies. A study has shown that Nanog null ES cells conserve their pluripotency characteristics and are capable to differentiate correctly confirming that LIF pathways can maintain pluripotency independently of Nanog, but the self-renewal capacities of such Nanog null ES cells seems to

be hindered (Chambers et al., 2007). Indeed, Chambers demonstrates that Nanog is required for primordial germ cells to prosecute the germ-cell development program beyond E11.5.

Moreover, it was shown that Nanog has different expression levels across ES colonies and can vary in each ES cell itself when in culture. This notion of a heterogeneous expression of Nanog reflects that ES cells are maintained in a condition where high levels of Nanog preserve pluripotency and lower levels give a window of opportunity for lineage commitment (Chambers et al., 2007). Indeed, an exogeneous control of Nanog expression can help produce a more homogeneous ES cell population (MacArthur et al., 2012). Furthermore, Abranches emphasizes more the role of Nanog and speculates that not only does it allow ES cells to explore their commitment opportunities, but also acts as an autonomous component of the pluripotency gene regulatory network, where it buffers the molecular heterogeneity and controls cellular decision-making in ES cells (Abranches et al., 2013, 2014).

At the transcriptional level, Nanog doesn't work alone (even though it might belong to differential signaling pathways), it rather cooperates with Oct4 in order to govern the ES cell state (Liang et al., 2008a; Loh et al., 2006). Loh et al. show that Nanog and Oct4 binding patterns overlap and depletion of either one of these factors influences a common important target of genes; conferring the interconnection between the two factors in order to control pluripotency, self-renewal and fate determination of ES cells.

e. Interplay between the core regulatory factors to regulate transcription

The role of the Oct4/Sox2/Nanog trio in the control of ES cell state resides in two mechanisms (Figure12): First, the three TFs co-occupy enhancer regions and positively control the transcription of genes necessary for pluripotency maintenance (Chen et al., 2008b; Loh et al., 2006); on the other hand they occupy repressed genes encoding for differentiation regulators by recruiting different repressing complexes to such gene sites (Bilodeau et al.,

2009; Loh et al., 2006; Pasini et al., 2008). Second, the core TFs are capable to regulate their own promoters forming an interconnected autoregulatory loop that provides positive feedback expression for pluripotency maintenance. The deregulation of one of the core TFs alters this loop and causes the cell entrance in the different differentiation programs (Boyer et al., 2006; Loh et al., 2006).

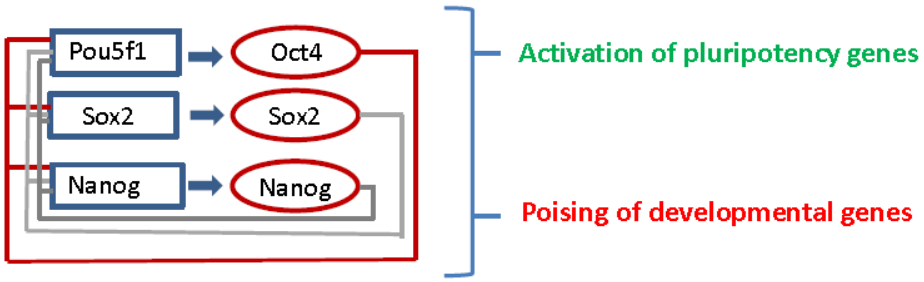


Figure 12 Oct4, Sox2 and Nanog collaborate to control their own promoters forming an autoregulatory loop. Together these core transcription factors function to activate pluripotency genes and inhibit developmental genes. Adapted from (Young, 2011).

2. The Expanded Transcriptional Regulatory Network

The core TFs Oct4, Sox2 and Nanog form the central unit of transcriptional control in ES cells. However, these factors are a part of a much bigger transcriptional regulatory network, composed of a large number of additional TFs, cofactors, chromatin regulators and non-coding RNAs.

a. Important transcription factors in ES cells

As mentioned in Chapter I, the main signaling pathways that control the ES cell state are the LIF, Wnt and BMP pathways. These external signaling cues have downstream internal transcription responses. Three main TFs play essential roles in communicating the various signaling cues into the core transcription network. They include, **STAT3** (LIF signaling), **Tcf3** (Wnt signaling) and **Smad1** (BMP signaling), these factors' importance was discussed

in Chapter I. Several studies have expanded gradually the transcription regulatory network responsible for the control of the ES cell state.

c-Myc:

Myc (Myelocytomatosis oncogene) belongs to a family of helix-loop-helix/leucine zipper TFs. It's potential role in ES cells was speculated through studies on human tumors where it was shown to delay differentiation and promote cell proliferation (Knoepfler et al., 2002; MacLean-Hunter et al., 1994; Pelengaris et al., 1999; Schreiner et al., 2001). Further studies showed that high c-Myc expression is required for the ES cell self-renewal through downstream the LIF/STAT3 signaling pathway (the overexpression of c-Myc replaces the need for LIF supplement) and inhibiting c-Myc causes ES cell differentiation (Cartwright et al., 2005). c-Myc was shown to control the transcription in ES cells through its capacity to bind E box sequences at core promoters and stimulating RNA polymerase II pause release (Rahl et al., 2010). Moreover, c-Myc was shown to be mostly bound on the regulatory elements of transcribed genes along with the core TFs (Young, 2011). c-Myc is one of the TFs essential in reprogramming into induced pluripotent stem cells (Takahashi and Yamanaka, 2006).

Esrrb:

The estrogen related receptor beta is an orphan nuclear receptor shown to be a part of the core pluripotency network, where it was shown to interact with Oct4 (van den Berg et al., 2010a; Chen et al., 2008b; Loh et al., 2006). The importance of Essrb in self-renewal was revealed by RNA interference studies where it was shown to be essential in ES cell self-renewal and its absence triggers ES cell differentiation (Ivanova et al., 2006; Loh et al., 2006). Interestingly, Essrb overexpression was shown to substitute for a short-term LIF requirement (Zhang et al., 2008).

Furthermore, Essrb was shown to regulate Nanog's activity where it mediates the positive regulatory effect of Oct4 (van den Berg et al., 2008). Moreover, recent studies have shown that Essrb overexpression can even substitute Nanog function in ES cells and its deletion causes severe impaired self-renewal (Festuccia et al., 2012). However, the derivation of Essrb null ES is possible (like Nanog null ES cells), in contrast to the absolute requirement of Oct4, this reflects the relative importance of such factors, where Essrb, like Nanog seem to help more in the fine tuning of the gene expression in ES cells.

Sall4:

Sall4 belongs to the Spalt family of zinc-finger TFs, mutations in the *Sall4* gene cause numerous developmental defects (Al-Baradie et al., 2002; Kohlhase et al., 2002). *Sall4* null mice die shortly after implantation; however these mice do not present ICM defects and ES cells isolated from such ICM retain their pluripotency with a slower proliferation rate (Sakaki-Yumoto et al., 2006).

The role of Sall4 was studied in several reports. Sall4 RNA interference experiments have demonstrated that Sall4 depleted ES cells differentiate and fail to maintain their self-renewal on feeder free culture conditions (Zhang et al., 2006). Furthermore, Zhang et al show that Sall4 binds to *Pou5f1* regulatory elements where it affects Oct4 expression in a dosage-dependent manner. Additional studies also show that Sall4 binds *Nanog* regulatory elements and Nanog gene targets (Wu et al., 2006), conferring its involvement in the core transcription circuitry. Nevertheless, the absolute necessity of Sall4 in ES cell pluripotency is debated. As a study shows that Sall4 null ES cells retain their pluripotency even on feeder-free culture conditions (Yuri et al., 2009). This study suggests that Sall4 acts as a stabilizer of the ES cell state by repressing trophoectoderm lineage differentiation.

Tbx3:

The T-box transcription factor family was shown to be important in a variety of developmental processes (Miller et al., 2008). Tbx3 is expressed early in the mouse inner cell mass, and it is later expressed in extra embryonic endoderm cells (Chapman et al., 1996). In ES cells, Tbx3 was reported to be essential for ES cell self-renewal (Ivanova et al., 2006), and the continuous expression of Tbx3 allows the maintenance of ES cell in an undifferentiated state in the absence of LIF (Niwa et al., 2009). Interestingly, Niwa et al. show that Tbx3 is an upstream regulator of *Nanog* gene along with Klf4 where it boosts the expression of core pluripotency genes (Figure 13) without being essential in pluripotency maintenance. Indeed, Tbx3 was reported to improve the germ-line competency of induced pluripotent stem cells (Han et al., 2010).

Klf4:

The Krüppel-like factor (Klf) family is an evolutionarily conserved family of zinc finger TFs that play a role in different biological processes, including proliferation, differentiation, development and apoptosis (McConnell et al., 2007). Klf4 is a member of the Klf family, studies have shown to interact with Oct4 in order to activate Oct4/Sox2 target genes (Nakatake et al., 2006). Klf4 is one of the TFs used in reprogramming into induced pluripotent stem cells (Takahashi and Yamanaka, 2006).

The overexpression of Klf4 prevents differentiation of ES cells, emphasizing more its role in self-renewal (Li et al., 2005). However, Klf4 depletion does not seem to effect on ES cell morphology or self-renewal (Jiang et al., 2008; Nakatake et al., 2006), due to the presence of other Klf proteins, mainly Klf2 and Klf5 with redundant functions in ES cells. Indeed, the depletion of the three factors caused the differentiation of ES cells (Jiang et al., 2008).

Klf2/4/5 TFs were also shown to be bound to regulatory elements of key pluripotency genes such as *Pou5f1*, *Sox2*, *Nanog* and *Essrb*.

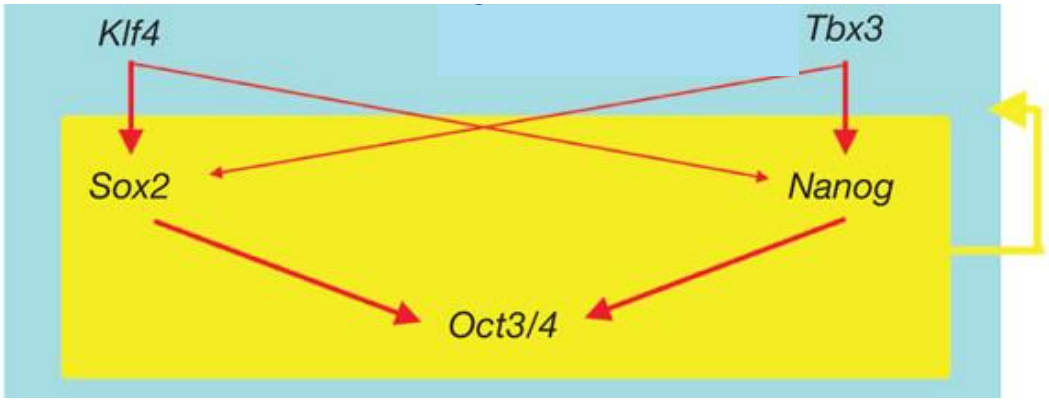


Figure 13 Transcription control by Klf4 and Tbx3. Klf4 and Tbx3 mainly activate Sox2 and Nanog, respectively, and maintain expression of Oct3/4. Transcription of all these transcription factors is positively regulated by Oct3/4, Sox2 and Nanog (Niwa et al., 2009).

Foxd3:

Foxd3 belongs to the family of fork head box TFs. Initially, it was solely identified in ES cell and their malignant progenitors (Sutton et al., 1996). Later on, it was demonstrated that Foxd3 is also expressed during embryogenesis in the epiblast and even the neural crest (Dottori et al., 2001; Hromas et al., 1999; Labosky and Kaestner, 1998). Foxd3 knockdown mice die early during embryogenesis with a loss of the epiblast; moreover Foxd3 knockdown cells fail to proliferate and give a normal ICM (Hanna et al., 2002).

Zfx:

Zfx is a zinc finger TFs of the Zfy family, a family highly conserved in vertebrates (Schneider-Gädicke et al., 1989). It was shown to be required in ES cell renewal but not essential for pluripotency, where Zfx deletion caused impaired cell proliferation and increase apoptosis but did not influence the differentiation potential (Galan-Caridad et al., 2007). Interestingly, Zfx was shown to upregulate the expression of c-Myc in order to enhance self-renewal (Fang et al., 2014).

Ronin:

Ronin, referring to the master-less Japanese samurai, is a zinc finger transcription factor. This factor was shown not to have a direct relationship with the core TFs and it seemed to be able to partially override the necessity for Oct4 (Dejosez et al., 2008). Moreover, Ronin independency from the core transcription network was further demonstrated with its persistent expression upon knockdown of *Oct4, Sox2* and *Nanog.*

Ronin knockdown causes ES cell death, probably due to the activation of repressed genes simultaneously or due to an unknown effect on apoptosis pathways. Moreover, Ronin knock down during embryonic development show a common phenotype with Oct4 deletion, where there is failure of ICM formation and embryo lethality (Dejosez et al., 2008).

Dax1:

Dax1 (DSS-AHC on X chromosome gene) is an atypical orphan nuclear receptor that has been shown to play a role in ES cell pluripotency (Kelly et al., 2010; Niakan et al., 2006; Sun et al., 2009). *DAX1* gene knockdown caused mouse ES cell differentiation (Niakan et al., 2006). It was demonstrated that Dax1 interacts with Esrrb and represses its function (Uranishi et al., 2013). Moreover, Dax1 was shown to regulate in a negative way the expression of Oct4, where it seems to fine tune the expression of Oct4 in order to maintain ES cell pluripotency (Sun et al., 2009).

Other transcription factors with potential roles in ES cells:

Several core pluripotency factors interactome studies revealed a large number of involved TFs with potential roles in the regulation of the ES cell state (van den Berg et al., 2010a; Pardo et al., 2010; Wang et al., 2006a). Some of these factors are: Rex1, Rif1, Nac1, Tcfcp2l1, Sox18,

and Zfp281 shown to be a part of the ES cell transcriptional regulatory network with potential roles in the optimal maintenance of the ES cell state.

b. Additional actors important in ES cell self-renewal

Transcription Cofactors:

Cofactors are capable to interact selectively and non-covalently with TFs and the basal transcription machinery in order to regulate transcription. They can either activate (coactivators) or repress (corepressors) gene transcription. Cofactors generally do not bind DNA, but rather assure the protein-protein interactions between TFs and the transcription machinery. They are expressed in all cell types, but ES cells seem to be quite sensitive to reduced levels of such cofactors (Fazzio and Panning, 2010; Kagey et al., 2010).

Mediator is one of the most important coactivators of transcription. In ES cells, the mediator was shown to physically link Oct4/Sox2/Nanog bound enhancers to the promoters of active genes (Kagey et al., 2010). This physical link is assured by a second cofactor, the cohesin (Figure 14). **Cohesin** was shown to assure the DNA looping needed to approach regulatory elements and activate transcription in ES cells (Kagey et al., 2010). ChIP-seq data demonstrate the co-occupancy of mediator and cohesion with the core TFs at regulatory elements (Young, 2011).

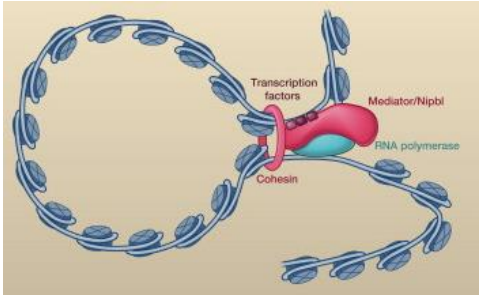


Figure 14 Model for DNA looping by mediator and cohesion (Young, 2011)

The proper chromatin structure and condensation is assured by an additional cofactor, the condensin. **Condensin** complexes were shown to be required in ES cells where the deletion of such complexes causes various phenotypic alterations at the level of the chromatin structure (Fazzio and Panning, 2010).

Corepressors were also found to play a role in the maintenance of ES cells. **Cnot3** (Ccr4-Not) and **Trim28** (Tripartite motif-containing 28) are two corepressors shown to co-occupy promoter regions with important TFs in ES cells such as c-Myc and Zfx (Hu et al., 2009a). The down-regulation of Cnot3 and Trim28 causes ES cell differentiation into trophectoderm and primitive ectoderm respectively, conferring their role in self-renewal. Moreover, Hu et al. demonstrate that c-Myc/Zfx/Cnot3/Trim28 form a distinct pluripotency module from the Oct4/Sox2/Nanog module, indicating the increased transcriptional regulatory network complexity in ES cells.

Chromatin regulators:

Many chromatin regulators were shown to play important roles in ES cells (Fazzio and Panning, 2010; Kagey et al., 2010; Leeb et al., 2010; Meissner, 2010; Niwa, 2007). Chromatin regulators can be divided into four groups: Cohesin/condensing complexes (discussed earlier), histone-modifying enzymes, ATP-dependent chromatin remodeling complexes and DNA methyltransferases (Young, 2011).

Histone-modifying enzymes encompass a significant number of different complexes that participate in either the activation or repression of transcription in ES cells. They include Polycomb complexes, SetDB1, Tip60 and TrxG proteins, the function of these proteins will be detailed further in this chapter.

ATP-dependent chromatin remodeling factors were shown to regulate transcription, presumably through their capacity to modify nucleosome positioning and occupancy (Clapier

and Cairns, 2009a; Ho and Crabtree, 2010). Several remodeling factors were shown to be involved in the control of the ES cell state. The different remodeling complexes involved in ES cell transcriptional control will be discussed later in Chapter II and more widely in Chapter III.

DNA methylation does not seem to be essential in ES cell maintenance. Indeed, although DNA methyltransferases (Dnmts) are expressed in ES cells and about 60-80% of the CpG islands are methylated, ES cells can be established and maintained in the absence of Dnmts and DNA methylation (Meissner, 2010). However, DNA methylation becomes required during differentiation, as Dnmt-deficient ES cells fail to properly differentiate (Jackson et al., 2004).

Noncoding RNAs:

miRNAs

Evidence have shown that several microRNAs (miRNAs) play essential roles during developmental, where they were demonstrated to regulate the expression of a large number of genes (Farh et al., 2005). A group of miRNAs was shown to play a role in the control of the ES cell identity (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2008). Furthermore, these miRNA polycistrons where shown to be controlled by the core TFs Oct4/Sox2/Nanog (Marson et al., 2008a). Marson et al. demonstrate that the core TFs regulate the transcription of miRNAs in ES cells. miRNA genes involved in ES cell maintenance and proliferation are activated, where these miRNAs (mainly the most abundant mir 290-295 cluster) fine-tune the expression of the ES TFs. On the other hand, Oct4/Sox2/Nanog occupy the genes of developmental miRNAs along with repressing complexes poising their expression (Marson et al., 2008b).

lncRNAs:

Recent studied have revealed the role of long non-coding RNAs (lncRNAs) in the control of the ES cell state (Guttman et al., 2009; Khalil et al., 2009). It was shown that a set of lncRNA genes are bound by the core TFs, and the downregulation of such lncRNAs results in dramatic effects on the expression of the core circuitry indicating a potential feedback loop created by lncRNAs (Sheik Mohamed et al., 2010). On the other hand; lncRNAs were also shown to be involved during differentiation. A lncRNA called Xist is responsible for X chromosome imprinting during development and ES cell lineage commitment (Navarro and Avner, 2009). Moreover, lncRNAs where shown to be involved in later developmental steps, inducing the expression of lineage associated genes (Ng et al., 2012). Interestingly, the expression of lncRNAs was shown to be correlated with the development potential of induced pluripotent stem cells (Liu et al., 2010).

3. Pluripotency transcription factors and iPS cell generation

a. Scientific milestones leading to induced pluripotent stem cell generation

Reprogramming of somatic cells into pluripotent embryonic stem cells first started using nuclear transfer (NT) approaches. The early experiments in amphibians and later in mammals showed that terminally differentiated somatic cells were able to generate cloned animals (Briggs and King, 1952; Gurdon, 1962; Wilmut et al., 1997). This ability of a somatic cell to revert to an early embryonic state rose questions about the nature of genome changes upon differentiation, introducing the probable role of epigenetics (Hochedlinger and Jaenisch, 2002). Moreover, this capacity of somatic cells to dedifferentiate in oocyte conditions led to the speculation of the presence of specific factors in the oocyte that are capable to induce pluripotency.

Another strategy to induce the reprogramming is fusing adult somatic cells with ES cells in culture generating tetraploid hybrid cells (Cowan et al., 2005; Tada et al., 1997, 2001). These observations indicated that there are some nuclear factors present in ES cells (as in oocytes) that are responsible for this reprogramming mechanism (Do and Schöler, 2004; Egli et al., 2007). More recent studies have demonstrated that the overexpression of pluripotency factors such as Nanog enhances the formation of the reprogrammed hybrids about 200 folds (Silva et al., 2006). Indeed, previous work have demonstrated that the expression of certain TFs in certain mature terminally differentiated cells is sufficient to reprogram them into certain adult stem cell progenitors; by the same manner, the deletion of certain TFs can lead to reprogramming to multipotent progenitors (Davis et al., 1987; Laiosa et al., 2006; Nutt et al., 1999). These observations have led to the rational thinking that the overexpression of ES cells TFs in somatic cells could lead to their stable reprogramming.

b. Induced pluripotent stem cell generation from mouse and human somatic cells

The first attempt to generate induced pluripotent stem cells (iPS cells) from somatic mouse cells was conducted by Takahashi and Yamanaka. Experiments were conducted on mouse fibroblasts, where 24 ES cell TFs were introduced with retroviral vectors and reprogrammed cells were drug selected for the ES cell-specific gene *Fbx15* (Takahashi and Yamanaka, 2006). In those experiments, Takahashi and Yamanaka were capable of identifying the 4 TFs Oct4, Sox2, Klf4 and c-Myc (the OSKM factors) sufficient to reprogram mouse fibroblasts into iPS cells. This first generation of iPS cells was similar to ES cells but not identical, as transcriptional and epigenetic patterns appeared to be only partially reset and such cells could not form mouse chimeras when injected into blastocysts (partially reprogrammed cells).

Later studies, gave rise to the second generation of iPS cells from mouse fibroblasts. The iPS cell drug selection was based this time on the usage of essential pluripotency genes such as *Pouf51* (Oct4) and *Nanog* (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007). This

second generation of ES cells with transcriptional and epigenetic patterns highly similar to those in ES cells and can form chimeric mice. The different steps and the molecular changes involved in the direct reprogramming to iPS cells are demonstrated in Figure 15.

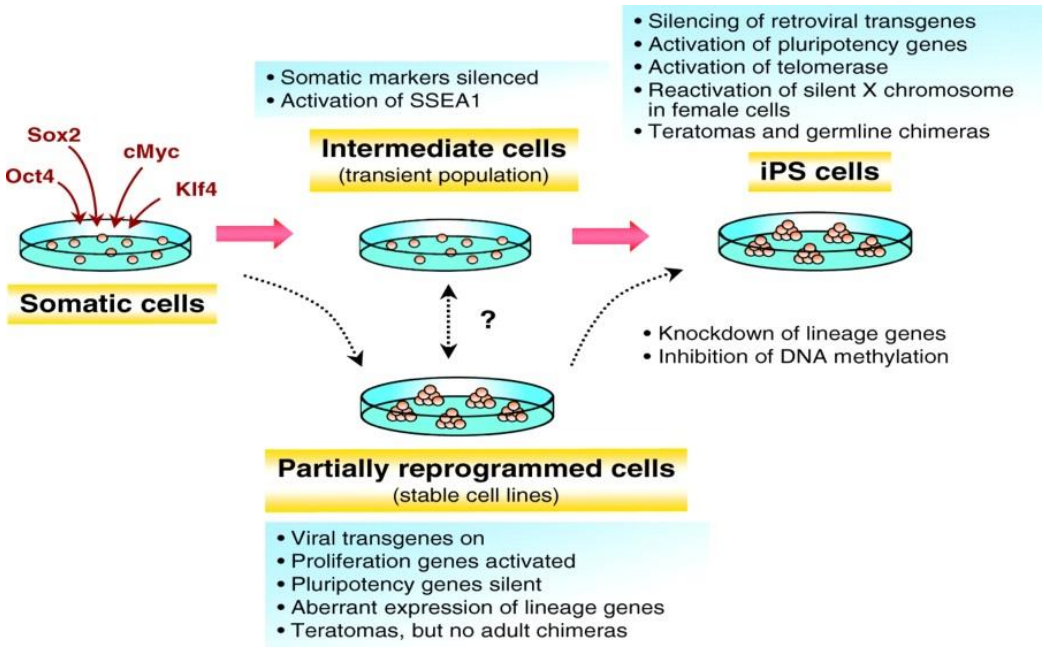


Figure 15 Steps involved in direct reprogramming to pluripotency. The starting, intermediate and end stages of reprogramming to pluripotency that can be identified during the generation of iPS cells (Hochedlinger and Plath, 2009)

Since the iPS cell derivation from mouse fibroblasts, a dozen of studies have been capable to derivate iPS cells from various cell types including the blood, liver, stomach, pancreas, brain, intestine and adrenals. Furthermore, the generation of human iPS cells was also successful from fibroblasts (Lowry et al., 2008; Park et al., 2008; Takahashi et al., 2007) and keratinocytes (Aasen et al., 2008; Maherali et al., 2007) using the OSKM cocktail or a variant combination of factors such as the core TFs Oct4/Sox2/Nanog with a miRNA called LIN28 (shown to enhance c-Myc function) (Yu et al., 2007).

B. Chromatin Modifying Enzymes and the Control of the ES Cell State

Chromatin, DNA packaged with histones, assures cell-specific gene expression. The chromatin structure is under the control of different chromatin modifying enzymes. Chromatin modifying enzymes fall into two major groups: Histone modifying enzymes that remove or add covalent histone modifications to histone tails and ATP-dependent chromatin remodeling factors that use the energy of ATP to modify the nucleosomal chromatin architecture. Several proteomic and loss-of-function RNA interference studies were conducted to reveal the participation of such enzymes in the transcriptional regulatory network in ES cells.

1. Proteomic interactome studies reveal numerous chromatin modifying enzymes as a part of the transcriptional regulatory network

The interactome of the core transcription regulatory network has been the subject of a number of studies interested in revealing the different actors that play a role in the regulation of the ES cell state.

A number of studies show the presence of chromatin modifying enzymes as part of the core transcription factor proteomic interactome (Liang et al., 2008a; Pardo et al., 2010; van den Berg et al., 2010b).

A FLAG-affinity based mass spectrometry study revealed a great number of Oct4 interacting proteins (van den Berg et al., 2010b). Van den Berg et al. presents an Oct4 interactome that identified a number of TFs and chromatin modifying complexes with previously known roles in the regulation of ES cell state, in addition to a series of newly identified actors. Primary immune-precipitation of Oct4 partners revealed a list of about 50 proteins, including about 22

TFs, some of them with known roles in ES cells (such as Sall4, Essrb, Klf5 and Sox2); moreover, a number of chromatin modifying enzymes (histone modifying enzymes and chromatin remodeling factors) were identified with the majority of these proteins found to be essential during development. This set included the various subunits of the NuRD repressive complex (including the histone deacetylases Hdac1/2 and the chromatin remodeling factor Chd4), members of the SNF family of chromatin remodeling factors (Brg1, also known as Smarca4), subunits of the repressive Polycomb complex (PRC1), members of the LSD1 histone demethylase complex and subunits of the Trrap/p400 complex including the chromatin remodeling factor Ep400. Van den Berg et al. further purify the four different TFs Sall4, Essrb, Dax1 and Tcfcp2l1 shown to be partners in the Oct4 proteomic interactome, which revealed common and unique sets of interacting proteins. The combined interactomes from the five TFs resulted in a dense Oct4 regulatory network of about 160 different partners (Figure 16). Interestingly, Nanog was not identified in this network although several previous studies showed Nanog interactions with pluripotency TFs (including Sall4 and Oct4) (Liang et al., 2008a; Wang et al., 2006b; Wu et al., 2006), maybe due to the fact it might be hard to detect by mass spectrometry (resistance to digestion).

Similarly a second Oct4 proteomic interactome study using tagged Oct4 followed by mass spectrometry revealed a broader interactome (Pardo et al., 2010). Mass spectrometry results identified 92 proteins in the tagged Oct4 purifications (Figure 17A). A considerable number of these factors were also identified in van den Berg et al. interactome. However, Pardo et al. identified Nanog in the Oct4 partners and an additional series of chromatin remodeling factors such as INO80, Smarca5, Chd1. Moreover, about 80% of the Oct4 interacting proteins were shown to be essential in development, as the deletion of these proteins is lethal (Figure 17B).

Analysis of Nanog proteomic interactome also revealed the association of chromatin modifying enzymes with this core TFs (such as members of the NuRD complex and the SNF2 remodeler Brg1) (Liang et al., 2008b).

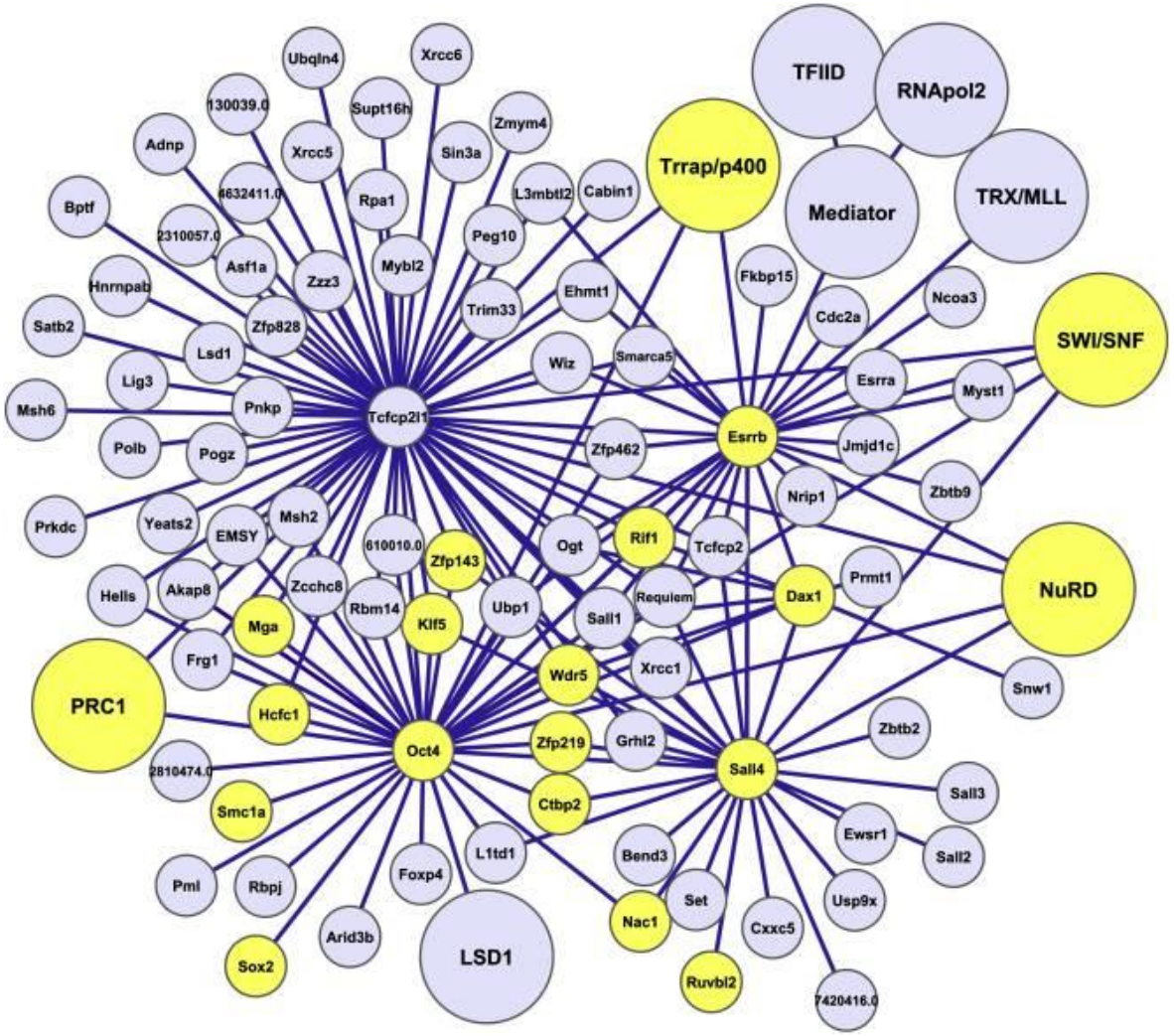


Figure 16 Protein Interaction Network of Oct4 and Its Associated Proteins Sall4, Dax1, Tcfcp2l1, and Esrrb (van den Berg et al., 2010b)

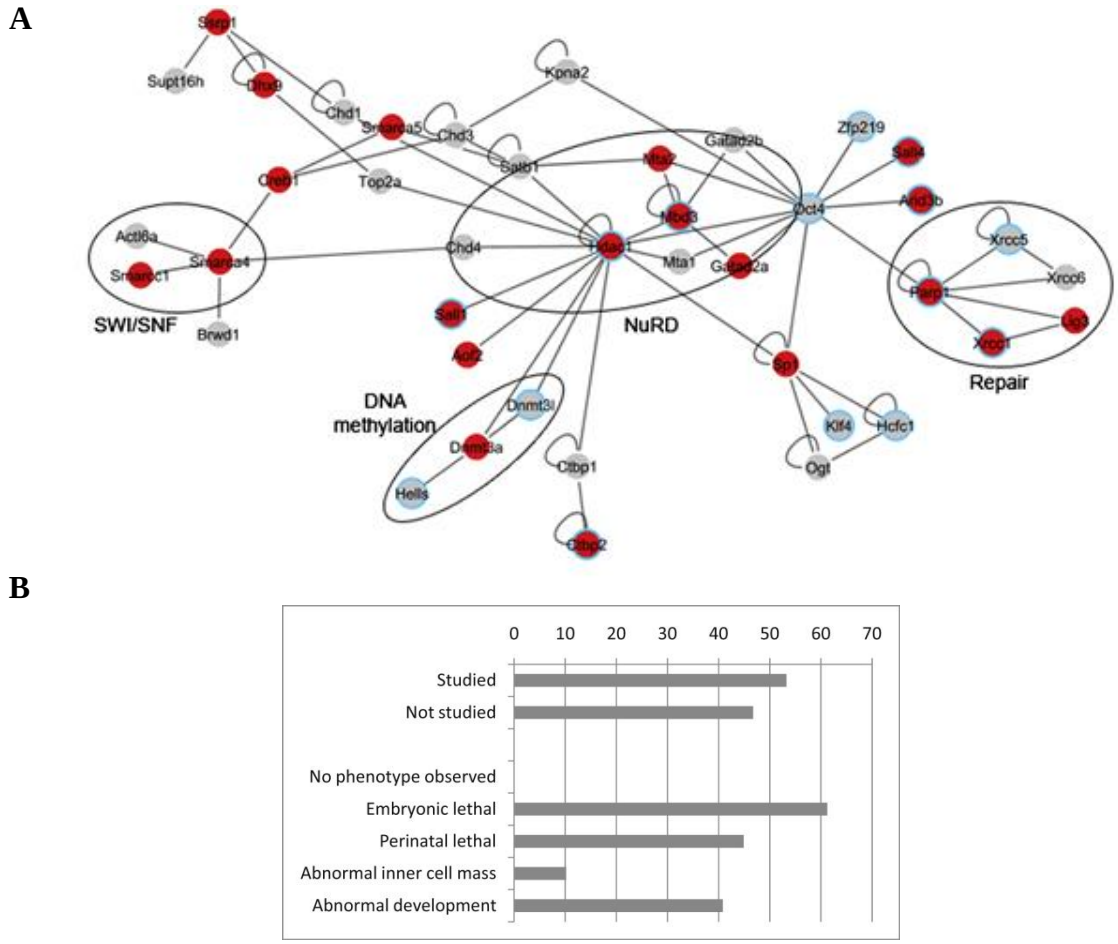


Figure 17 The Oct4 network and its importance during development. A) Network of protein-protein interactions within the Oct4 dataset. Blue circles are proteins downregulated upon ES cell differentiation. Red fill indicates proteins whose absence results in embryonic lethality in the mouse B) Percentage Distribution of Phenotypes Caused by Mutations in the Genes Encoding Oct4-Associated Proteins (Pardo et al., 2010).

2. RNA interference-based screens for pluripotency controlling factors reveal a series of important chromatin remodeling enzymes

ATP-dependent chromatin remodeling factors or remodelers were shown to be essential during development. Several studies have shown that mutations in key remodelers such as Ep400 (also known as (p400) and Brg1 (Smarca4) are lethal in preimplantation embryos (Bultman et al., 2000; Gorrini et al., 2007). Several RNA interference (RNAi)-based screens identified a series of chromatin remodelers as regulators of the ES cell identity (Fazzio et al., 2008; Gaspar-Maia et al., 2009).

In one study, Fazzio et al. show the importance of the chromatin regulatory complex Tip60/p400 in the maintenance of ES cell features (Fazzio et al., 2008). Tip60/p400 has histone acetyltransferase and nucleosome remodeling functions and it was shown that mutations in *Tip60* or *Trrp* (a component of the Tip60/p400 complex) cause preimplantation lethality (Gorrini et al., 2007; Herceg et al., 2001). In the goal of identifying chromatin regulatory complexes involved in ES cell maintenance and differentiation, Fazzio et al. performed a large-scale RNAi-based screen of about 1008 genes encoding chromatin proteins. They identified 68 genes with phenotypes resulting from the knock-down. Among the 68 targets, subunits of the Tip60/p400 complex were revealed to be important in ES cells. The down regulation of the ATP-dependent chromatin remodeling factor p400 (Ep400) induces a change in ES cell morphology associated to loss of pluripotency. In addition, the KD of Ep400 effected the expression of about 802 genes, with about 128 upregulated and 674 down regulated genes conferring a rather repressive role of Ep400. Indeed, most of the upregulated genes are key developmental loci. However, the expression of key transcriptional factors was not changed suggesting a role of Ep400 rather downstream the ES core TFs. Interestingly, Nanog knock down showed an overlap in the transcriptomic deregulation profile with Ep400 suggesting a mutual regulation of a set of genes. Moreover, a strong correlation between H3K4me3 and Ep400 binding was observed both at active and silent genes (presenting the H3K27me3: bivalent genes) with highest binding at the most active genes. In addition, Nanog was shown to be indirectly involved in the binding of Ep400 to its target genes suggesting that the Tip60/p400 complex to be an important effector of Nanog transcriptional repression.

A second study reveals the role of a different chromatin remodeling factor in the maintenance of the ES cell profile. RNAi screening of 41 candidate factors in the regulation of pluripotency identified Chd1 (Chromodomain1), an ATP-dependent chromatin remodeling factor as an important factor in ES cell pluripotency and self-renewal (Gaspar-Maia et al.,

2009). The authors of this study have shown that Chd1 knock down in ES cells caused defects in cell expansion and a lower activity of the Oct4 promoter, concordant with recent data where Chd1 was identified as an Oct4 interacting partner (Pardo et al., 2010). However, Chd1 knock down resulted in a small number of downregulated genes and a more important upregulation of genes involved in differentiation processes. Indeed, knock down ES cells for Chd1 show increased propensity for abnormal neural differentiation coupled with a loss of endoderm formation capacity. However, ES with low levels of Chd1 keep their global ES transcriptome, suggesting that such levels of Chd1 are sufficient to maintain an appropriate ES cell state. Furthermore, Chd1 was shown to be coupled with active transcription (Sims et al., 2007), where it was demonstrated to be coupled with euchromatin by binding to H3K4me3 not only in ES cells but also in differentiated cells. Similarly, Gaspar-Maia et al. show that Chd1 knockdown causes the increase in the number of heterochromatin marks such as H3K9me3. To sum up, Chd1 was shown to be required for open chromatin in ES cells and for heterochromatin-poor pluripotent cell state maintenance.

Several additional studies that discussed the global transcription implication of chromatin remodeling factors in ES cell maintenance (Efroni et al., 2008; Ho et al., 2011a) revealed that the depletion by RNAi of Brg1 (Smarca4) caused both proliferation and differentiation deficient phenotypes; on the other hand the depletion of a second remodeler (Chd1l), caused severe proliferative defects but had no effect on the differentiation capacities.

C. The Epigenetic Regulatory Landscape of the ES Genome

The chromatin in ES stem cells was suggested to be in a hyper-dynamic state, as a selection of chromatin proteins were found more loosely bound to genomic DNA in ES cells compared to *in vitro* differentiated ES cells (Meshorer et al., 2006b). This particular chromatin state might

contribute to the transcriptional levels of the different genes essential in the conservation of pluripotency and self-renewal. However, silencing of developmental genes to prevent their inappropriate expression in ES cell should also be achieved despite this more dynamic state of the chromatin. It is currently unclear how the mechanisms of transcriptional control might be affected by the particular nature of ES cell chromatin. In this last part of this chapter, I will discuss the current knowledge on the mechanisms that control transcription in ES cells.

1. Promoters in ES cells

a. Promoters and H3K4me3

Promoters are regulatory genome regions that initiate transcription of their target genes. The identification of promoters is usually associated with a particular epigenetic mark which is the trimethylation of the lysine 4 of H3 histone (H3K4me3). The activating Trithorax group of chromatin modifying enzymes is responsible for the deposition of this mark.

Epigenetic control by the Trithorax (TrxG) group and its counterpart repressive group Polycomb (PcG) was first discovered in *Drosophila melanogaster* as activators and repressors of Hox genes, which are a group of TFs that play a role in the cell specification of segmented animals (Brock and Fisher, 2005; Ringrose and Paro, 2004; Simon and Tamkun, 2002). The attribution of the H3K4 methyltransferase activity to TrxG was first given after the identification of the TrxG homologue Ash2 in the yeast Set1 H3K4 methyltransferase complex (Roguev et al., 2001).

Similarly a TrxG orthologue was identified in mammals (Mll1) (Milne et al., 2010). There are six Set1/Trithorax like H3K4me3 methyltransferases identified in mammals (Glaser et al., 2006), all present in similar complexes based on the Set1C/Ash2 scaffold (Ruthenburg et al., 2007; Yokoyama et al., 2004). They include Mll1, Mll2, Mll3, Mll4 (which are Trithorax-like) and Set1A and Set1B (which are Set1 complex-like).

In yeast, deposition of H3K4me3 is associated with active genes (Ng et al., 2003; Santos-Rosa et al., 2002); on the other hand, in mammals the mark is associated with both active and inactive promoters, however at distinct levels (Barski et al., 2007; Guenther et al., 2007). Moreover, it was shown that the deposition of H3K4me3 mark is strongly correlated with the presence of CpG rich islands at promoter regions : (Guenther et al., 2007; Mikkelsen et al., 2007). Initial studies have identified CxxC zinc finger proteins that bind to CpG islands (Lee et al., 2001). Cfp1 or CxxC-finger protein one was shown to be a part of both Set1A and Set1B complexes (Lee and Skalnik, 2005). Similarly Mll proteins also present similar CxxC zinc finger domains (Bach et al., 2009; Birke et al., 2002).

The deposition of the H3K4me3 histone modification in ES cells was intensively investigated in two recent studies (Clouaire et al., 2012; Denissov et al., 2014).

In a genome-wide study Clouaire et al. demonstrate the importance of Set1 complex, particularly the Cfp1 CxxC zinc finger protein, in the deposition of the H3K4me3 mark on active promoters. The absence of Cfp1 in ES cells does not affect self-renewal; however such cells cannot differentiate (Carlone et al., 2005) and somatic cells deficient in Cfp1 have detrimental phenotypes (Thomson et al., 2010; Young and Skalnik, 2007). However, upon Cfp1 deletion in ES cells, there is a wide loss of H3K4me3 at active genes with no effect on poised genes (Clouaire et al., 2012). This loss of H3K4me3 at active genes was coupled with an aberrant reorganization of H3K4me3 on different regulatory elements such as enhancers and CTCF binding sites, suggesting a probable role of Cfp1 in the targeted recruitment of the Set1 methyltransferase complex onto active promoters with rich unmethylated CpG islands. Indeed, a mutant Cfp1 that does not recognize CpG islands is not capable of inhibiting the aberrant Set1 binding at regulatory elements.

The authors further speculate about the real importance of the H3K4me3 mark in activating transcription, as the aberrant H3K4me3 deposition in ES cells did not seem to have an effect on the transcription. However, this could be due to the non-total depletion of this mark, suggesting a redundancy in the function of different H3K4me3 methyltransferases in the global deposition of H3K4me3 on active and poised promoters. Indeed, the identification of Mll1/2 importance early during development (Andreu-Vieyra et al., 2010; Glaser et al., 2006) and the dominance of Set1 in later stages prove the various roles CxxC proteins play in determining the chromatin landscape (Ardehali et al., 2011). Interestingly, the depletion of subunits of the Set1 and Mll1/2 complexes causes the decrease in H3K4me3 at both active and poised promoters (Ang et al., 2011; Jiang et al., 2011), suggesting an interplay between the different methyltrasnferases in the deposition of the H3K4me3 mark in the ES cell genome and a probable role of Mll1/2 methyltransferases in the deposition of H3K4me3 on poised promoters.

The role of Mll1/2 proteins, more particularly Mll2, was further investigated (Denissov et al., 2014). Mll2 as mentioned previously was shown to be required during oogenesis where it seems to be the major H3K4me3 methyltransferase (Andreu-Vieyra et al., 2010). Denissov et al. demonstrate that Mll2 is required for the deposition of H3K4me3 on poised promoters (bivalent genes with the both the activating H3K4me3 and H3K27me3 marks, discussed later in this chapter). CHIP-seq analysis of genome-wide H3K4me3 upon Mll2 depletion shows loss of this mark at a set of bivalent genes but not at active genes. However, Mll2 binds both active and poised promoters, but its depletion only affects H3K4me3 at poised ones. The depletion of H3K4me3 at the set of bivalent genes was not concordant with a loss of differentiation gene activation, with only a small set of influenced genes. Denissov et al. propose a variant hypothesis about the recruitment of Set1 complex on active genes. They propose that Mll2 with its CxxC domain defines CpG rich islands at transcription start sites

(TSS) in active and bivalent promoters, and further loss of the H3K27me3 repressive mark and recruitment of Mll1 and the Set1 complex is required in order to activate genes (Figure 18).

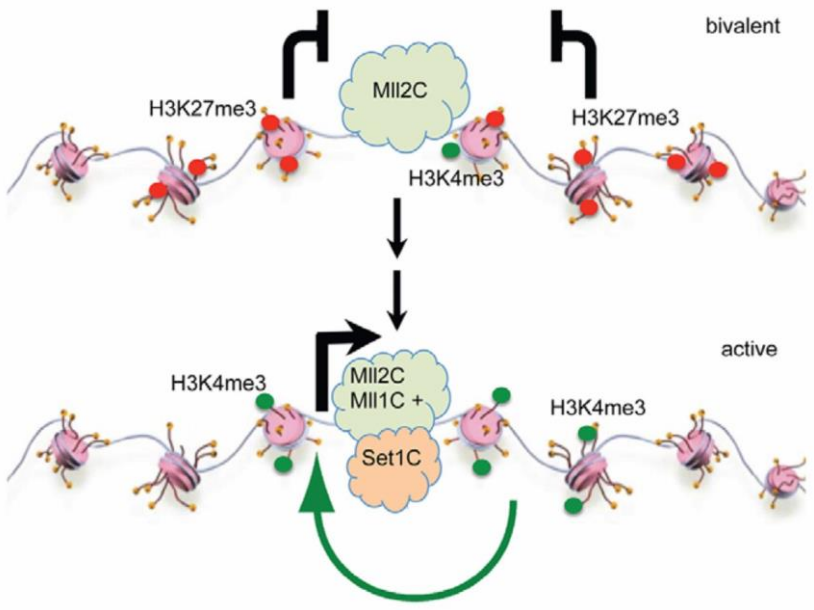


Figure 18 Model for the relationships between Mll and Set1 complexes on bivalent and active promoters (Denissov et al., 2014)

b. Bivalent genes and Polycomb complex

The Polycomb group proteins (PcG) were shown to play a role in the silencing of a wide range of genes during the different developmental stages (Simon and Kingston, 2013). As mentioned previously, this group was initially discovered in *Drosophila Melanogaster*. PcG proteins form two distinct complexes, PRC1 (Polycomb Repressive Complex 1) and PRC2 (Margueron and Reinberg, 2011a; Müller and Verrijzer, 2009; Simon and Kingston, 2013). PRC2 includes Ezh2, which catalyzes the deposition of H3K27me3 (Cao et al., 2002; Kuzmichev et al., 2002; Müller et al., 2002). PRC1 complex is responsible for the ubiquitylation of histone H2A on K119 (Cao et al., 2005) and the nonenzymatic compaction of polynucleosomes (Francis et al., 2004).

In mammals PRC2 is composed of the core subunits EZH2, EED, SUZ12 and RbAp48 (Margueron and Reinberg, 2011b). EZH2 is responsible for the catalytic activity of PRC2 and catalyzes the di/trimethylation of H3K27, but the other two core components SUZ12 and EED were also shown to be necessary for the repressive role of PRC2 (Cao and Zhang, 2004; Pasini et al., 2007). PRC2 components are usually recruited near promoters of repressed genes (Cao and Zhang, 2004; Kimura et al., 2004; Kirmizis et al., 2004); in addition to their important role in the inactivation of the X chromosome during development (Plath et al., 2003; Umlauf et al., 2004). The recruitment of PRC2 subunits to the genome is still a matter of debate. It might be achieved in part similarly to TrxG subunits, by the recognition of CpG rich islands (Deaton and Bird, 2011; Ku et al., 2008).

PRC1 complexes are usually formed around a RING1/2 subunit with the binding six PcG RING finger proteins (PCGF). Such heterodimers form a core unit that is responsible for the ubiquitination of the lysine 119 of histone 2A (H2AK119ub) (Schwartz and Pirrotta, 2013). The canonical PRC1 subunits have chromobox-containing (CBX) subunits that are capable of recognizing the H3K27me3. In addition, several non-canonical PRC1 variants have different roles like a stronger ubiquitination activity or capacity to demethylate the transcription elongation histone marks H3K36me3, necessary to repress transcription (Farcas et al., 2012; He et al., 2013).

The mechanism by which the two PcG groups PRC1 and PRC2 cooperate in order to repress transcription has been debated in recent studies. Initially it was speculated that PRC2 is recruited to CpG islands of repressed genes and later on PRC1 recognizes the H3K27me3 deposited by PRC2, adds the H2AK119ub thus impedes RNA Polymerase activity and further compacts the chromatin (Stock et al., 2007; Zhou et al., 2008). Contrary to this hypothesis, recent studies inverse the recruitment steps of PRC1 and PRC2 (Blackledge et al., 2014; Cooper et al., 2014; Schwartz and Pirrotta, 2013). They argue that the genome peaks of

H3K27me3 form a broad signal while H2AK119ub give sharp well-defined peaks. In these studies they rather speculate that the first step is the recruitment of the PRC1 complex with its non-canonical components that recognize CpG islands with defined H3K27me3 (through the CBX subunits), this step recruits the rest of the PRC1 complex (canonical subunits) that add the H2AK119ub mark. Later on, PRC2 recognizes this mark through one of its two non-core subunits JARID2 or AEBP2, this causes a broad deposition of the H3K27me3 mark nearby repressed promoter regions (Kalb et al., 2014). The authors further consider the presence of a self-reinforcing loop where the ubiquitination stimulates more PRC1 and PRC2 binding, where PRC2 methyltransferase activity recruits again more the PRC1 complex. EZH1 of the PRC2 complex is responsible for the deposition of the H3K27me3 repressive mark on bivalent domains (Shen et al., 2008).

The notion of bivalent promoters was first described by Berstein et al. The authors describe the presence of a particular histone modification pattern in ES cells (Bernstein et al., 2006) on the promoter region, with both H3K4me3 activating mark and H3K27me3 repressive mark. They show that these bivalent domains are specific for ES cells and are present at promoter regions of developmental genes that are poised (kept in an inactive state, ready to be activated at specific cues) for expression, while differentiated cells present active or repressive histone modifications depending on the cell type (Figure 19). The presence of bivalent domains was also found in other cell types such as multi-potent neural, hematopoietic stem cells and embryonic fibroblasts (Barski et al., 2007; Cui et al., 2009; Mikkelsen et al., 2007). Furthermore, they demonstrate that genes with bivalent promoters have low transcription levels. Interestingly bivalent domains co-localize with the core TFs, where these factors aid in the maintenance of the poised state.

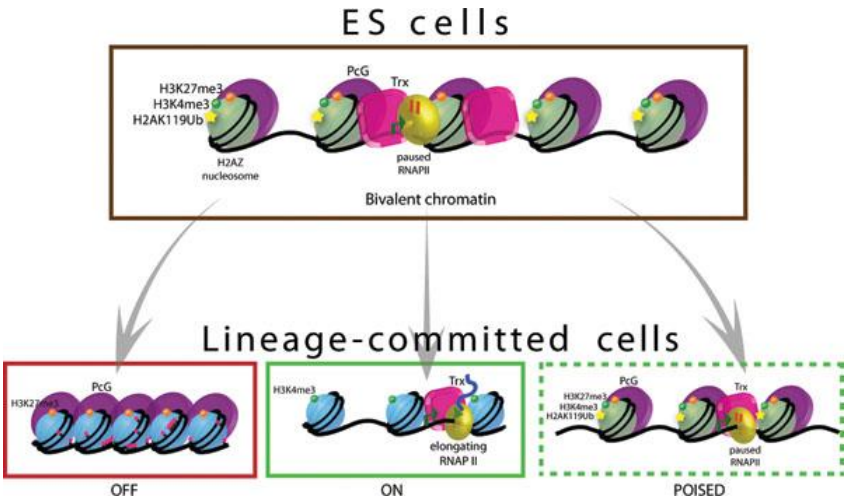


Figure 19 Bivalent chromatin domains mark the promoters of developmentally important genes in pluripotent ES cells. PRC2 and TrxG proteins catalyze the tri-methylation of histone H3 on lysine 27 and 4, respectively. PRC1 is also recruited to many of these genes t to add the H2AK119ub mark (Boyer, 2009)

c. CpG content and promoter state in ES cells

As discussed earlier, the deposition of both the H3K4me3 activating and the H3K27me3 repressive marks is assured by the PcG/TrxG duo. The recognition of the promoter regions by subunits of both complexes is likely done through the recognition of CpG regions.

An interesting study highlights the link between CpG composition in promoters and the nature of genes in both ES cells and lineage-commited cells (Mikkelsen et al., 2007). The authors describe two types of promoters, high CpG promoters (HCP) and low CpG promoters (LCP). They show that CpG high promoters are usually linked to housekeeping and developmental gene, while low CpG promoters are associated to more tissue-specific specialized genes. HCP in ES cells are mostly marked by H3K4me3, where this trimethylation level is directly proportional to gene expression level. However; not all genes with H3K4me3 are expressed, a certain percentage present the repressing H3K27me3 marks (bivalent genes), that keep one or both marks during differentiation. Moreover, the authors identify that H3K4me3 monovalent promoters are most of the time linked to housekeeping genes that are always expressed. Bivalent genes associate with development genes, TFs and other factors that require a more

complex transcriptional control. On the other hand, LCP show little H3K4me3 and almost no H3K27me3. These promoters are usually inactive by default, while the small percentage of genes associated with H3K4me3 usually belong to very specialized tissue-specific genes and could be activated in a selective manner according to the cell type.

2. Distal regulatory sequences

In addition to the promoters, which are the proximal regulatory elements of each gene, the genome includes a larger number of distal regulatory elements (enhancers and insulators) that play an essential role in the control of transcription.

a. Enhancers and super-enhancers

Enhancers are identified as cis-regulatory elements that control transcription in a positive manner (Blackwood and Kadonaga, 1998; Pennacchio et al., 2013). Enhancers are capable to exert this transcriptional control on target genes independent of genomic distance or orientation (Visel et al., 2009). In addition, they do not necessarily control the most proximal genes; in fact, they can even control more than one gene (Mohrs et al., 2001). Adding to the complexity of enhancer identification, is that they can be found anywhere in the genome (intergenic region or even within introns).

Enhancer identification was primarily conducted by individual characterization of disease-related distal regulatory elements (Bulger and Groudine, 1999) and on a larger scale by conservational comparative genomics (Boffelli et al., 2004). Later on, genome-wide studies have permitted the identification of thousands of putative enhancers across different cell-types. As in the case of promoters (active or bivalent), enhancer identification is also based on histone modification signatures. The initial histone modification used in genome-wide studies to identify enhancers is the monomethylation of lysine 4 of Histone H3 (H3K4me1) where the enhancer patterns obtained were cell-type specific (Ghisletti et al., 2010; Heintzman et al.,

2009; Heinz et al., 2010; Kim et al., 2010). Other studies also noticed an enrichment of the histone acetyltransferase p300 on enhancer regions (Visel et al., 2009).

The usage of the H3K4me1 mark alone is not sufficient to identify active enhancers, as a number of identified putative enhancers according to H3K4me1 did not show any activity. This phenomena lead to the identification of the different enhancer states, with active or poised enhancers dependent on the cell-type. Active enhancers were shown to carry in addition to the H3K4me1 mark, the H3K27 acetylation (H3K27ac) (Creyghton et al., 2010). Creyghton el al. demonstrated the presence of about 135000 putative enhancers in ES cells and a number of differentiated cells based on the double histone marks. Further on, they demonstrate the variation in the enhancer state upon differentiation of ES cells, where some poised/inactive enhancers become active and vice-versa in order to sustain the new cell state. In ES cells enhancer elements were also shown to be bound by the core TFs Oct4/Sox2/Nanog. Interestingly, these factors bind both active and poised, conferring a role of the core regulatory network in the both the maintenance of the pluripotency/self-renewal enhancers and the repression of developmental enhancers.

Later studies have identified the presence of multiple classes of enhancers. In one study, the ES cell/ differentiation system was explored in order to better understand the different enhancer states throughout development (Zentner et al., 2011). Using CHIP-seq and DNase hypersensitivity analysis the authors identified; in addition, to the active (H3K4me1/H3K4ac) and poised (only H3K4me1) enhancers, the H3K27me3 mark enriched on certain enhancers. So the authors reclassify enhancers into three major classes: Active (H3K4me1/H3K27ac), intermediate (H3K4me1) and poised (H3K4me1/H3K27me3 or H3K9me3). Active enhancers were further sub-classified according to their transcriptional level (using the H3K36me3 transcriptional elongation mark), with the most active enhancers with the highest eRNA transcripts. Upon the differentiation of mouse ES cells, the enhancer classes resolved to

different enhancer states depending on the developmental state (Zentner et al., 2011). Active enhancers retained their active state, became neutral (no mark) or lost H3K4me1 and acquired H3K27me3. Poised enhancers might acquire H3K27ac to become active, lose all marks to become neutral or lose H3K4me3 and retain H3K27me3. On the other hand, intermediate enhancers showed transition between the neutral and poised state but failed to lose their H3K4me1 or acquire H3K27me3.

In a more recent study, a special type of enhancer clusters was identified. Those enhancer clusters were called super-enhancers (Whyte et al., 2013). In ES cells these super enhancer clusters were shown to be occupied with five essential TFs (Oct4, Sox2, Nanog, Klf4 and Essrb) in addition to the mediator (facilitator of enhancer/promoter interactions). However, the occupancy of the mediator was much higher in such super enhancers (SEs) than in classical or typical enhancer (TEs). Whyte et al. further demonstrate that in addition to the high Mediator occupancy, SEs span much larger DNA regions (as much as 50kb) than TEs (Figure 20A). Furthermore, SEs presented higher levels of H3K27ac/H3K4me1 and higher occupancy by Klf4 and Essrb (not the case for Oct4/Sox2/Nanog) (Figure 20B). In ES cells SE were shown to be associated with important ES cell key genes, including the core transcription factor genes, DNA-modifying enzymes, miRNA where they play a role in driving a high-level expression. On the other hand, SEs were found to be associated with very few housekeeping genes.

Interestingly, SEs were not only detected in ES cells, different cell types also present SEs that drive the expression of cell identity genes (Hnisz et al., 2013; Whyte et al., 2013). In addition, SEs variations were detected in different diseases. For instance, cancerous cells can generate SEs at oncogenes and other genes for tumor sustaining (Hnisz et al., 2013).

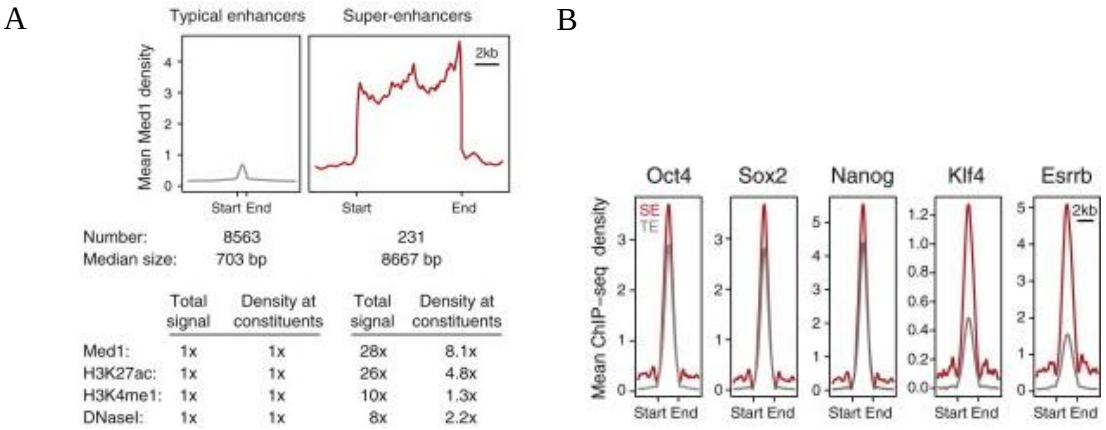


Figure 20 Characteristics of typical enhancers and super enhancers in ES cells. A) Mediator ChIP-seq density and ChIP-seq fold difference for enhancer features Mediator, H3K27ac, H3K4me1, and DNaseI hypersensitivity at super-enhancers versus typical enhancers. B) Oct4, Sox2, Nanog, Klf4, and Esrrb ChIP-seq density across the consti constituent enhancers within the 8,563 typical enhancers and the 231 super-enhancer regions. (Whyte et al., 2013)

b. CTCF-binding sites and insulators

The genome in eukaryotes is organized in a three-dimensional way, with packaging at different levels, starting from nucleosomes ending with the chromosomes. Gene expression is highly dependent on this genome architecture that provides spatial-temporal control of transcription. CTCF or CCCTC-binding factor was shown to be one of the architectural proteins responsible for the DNA looping of two sequences resulting in various functional outcomes depending on the nature of these sequences and their bound proteins. CTCF contains a highly conserved DNA-binding domain with 11 zinc fingers (Ohlsson et al., 2001); it binds to a high number of sites in the mammalian genome (~55000-65000 sites) (Chen et al., 2012) and normally targets linker regions with well positioned nucleosomes (Cuddapah et al., 2009). CTCF is the best characterized insulator (capacity to interfere with the interaction between regulatory sequences/buffering transgenes from the heterochromatin effect) protein in vertebrates. CTCF was first identified as a transcription factor that activates or represses gene expression in heterologous reporter assays (Baniahmad et al., 1990; Lobanenkov et al., 1990).

The sole role of CTCF as an insulator was debated in later studies. A recent review even replaces the name of insulator by rather architectural protein (Ong and Corces, 2014). CTCF-binding sites were identified in various genomic regions (intergenic, promoter-proximal and intragenic) (Chen et al., 2012, 2008b). CTCF was suggested to play a role in tethering distant enhancers to their promoters, where it was found to bind active enhancers; indeed promoters and enhancer elements were observed to be enriched with CTCF binding sites (Sanyal et al., 2012; Shen et al., 2012). In mouse ES cells, a study further shows the importance of CTCF in promoter-enhancer interactions (Liu et al., 2011), where it was shown to bind to an endodermal differentiation factor (TAF3) along with cohesion. Indeed, in mouse ES cells genome-wide analysis using ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing), revealed CTCF-mediated promoter/enhancer interaction for the establishment of 'transcription factories' at certain genes (Handoko et al., 2011).

The higher eukaryotes genome is organized into topologically associating domains (TADs). These TADs are defined by the high frequency of DNA interaction within such domains (intra-domain interactions); however inter-TAD interactions are very low. This topological domain phenomenon may account for the two different roles assured by CTCF as an insulator or an activator. CTCF-binding site were found at TAD boundaries (15% of all CTCF-binding sites), whereas the rest 85% were shown to be located within TAD domains (Dixon et al., 2012). The presence of CTCF at TAD boundaries may confer to its enhancer insulator properties between different TADs; on the other hand, intra-TAD interactions of promoter/enhancer elements are facilitated by CTCF proteins found with each TAD domain (Figure 21). Moreover, CTCF was shown to create insulated regions of promoter/super-enhance interactions in ES cells, where with cohesion, CTCF is able to create an insulated SE domain (Figure 22) (Dowen et al., 2014). Dowen et al. also demonstrate the role of CTCF in insulating Polycomb domains for the correct silencing of developmental genes in ES cells

(Figure 22). This insulation activity is essential in order to inhibit any inappropriate gene activation/repression in nearby domains.

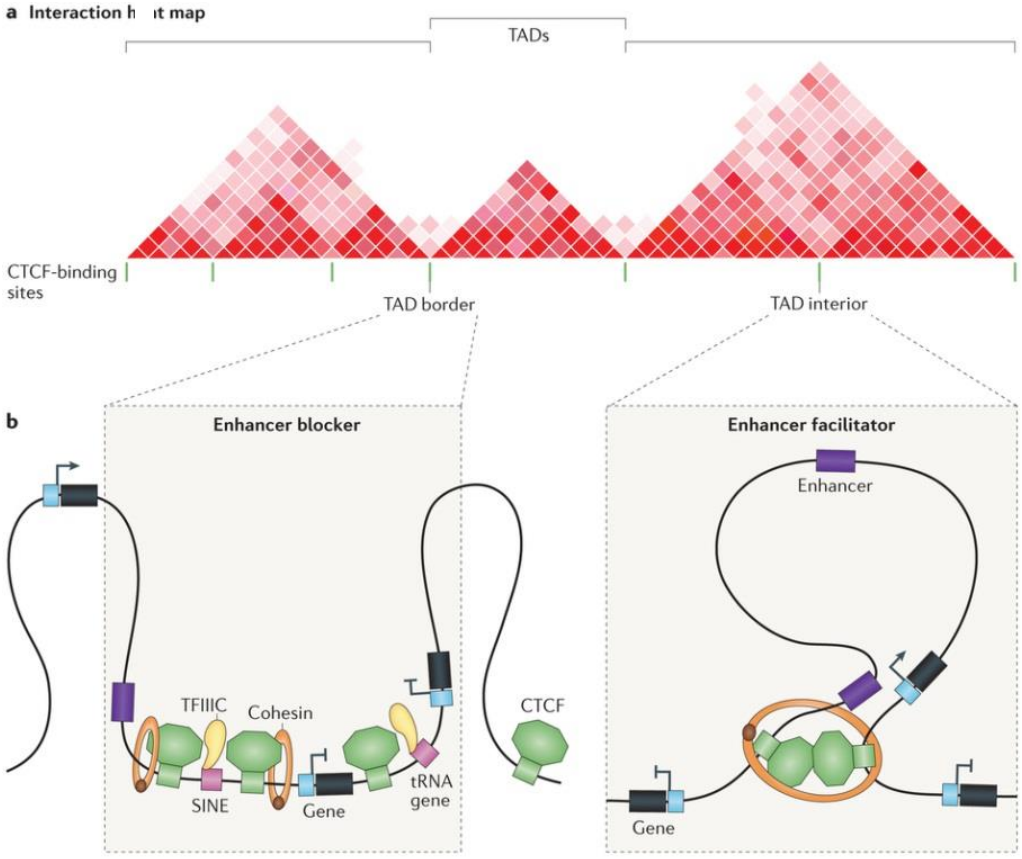


Figure 21 The role of CTCF as an enhancer blocker or facilitator (Ong and Corces, 2014)

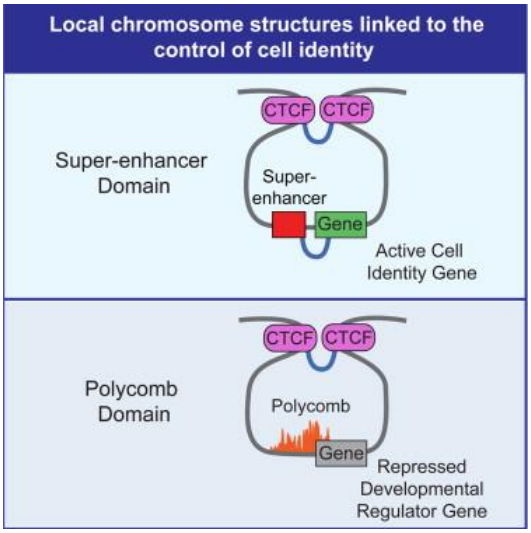


Figure 22 The role of CTCF in the creation of both super enhancer domains (SD) and polycomb insulated domains (PD) in ES cells (Dowen et al., 2014)

Chapter III. ATP-dependent Chromatin Remodeling Factors and Pluripotency

ATP-dependent chromatin remodeling factors or remodelers belong to the class of chromatin modifying enzymes. By using the energy of ATP, they can alter the chromatin structure by acting on the chromatin building block, the nucleosomes. Remodelers were shown to have important roles in different cellular mechanisms such as DNA repair, replication and more importantly in the transcriptional control of genes.

A. The Snf2 family of ATP-dependent chromatin remodeling factors: Classification

The classification and nomenclature of ATP-dependent chromatin factors have been debated in several studies. ATP-dependent chromatin remodeling enzymes are usually referred to as Snf2 or SWI/SNF related enzymes; this nomenclature originated from the first discoveries on ATP-dependent chromatin remodeling mechanisms of the SWI/SNF (Switch/sucrose non-fermenting) complex observed in the yeast *Saccharomyces cerevisiae* (Côté et al., 1994).

The Snf2 family of ATP-dependent chromatin remodeling factors contains a domain homologous to the yeast Snf2 domain (Laurent et al., 1992). This Snf2 domain has seven conserved helicase-related sequence motifs that allows the further classification of the Snf2 family as a part of the SF2 superfamily encompassing helicase-like proteins without a helicase detected activity and with ATP-dependent remodeling functions (detailed later) (Eisen et al., 1995; Flaus et al., 2006). Flaus et al. propose an interesting sub-classification of the Snf2 family into 24 subfamilies, based on multiple sequence alignment of the helicase-like regions. These 24 subfamilies can be further regrouped into six groups (Rad5/6-like, Swr1-like, Snf2-like, Rad 54-like, SSO1653-like and the distant SMARCAL1) based on specific functional

domain similarities (Figure 23). To avoid any confusion it is worth mentioning that the Snf2 family encompasses a group named Snf2-like composed of 7 subfamilies including the Snf2 subfamily, which encompasses the yeast Snf2 (*Drosophila melanogaster* Brahma and mammalian Brg1). The different subfamilies, their archetypal organism, the different nomenclature and members are represented in Table2.

An alternative classification exists for the ATP-dependent chromatin remodeling factors. This classification simplifies the observed diversity of Snf2 remodelers and partitions them into 4 families which are SWI/SNF, ISW1, CHD and INO80 (Clapier and Cairns, 2009b; Hargreaves and Crabtree, 2011). However, such classification does not take into consideration the diversity within the CHD family and fails to correctly categorize remodelers such as Alc1 and Atrx. To conclude, the classification of ATP-dependent remodelers is not an easy task, taking into account the diversity of the flanking domains and the absence of robust domain-finding tools necessary to assign many sequences.

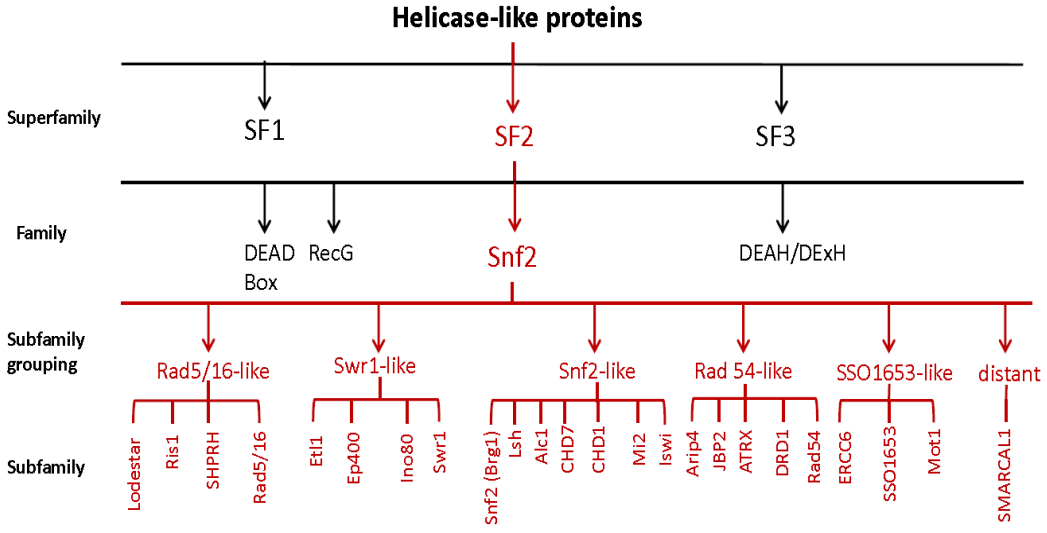


Figure 23 Schematic diagram of the hierarchical classification of ATP-dependent helicase-like proteins. Adapted from (Flaus et al., 2006).

Subfamily	Archetype gene	Members and their variant nomenclature
Snf2	S. Cerevisiae (Snf2)	Yeast Snf2/ Drosophila Brahma/mammalian **BRG1**(SMARCA4)
Iswi	D. melanogaster Iswi	SMARCA1 (Snf2l), **SMARCA5** (Snf2h)
Lsh	M. musculus Hells	HELLS (SMARCA6,/Lsh)
ALC1	H. Sapiens CHD1L	ALC1**(CHD1L**/SNF2P)
Chd1	M. musculus CHD1	CHD1, CHD2
Mi-2	H. Sapiens CHD3	CHD3 (Mi2α/ZFH), **CHD4** (Mi2β), **CHD5** (Kiaa0444)
CHD7	H. Sapiens CHD7	CHD6, CHD7, CHD8, CHD9
Swr1	S. Cerevisiae (SWR1)	SRCAP (domino/PIE1)
Ep400	H. sapiens EP400	**EP400** (Domino)
Ino80	S. Cerevisiae (INO80)	INO80
Etl1	M.. musculus (Smarcad1)	SMARCAD1
Rad54	S. Cerevisiae (Rad54)	RAD54
Atrx	H. sapiens (Atrx)	Atrx
Arip4	M.. musculus (Srisnf21)	ARIP4
DRD1	Arabidopsis thaliana (DRD1)	DRD1
JBP2	T. brucei (JBP2)	JBP2
Rad5/16	S. Cerevisiae (RAD5, RAD16)	**SMARCA3** (HLTF)
Ris1	S. Cerevisiae (RIS1)	RIS1
Lodstar	D. melanogaster (Lodestar)	Lodestar (HuF2)
SHPRH	H. sapiens (SHPRH)	SHRPH (YLR247C)
Mot1	S. Cerevisiae (MOT1)	MOT1 (BTAF1)
ERCC6	H. sapiens (ERCC6)	ERCC6 (RAD26L)
SSO1653	S. solfataricus (SSO1653)	SSO1653 (SsoRad54like)
SMARCAL1	H. sapiens (SMARCAL1)	SMARCAL1 (DAAD)

Table 2 The different Snf2 subfamilies, the archetype organism and the different constituting members. Subfamilies and members marked in bold letters are of particular interest in the lab and thesis project. Adapted from (Flaus et al., 2006).

B. The Snf2 family of ATP-dependent chromatin remodeling factors: Mode of action

1. The chromatin state

Chromatin is identified as a combination of DNA, RNA and proteins. This macromolecular structure is essential for DNA compaction and the correct cellular state maintenance (replication, repair and transcription).

In eukaryotes, the 1 meter DNA fiber is subjected to different levels of packaging in order to fit in the nucleus. The basic building block of chromatin is the nucleosome (Kornberg, 1974).

The nucleosome consists of 146bp of DNA rapped around the histone octamer that comprises a homodimer of each of H2A, H2B, H3 and H4 histones. This packaging level represents the well-known “beads on a string” chromatin structure that is further compacted to form the 30nm fiber thought to be stabilized by H1 linker histone (Felsenfeld and Groudine, 2003a). Nevertheless, this compaction level represents just a 50 fold of a 5000 fold required level to attain the chromosome state (Figure 24); this sheds the light on various other yet undiscovered mechanisms interfering in chromatin compaction (Hargreaves and Crabtree, 2011).

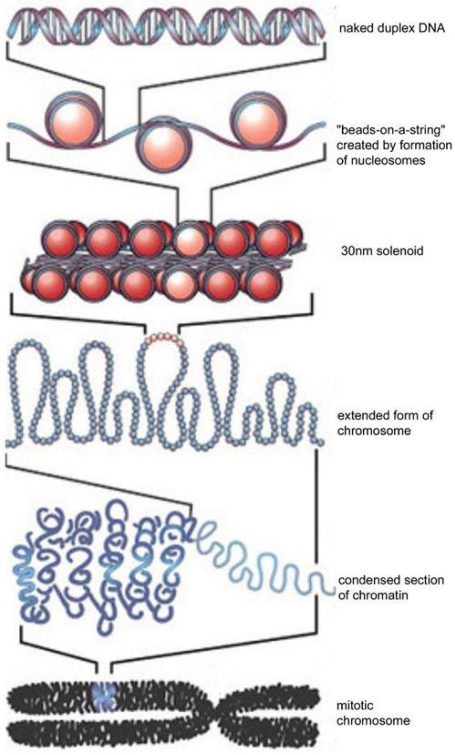


Figure 24 The different compaction levels of the chromatin (Felsenfeld and Groudine, 2003b)

The extensive DNA compaction, on the other hand, should not hinder the accessibility to insure the appropriate cellular processes such as gene transcription. As mentioned previously two types of enzymes assure the proper DNA accessibility, histone modifying enzymes and chromatin remodeling factors. Cells require remodeling factors in order to package the genome and at the same time assure a regulated DNA accessibility (Clapier and Cairns, 2009b).

2. Mode of action of ATP-dependent chromatin remodeling enzymes

ATP-dependent chromatin remodeling enzymes or factors (remodelers), and as their name implies, are capable to remodel the chromatin by using the energy of ATP. Nucleosome or chromatin remodeling factors act at different levels of nucleosome organization by a sliding mechanism (Figure 25). First, some remodelers can modify nucleosome spacing; second, they assure the phasing of the nucleosomal arrays with respect to a barrier, which can be DNA-bound protein, or the so called nucleosome free regions (NFR), which are typically found at eukaryotic promoter elements. Last, the nucleosome sliding activity of remodelers assures a proper regulation of the DNA sequence accessibility (reviewed in (Mueller-Planitz et al., 2013)).

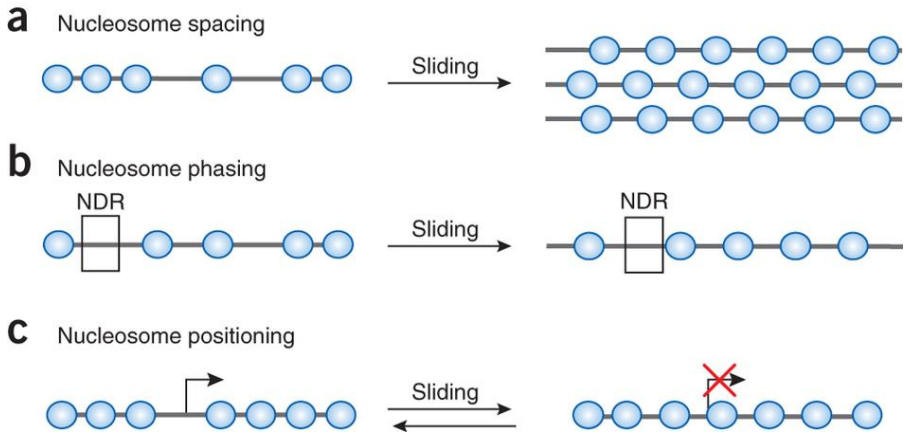


Figure 25 The outcomes of nucleosome sliding on nucleosome organization (Mueller-Planitz et al., 2013)

Several mechanisms were proposed in order to explain how remodelers interfere in the DNA accessibility. By using the energy of ATP, remodelers can expose a DNA binding domain by several mechanisms. Remodelers can expose the DNA regulatory domain either by repositioning or ejecting a nucleosome, or by causing the unwrapping of a DNA filament. Another proposed mechanism is by altering the histone composition of nucleosomes, where the original histone dimers can be replaced by histone variants, or eventually a dimer ejection can take place in order to assure proper DNA accessibility (Clapier and Cairns, 2009b). Figure 26 illustrates the different outcomes of chromatin remodeling.

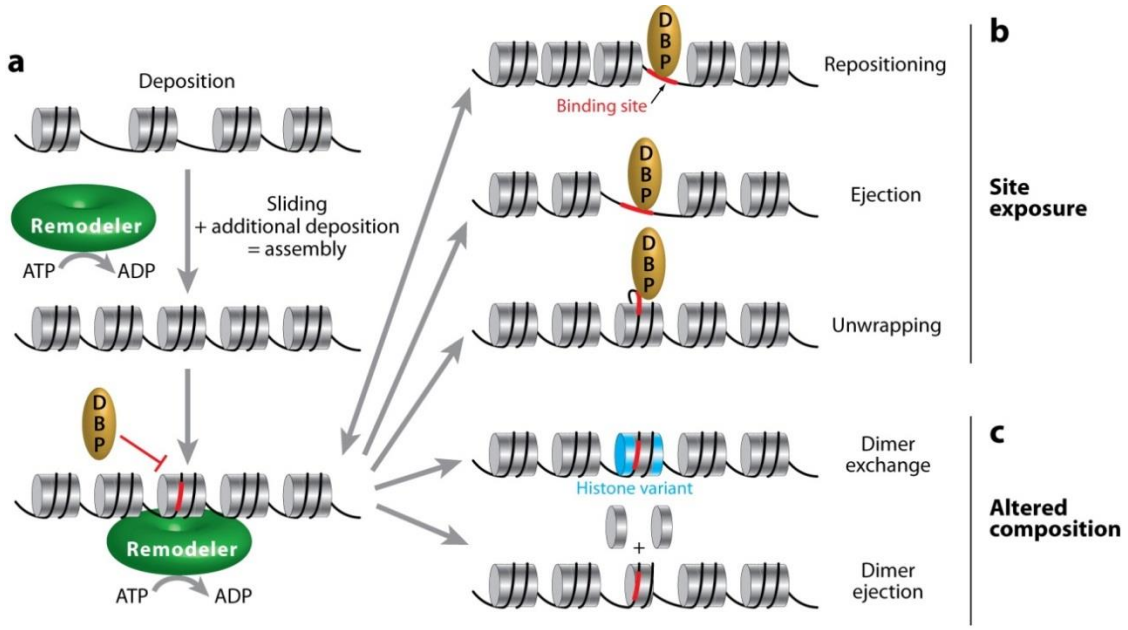


Figure 26 Mode of action of ATP-dependent chromatin remodeling factors (Clapier and Cairns, 2009)

In addition to the central ATPase catalytic domain of remodelers, several other accessory domains are present. Such domains play various roles in covalent histone modification recognition, regulation of the action of the central ATPase domain, and in the interaction with other chromatin remodelers and TFs. So what are the different mechanistic features of ATP-dependent chromatin remodeling factors?

The ATPase motor of Snf2 ATP-dependent chromatin remodeling factors presents a DNA 3'>5' translocase activity that helps moving on the DNA helix and drawing DNA from one entry/exit site and pump it in a directional wave (Lavelle et al., 2011). Moreover, the remodeler binding to nucleosomes eventually causes DNA/Histone distortion and loosening that might allow for the later DNA translocase activity for required for the DNA bulge creation (Lorch et al., 2010).

On the other hand the accessory domains (at the C and N termini) of each remodeler contain regulatory regions. For instance, the N-terminal domain in the CHD protein CHD1 changes its conformation and facilitates the binding of the ATPase catalytic region to promote correct remodeling activity (McKnight et al., 2011). The action of such accessory domains is to

provide the appropriate specificity for nucleosomes and to inhibit ATP hydrolysis in the absence of an adequate substrate (Mueller-Planitz et al., 2013). Moreover, some Snf2 family remodelers present domains such as SANT and SLIDE DNA binding domains (Ryan and Owen-Hughes, 2011) that recognize the nucleosomal entry site and its adjacent linker. Such domains improve ATPase catalysis activity by the remodeling factor through anchoring it on DNA and increasing the chances for the exertion of a translocase activity (Dang and Bartholomew, 2007; Yamada et al., 2011).

To conclude, the sequence diversity of Snf2 family ATP-dependent chromatin factors suggests different and specific functions exerted by the ATPase catalytic domain (Flaus and Owen-Hughes, 2011).

C. The functional output of ATP-dependent chromatin remodeling factors on nucleosome positioning

1. Nucleosome positioning in the genome

As previously discussed, nucleosome chemical histone modifications and histone composition play an essential role in gene transcription regulation. A third essential aspect adds to the complexity of the genome access control, which is the nucleosome positioning across the genome. So how are nucleosomes positioned across the genome?

It has been long debated on whether the deposition of nucleosomes across the genome occurs in a random or rather a defined manner. Genome-wide mapping of nucleosome positioning has deepened our understanding of nucleosome deposition across the genome. One of the first organisms where a CHIP-seq based high-resolution nucleosome mapping was conducted was the yeast *Saccharomyces cerevisiae* (Albert et al., 2007). In later studies, the nucleosome positioning of other organisms such as *Drosophila melanogaster* and humans have been published (Barski et al., 2007; Mavrich et al., 2008).

The exploration of such nucleosome positioning maps revealed a Gaussian distribution of nucleosomes around specific gene loci (reviewed in (Jiang and Pugh, 2009)). In such distributions, nucleosomes show to adapt preferential positions, a phenomenon referred to as nucleosome phasing. These positioned nucleosomes show equal spacing from each other with short linker parts between them. The linker length can vary; long stretches of linkers with no nucleosomes are referred to as nucleosome free regions (NFRs). It was demonstrated that NFRs presence links the involvement of nucleosome positioning and organization in gene regulation (Jiang and Pugh, 2009).

The nucleosome organization on genes in yeast showed the presence of phased nucleosomes around the TSS (transcription start site), with more random positioning further away (reviewed in (Jiang and Pugh, 2009)). The -1 nucleosome is located upstream the TSS and plays a role in the promoter accessibility. A 5' NFR is located just downstream the -1 nucleosome that is followed directly by a TSS. The +1 nucleosome is located after the TSS and showed to have the tightest position (highest phasing). The +1 nucleosome usually contains histone variants and histone tail modifications necessary for rapid eviction during transcription initiation. The more downstream nucleosomes start showing a less phased organization that continues covering the gene body until reaching a 3' NFR where transcription termination takes place (Figure 27).

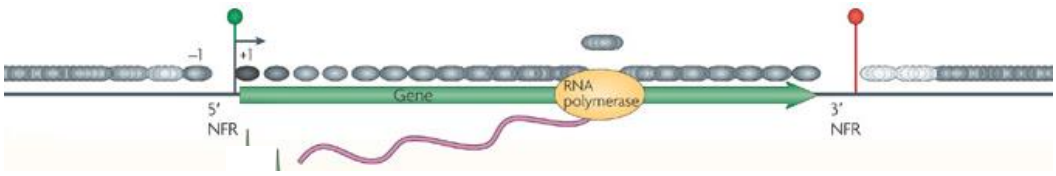


Figure 27 Nucleosome organization on yeast genes

Moreover, in mammals nucleosome distribution was shown to change during differentiation (Teif et al., 2012). Teif et al. show that distinct nucleosome positioning profiles at important functional regulatory sequences are observed between different cell types (ES cells, NP cells

and Mouse fibroblasts). Moreover, the authors observed an increase in the NRL (Nucleosome Repeat Length) during differentiation, confirming the necessity of chromatin structure variations to assure the correct cell state. This observation also is concordant with the different studies that demonstrate that ES cells have a less compact chromatin compared to their differentiated counterparts.

Several studies tried to explain the origins of nucleosome positioning (Gupta et al., 2008; Ioshikhes et al., 2006; Rando and Ahmad, 2007). It was proposed that nucleosome positioning across the genome could be a combination of a dual mechanism of independent and statistical positioning where nucleosomes are in a dynamic state (Jiang and Pugh, 2009). In this model, the presence of favorable sequences and unfavorable linker regions (NFRs) for nucleosome deposition creates this thermodynamic state responsible for the correct balance between transcription permissiveness and DNA inaccessibility. So how is the access to DNA regulated?

Several studies proposed a non-catalytic site exposure of DNA on the nucleosome surface due to the presence of thermal DNA fluctuations (Polach and Widom, 1995, 1996). However, the access to deeper sequences within the nucleosome core requires the interference of chromatin remodeling enzymes.

2. ATP-dependent nucleosome remodeling enzymes and nucleosome positioning

The dynamic organization of the nucleosomes needs the presence of ATP-dependent chromatin remodeling enzymes. A number of studies have highlighted the importance of such enzymes in the nucleosomes positioning in both the goal of facilitating DNA accessibility and hindering the regulatory sequences inaccessible (Morris et al., 2014; Ramirez-Carrozzi et al., 2009; Yen et al., 2012).

Several studies proposed a non-catalytic site exposure of DNA on the nucleosome surface due to the presence of thermal DNA fluctuations (Polach and Widom, 1995, 1996). However, the access to deeper sequences within the nucleosome core requires the interference of chromatin remodeling enzymes.

It was shown that inside genes, ATP-dependent remodelers pack nucleosomes into arrays with a specific distance from the TSS, leaving a 3' NFR upstream and a 5'NFR downstream (Morris et al., 2014). MNase-CHIP experiments on a selection of Snf2 ATP-dependent chromatin remodeling factors in yeast showed the preferential binding of certain remodelers to specific nucleosome positions (Yen et al., 2012). According to Yen et al., ATP-dependent remodelers can be separated into three classes according to the nucleosomes they interact with: Remodelers that mostly bind to the +1 nucleosome, remodelers that bind to nucleosomes in coding regions and are depleted on the +1 nucleosome and remodelers that bind NFR flanking nucleosomes. The authors further suggest that remodelers show an activation or inhibition of transcription characteristics depending on the directionality of nucleosome positioning according to the NFR. Remodelers that move nucleosomes towards NFR are generally considered repressive and remodelers that on the contrary move nucleosomes away from the NFR show transcription activation properties.

A complementary study focused on the genome-wide positions of three essential Snf2 chromatin remodelers (Brg1, Chd4 and Snf2h) (Morris et al., 2014). Similar distribution patterns on regulatory regions were observed for the three remodelers with specific slight remodeler-linked preferences, an observation already reported where remodelers showed similar gene regulatory profiles (Yen et al., 2012). The authors propose that this remodeler co-localization might be due to a sequential binding events rather than direct interaction. Loss-of-function of each remodeler followed by DHS profiling (DNase Hypersensitive Sites) showed a change in the DHS profile of the genome, with both lost and newly acquired DHS upon the

depletion of each of the three remodelers (Morris et al., 2014). This observation confirms the double role of remodelers in both opening and closing the chromatin. So such a co-localization might reflect a cooperative mechanism where remodelers act together in order to position nucleosomes. Interestingly, recent studies have also highlighted the antagonistic remodeling events at specific sites and during specific processes especially for the two remodelers Brg1 (Snf2) and Chd4 (Curtis and Griffin, 2012; Gao et al., 2009; Ramirez-Carrozzi et al., 2006).

To conclude, it remains essential to understand the way remodelers are recruited to their specific sites. Such recruitment might be assured due to the presence of specific motifs for transcription factor clustering that attracts remodelers to regulatory element sites or it might be due to the presence of certain histone modifications (remodelers were shown to possess bromo/chromo domains that could recognize such modifications) (Hassan et al., 2002; Sims et al., 2005; Won et al., 2009). Another study links CpG rich promoters to remodeler independency, where such promoters are readily unstable and do not require remodeler interference (Ramirez-Carrozzi et al., 2009); on the other hand, their non-CpG counterparts require remodeling activity for transcription (where a more regulated gene expression is required).

D. Chromatin Remodeling Complexes Function in ES cells

ATP-dependent chromatin remodeling factors were shown to play important roles in the regulation of the particular chromatin landscape in ES cells. Most remodelers are a part of large complexes with variable subunit compositions. So what are the to-date remodeling complexes shown to have a role in ES cell state control?

1. The NuRD complex (encompassing Chd4)

The NuRD complex (Nucleosome Remodeling and Deacetylase complex) is mainly composed of the core Mbd3 (Methyl-CpG binding domain protein 3) responsible for complex assembly (Kaji et al., 2006; Zhang et al., 1999), histone deacetylase subunits (HDAC1/2) and ATP-dependent chromatin remodeling subunits encompassing the Snf2 family remodelers Chd3 or Chd4 (also called Mi-2α and Mi-2β, respectively) (Lai and Wade, 2011; McDonel et al., 2009) responsible for the enzymatic activity of the complex.

NuRD is characterized by its repressive functions through its capacity to deacetylate H3K27ac, where it was shown to trigger the recruitment of the PRC2 repressive complexes and the eventual deposition of the H3K27me3 repressive mark (Reynolds et al., 2012a). In ES cells, the Mbd3 subunit of NuRD was shown to be essential for differentiation rather than the maintenance of the self-renewal state (Kaji et al., 2006). Through the analysis of Mbd3 loss-of-function, the NuRD complex was shown to suppress the expression level of pluripotency genes to a certain level in order to permit the exit from the self-renewal state for lineage commitment (Reynolds et al., 2012b) (Figure 28). On the other hand, a more recent study (also based on Mbd3) shows that the NuRD complex can play a positive facilitating role during the reprogramming to iPS cells of certain cell types (Santos et al., 2014). Interestingly, components of the NuRD complex were found to be a part of the Oct4 interactome, suggesting a role in the transcriptional regulation of the ES cell genome (van den Berg et al., 2010a; Pardo et al., 2010). More particularly, Chd4 (Mi-2β) was detected at promoters and gene bodies of pluripotency associated genes in ES cells (Reynolds et al., 2012b) .In addition to the central ATPase/Helicase unit, it consists of accessory regulatory domains that confer its

DNA/nucleosome binding activity (PHD fingers and chromodomains that recognize DNA and H3K4 methylation).

Furthermore, Chd4 was shown to associate with the PRC2 component Ezh2, where it plays a dual role during neural differentiation (Sparmann et al., 2013). At the early neurogenesis stage, Chd4 recruits Ezh2 on astrogenic genes to repress astrogenesis; on the other hand, at the end of neurogenesis Chd4 localizes Ezh2 on neurogenic genes to repress neurogenesis and start astrogenesis. This dual role of a chromatin remodeling factor such as Chd4 illustrates how the same remodeling factor can be involved in different mechanisms during development.

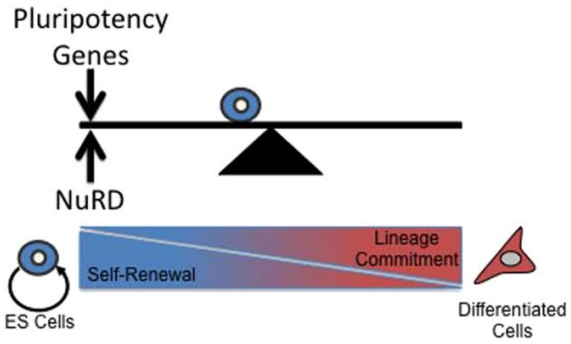


Figure 28 The role of NuRD in ES cells. NuRD balances the expression of pluripotency genes and differentiation genes in order to permit lineage commitment under the appropriate cues (Reynolds et al., 2012b)

2. INO80 complex

The INO80 complex is a large complex composed of different subunits including Ino80 and Snf2 remodeler responsible for the complex's catalytic activity (Shen et al., 2000; Tosi et al., 2013). The INO80 complex was shown to play an important role in the regulation of the distribution of the histone variant H2A.Z (Papamichos-Chronakis et al., 2011). This histone variant was shown to be particularly enriched at active and bivalent promoters in ES cells (Ku et al., 2012). Moreover, several RNAi screens identified Ino80 as a self-renewal regulator (Fazzio et al., 2008; Hu et al., 2009b).

A more recent study highlights the special importance of Ino80 in the control of the ES self-renewal (Wang et al., 2014). Wang et al. demonstrate that Ino80 exerts a direct regulation of the core pluripotency circuitry, even forming a kind of auto-regulatory loop with the master TFs. Furthermore, the authors show that Ino80 occupies pluripotency gene proximal regions where it interferes in the activation of such genes by preserving an open chromatin state facilitating the recruitment of Mediator and Pol II to such regions. On the other hand, Ino80 was not detected on differentiation genes, suggesting a major role in ES cell self-renewal gene transcriptional control but not in the repression of differentiation genes. Microarray transcriptomic analysis comparison in Ino80 depleted cells with ES cells depleted for any of the master TFs (Oct4/Sox2/Nanog) showed common deregulated gene profiles, further insisting on the role of Ino80 in the maintenance of the ES gene program within the core circuitry (Wang et al., 2014).

3. esBAF complex (encompassing Brg1)

BAF complexes have been show to play important roles in mammalian development (Chi et al., 2003; de la Serna et al., 2001; Lessard et al., 2007). This complex is composed of about 11 subunits encompassing the Snf2 family ATP-dependent chromatin remodeling family with complex cell-specific diversity (Olave et al., 2002; Wang et al., 1996). In ES cells, esBAF is the specialized complex shown to be important in ES cell state maintenance (Ho et al., 2009a). Ho et al. demonstrate that the knockdown of the catalytic subunit of esBAF (Brg1 or Smarca4) causes proliferative problems in ES cells represented by flattened colonies. However, the loss of the expression of the core pluripotency factors is not observed until about ten cell passages. Moreover, the authors show that Brg1 depleted cells were differentiation deficient, conferring the probable role of Brg1 in pluripotency as well.

Genome-wide analysis of Brg1 distribution using CHIP-seq revealed the colocalization of Brg1with the target genes of the core transcriptional regulatory network (Ho et al., 2009b). In addition, Ho et al. detect the colocalization of Brg1 with STAT3 and Smad1 gene targets. Moreover, the authors demonstrate that Brg1 can oppose the repressive effect of PcG. Interestingly, Brg1 acute depletion resulted in an increased expression in the core pluripotency TFs; in addition, the analysis of Brg1 gene targets showed that it works both in coordination and in opposition to the Oct4/Sox2/Nanog trio. Ho et al. speculate that Brg1 is required for the fine-tuning of the expression of the ES cell TFs in order to maintain a stable ES call state.

A third study conducted by Ho et al. further illustrates the role of Brg1 in pluripotency maintenance through the establishment an accessible chromatin at STAT3 targets and opposing PcG action (Ho et al., 2011b). Strikingly, the authors demonstrate that Brg1 also acts synergistically with PcG at certain HOX gene loci. This gives Brg1 a dual role, where it can act synergistically or antagonistically with repressive complexes in order to maintain the ES cell state. Figure 29 illustrates the different roles of esBAF in ES cells.

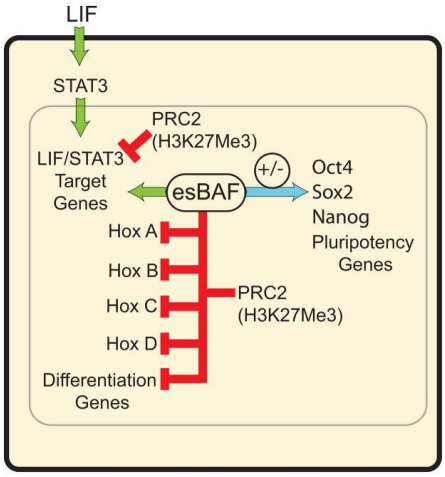


Figure 29 The different control mechanisms exerted by the esBAF complex in order to maintain ES pluripotency and self-renewal (Ho et al., 2011b)

The esBAF complex was also shown to facilitate the reprogramming to iPS cells by enhancing the binding of Oct4 to its target genes (Singhal et al., 2010) where an overexpression of esBAF subunits replaces the requirement for c-Myc.

4. Chd1

Chd1 is an Snf2 ATP-dependent chromatin remodeling factor belonging to the Snf-like group. Chd1 contains two chromodomains responsible for H3K4 methylation recognition (Sims et al., 2005) Chd1 is known to play roles in active transcription in the different eukaryotic organisms (Simic et al., 2003; Sims et al., 2007; Stokes et al., 1996). Core transcription factor distribution revealed their binding on Chd1 regulatory elements (Chen et al., 2008a).

Chd1 was demonstrated to be required for the maintenance of the open chromatin and the pluripotency of ES cells (Gaspar-Maia et al., 2009.). Gaspar-Maia et al. shown that the RNAi depletion of Chd1 affects the expansion of the ES cell, decreases clonogenic capacity and decrease in the Oct4 levels. Moreover, Chd1 is described to be a euchromatin protein (association with H3K4me3) that opposes the effect of heterochromatin formation and preserves the open chromatin state of ES cells. In addition, Chd1 was shown to increase the efficiency of reprogramming to iPS cells (Gaspar-Maia et al., 2009). However, Chd1 RNAi ES cells remain undifferentiated elucidating a rather complementary/redundant role for Chd1in the control of the ES cell state.

5. Ep400

The Ep400/Tip60 complex is composed of Ep400, an Snf2 ATP-dependent chromatin remodeling factor and Tip60 a histone acetyltransferase (HAT). An RNAi screen discussed previously in Chapter II revealed the importance of Ep400 in ES cells (Fazzio et al., 2008). Fazzio et al. showed that depletion of the Ep400/Tip60 complex strongly affected self-

renewal and pluripotency. In addition, the KD of Ep400 effected the expression of about 802 genes, where most of the upregulated genes are key developmental loci. Interestingly, Nanog knock down showed an overlap in the transcriptomic deregulation profile with Ep400 suggesting a mutual regulation of a set of genes. Moreover, a strong correlation between H3K4me3 and Ep400 binding was observed both at active and silent genes (presenting the H3K27me3: bivalent genes) with highest binding at the most active genes. In addition, Nanog was shown to be indirectly involved in the binding of Ep400 to its target genes suggesting that the Tip60/p400 complex is important effector of Nanog transcriptional repression.

6. Atrx

Atrx (alpha thalassemia/mental retardation syndrome X-linked) was shown to be localized on pericentric heterochromatin regions and nuclear bodies in mouse primary embryonic fibroblasts (Ishov et al., 2004). In addition, Atrx was found to bind telomeric regions in ES cells during interphase and was demonstrated to interact with H3.3 histone variant to insure its proper deposition at telomeres (Wong et al., 2010). However, a very recent study demonstrates that Atrx's role in H3.3 deposition is not exclusive to pericentric heterochromatin and telomeres but also occurs at heterochromatic repeats across the genome (Voon et al., 2015) . Voon et al. show that ATRX/H3.3 specifically localizes to silence imprinted alleles in mouse ES cells. Moreover, Atrx plays an essential role in directing the binding of the repressive PRC2 complex to Xist RNA causing X chromosome inactivation (Sarma et al., 2014). This role in X chromosome inactivation highlights the importance of Atrx in ES cell differentiation where imprinting is a crucial process. The role of Atrx in ES cell state maintenance is less studied.

7. Other remodelers (Smarca5, Smarcad1, Chd1l)

A number of other remodelers were less studied in the context of their role in ES cells but merit to be better analyzed.

a. Smarca5

Smarca5 was shown to be required during the preimplantation stage and ES cell deficient for Smarca5 fail to differentiate (Stopka and Skoultchi, 2003). Moreover, Smarca5 was shown to be a part of the Oct4 proteomic network in ES cells (van den Berg et al., 2010a).

b. Smarcad1

Few studies have examined the role of this remodeler. An RNAi and gene expression profiling screen identified Smarcad1 to contribute to ES cell state preservation (Hong et al., 2009). Moreover, a genome wide Smarcad1 localization in embryonic carcinoma cells revealed its binding to active gene transcription start site conferring its probable role also in ES cells (Okazaki et al., 2008).

c. Alc1 (Chd1l)

Chd1l was shown to be a proteomic partner of Oct4 (van den Berg et al., 2010a). The depletion of Chd1l in ES cells causes severe proliferative problems suggesting its importance in self-renewal (Efroni et al., 2008). Interestingly, Chd1l contains a domain that recognizes PAR or poly-ADP ribose which is a posttranslational modification added to nuclear receptor proteins (Mohrmann and Verrijzer, 2005). This posttranslational modification was shown to be important during the initial steps of reprogramming to iPS cells, where PARP1 protein was shown to interact with Chd1 through its PAR domain (Jiang et al., 2015). Moreover, CHIP analysis suggest that PARP1 and Chd1l co-occupy key pluripotency gene loci (Jiang et al., 2015).

Background and Objectives

Several studies have shown the importance of a number of ATP-dependent chromatin remodeling factors in the control of the ES cell state and differentiation. However, no global view of how these factors are recruited onto the mammalian genome was conducted.

In this work, we try to deal with several aspects regarding the role of Snf2 family of ATP-dependent chromatin remodeling factors in the transcriptional regulation of the mouse ES cell genome. Primarly, we focus on the distribution of remodelers in the genome, more precisely, on proximal (promoters) and distal regulatory (enhancers and CTCF-binding sites) DNA elements. Next, we try to understand whether remodelers are differentially recruited to specific target genes and how they integrate promoter nucleosomal architecture to regulate ES cell transcription programs. Further on, we analyze the extent by which chromatin remodeling factors contribute to the transcription networks that control ES cell pluripotency and self renewal.

We conducted a genome-wide analysis of the function of a selection of thirteen chromatin remodeling factors in the transcriptional control of the mouse ES cell state using mainly a double experimental strategy. First a ChIP-seq approach was used to identify the binding profiles of remodelers across the regulatory elements of the genome. Next, we performed loss-of function transcriptional analysis to reveal specific functional roles of remodelers in the ES cell transcriptional regulation.

Results

Part I

Article 1: Genome-wide nucleosome specificity and function of chromatin remodellers in embryonic stem cells

(article in review)

Maud de Dieuleveult[1]*, Kuangyu Yen[2,3]*, Isabelle Hmitou[1]*, Arnaud Depaux[1]*, Fayçal Boussouar[1,4], Daria Bou Dargham[1], Sylvie Jounier[1], Hélène Humbertclaude[1], Florence Ribierre[5], Céline Baulard[5], Céline Keime[6], Lucie Carrière[1], Soizick Berlivet[1], Marta Gut[7], Ivo Gut[7], Michel Werner[1], Jean-François Deleuze[5], Robert Olaso[5], Jean-Christophe Aude[1], Sophie Chantalat[5], B. Franklin Pugh[2], Matthieu Gérard[1]

* These authors contributed equally to this work

Correspondence and requests for materials should be addressed to M.G. (matthieu.gerard@cea.fr), K.Y. (kuangyuyen@gmail.com), or B.F.P. (bfp2@psu.edu)

Affiliations:

[1]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, CEN Saclay, 91191 Gif-sur-Yvette, France

[2]Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

[3]Present address: Department of Cell Biology, Southern Medical University, Guangzhou 510515, PR China

[4]Present address : INSERM, U823, Université Joseph Fourier - Grenoble 1, Institut Albert Bonniot, Faculté de Médecine, La Tronche, France

[5]Centre National de Génotypage, Institut de Génomique, CEA, Evry, France

[6]IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), INSERM, U964, CNRS, UMR7104, Université de Strasbourg, 67404 Illkirch, France

[7]Centre Nacional D'Anàlisi Genòmica, Barcelona, Spain

Summary

How various ATP-dependent chromatin remodellers bind to nucleosomes to regulate transcription is not well defined in mammalian cells. Here, we present genome-wide remodeller-interacting nucleosome profiles for Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Brg1 and Ep400 in mouse embryonic stem (ES) cells. These remodellers bind to nucleosomes at specific positions, either at one or both nucleosomes that flank each side of nucleosome-free promoter regions (NFRs), at enhancer elements, or within gene bodies. Surprisingly, large NFRs that extend downstream of transcriptional start sites are nevertheless chromatinized with non-nucleosomal histone modifications and variants. Thus, RNA polymerase II must navigate several hundred bp of noncanonical chromatin at the 5' ends of these genes. At promoters, bidirectional transcription commonly initiates on the flanks of remodeller-bound nucleosomes that reside at NFR boundaries. Transcriptome analysis upon remodeller depletion reveals reciprocal mechanisms of transcriptional regulation by remodellers. At active genes, certain remodellers are positive regulators of transcription, whereas others act as repressors. At bivalent genes, which are bound by repressive Polycomb complexes, the same remodellers act in the opposite way. Together, these findings reveal how remodellers integrate promoter nucleosomal architecture to regulate ES cell transcription programs.

Main Text

Embryonic stem (ES) cell pluripotency and differentiation are controlled by the coordinated action of multiple epigenetic factors that affect the structure of chromatin and regulate gene expression. Chromatin remodelling, mediated by the Snf2 family of ATP-dependent chromatin remodellers, is a key factor in these processes[1-5]. These remodellers are classified in 24 subfamilies that regulate transcription, DNA repair and DNA replication[6-9]. In yeast, remodellers interact with specific nucleosome positions at promoters and induce directional sliding of nucleosomes[10,11]. Remodellers also catalyze nucleosome destabilization and eviction at DNA regulatory elements[12-14]. These activities regulate access to DNA for transcription activators, as well as for the general TFs and RNA polymerase (pol) II, ultimately allowing or preventing the formation of the preinitiation complex (PIC) at promoters[15-17]. Subsequent transcription by pol II occurs in two phases. The first phase occurs at the 5' ends of genes and is marked by serine 5 phosphorylation (S5ph) of pol II, histone H3 lysine 4 trimethylation (H3K4me3), and transcriptional pausing of pol II[18-21]. The second phase occurs downstream in the 3' direction, and is marked by pol II S2ph, H3K36me3, and a lack of pausing[18-21].

In mouse ES cells, several remodellers are essential for the regulation of pluripotency and differentiation[1-5]. However, it is unclear whether mammalian remodellers target specific nucleosomes to regulate transcription. To explore this, we applied the genome-wide factor-nucleosome interaction assay[10,22] (remodeller MNase-seq) to ES cells. We first engineered an affinity tag onto individual remodellers at their endogenous loci. Proteins were then formaldehyde-crosslinked to chromatin *in vivo*, MNase digested to release individual nucleosomes, then immunoprecipitated sequentially with two distinct antibodies against the tag, using a tandem affinity protocol (Extended Data Fig. 1a). DNA fragments immunoprecipitated with each remodeller were then identified by deep sequencing.

Positionally-linked and positionally-independent remodellers

Since nucleosomes in cultured cell populations tend to have fuzzy positioning (not well resolved from each other), we began our analysis by examining 500 bp of remodeller-enriched genomic intervals, then subsequently focused on individual nucleosome positions. We employed Pearson Correlation to examine the connections between the remodellers, pol II, selected histone marks, and TFs (**Fig. 1a**). We focused on DNase I hypersensitive sites (DHSs), which reveal promoters and distal regulatory elements such as enhancers[23]. High correlation scores were observed among the remodellers Brg1, Ep400, Chd1, Chd4, Chd6 and Chd8, as well as with the mediator (Med1), suggesting that these factors tend to occupy the same genomic regions in ES cells.

In contrast, Chd9 and Chd2 generally did not occupy the same nucleosomes as these other remodellers, nor with each other (**Fig. 1a**), although they did interact with nucleosomes that Chd1 interacted with. Chd2 was the only remodeller associated with the transcription elongation-associated histone mark H3K36me3, and thus may be enriched in gene bodies and depleted near transcription start sites. One potential reconciliation of these results is that Chd9 and Chd2 target different nucleosomes on the same genes (e.g. 5' versus the inside of transcription units), with Chd1 extending across both types.

Ep400, Chd8, and Chd1 had correlated occupancies with pol II ser5ph (**Fig. 1a**), indicating that they may be present at the 5' end of genes that are transcriptionally active and/or paused. However, when we examined the pluripotency-associated transcription factor (pTF) Oct4/Pou5f1, which is active in ES cells, it tended to be associated with Brg1 and Chd4. This contrast suggests that stem cell-specific TFs work with a largely distinct cadre of remodellers compared to the general transcription machinery.

A specific analysis of active promoter regions within the DHSs revealed that most remodellers were correlated to varying degrees with components of the general transcription machinery, including pol II S5ph and TBP (**Fig. 1b**). A similar focus on enhancers confirmed the tight association of Brg1 and Chd4 with pTFs Oct4/Pou5f1, Sox2 and Nanog (Extended Data Fig. 2). Inspection of remodeller distributions at individual genes confirmed their widespread binding to promoters and enhancers, and specifically for Chd2, within transcribed units (**Fig. 1c**).

In this study, we chose to focus our analysis on remodeller function at promoters. To visualize in more detail where remodellers bind to around promoter regions, we examined their distribution around transcriptional start sites (TSSs). Locations were then sorted according to the local H3K4me3 signal, which is a mark of transcriptional activity (**Fig. 1d**). Remarkably, some remodellers like Brg1, Chd4, Chd6 bound similarly to all active genes, regardless of their H3K4me3/transcription level, while others, such as Chd1, Chd2, Chd9 and Ep400, were tightly linked to H3K4me3/transcription levels (**Fig. 1d** and Extended Data Fig. 3). Chd8 had a distribution pattern intermediate between these two groups.

Chd1 and Chd2, which are both related to *S. cerevisiae* (yeast) Chd1, showed strikingly different distributions. Whereas Chd1 is present near the 5' ends of genes, Chd2 enrichment pattern starts a few hundred base pairs downstream of the TSS, encompasses the entire transcription unit, and decreases to background levels at the transcription termination site (**Fig. 1c, d**). This is consistent with how yeast Chd1 works[24,25], and thus mammalian Chd2 and yeast Chd1 may be functionally equivalent.

Promoters are embedded in noncanonical chromatin, which is bordered by remodeller-bound nucleosomes

Since our assay detects primarily remodellers that are bound to nucleosomes, we next investigated more closely the relationship between individual remodeller-bound nucleosomes and promoter regions. We took advantage of a published dataset[26] to generate a reference nucleosome map for ES cells. Their use of MNase and size-selection ensured that native structures of fully assembled nucleosome were being reported. We first delimited for each gene their 5' nucleosome free region (NFR), which is a hallmark of promoters[27]. We aligned nucleosomes to the midpoint of these NFRs, which we define here as the half-way distance between the two MNase-defined nucleosomes that flank an annotated TSS. Nucleosomal patterns were then sorted by NFR length (**Fig. 2a**). We observed that Brg1, Chd1, Chd4, Chd6, Chd8 and Ep400 targeted specific nucleosomes that bounded both sides of the NFR. These nucleosomes also represent the boundaries of CpG islands (CpGIs), as reported previously[28] (**Fig. 2a)**. A comparison of our Chd4 dataset with a ChIP-seq dataset previously obtained using sonicated chromatin[29] showed that the tandem MNase ChIP-seq allows a far better resolution of the nucleosome positions bound by this remodeller and further validates this experimental approach (Extended Data Fig. 1d). Moreover, the sonication-based method demonstrates that Chd4 is not bound within NFRs in a non-nucleosomal manner.

The pattern of remodeller-bound nucleosomes contrasted with TSSs, which generally stayed towards the upstream side of the NFR (i.e., distal to gene bodies) (**Fig. 2a**). GRO-seq transcripts[30], which measure the location and direction of engaged pol II, tracked with annotated TSSs, and revealed previously described divergent noncoding transcripts[20,31] located at a fixed distance upstream of the TSS.

When DHSs were examined, they matched the more narrow positions of annotated TSSs, rather than reflecting the dimensions of NFRs. DHSs may represent relatively open accessible promoter/enhancer regions within NFRs since they are digested by both DNase I and MNase. (The former releases accessible DNA fragments generating a positive signal, whereas the latter destroys them thereby generating a negative signal.). The rest of the NFR may contain chromatin (being DNase I resistant), but may not be forming canonical nucleosomal structures (being MNase sensitive). We also probed these promoter regions using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements) method[32], and observed that the FAIRE signal was concentrated at the level of DHSs, but absent in the 3' part of the NFR (**Fig. 2b**). Thus, accessible promoter regions may be of fixed size and embedded on the upstream edge of islands of remarkably noncanonical chromatin.

To test this idea, we examined the distribution of histone marks and variants[33,34] in NFRs, measured by standard ChIP-seq. Remarkably, histone variant H3.3 was enriched both at nucleosomes flanking NFRs and inside NFRs, whereas H2A.Z, as well as histone marks H3K4me3 and H3K27ac were enriched primarily on the downstream half of NFRs (**Fig. 2b**). The CpG-rich NFR regions immediately downstream of the TSS are thus located within noncanonical chromatin containing H3.3 and H2A.Z.

To identify more precisely the nucleosome organization of remodellers surrounding NFRs, we divided genes into two different classes based on the length of the promoter NFR: class 1 (short NFR) and class 2 (large NFR), (**Fig. 2a**). Each of these two main classes was further subdivided into two subclasses: class 1a (NFR < 15 bp), class 1b (15-115 bp NFR), class 2a (116-504 bp), class 2b (505-1500 bp). At the short class 1 NFRs, Ep400 and Chd4 crosslinked predominantly to the two nucleosomes flanking the TSS, at positions -1 and +1 (**Fig. 2a** and **3**). Chd6, Chd8 and Brg1 interacted predominantly with the +1 position, and at lower levels with -1 and -2. Chd1 was also enriched at +1, and had a diffuse distribution on several

additional nucleosomes on both sides of the TSS, but had diminished interactions at the -1 position (**Fig. 2a** and **3**). Thus, the first nucleosome (+1 position) encountered by pol II after release from the pause state is one that is highly enriched with remodellers. These remodellers might play a role in the passage of pol II through these nucleosome barriers.

At the larger class 2 NFRs, Ep400 and Chd4 were preferentially bound to -1, and relatively less to +1 position. More strikingly, Chd6, Chd8 and Brg1 had shifted from their preferential binding to +1 at short NFRs toward a predominant enrichment at -1 position at large NFRs. Thus, mammalian remodellers interact with nucleosomes in a position-specific manner, with a distribution pattern adapting to NFR size.

To more closely explore the relation between the binding of remodellers to specific nucleosomes and to transcription initiation, we compared their positioning with transcriptionally engaged pol II (defined by GRO-seq) (**Fig. 2a** and **Fig. 3**). We observed that where NFRs were small, pol II that was initiated/paused in the sense direction was flanked by remodeller-interacting nucleosomes -1 and +1. However, antisense pol II accumulated as a peak present upstream of the -1 position. Thus, remodeler-bound -1 nucleosomes are boundaries between sense and antisense pol II (**Fig. 3**, top panels). Intriguingly, the peak of DHSs coincided with the -1 position in the short NFR promoter category. As the TSS either overlaps or is very close to the -1 position, the Ep400 and Chd4 remodellers that bind this nucleosome might be required for its destabilization, ejection or repositioning to regulate sense and antisense preinitiation complex (PIC) formation.

In contrast to short NFRs, large NFR promoters had DHS peaks and the midpoint between sense and antisense pol II residing downstream of -1, near the 5’ part of the NFR. This is precisely at the location where we detect a minor population of remodeler-bound nucleosomes (asterisk in **Fig. 3**). It is thus remarkable that for all categories of promoters, bidirectional

transcription initiates on either side of remodeller-bound nucleosomes. These nucleosomes are highly accessible, in that they are readily released by DNase I and correspond to the peak of DHSs (**Fig. 2a** and **Fig. 3**).

Control of ES cell transcriptome by remodellers

While virtually all transcribed genes in ES cells are marked by H3K4me3 at the promoter, a subset of less active genes is marked by a combination of H3K4me3 and H3K27me3, defining them as bivalent[35-37] (**Fig. 1d**). Bivalent genes were bound by remodellers with three distinct patterns: i) Chd4, Chd6 and Brg1 bound these genes with similar intensities as non-bivalent H3K4me3-only (hereafter referred to as H3K4me3) genes, ii) Ep400 and Chd8 showed substantially lower enrichments at bivalent promoters, but were nevertheless bound to most (> 95 %) of them, iii) Chd1 and Chd9 were bound at low levels to fewer than half of the total number of bivalent promoters (Extended Data Fig. 3 and 4). To investigate how the remodellers bound at these distinct classes of genes are involved in transcription regulation, we profiled mRNA expression in ES cells depleted of each remodeller using shRNA vectors (Extended Data Fig. 1a and 5). For Brg1, we took advantage of a published dataset obtained using cells depleted of Brg1 by genetic ablation[38].

We observed that Chd4, Ep400 and Brg1, among the tested remodellers, were the most required for transcriptional regulation in ES cells, in both the H3K4me3 and bivalent categories (**Fig. 4**). Ep400 and Chd4 were primarily involved in transcriptional activation of H3K4me3 promoters, whereas Brg1 showed a preferential involvement in transcription repression in this category (**Fig. 4a**). At bivalent promoters, Ep400 and Chd4 were mostly involved in transcriptional repression (**Fig. 4b**). In contrast, the dominant activity of Brg1 at bivalent genes was to counteract transcriptional repression, underlining a previously identified anti-silencing function associated with this remodeller[38].

Loss-of-function of the other remodellers resulted in more limited changes in gene expression. Chd1 depletion triggered the down-regulation of genes with H3K4me3 promoters, whereas bivalent genes were much less affected (**Fig. 4a**). Loss of Chd6 and Chd8 function resulted each in similar numbers of up- and down-regulated genes with H3K4me3 promoters, showing that they are equally involved in positive and negative modulation of this category. However, bivalent target genes were mostly up-regulated, suggesting that Chd6 and Chd8 are principally transcription repressors at these loci. We validated the results of this analysis by RT-qPCR using two different shRNA vectors for each remodeller (Extended Data Fig. 6 and Methods).

Relationships of NFR size and CpG islands with remodeller function

CpGIs were previously proposed to be a major determinant of remodeller requirements in transcriptional control[39]. We wondered how NFR length and CpG content could influence remodeller-mediated transcription regulation of genes with H3K4me3 and bivalent promoters. We thus compared the percentages of genes either down- or up-regulated by loss of function of each remodeller in the following two groups: 1) NFR length classes subdivided into H3K4me3 and bivalent subclasses (**Fig. 5a**); 2) mouse genes divided into four quartiles based on GC/CpG content at promoters (GC/CpG class 1, 2, 3, 4: low, moderate, intermediate, high, see Extended Data Fig. 7), then further subdivided into H3K4me3 and bivalent subclasses (**Fig. 5b**).

We focused our analysis on Chd4, Ep400 and Brg1, which, among the studied remodellers, are the most involved in transcriptional control. At H3K4me3 promoters, Chd4 and Ep400 were the most frequently required for the transcriptional activation of genes with short NFRs, suggesting that these remodellers stimulate transcription by destabilizing or repositioning nucleosomes at these densely occupied promoters (**Fig. 5a**). Accordingly, Chd4 function was

more often essential for the transcriptional activation of genes with CpG-poor promoters than those with CpG-rich promoters (**Fig. 5b**).

At bivalent promoters, at which Chd4 and Ep400 are potent transcription repressors, the repression of large NFR/high CpG content promoters was more frequently dependent on Chd4 and Ep400 function, compared to short NFR/low-CpG content promoters (**Fig. 5a,b**). These two remodellers might be involved in the stabilization or repositioning of labile nucleosomal particles within the NFR.

In contrast to Chd4 and Ep400, genes having short NFRs and/or low CpG content were relatively repressed by Brg1, and those having long NFRs and/or high CpG content were relatively activated, regardless of whether they were of the H3K4me3 class or bivalent (**Fig. 5a,b**). These regulatory properties of Brg1 in ES cells contrast with those previously reported for Brg1 and Brm in the regulation of inflammation-induced genes in macrophages[39]. At those genes, Brg1/Brm function was preferentially required for transcriptional activation at a subset having non-CpGI promoters, and thus short NFRs with dense nucleosome occupancy. In ES cells, an equivalent function in remodeller-dependent transcription activation of H3K4me3, non-CpGI promoters is supported by Ep400 and Chd4, but not by Brg1.

Conclusion

The results presented here paint three contrasting stereotypes of remodeller control of gene expression in mouse ES cells (**Fig. 5c**), although not all genes fall into these stereotypes.

First, at active (H3K4me3-only) genes with short NFRs, the upstream (-1) and downstream (+1) flanking nucleosomes are engaged with positive-acting Chd4 and Ep400. The downstream +1 nucleosome is engaged with negatively acting Brg1, as well as Chd1, Chd6

and Chd8. Further downstream, where transcription-linked H3K36me3 is found, so too is Chd2. Conceivably, Chd2 may be utilizing the H3K36me3 mark to organize nucleosomes analogously to Chd1 or Isw1b in budding yeast[25].

The second type of organization is at active genes with large NFRs, in which the TSS is present in a region of low nucleosomal occupancy, within the 5' part of a CpGI. While NFRs are larger, they are nevertheless mostly chromatinized. There, Chd4 and Ep400, which are engaged with the upstream -1 nucleosome, are less frequently required for transcription activation than at short NFR promoters. However, Brg1, which is preferentially engaged at -1 position at these large NFRs active promoters, is now mostly involved in transcriptional activation.

The third stereotypic organization is at bivalent genes, which often have a large NFR and CpGI, and which are epigenetically marked by Polycomb Repressive Complex (PRC) 2. Since repression of transcription by PRC2 is only effective in differentiating, but not in self-renewing ES cells[40], a possible function of Chd4 and Ep400 might be to keep bivalent genes repressed in ES cells until differentiation initiates. In contrast, Brg1 also predominates at -1 nucleosomes at these promoters, and is activating rather than inhibiting. This function of Brg1 might be required to keep a certain level of transcription at bivalent genes, preventing a fully repressed status.

Thus, we have two types of reciprocal relationships among chromatin remodellers at non-bivalent and bivalent gene classes: 1) An activating remodeller in one class is an inhibitor remodeller in the other class; 2) Within the same class, an activating remodeller can be counteracted by an inhibitor remodeller.

Taken together, the work presented here reveals that the various remodellers in mouse ES cells work together. They do so in a position-specific manner, organizing nucleosomes and

regulating gene expression either positively or negatively depending on nucleosomal/chromatin architecture flanking the core promoters.

METHODS

Full Methods are available in the Supplementary Information section.

References

1 Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5187-5191, doi:10.1073/pnas.0812888106 (2009).

2 Fazzio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134**, 162-174, doi:10.1016/j.cell.2008.05.031 (2008).

3 Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863-868, doi:10.1038/nature08212 (2009).

4 Reynolds, N. *et al.* NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell stem cell* **10**, 583-594, doi:10.1016/j.stem.2012.02.020 (2012).

5 Wang, L. *et al.* INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell stem cell* **14**, 575-591, doi:10.1016/j.stem.2014.02.013 (2014).

6 Flaus, A., Martin, D. M., Barton, G. J. & Owen-Hughes, T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic acids research* **34**, 2887-2905, doi:10.1093/nar/gkl295 (2006).

7 Becker, P. B. & Workman, J. L. Nucleosome remodeling and epigenetics. *Cold Spring Harbor perspectives in biology* **5**, doi:10.1101/cshperspect.a017905 (2013).

8 Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annual review of biochemistry* **78**, 273-304, doi:10.1146/annurev.biochem.77.062706.153223 (2009).

9 Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490-503, doi:10.1016/j.cell.2013.07.011 (2013).

10 Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T. & Pugh, B. F. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**, 1461-1473, doi:10.1016/j.cell.2012.04.036 (2012).

11 Gkikopoulos, T. *et al.* A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* **333**, 1758-1760, doi:10.1126/science.1206097 (2011).

12 Badis, G. *et al.* A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Molecular cell* **32**, 878-887, doi:10.1016/j.molcel.2008.11.020 (2008).

13 Ehrensberger, A. H. & Kornberg, R. D. Isolation of an activator-dependent, promoter-specific chromatin remodeling factor. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10115-10120, doi:10.1073/pnas.1101449108 (2011).

14 Gutierrez, J. L., Chandy, M., Carrozza, M. J. & Workman, J. L. Activation domains drive nucleosome eviction by SWI/SNF. *The EMBO journal* **26**, 730-740, doi:10.1038/sj.emboj.7601524 (2007).

15 Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193-198, doi:10.1038/nature08450 (2009).

16 Smith, C. L. & Peterson, C. L. ATP-dependent chromatin remodeling. *Current topics in developmental biology* **65**, 115-148, doi:10.1016/S0070-2153(04)65004-6 (2005).

17 Mueller-Planitz, F., Klinker, H. & Becker, P. B. Nucleosome sliding mechanisms: new twists in a looped history. *Nature structural & molecular biology* **20**, 1026-1032, doi:10.1038/nsmb.2648 (2013).

18 Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88, doi:10.1016/j.cell.2007.05.042 (2007).

19 Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & development* **26**, 2119-2137, doi:10.1101/gad.200303.112 (2012).

20 Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848, doi:10.1126/science.1162228 (2008).

21 Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445, doi:10.1016/j.cell.2010.03.030 (2010).

22 Koerber, R. T., Rhee, H. S., Jiang, C. & Pugh, B. F. Interaction of transcriptional regulators with specific nucleosomes across the Saccharomyces genome. *Molecular cell* **35**, 889-902, doi:10.1016/j.molcel.2009.09.011 (2009).

23 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).

24 Simic, R. *et al.* Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *The EMBO journal* **22**, 1846-1856, doi:10.1093/emboj/cdg179 (2003).

25 Smolle, M. *et al.* Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nature structural & molecular biology* **19**, 884-892, doi:10.1038/nsmb.2312 (2012).

26 Teif, V. B. *et al.* Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology* **19**, 1185-1192, doi:10.1038/nsmb.2419 (2012).

27 Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews. Genetics* **10**, 161-172, doi:10.1038/nrg2522 (2009).

28 Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome research* **22**, 2399-2408, doi:10.1101/gr.138776.112 (2012).

29 Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221-225, doi:10.1038/nature10805 (2012).

30 Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & development* **25**, 742-754, doi:10.1101/gad.2005511 (2011).

31 Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849-1851, doi:10.1126/science.1162253 (2008).

32 Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* **17**, 877-885, doi:10.1101/gr.5533506 (2007).

33 Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature genetics* **41**, 941-945, doi:10.1038/ng.409 (2009).

34 Weber, C. M. & Henikoff, S. Histone variants: dynamic punctuation in transcription. *Genes & development* **28**, 672-682, doi:10.1101/gad.238873.114 (2014).

35 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).

36 Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353, doi:10.1038/nature04733 (2006).

37 Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature cell biology* **8**, 532-538, doi:10.1038/ncb1403 (2006).

38 Ho, L. *et al.* esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nature cell biology* **13**, 903-913, doi:10.1038/ncb2285 (2011).

39 Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114-128, doi:10.1016/j.cell.2009.04.020 (2009).

40 Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Molecular cell* **55**, 347-360, doi:10.1016/j.molcel.2014.06.005 (2014).

Figure LEGENDS

Figure 1. Binding of ATP-dependent remodellers to cis-acting regulatory sequences on the ES cell genome

a, Heat map representing Pearson correlations between remodellers, DNase I hypersensitivity (DNase-seq), histone marks, Oct4 (Pou5f1), mediator (Med1) and Pol II S5ph at 138,582 genomic regions centred on DHSs. **b**, Similar as in (a) but for 16,300 promoter-like, H3K4me3-, TBP- and Pol II S5ph-positive DHSs. **c**, Binding profiles at a representative locus. Counts indicate reads per 10 millions. Promoters and enhancers are highlighted by blue and yellow squares, respectively. **d**, Heat-map representation of ChIP-seq binding for remodellers at 14,623 RefSeq promoters, rank-ordered from highest to lowest H3K4me3 deposition level. Color intensity represents sequencing tag counts. For each promoter, remodeller occupancy is indicated within a 10 kb window centred on the TSS. All promoters are transcribed from left to right. RNA expression level of the corresponding genes is indicated on the right (color intensity reflects expression level).

Figure 2. Remodellers target specific nucleosomes on both sides of promoter NFRs

Remodeller-bound nucleosomal tags were aligned to the promoters of 14,623 RefSeq genes rank-ordered from smallest to largest NFR length. Remodeller occupancy and other features are indicated within a 4 kb window centred on mid-NFR. **a**, Panels show from left to right : reference nucleosome map in grey with positions of TSSs in green, presence of CpG islands, DNase I hypersensibility (red) superposed with reference nucleosome map, pol II (GRO-seq signal) engaged in sense (blue) and antisense (red) transcription, and the tandem ChIP

MNase-seq signal for each remodeller. Color intensity represents sequencing tag counts. Chd2 did not crosslink with any particular nucleosome position in the promoter regions. Chd9 exibited a restricted number of specific binding sites at promoters, and thus shows no clear preference for particular nucleosome positions. The delimitation of the four subclasses of promoters defined by NFR length are indicated by dashed lines. **b**, same as in (a) for histone variant H3.3, histone marks H3K4me3, H3K27me3, H3K27ac, H2A.Z, and FAIRE-seq signal.

Figure 3. Bidirectional transcription initiates on either side of remodeller-bound nucleosomes

Average binding profiles of remodellers at active (H3K4me3-only) promoters with short NFRs (left, subclass 1b from **Fig. 2**) and large NFRs (right, subclass 2b). The top panels show in grey the reference nucleosome map (derived from MNase-seq data). Tags from reference nucleosomes, remodeller-interacting nucleosomes and GRO-seq were aligned to -1 and +1 nucleosome dyad positions, and plotted upstream or downstream of these reference nucleosomes. Distances (bp) are indicated from -1 and +1 dyad positions. A gap in the NFR was introduced to normalize variations in NFR length inside each class. Sense and antisense pol II, which are defined by GRO-seq signal, are indicated as blue and red dashed lines, respectively. Asterisks point to remodeller-bound nucleosomes that were not detected in the reference ES cell map.

Figure 4. Remodellers differentially regulate active versus bivalent genes

The transcriptome of ES cells depleted of each remodeller was analyzed by microarray hybridization. The number of genes either down-regulated (dark grey) or up-regulated (light

grey) following the depletion of each remodeller was defined for genes transcribed from H3K4me3 promoters (a), and for genes with bivalent promoters (b), using a 1.5 fold-change threshold. Statistical analysis was performed using a 2-sample test for equality of proportions with continuity correction. * P < 0.05, ** P < 0.01, *** P < 0.001.

Figure 5. Remodellers distinctively control the transcription of genes having either large NFR, CpGI-rich, or small NFR, CpGI-poor promoters

a, **b**. The percentages of genes down-regulated and up-regulated by remodeller depletion in ES cells were determined for each of the indicated subgroups. **a**, Genes were divided in 4 subclasses based on NFR length at the promoter, as defined in **Fig. 2**. Each of these subclasses was further divided in two categories with H3K4me3 or bivalent promoters. Statistical analysis of small NFRs (class 1a) versus large NFRs (class 2b) is shown in black and grey for genes down- and up-regulated by remodeller depletion, respectively. Statistical analysis was performed as in Fig. 4. * P < 0.05, ** P < 0.01, *** P < 0.001. **b**, Genes were divided in 4 classes based on GC/CpG content at the promoter (Extended Data Fig. 7), and analyzed as in (a). Statistical analysis was carried out to compare GC/CpG class 1, which is composed of genes with CpGI-poor promoters, and CpG class 4, which is CpGI-rich. **c**, Model illustrating the binding patterns and transcriptional control exerted by remodellers at H3K4me3 promoters with either short or large NFRs, and at bivalent genes. Several of the Chd-named remodellers are designated by their number only.

Acknowledgements

We thank J.C. Andrau, S. Ravens, L. Tora and M. de Chaldée for critical reading, J.B. Charbonnier's team for sharing material, A. Krebs, T. Ye, I. Davidson, R. Guerois and F Ochsenbein for discussions, A. Martel, K. H. (Sol) Han and B. Park for bioinformatic support. This work was supported by the CEA, and by grants from the Association pour la Recherche sur le Cancer (grant 3164), the Agence Nationale de la Recherche (ANR-05-BLAN-0396) and from National Institutes of Health (grant HG004160, to B.F.P.)

Figure 1

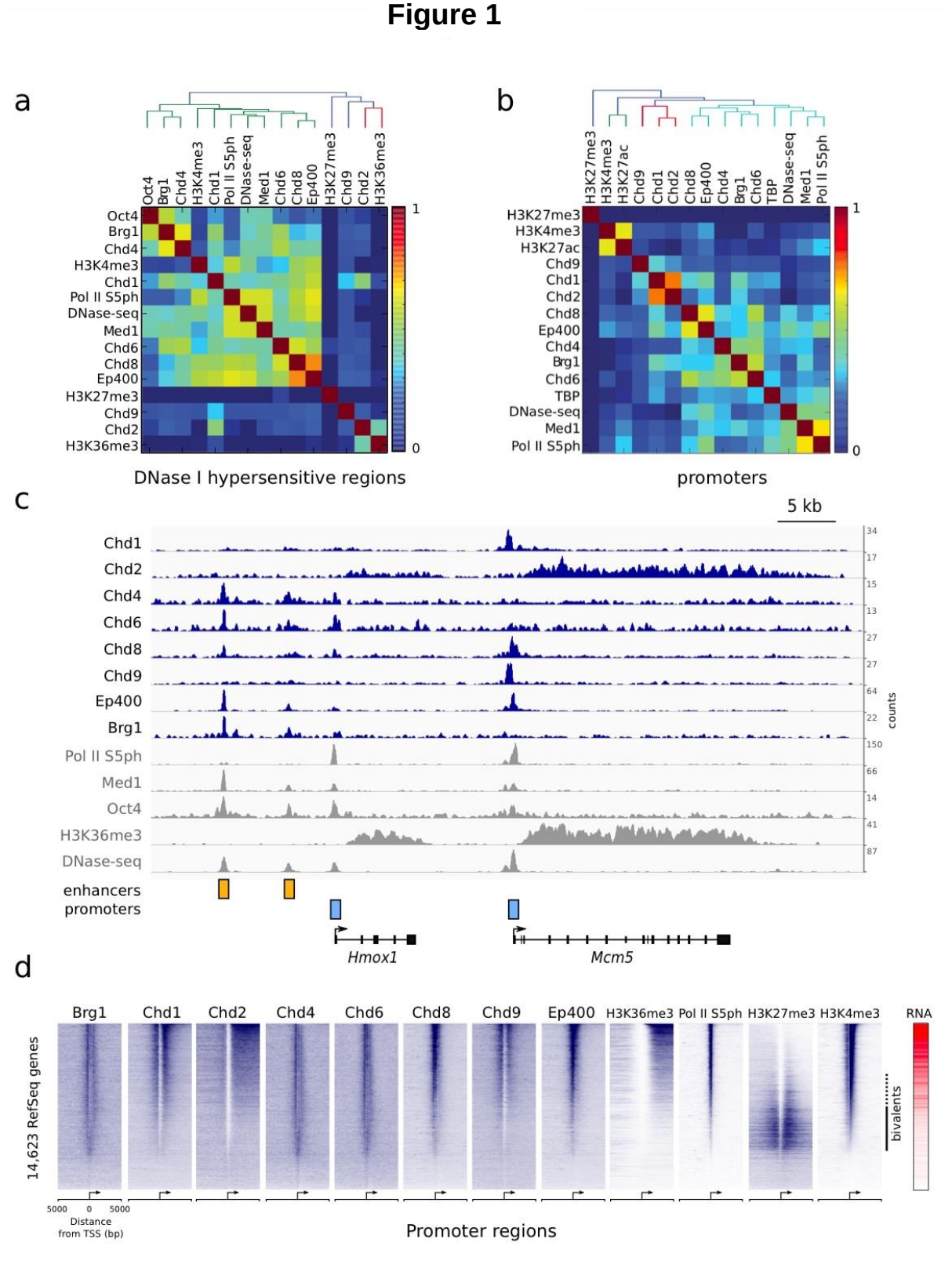

a
Oct4
Brg1
Chd4
H3K4me3
Chd1
Pol II S5ph
DNase-seq
Med1
Chd6
Chd8
Ep400
H3K27me3
Chd9
Chd2
H3K36me3
1
0
DNase I hypersensitive regions
b
H3K27me3
H3K4me3
H3K27ac
Chd9
Chd1
Chd2
Chd8
Ep400
Chd4
Brg1
Chd6
TBP
DNase-seq
Med1
Pol II S5ph
promoters
c
5 kb
counts
enhancers
promoters
Hmox1
Mcm5
d
14,623 RefSeq genes
bivalents
RNA
5000
0
5000
Distance from TSS (bp)
Promoter regions

Figure 2

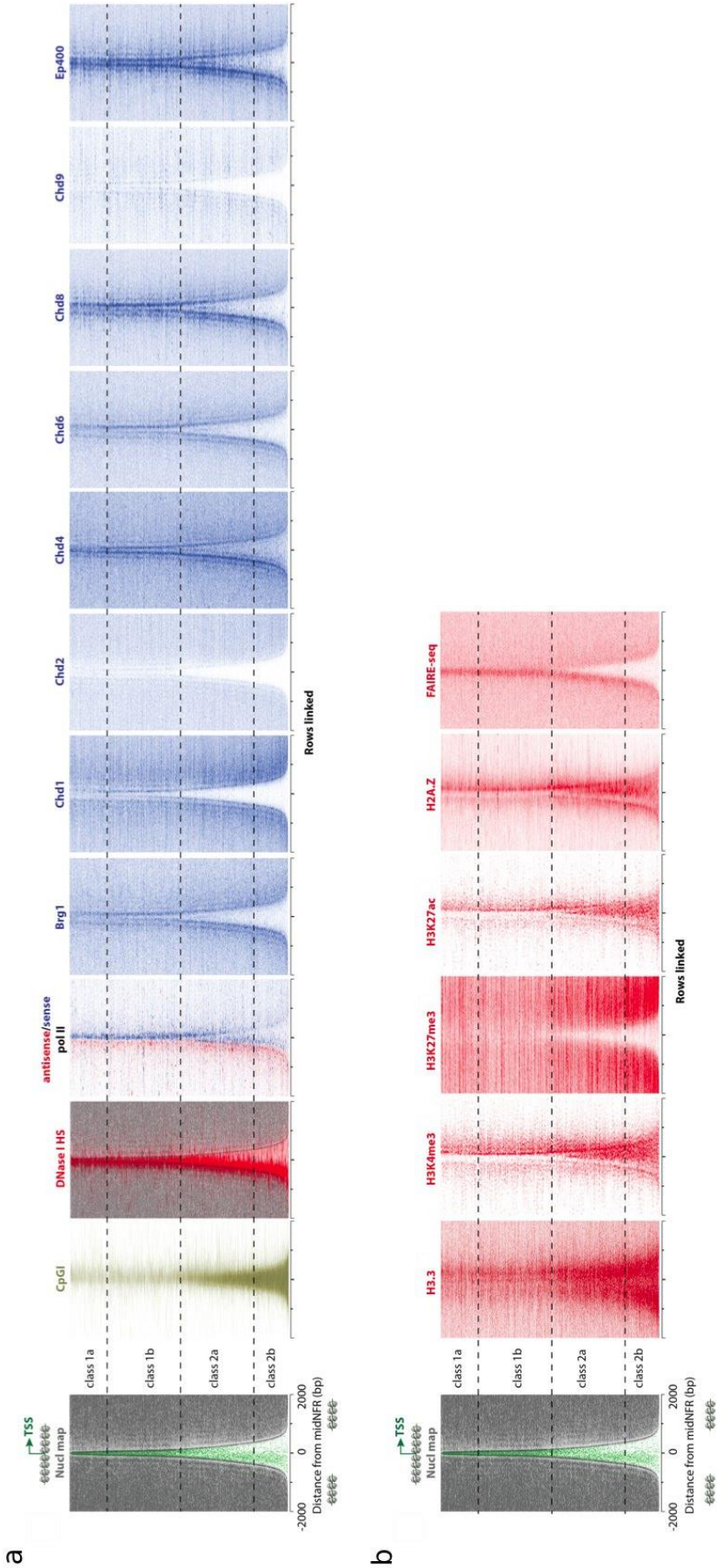

a
TSS
Nucl map
class 1a
class 1b
class 2a
class 2b
-2000
0
2000
Distance from midNFR (bp)
CpGI
DNase I HS
antisense/sense pol II
Brg1
Chd1
Chd2
Chd4
Chd6
Chd8
Chd9
Ep400
Rows linked
b
TSS
Nucl map
class 1a
class 1b
class 2a
class 2b
-2000
0
2000
Distance from midNFR (bp)
H3.3
H3K4me3
H3K27me3
H3K27ac
H2A.Z
FAIRE-seq
Rows linked

Figure 3

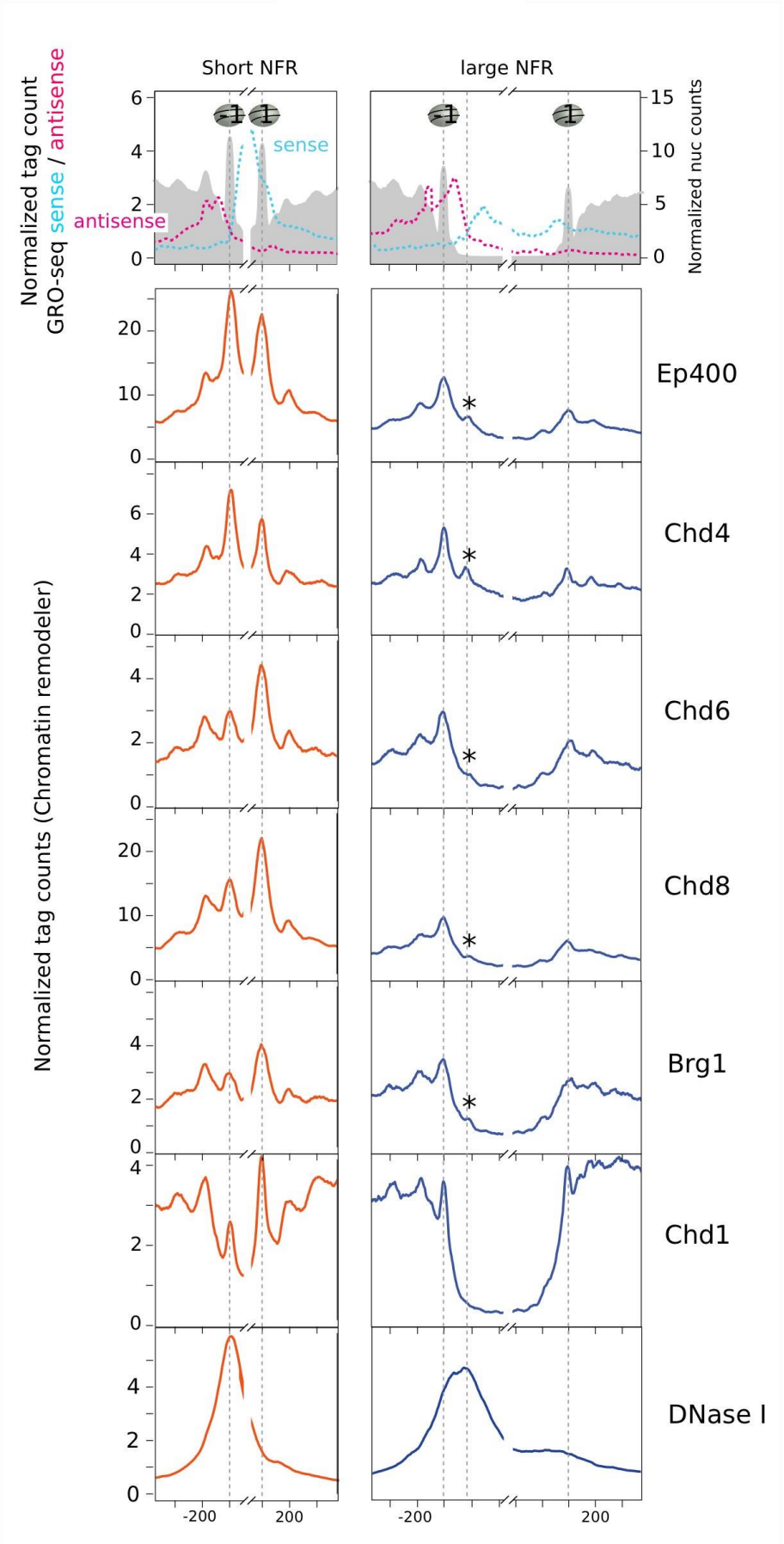

Short NFR
large NFR
Normalized tag count
GRO-seq sense / antisense
sense
antisense
Normalized nuc counts
Normalized tag counts (Chromatin remodeler)
Ep400
Chd4
Chd6
Chd8
Brg1
Chd1
DNase I
-200
200

Figure 4

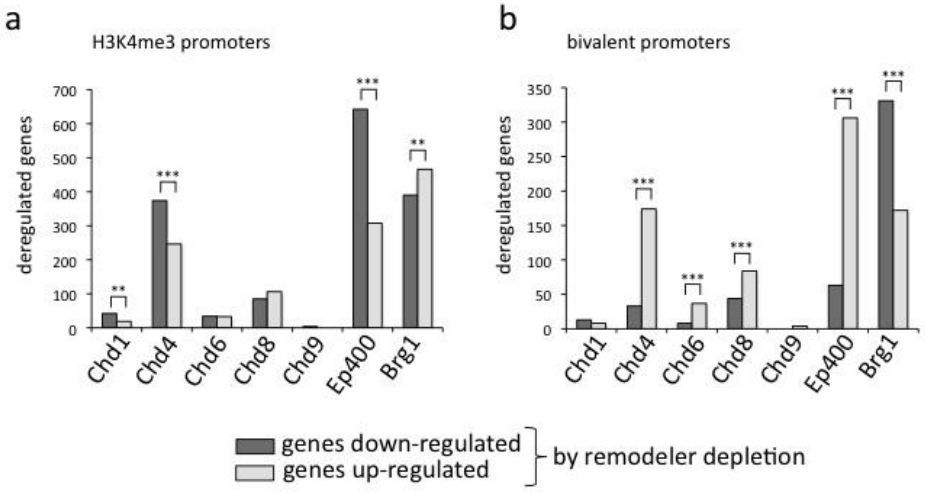
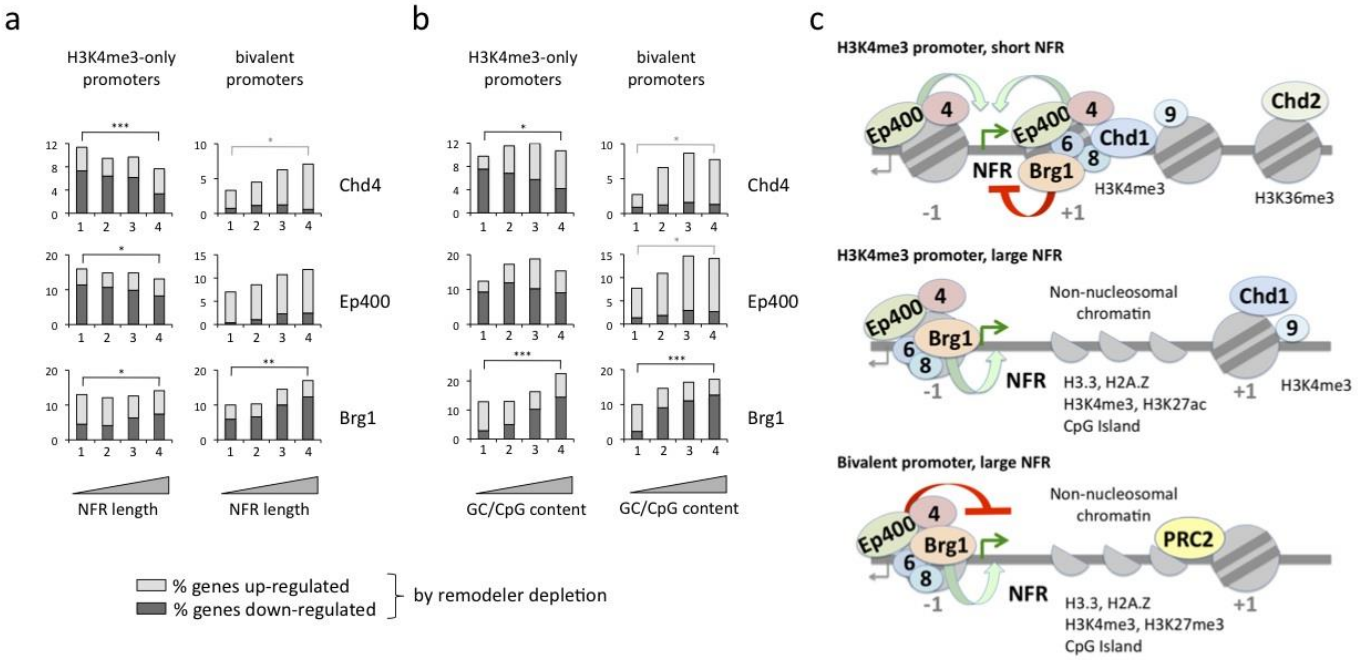

a
H3K4me3 promoters
deregulated genes
700
600
500
400
300
200
100
0
**
***
***
**
Chd1
Chd4
Chd6
Chd8
Chd9
Ep400
Brg1
b
bivalent promoters
deregulated genes
350
300
250
200
150
100
50
0
***
***
***
***
Chd1
Chd4
Chd6
Chd8
Chd9
Ep400
Brg1
genes down-regulated
genes up-regulated
by remodeler depletion

Figure 5


a
H3K4me3-only promoters
bivalent promoters
Chd4
Ep400
Brg1
NFR length
NFR length
b
H3K4me3-only promoters
bivalent promoters
Chd4
Ep400
Brg1
GC/CpG content
GC/CpG content
% genes up-regulated
% genes down-regulated
by remodeler depletion
c
H3K4me3 promoter, short NFR
Ep400
Chd1
Chd2
NFR
Brg1
H3K4me3
H3K36me3
-1
+1
H3K4me3 promoter, large NFR
Non-nucleosomal chromatin
Ep400
Brg1
Chd1
NFR
H3.3, H2A.Z
H3K4me3, H3K27ac
CpG Island
H3K4me3
Bivalent promoter, large NFR
Non-nucleosomal chromatin
Ep400
Brg1
PRC2
NFR
H3.3, H2A.Z
H3K4me3, H3K27me3
CpG Island

SUPPLEMENTARY INFORMATION

EXTENDED DATA Figure LEGENDS

Extended Data Figure 1: MNase-ChIP-seq of tagged ATP-dependent remodellers in ES cells

a, Experimental strategy. Using homologous recombination in ES cells, a sequence encoding a combination of FLAG and hemagglutinin (HA) epitopes was introduced at the 3' end of the coding sequence of the genes encoding the catalytic subunit of each remodeller. After *in vivo* crosslinking, chromatin was prepared and fragmented to mononucleosomes by MNase. Remodeller-bound mononucleosomes were isolated using a double immunoaffinity procedure. Immunopurification efficiency was assessed by Western Blotting. Deep sequencing of the DNA from purified nucleosomes allowed the mapping of remodeller-bound nucleosomes across the mouse genome. The same tagged ES cell lines were used for shRNA-mediated depletion of remodellers and transcriptome analysis. **b**, Scheme of the remodellers tagged in this study, with known domains: Chromo, (Chromatin organization modifier) domain; PHD, (Plant Homeo Domain) finger; BRK domain; Bromo, bromodomain; Helicase, Helicase conserved C-terminal domain; SNF2, SNF2 family N-terminal domain, which contains the ATP-binding domain. **c**, Tagging of remodellers does not alter ES cell pluripotency. ES cell colonies tagged for the indicated remodellers were assayed for alkaline phosphatase activity. Colonies fully positive, partially positive, and negative for alkaline phosphatase activity were scored as undifferentiated, mixed, and differentiated, respectively. Results show the average and s.d. of the percentages of ES cell colonies belonging to each category in three independent experiments. **d**, Comparison of MNase ChIP-seq and sonication ChIP-seq for Chd4. The left panel shows the reference nucleosome map of 14,623 RefSeq genes, rank-ordered from smallest to largest NFR length, as in **Fig. 2**. On the right are

compared the distribution patterns obtained for Chd4 either by MNase ChIP-seq, with chromatin prepared from Chd4-tagged ES cells, or by ChIP-seq with sonicated chromatin (dataset accession number: GSM687284).

Extended Data Figure 2: Remodellers at enhancer elements

Heat map representing Pearson correlations between remodellers, DNase I hypersensitivity (DNase), histone marks, Oct4 (Pou5f1), Sox2, Nanog and the mediator (Med1) at 25,547 enhancer-like DHSs.

Extended Data Figure 3: Relation between remodeller enrichment at promoters and RNA expression level

Average binding profile of remodellers at promoters, divided in four quartiles based on RNA expression level of the corresponding genes. All promoters are transcribed from left to right. Promoter binding intensity of Chd1, Chd2, Chd9 and Ep400 at H3K4me3 promoters was correlated with RNA expression (see Methods). Accordingly, binding of these remodellers to bivalent promoters, which are transcribed at lower levels, showed a significant reduction compared to H3K4me3 promoters. In contrast, Chd4, Chd6, and Brg1 enrichment at promoters showed little correlation with the transcription level of the corresponding genes, and was only slightly lower at bivalent, compared to H3K4me3 promoters.

Extended Data Figure 4: Nucleosome targeting by remodellers at H3K4me3-only and bivalent promoters

a, Remodeller-bound nucleosomal tags were aligned to the promoters of 6,481 active (H3K4me3 promoters) genes, rank-ordered from smallest to largest NFR length. Reference nucleosome map, remodeller occupancy and the other indicated features are shown as in Fig. 2. **b**, same as in (a) for the promoters of 3,411 bivalent genes.

Extended Data Figure 5: Western blot analysis of remodeller depletion by shRNA for transcriptome analysis

ES cells tagged for each remodeller were transfected with the corresponding shRNA vector, or a control plasmid. After puromycin selection, ES cells were collected for RNA preparation and Western blot analysis. Three independent experiments were performed for each remodeller. Remodeller depletion was assessed using antibodies against FLAG or HA epitopes. Loading control: Gapdh.

Extended Data Figure 6: Representative examples of genes regulated by chromatin remodellers in ES cells.

Remodellers and histone marks enrichment profiles are shown as indicated on the left of each panel. A control ChIP profile, obtained with untagged ES cells, is shown for comparison. Scores indicate reads per 10 millions. On the right of each panel are shown the results of RT-qPCR analysis that indicate how the RNA expression level of the corresponding genes is affected by remodeller depletion in ES cells. Two distinct shRNA vectors (shRNA1 and shRNA2, see Methods) were used for each remodeller. Scores on the y axis indicate the relative expression of the indicated genes compared to reference genes.

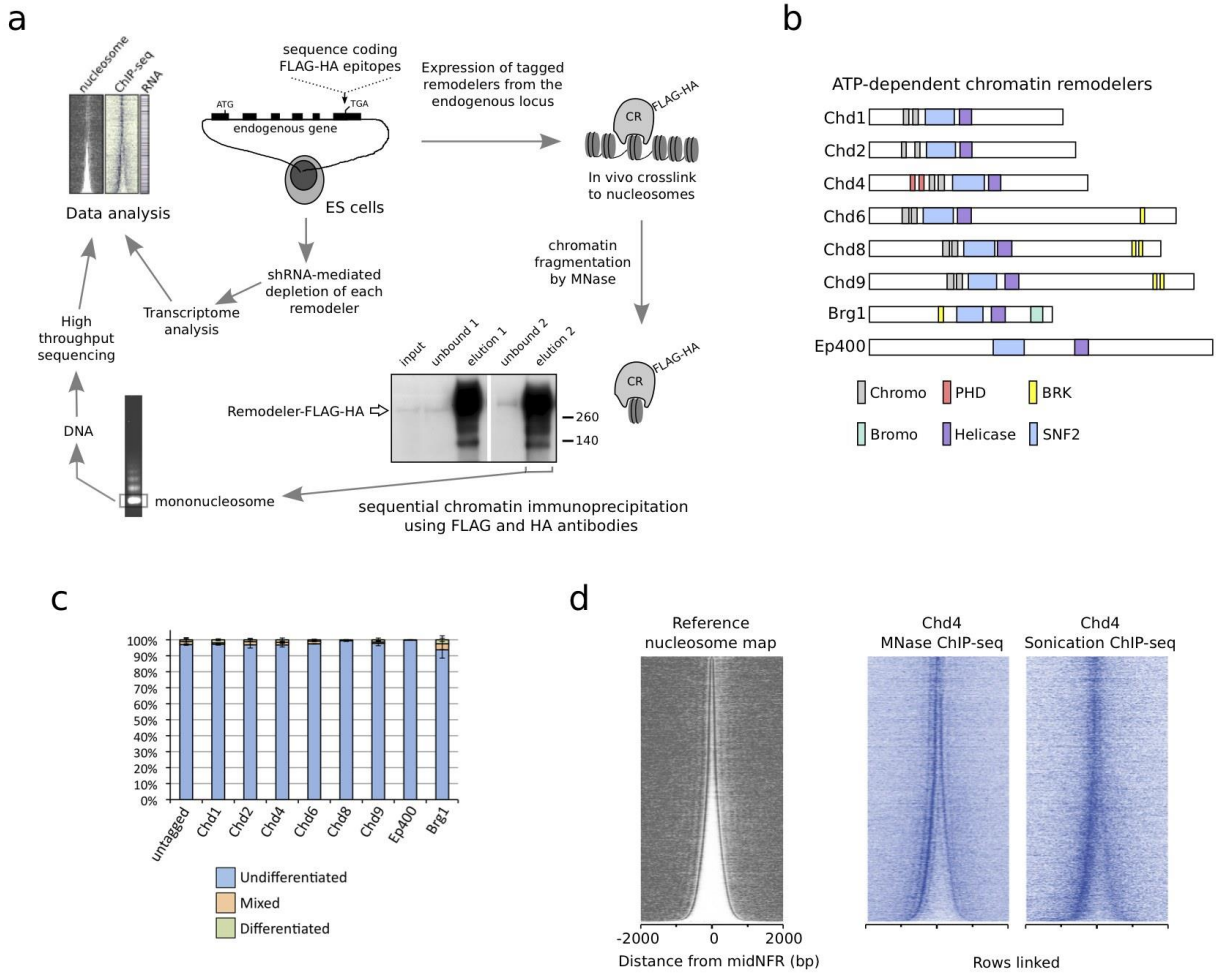
Extended Data Figure 7: Remodeller distribution at promoters in relation to GC/CpG content

Heat-map representation of ChIP-seq binding for remodellers and transcription landmarks at 6317 promoters, ordered by increasing GC content, as in (Fenouil et al, 2012), and centred on mid-NFR. Promoters in close proximity to other annotations were excluded from the analysis. All promoters are transcribed from left to right. Colour intensity represents sequencing tag counts. The limits of the four classes of increasing GC content are indicated by dashed lines.

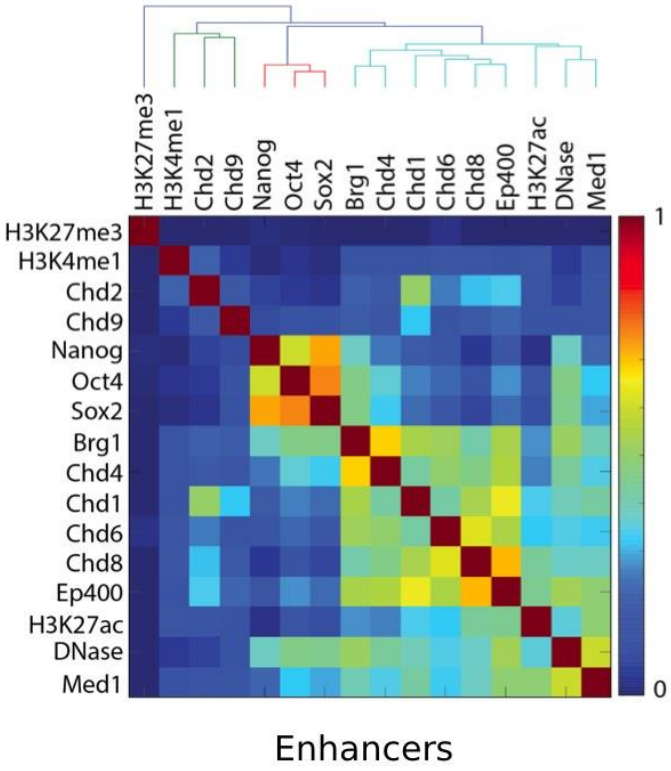
Extended Data Figure 8: Quality control of ChIP-seq replicate experiments

Heat-maps showing each chromatin remodeller bound to its most enriched genomic binding sites in ES cells. These genomic regions were defined for all remodellers, except Chd2, by peak calling using SICER software (see Methods), and were ordered from most to less binding. For Chd2, the 2,500 genes the most bound by Chd2 are presented in the heat-map. Two independent experiments, replicate 1 and 2, are presented for each remodeller, in comparison with a control ChIP-seq experiment, realized with chromatin from untagged ES cells, and an unrelated dataset to control the specificity of each pattern. Rows are linked among heat maps that are adjacent to each other along the horizontal axis. For Chd4 and Brg1, publicly available datasets (GSM687284 and GSM359413, respectively) obtained by ChIP-seq with sonication-fragmented chromatin are shown for comparison. Distances are indicated in kb from the centre of each genomic region, or from the TSS for Chd2.

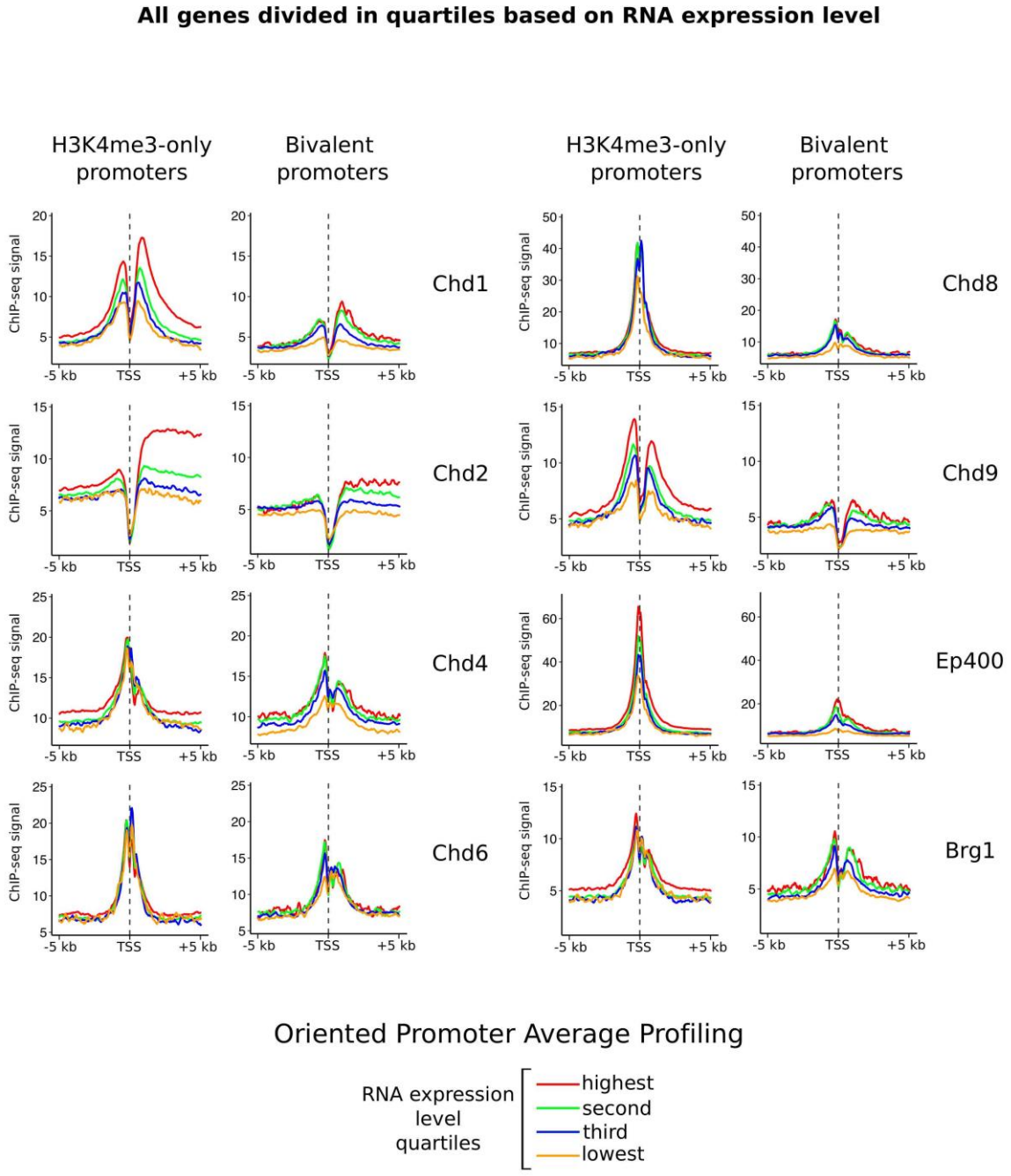
Extended Data Figure 1

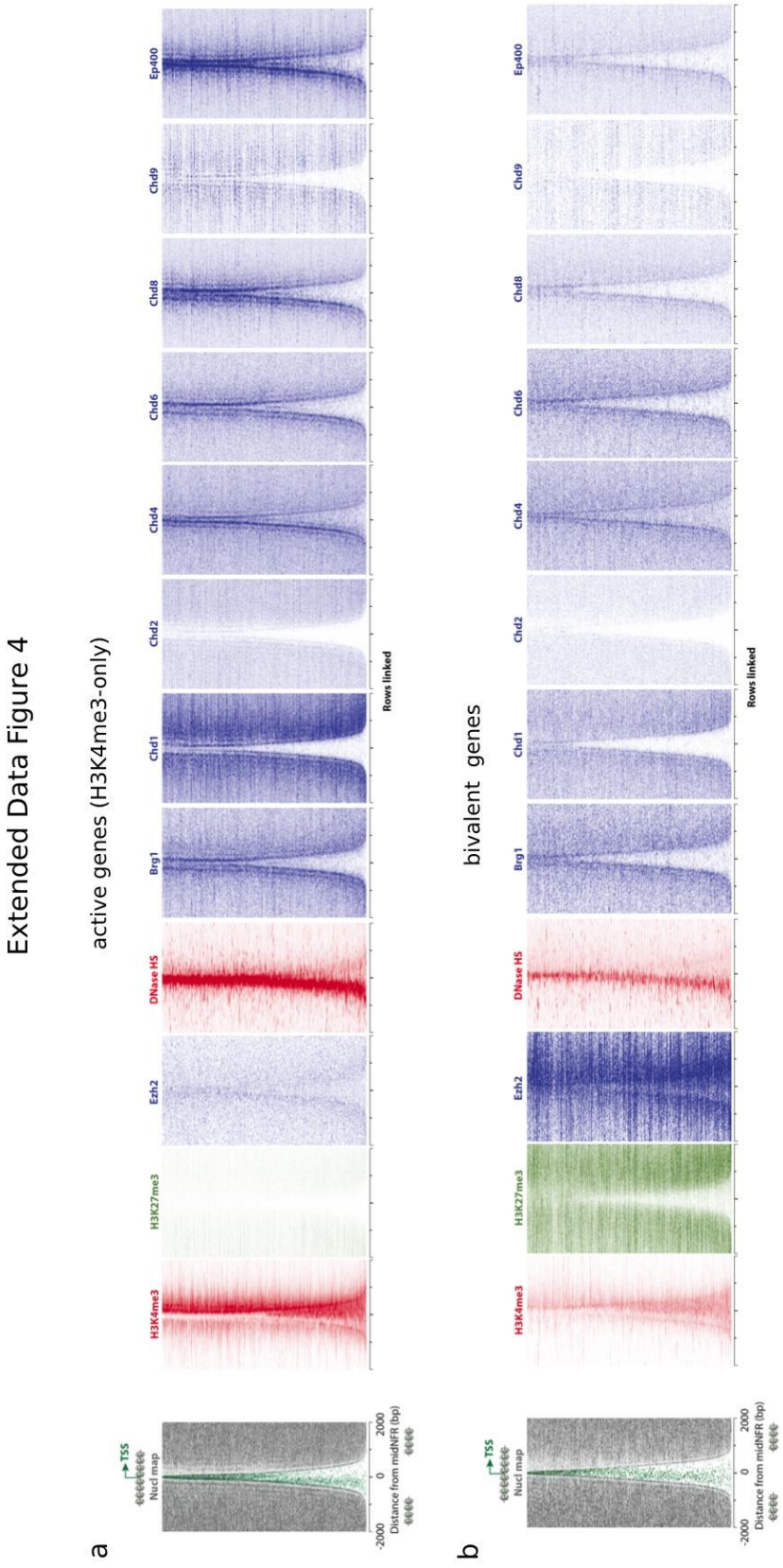


Extended Data Figure 2

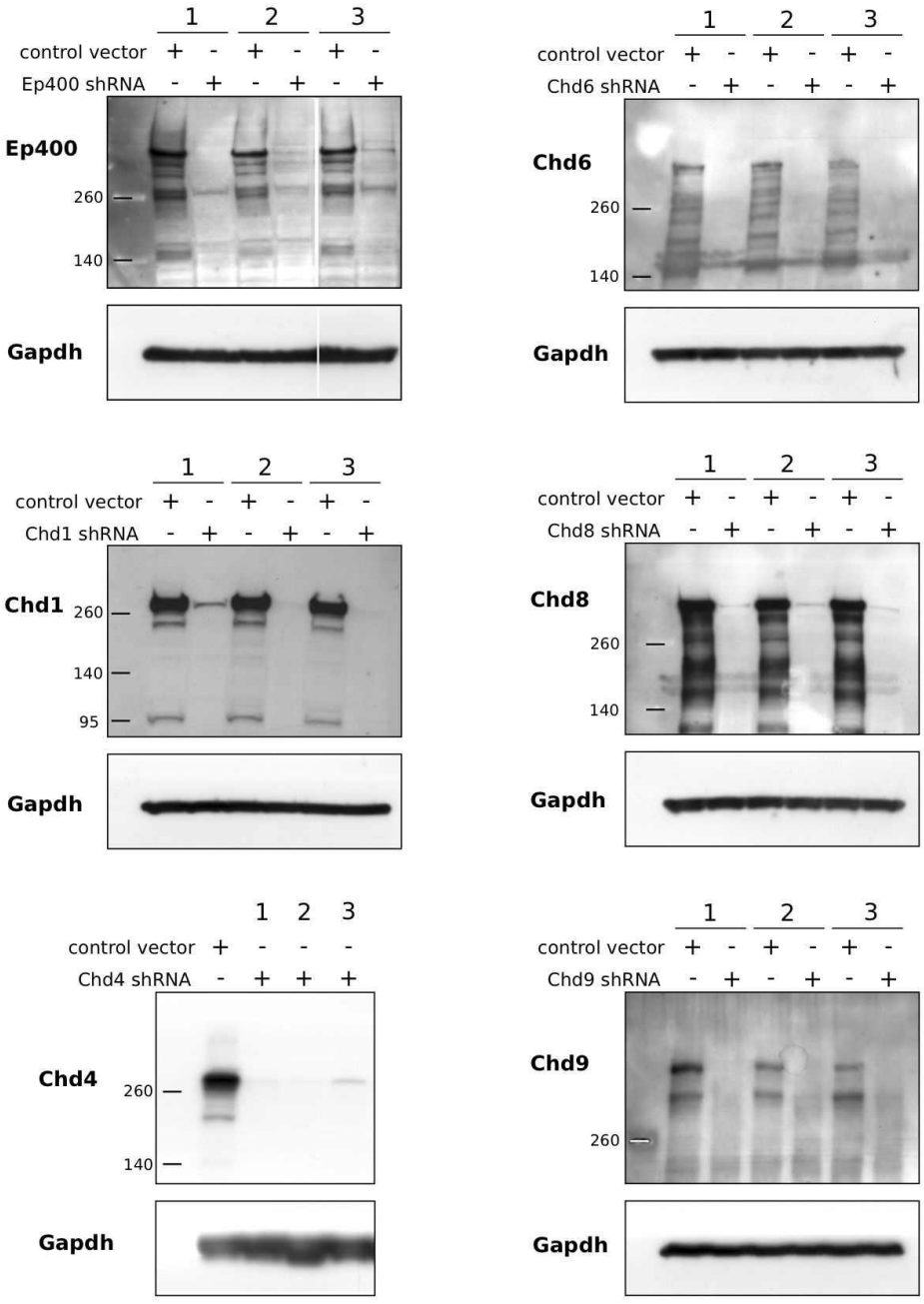


Extended Data Figure 3

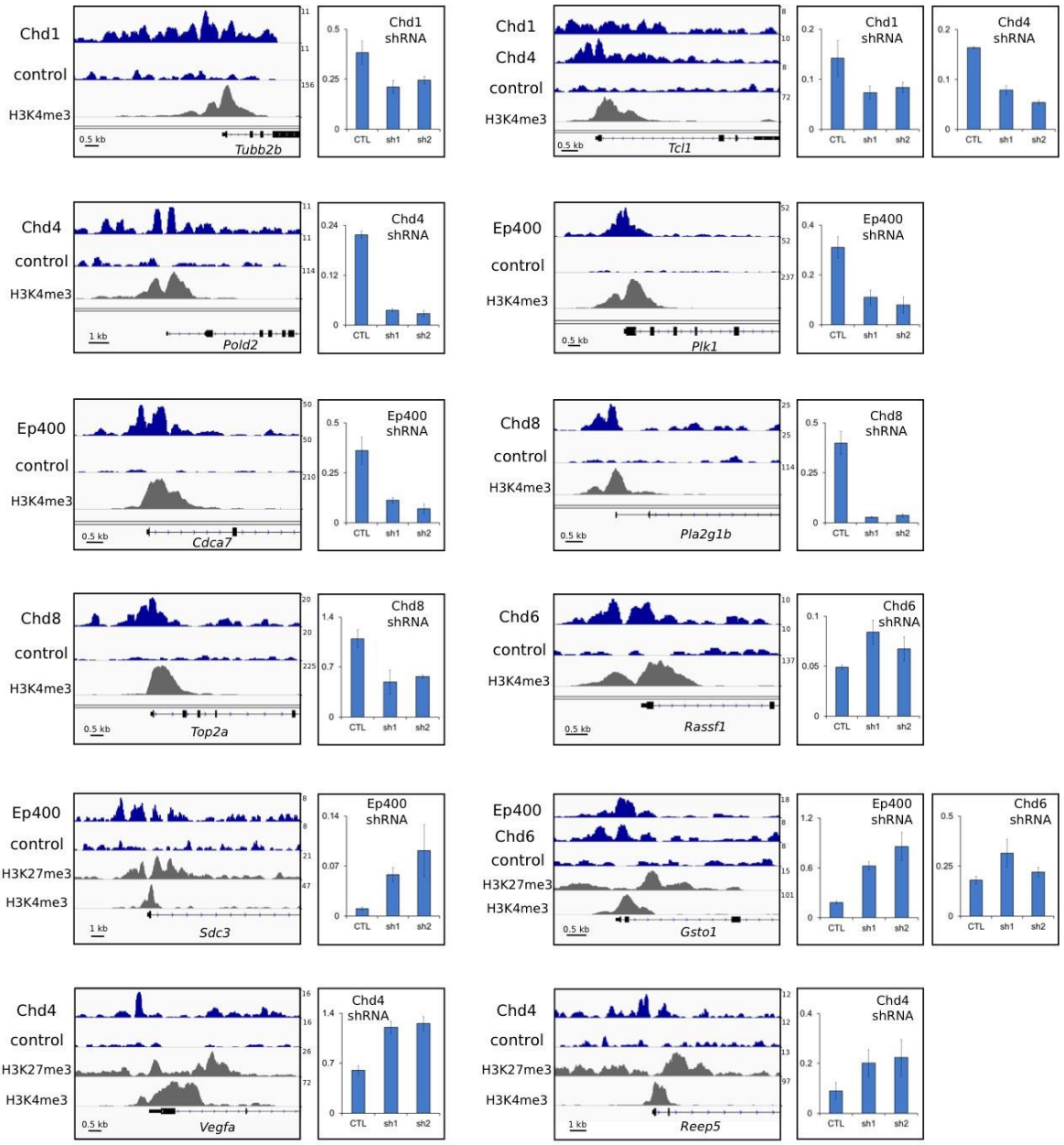


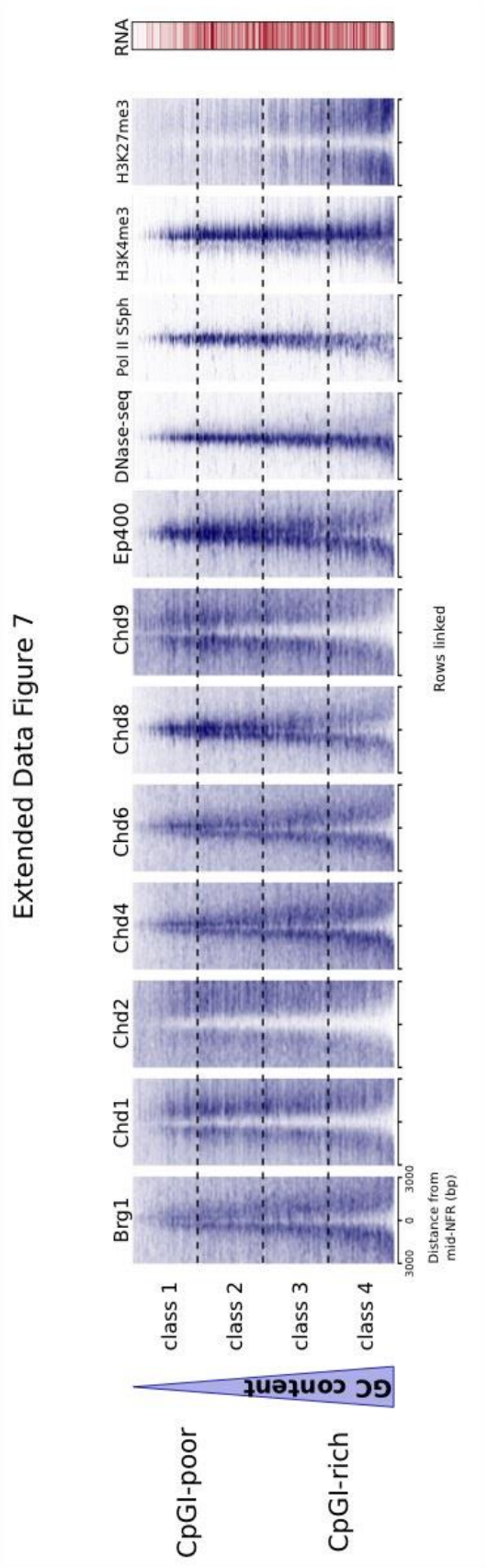

Extended Data Figure 4
active genes (H3K4me3-only)
bivalent genes
a
b
TSS
Nucl map
-2000
0
2000
Distance from midNFR (bp)
H3K4me3
H3K27me3
Ezh2
DNase HS
Brg1
Chd1
Chd2
Chd4
Chd6
Chd8
Chd9
Ep400
Rows linked

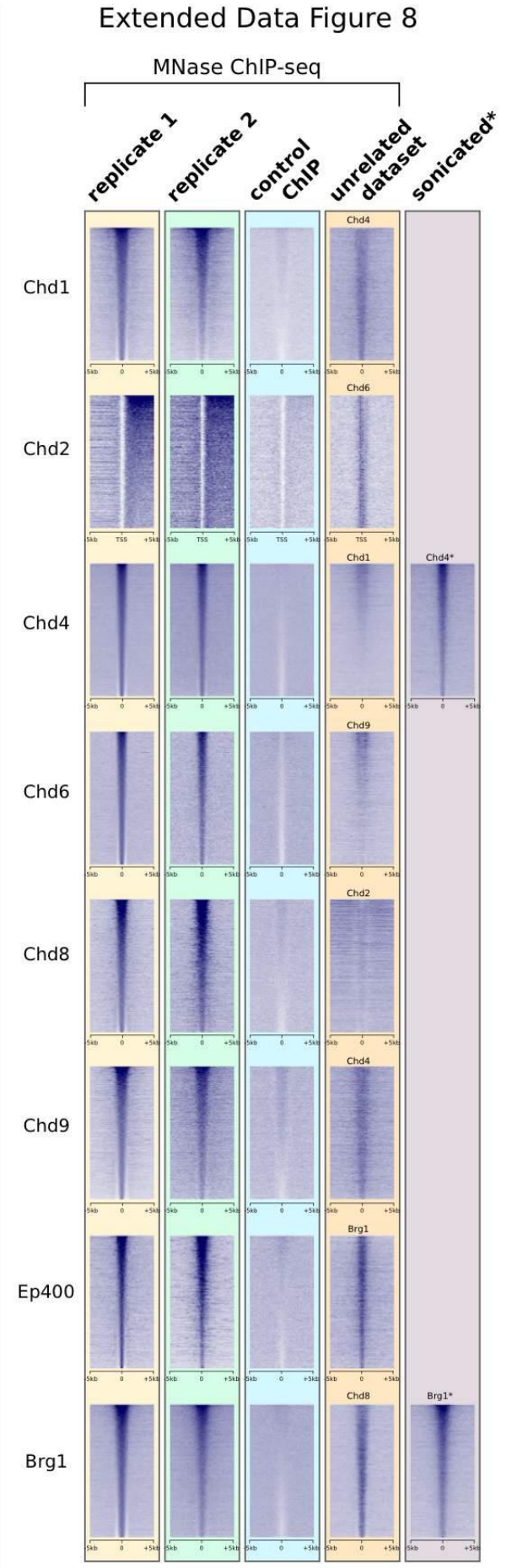
Extended Data Figure 5



Extended Data Figure 6




Extended Data Figure 7
RNA
H3K27me3
H3K4me3
Pol II S5ph
DNase-seq
Ep400
Chd9
Chd8
Chd6
Chd4
Chd2
Chd1
Brg1
Rows linked
Distance from mid-NFR (bp)
3000
0
3000
class 1
class 2
class 3
class 4
GC content
CpGI-poor
CpGI-rich


Extended Data Figure 8
MNase ChIP-seq
replicate 1
replicate 2
control ChIP
unrelated dataset
sonicated*
Chd1
Chd2
Chd4
Chd6
Chd8
Chd9
Ep400
Brg1
Chd4
Chd6
Chd1
Chd4*
Chd9
Chd2
Chd4
Brg1
Chd8
Brg1*
5kb
TSS
+5kb

Extended Data Tables

Extended Data Table 1. Sequencing experiments characteristics

TAP-tag	Cell line	Replicate experiment	Fixation method	Number of mapped reads
none	Untagged 46C	1	Formaldehyde	22 050 827
		2	Formaldehyde	8 298 397
(FLAG)3-TEV-HA	Chd1	1	DSG + formaldehyde	18 808 113
		2	Formaldehyde	10 639 876
(FLAG)3-TEV-HA	Chd2	1	Formaldehyde	23 221 972
		2	Formaldehyde	20 533 722
(FLAG)3-TEV-HA	Chd4	1	Formaldehyde	29 249 943
		2	Formaldehyde	22 091 884
(FLAG)3-TEV-HA	Chd6	1	Formaldehyde	27 824 849
		2	Formaldehyde	23 474 895
(FLAG)3-TEV-HA	Chd8	1	Formaldehyde	23 796 293
		2	Formaldehyde	11 708 086
6His-FLAG-HA	Chd9	1	DSG + formaldehyde	17 325 960
		2	Formaldehyde	20 139 275
(FLAG)3-TEV-HA	Ep400	1	Formaldehyde	23 999 045
		2	Formaldehyde	20 920 354
(FLAG)3-TEV-HA	Brg1	1	Formaldehyde	31 222 065
		2	Formaldehyde	11 378 387

Extended Data Table 2: Primers used in RT-qPCR experiments

	primer_up_5'-3' sequence	primer_down_5'-3' sequence	final concentration
Tubb2	CAAGACAGCAAGAAGCTAA	TGCTCATCACTGATGACCT	200nM
Pold2	GTACGCCCATATTTATGCCA	GCTGCAGCTCACACAACTTC	500nM
Cdca7	TTGTTCTCTGGACGACATT	CTGGTAAGAGGTCGAGTTCT	200nM
Top2a	GAAAGTAACGTTGATGTCCCT	CCGAGTCTAAATCCACTGTGA	500nM
Sdc3	GCCCAGCTCCCTCAGAAGA	GACCAGGAAGGCAGCGAAGA	200nM
Vegfa	ACCAAAGCCAGCACATAGGA	CGCCTTGGCTTGTCACA	500nM
Tcl1	GCTGCGACTCCATGTATTG	GGGTTATTCATCGTTGGACT	200nM
Plk1	GGACAGCGACTTTGTATTTGT	TATCTGTCGCAGGTAGTAGC	200nM
Pla2g1b	CTACAGTCAGGCCAAGAAG	CAGATGAAGTCCTCGCATTT	200nM
Rassf1	GGAGCTCGAGCAGTACTTCA	GCTTGAGAAAGATCGGGTGT	500nM
Gsto1	CAAGGGAATCCGGCACGAA	TCTCCAGAACCGGCACCAGC	200nM
Reep5	TTCGGATACCCAGCCTACAT	GGAGAAGAATTCGGCAATG	500nM
Nmt1	GTCTGAAAGCTGCGTACTCC	CAGTGCATTGAATACGTCGAA	500nM
Actb	GCTTCTTTGCAGCTCCTT	GACGACCAGCGCAGCGATA	500nM
Ddb1	CCGCATTGAAGTCCAAGATAG	GTCTCGTGAGGAGCCGTACT	200nM

SUPPLEMENTARY METHODS

Cell line and ES cell culture

Mouse 46C ES cells have been described previously[41]. 46C ES cells and their tagged derivatives were cultured at 37°C, 5% CO2, on mitomycin C-inactivated mouse embryonic fibroblasts, in DMEM (Sigma) with 15% foetal bovine serum (Invitrogen), L-Glutamine (Invitrogen), MEM non-essential amino acids (Invitrogen), pen/strep (Invitrogen), 2-mercaptoethanol (Sigma), and a saturating amount of leukemia inhibitory factor (LIF), as described in reference[42].

Knock-in of a TAP-tag in the genes encoding the remodellers through homologous recombination in ES cells

The recombineering technique[43] was adapted to construct all targeting vectors for homologous recombination in ES cells. Retrieval vectors were obtained by combining 5' miniarm (NotI/SpeI), 3' miniarm (SpeI/BamHI) and the plasmid PL253 (NotI/BamHI). SW102 cells containing a BAC encompassing the C terminal part of the gene encoding the remodeller, were electroporated with the SpeI-linearized retrieval vector. This allowed the subcloning of genomic fragments of approximately 10 kb comprising the last exon of the gene encoding each remodeller. The next step was the insertion of a TAP-tag into the subcloned DNA, immediately 3' to coding sequence. The TAP-tag was (FLAG)3-TEV-HA for Chd1, Chd2, Chd4, Chd6, Chd8, Ep400, Brg1, and 6His-FLAG-HA for Chd9. We first inserted the TAP-tag and an AscI site into the PL452 vector, in order to clone 5' homology arms as SalI/AscI fragments into the PL452TAP-tag vector. 46C ES cells were electroporated with NotI-linearized targeting constructs and selected with G418. In all cases, G418-positive clones were screened by Southern blot. Details on the Southern genotyping strategy, as well as sequences of primers and plasmids used in this study are available upon request. Correctly

targeted ES cell clones were karyotyped, and the expression of each tagged remodeller was controlled by western blot analysis, using antibodies against FLAG and HA epitopes (see Extended Data Fig. 5). We also verified by immunofluorescence, using antibodies against FLAG and HA epitopes, that each tagged remodeller was properly localized in the nucleus of ES cells (data not shown).

Assessment of pluripotency in tagged ES cell line

ES cell lines expressing a tagged remodeller were all indistinguishable in culture from their mother cell line (46C). Pluripotency of tagged ES cell lines was verified by detecting alkaline phosphatase activity on ES cell colonies five days after plating, using the Millipore alkaline detection kit, following manufacturer's instructions. Three independent experiments were performed for each tagged ES cell line. The results of these experiments are shown in Extended Data Fig. 1c. In addition, we verified by immunofluorescence that expression of the pluripotency-associated transcription factor Oct4/Pou5f1 was uniform in each tagged ES cell line (data not shown).

Antibodies

Antibodies used for western blot were as follows:

Monoclonal antibodies anti-HA (H7, Sigma H3663) and anti-FLAG (M2, Sigma F1804), anti-GAPDH (Abcam ab9485).

Antibodies used for immunofluorescence : Oct4/Pou5f1 (Abcam ab9857), monoclonal antibodies anti-HA (HA.11, Covance MMS-101P), and anti-FLAG (M2, Sigma F1804).

Tandem affinity purification of remodeller-nucleosome complexes

ES cells were fixed either with formaldehyde, or with a combination of disuccinimidyl glutarate (DSG) and formaldehyde, then permeabilized with IGEPAL, and incubated with micrococcal nuclease (MNase) in order to fragment the genome into mononucleosomes. This nucleosome preparation was next incubated with agarose beads coupled with an antibody anti-HA or anti-FLAG. Anti-HA-agarose (ref. A2095) and anti-FLAG-agarose (ref. A2220) beads were purchased from Sigma. After a series of washes, tagged-remodeller-nucleosome complexes were eluted, either by TEV protease cleavage or by peptide competition. The eluted complexes were then subjected to a second immunopurification step, using beads coupled to the antibody specific of the second HA or FLAG epitope. After elution, DNA was extracted from the highly purified mononucleosome fraction, and processed for high-throughput sequencing. As a negative control, chromatin from untagged ES cells was subjected to the same protocol to define background signal. A detailed version of this protocol is available on the protocol exchange website: http://dx.doi.org/10.1038/protex.2014.040.

High-throughput sequencing of tandem ChIP samples

After crosslink reversion, phenol-chloroform extraction and ethanol precipitation, the DNA from remodeller-nucleosome complexes was quantified using the picogreen method (Invitrogen) or by running 1/20e of the ChIP material on a High sensitivity DNA chip on a 2100 Bioanalyzer (Agilent, USA). 5 to 10 ng of ChIP DNA were used for library preparation according to the Illumina ChIP-seq protocol (ChIP-seq sample preparation kit). Following end-repair and adapter ligation, fragments were size-selected on an agarose gel in order to purify genomic DNA fragments between 140 and 180 bp. Purified fragments were next amplified (18 cycles) and verified on a 2100 Bioanalyzer before clustering and single-read sequencing on a Genome Analyzer (GA) or GA II (Illumina), according to manufacturer's instructions. Sequencing experiments characteristics are shown in Extended Data Table 1.

Analysis of gene expression in 46C ES cells by RNA-seq

46C ES cells were amplified on feeder cells except for the last passage, at which cells were plated onto 60-mm dishes coated with gelatine, and grown to 70% confluence in D15 medium with LIF. Total RNA was extracted using an RNeasy Kit (Qiagen), and depleted of ribosomal RNA using the eukaryotic Ribominus kit (Invitrogen). RNA quality was verified on a 2100 Bioanalyzer. Library preparation was performed using the Illumina total RNAseq sample preparation kit according to manufacturer's instructions. After RNA fragmentation, reverse transcription and PCR amplification, single-read sequencing of 75 bp tags was carried out on a GA.

In order to keep only sequences of good quality we retained the first 40 bp of each read and discarded all sequences with more than 10% of bases having a quality score below 20, using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Mapping of these sequences onto the mm9 assembly of mouse genome and RPKM computation were then performed using ERANGE v3.1.0 [44] and bowtie v0.12.0 [45]. Briefly, a splice file was created with UCSC known genes and maxBorder=36. Then an expanded genome containing genomic and splice-spanning sequences was created using bowtie-build and bowtie was used to map the reads onto this expanded genome. Then the ERANGE runStandardAnalysis.sh script was used to compute RPKM values following steps previously described[44], using a consolidation radius of 20kb.

RNA preparation from ES cells depleted of each remodeller by shRNA

We used the pHYPER shRNA vector for remodeller depletion in ES cells, as previously described[46]. shRNA design was performed using DESIR software (http://biodev.extra.cea.fr/DSIR/DSIR.html). Below are listed the shRNA selected for each

remodeller. The sense strand sequence is given; the rest of the shRNA sequence is as described[46].

Chd1 shRNA 1: GCAAAGACGGCGACTAGAAGA

Chd1 shRNA 2: GACAGTGCTTAATCAAGATCG

Chd4 shRNA 1: GGACGACGATTTAGATGTAGA

Chd4 shRNA 2: GCTGACGTCTTCAAGAATATG

Chd6 shRNA 1: GTACTATCGTGCTATCCTAGA

Chd6 shRNA 2: CAGTCAGAACCCACAATAACT

Chd8 shRNA 1: GCAGTTACACTGACGTCTACA

Chd8 shRNA 2: GACTTTCTGTACCGCTCAAGA

Chd9 shRNA 1: TATACCAATTGAACAAGAGCC

Chd9 shRNA 2: AGTTAAAGTCTACAGATTAGT

Ep400 shRNA 1: GGTAAAGAGTCCAGATTAAAG

Ep400 shRNA 2: GGTCCACACTCAACAACGAGC

Each shRNA was transfected in its corresponding tagged ES cell line, in order to follow remodeller depletion by Western blotting using antibodies against Flag or HA epitopes (Extended Data Fig. 5).

The pHYPER shRNA vectors were transfected in ES cell by electroporation, using an Amaxa nucleofector (Lonza). 24 h after transfection, puromycin (2 μg/ml) selection was applied for an additional 48 h period, before cell collection and RNA preparation, except for Chd4, for

which cells were collected after 30 h of selection. Total RNA was extracted using an RNeasy Kit (Qiagen). Total RNA yield was determined using a NanoDrop ND-100 (Labtech). Total RNA profiles were recorded using a Bioanalyzer 2100 (Agilent). For each remodeller, RNA was prepared from three independent transfection experiments, and processed for transcriptome analysis.

Transcriptome analysis in remodeller depleted ES cells

cRNA was synthesized, amplified, and purified using the Illumina TotalPrep RNA Amplification Kit (Life Technologies) following Manufacturer's instructions. Briefly, 200 ng of RNA were used to prepare double-stranded cDNA using a T7 oligo (dT) primer. Second strand synthesis was followed by *in vitro* transcription in the presence of biotinylated nucleotides. cRNA samples were hybridized to the Illumina BeadChips Mouse WG-6v2.0 arrays. These BeadChips contain 45,281 unique 50-mer oligonucleotides in total, with hybridization to each probe assessed at 30 different beads on average. 26,822 probes (59%) are targeted at RefSeq transcripts and the remaining 18,459 (41 %) are for other transcripts. BeadChips were scanned on the Illumina iScan scanner using Illumina BeadScan image data acquisition software (version 2.3). Data were then normalized using the 'normalize quantiles' function in the GenomeStudio Software. Following analyses were done using Genespring software.

For Brg1, we used a previously published transcriptome dataset, in which loss of Brg1 function was obtained by genetic ablation[38]. All array analyses were undertaken using the Limma package from the R/Bioconductor software (R-Development-Core-Team, 2007). Microarray spot intensities were normalized using the RMA method as implemented in the R affy package. Normalized measures served to compute the log2-ratios for each gene between the wild-type strain and the Brg1 KO mutant. Then, to identify genes with a log2-ratios

significantly different between the mutant and wild- type strain, p-values were calculated for each gene using a moderated t-test. The moderated t-test applied here was based on an empirical Bayes analysis and was equivalent to shrinkage (or expansion) of the estimated sample variances towards a pooled estimate, resulting in a more stable inference. Finally, adjusted p-values were calculated using the Benjamini & Hochberg method and filtered using the thresholds of 0.05 for the p-value and 1.5 for the fold-change.

Quantitative RT-PCR

Random-primed reverse transcription was performed at 52°C in 20μl using Maxima First strand cDNA synthesis kit (Thermo Scientific) with 10 μg of total RNA isolated from ES cells (Qiagen), quantified with NanoDrop instrument (Thermo Scientific). Reverse transcription products were diluted 40-fold before use. Composition of quantitative PCR assay included 2.5 μl of the diluted RT reaction, 0.2 to 0.5 mM forward and reverse primers, and 1X Maxima SYBR Green qPCR Master Mix (Thermo Scientific). Reactions were performed in a 10μl total volume. Amplification was performed as follows: 2 min at 95°C, 40 cycles at 95°C for 15 sec and 60°C for 60 sec in the ABI/Prism 7900HT real-time PCR machine (Applied Biosystems). The real-time fluorescent data from quantitative PCR were analyzed with the Sequence Detection System 2.3 (Applied Biosystems). Each quantitative real-time PCR was performed using the set of primer pairs listed in Extended Data Table 2, validated for their specificity and efficiency of amplification. All reactions were performed in triplicates, using RNA prepared from three independent cell transfection experiments. Control reactions without enzyme were verified to be negative. Relative expression was calculated after normalization with three reference genes (*Actb*, *Nmt1* and *Ddb1*), validated for this study.

FAIRE-seq

FAIRE was performed as described[32] with modifications. 46C ES cells were amplified as described above for RNA preparation. Formaldehyde was added directly to the growth media (final concentration 1%), and cells were fixed for 5 min at room temperature. After quenching with glycine and several washes, cells were collected, resuspended in 500 µl of cold lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) and disrupted using glass beads for five 1-min sessions with 2 min incubations on ice between disruption sessions. Samples were then sonicated for 16 sessions of 1 min (30 sec on/30 sec off) using a bioruptor (Diagenode) at max intensity, at 4°C. After centrifugation, the supernatant was extracted twice with phenol-chloroform. The aqueous fractions were collected and pooled, and a final phenol-chloroform extraction was performed before DNA precipitation. Prior to sequencing, FAIRE DNA was analyzed and quantified by running 1/25e of the FAIRE material on a High sensitivity DNA chip on a 2100 Bioanalyzer (Agilent, USA). 20 ng of FAIRE DNA was used for library preparation according to the Illumina ChIP-seq protocol (ChIP-seq sample preparation kit). Single-read sequencing was performed on a Genome Analyzer II (Illumina) according to manufacturer's instructions.

Analysis of ChIP-seq datasets (Fig. 1d, Extended Data Fig. 7 and 8)

Short read alignment onto mouse mm9 genome was performed with Bowtie 0.12.7 with the followings settings : -a -m1 --best --strata -v2 -p3. Datasets were next converted to BED format files, and data analysis was performed using the seqMINER platform[47].

Peak calling was performed by running SICER software[48] with a control library obtained by applying the tandem ChIP protocol to untagged ES cells. The following settings were used: window size, 200 bp ; gap size, 200 bp ; FDR=1e-9.

In order to examine the distribution of remodellers at individual genes, we used WigMaker3 (default settings) to convert BED files into wig files, which were uploaded onto the IGV genome browser.

The reproducibility of ChIP-seq replicate experiments was assessed for each remodeller in Extended Data Fig. 8.

Lists of genes

The list of 14,623 genes used in Fig. 1 and 2 was obtained by filtering all mm9 RefSeq genes. We removed redundancies (that is, genes having the same start and end sites), unmappable genes and those with high ChIP-seq background, as well as genes shorter than 2 kb. The purpose of this last filtering step was to unambiguously distinguish the promoter region from the end of the genes in heat-maps. Transcription start sites (TSS) correspond to the location of the 5' RNA end.

Lists of genes with H3K4me3 and bivalent promoters: We first defined, among the 14,623 RefSeq genes, those with a promoter positive for H3K4me3 (accession number: GSM590111). Operating with the seqMINER platform, tag densities from this dataset were collected in a -500/+1,000 bp window around the TSS, and subjected to three successive rounds of k-means clustering, in order to remove all genes with a promoter negative for H3K4me3. We next conducted on this series of H3K4me3-positive promoters three successive rounds of k-means clustering, using several published datasets for H3K27me3. The genes with a promoter positive for H3K27me3 in four distinct H3K27me3 datasets (accession numbers: GSM590115, GSM590116, GSM307619 and GSM392046/GSM392047) were considered as bivalent. We eventually obtained a list of 6,481 genes with H3K4me3-only promoters, and a list of 3,411 bivalent genes.

For the generation of GC-content-based lists of promoters, we used the list of promoters defined in Fig. 3 of reference 28, that we crossed with the 14,623 promoter list, to obtain a list of 6,317 promoters rank ordered according to GC content, from least to most, shown in Extended Data Figure 7.

Pearson correlation analysis

We used DNaseI-Seq data from the Mouse ENCODE Consortium (GSM1004653) for the identification of DNase hypersensitive (DHS) regions in the mouse ES cell genome. DHS regions were defined using MACS 2.0 [49] (default setting), which resulted in the identification of 139,454 DHS regions. Each of these DHS regions was represented as a 500bp window (-250bp / +250bp) centred on the midpoint of the DHS peak. DHS regions overlapping with the blacklisted (high background signal) genomic areas (mm9) were removed, resulting in a final list of 138,582 DHS regions. Tags from each tested ChIP-seq dataset were summed up for each DHS region before pair-wise Pearson correlation comparison. The R^2 value from each pair-wise Pearson correlation was then visualized by heatmap (Fig. 1a).

Pearson correlation analysis at promoter-like DHS regions. Operating with the seqMINER platform, we retrieved, from the 138,582 DHS regions list, those positive for H3K4me3, TBP and Pol II S5ph. We obtained 16,300 promoter-like DHS regions befitting the criteria. Pair-wise Pearson correlation was performed and plotted (Fig. 1b) as described for Fig 1a.

Pearson correlation analysis at enhancer-like DHS regions. After having removed promoter-like DHS regions from the list of 138,582 DHS regions, we next retrieved those positive for Med1, Oct4, Sox2 and Nanog, following the criteria described for enhancer function in ES cells[50,51]. We obtained a list of 24,547 putative enhancer elements. Pair-wise Pearson correlation was performed and plotted (Extended data Fig.2) as described above.

Reference ES cell nucleosome map

Mouse ES nucleosomal tags were acquired from a published MNase-seq dataset[26] to make the reference map shown in Fig. 2. Reference nucleosomes were called using MACS 2.0 before assigning the first nucleosomes downstream and upstream of TSSs as +1 and -1, respectively. Regions between the associated +1 and -1 nucleosomes were defined as nucleosome free regions (NFR). ChIP-seq tags from remodeller-nucleosome interaction assays were first globally shifted as described[10]. GRO-seq tags[30] sharing the same or opposite orientation with the TSS were assigned as sense and antisense tags, respectively. The orientation of each NFR was arranged so that sense transcription proceeds to the right. ES nucleosomal tags[26], globally shifted tags from remodeller-nucleosome interaction assays (this current study), tags from DHS regions (Mouse ENCODE), GRO-seq oriented tags from transcriptionally engaged pol II and CpG islands (UCSC, mm9 build) were then aligned to the midpoint of each NFR. Promoter regions were then sorted by NFR length and visualized by JavaTree.

CpG island information was retrieved from UCSC (mm9 build) and assigned to the closest TSS by using bedtools. We noticed that promoters with large NFRs were mostly CpG island (CpGI)-rich, while those with small NFRs were globally CpGI-poor, in agreement with a previous report showing that CpGIs induce nucleosome exclusion[28].

Construction of the composite plot shown in Fig. 3

Tags from reference nucleosomes[26], remodeller-interacting nucleosomes (this study) and transcriptionally engaged pol II (GRO-seq)[30] were aligned to nucleosome (nuc) -1 and nuc +1 dyad positions. The direction of each -1/+1 dyad was assigned according to the orientation of its associated TSS, whose orientation was arranged so that the transcription proceeds to the right. After normalization with the amount of genes in the two different NFR subclasses, tags were plotted from 500 bp upstream or downstream of reference points (-1 or +1, respectively) until the midpoint of NFR.

Average binding profiles (Extended Data Fig. 3) :

Genes were rank ordered according to rpkm and divided in four quartiles (highest:Q4, second:Q3, third:Q2 and lowest:Q1). Operating with the k-means clustering function of seqMINER, genes in each quartile were further subdivided in H3K4me3-only and bivalent genes, as described above.

Using these lists of genes, tag densities from remodeller ChIP-seq datasets were collected in a window of –2kb/+2kb around the TSS, except for Chd2, for which densities were collected from the TSS until +4kb. Output tag density files were first analyzed using R software to establish average binding profiles. Statistical comparisons were performed between remodeller distributions at H3K4me3 promoters, to assess a significant increasing trend among distributions. Differences between successive pairs of quartiles (Q4 - Q3, Q3 - Q2, Q2 - Q1) were compared against a null distribution using a one side t-test.

The respective p values are reported for each remodeller:

Chd1, Q4 - Q3 p = 1.371138e-27 ; Q3 - Q2 p = 1.728126e-16 ; Q2 - Q1 p = 7.985217e-23.

Chd2, Q4 - Q3 p = 7.543473e-33 ; Q3 - Q2 p = 1.115223e-25 ; Q2 - Q1 p = 3.283427e-38.

Chd4, Q4 - Q3 p = 0.2094255 ; Q3 - Q2 p = 0.1081455 ; Q2 - Q1 p = 0.07202865.

Chd6, Q4 - Q3 p = 0.4168748 ; Q3 - Q2 p = 0.1534144 ; Q2 - Q1 p = 0.01138035.

Chd8, Q4 - Q3 p = 4.031959e-15 ; Q3 - Q2 p = 1.231527e-06 ; Q2 - Q1 p = 1.34455e-09.

Chd9, Q4 - Q3 p = 9.484578e-44 ; Q3 - Q2 p = 1.059783e-14 ; Q2 - Q1 p = 4.646352e-28.

Ep400, Q4 - Q3 p = 3.046796e-20 ; Q3 - Q2 p = 1.215304e-14 ; Q2 - Q1 p = 6.462667e-11.

Brg1, Q4 - Q3 p = 3.512021e-24 ; Q3 - Q2 p = 2.515217e-07 ; Q2 - Q1 p = 0.977422.

We concluded from this analysis that Chd1, Chd2, Chd9 and Ep400 binding at promoters is tightly linked to gene expression level. In contrast, Brg1, Chd4 and Chd6 deposition showed little correlation with gene expression level (statistical test failed for at least one comparison for these remodellers). Whilst statistical analysis of Chd8 distributions concluded to significant differences between quartiles, inspection of distributions in Extended Data Fig. 3 showed that Chd8 binding profile was intermediate between these two categories.

Statistical analysis of the differences in transcriptional activation and repression by remodelers (Fig. 4 and 5)

This analysis was performed using a 2-sample test for equality of proportions with continuity correction.

Accession numbers and references of the publicly available Datasets used in Fig. 1 and 2 and Extended Data Fig. 1, 4 and 7:

Brg1[1]: GSM359413

DNase-seq : GSM1014154

Ezh2[52]: GSM590132

GRO-seq[30]: GSM665994

H2A.Z[53,54]: GSM958501, DRP001103

H3.3[55]: GSM1386359

H3K27ac[56]: GSM594578

H3K27me3[52]: GSM590115

H3K36me3[52]: GSM590119

H3K4me3[52]: GSM590111

Med1[51] : GSM560347

Mi2b (Chd4)[29]: GSM687284

MNase-seq[26]: GSM1004653

Oct4/Pou5f1, Sox2, Nanog[57] : GSM1082340

Pol II S5ph[21]: GSM515662

TBP : GSM958503

Supplementary references

41 Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature biotechnology* **21**, 183-186, doi:10.1038/nbt780 (2003).

42 Tessarollo, L. Manipulating mouse embryonic stem cells. *Methods Mol Biol* **158**, 47-63, doi:10.1385/1-59259-220-1:47 (2001).

43 Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome research* **13**, 476-484, doi:10.1101/gr.749203 (2003).

44 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).

45 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

46 Berlivet, S., Houlard, M. & Gerard, M. Loss-of-function studies in mouse embryonic stem cells using the pHYPER shRNA plasmid vector. *Methods Mol Biol* **650**, 85-100, doi:10.1007/978-1-60761-769-3_7 (2010).

47 Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic acids research* **39**, e35, doi:10.1093/nar/gkq1287 (2011).

48 Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958, doi:10.1093/bioinformatics/btp340 (2009).

49 Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nature protocols* **7**, 1728-1740, doi:10.1038/nprot.2012.101 (2012).

50 Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117, doi:10.1016/j.cell.2008.04.043 (2008).

51 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).

52 Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604, doi:10.1016/j.cell.2012.03.026 (2012).

53 Hu, G. *et al.* H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell stem cell* **12**, 180-192, doi:10.1016/j.stem.2012.11.003 (2013).

54 Yukawa, M. *et al.* Genome-wide analysis of the chromatin composition of histone H2A and H3 variants in mouse embryonic stem cells. *PloS one* **9**, e92689, doi:10.1371/journal.pone.0092689 (2014).

55 Yildirim, O. *et al.* A system for genome-wide histone variant dynamics in ES cells reveals dynamic MacroH2A2 replacement at promoters. *PLoS genetics* **10**, e1004515, doi:10.1371/journal.pgen.1004515 (2014).

56 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

57 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

Part II:

Article 2: ATP-dependent Chromatin Remodeling Factors Target Regulatory Regions and Contribute to the Transcriptional Network in ES Cells

(Preliminary version)

Abstract

Embryonic stem (ES) cell pluripotency and self-renewal are controlled by a defined series of transcription factors that together control the ES cell transcriptional network. These factors, identified as pluripotency-associated transcription factors (pTFs) directly interact with chromatin modifying enzymes, including several members of the Snf2 family of ATP-dependent chromatin remodelers. Here, we analyzed how remodelers bind to ES cell DNA regulatory elements, and how they contribute to the regulation of the ES cell transcription network. We show that remodelers bind with specific patterns to enhancer elements, as well as to CTFC-binding sites and promoter elements. We observed that a subset of remodelers, including Chd4, Ep400, Brg1, Chd1 and Smarcad1, contribute to the regulation of genes controlled by super enhancers. We also analyzed how each remodeler contributes to the regulation of pTFs target genes. Together, this data provides a more detailed view of how remodelers regulate ES cell fate.

Introduction

Mammalian embryonic stem cells have the capacity to self-renew indefinitely, and to differentiate into almost all cell types of the body, a property known as pluripotency (Keller, 1995). Self-renewal and pluripotency are primarily controlled by a series of transcription factors, which together establish a self-perpetuating transcription regulatory network. This transcription network is controlled by a series of pluripotency-associated transcription factors (pTFs), including the core TFs Oct4 (*Pou5f1*), Sox2 and Nanog (Chambers and Smith, 2004; Niwa, 2007; Silva and Smith, 2008) . In addition, other important pTFs integrate the ES transcription network such as Esrrb, Tcfcp2l1, Klf4, Klf2, STAT3 and Smad1 (van den Berg et al., 2010a; Pardo et al., 2010). The pTFs STAT3 and Smad1 are the terminal effectors that integrate to the transcription network external stimuli from two major signaling pathways in ES cells, the leukemia inhibitory factor (LIF) and the transformation growth factor beta (TGFβ) signaling pathways (Chen et al., 2008a; Ying et al., 2003a; Zhang et al., 2010).

Pluripotency-associated TFs are DNA-binding proteins that bind to the ES cell genome at the level of promoters and distal regulatory elements to activate the transcription of genes required for ES cell phenotype, and repress differentiation-associated genes. Moreover, pTFs regulate their own promoters forming an interconnected autoregulatory loop that provides positive feedback expression for pluripotency maintenance. Several studies showed that pTFs physically interact with a large series of chromatin modifying enzymes, including several members of the Snf2 family of ATP-dependent chromatin remodeling factors (also called remodelers) (van den Berg et al., 2010b; Liang et al., 2008; Pardo et al., 2010).

Several members of the Snf2 family of remodelers were shown to be of specific importance for ES cell self-renewal and during differentiation (Fazzio et al., 2008; Gaspar-Maia et al.,

2009; Ho et al., 2011; Reynolds et al., 2012; Wang et al., 2014), indicating that this family of enzyme has a particular importance in ES cells.

The Snf2 family is composed of 24 subfamilies (Supplementary Figure 2) that share a conserved catalytic ATPase domain (Flaus et al., 2006). These factors are believed to play essential roles in modifying the chromatin landscape through their capacity to position nucleosomes and determine their occupancy throughout the genome, making the chromatin more or less accessible to DNA binding factors (Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011; Jiang and Pugh, 2009). Although several members of the Snf2 remodelers were shown to play essential function in the control of ES cell phenotype, there is so far no global, comparative view of how they work together to control ES cell fate.

Here we present a genome-wide comparative analysis of the function of a series of ATP-dependent chromatin remodelers (Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Brg1, Ep400, ATRX, Smarca3, Smarca5, Smarcad1 and Alc1) in ES cells. We first used a high performance ChIP-seq (Chromatin Immunoprecipitation followed by deep sequencing) strategy based on ES cells tagged for each factor to define the genomic binding profiles of each remodeler. Second, we performed transcriptome analysis in ES cells depleted of each remodeler to understand their role in transcription regulation. We specifically analyzed how each remodeler contributes to the transcription program controlled by pTFs. Integration of these data allowed us to better understand the function of remodelers in the transcriptional network of ES cells.

Results

Remodelers extensively bind cis-regulatory DNA elements in the ES cell genome

In order to map the genomic distribution of remodelers onto the ES cell genome, we applied a tandem ChIP-seq protocol using ES cells tagged at the gene encoding for the different remodelers, as described in article 1 (De Dieuleveult et al., 2015, in revision). To identify ES cell DNA regulatory elements, we extensively mapped DNase I hypersensitive sites (DHS) onto the ES cell genome using a publicly available ENCODE DNase-seq dataset. We identified 139,454 DHS sites across the ES cell genome, which we annotated using selected ChIP-seq datasets for pTFs and histone modifications. We were able to detect 24,357 canonical enhancer-like elements, based on high Oct4/Sox2/Nanog binding, combined with Mediator and cohesin occupancy (Chen 2008, Kagey 2010), 39,462 non-canonical enhancer-like elements (having variable Oct4/Sox2/Nanog signals and thus differing from canonical enhancers) and 20,182 CTCF binding sites, identified by the presence of bound CTCF and cohesin (Smc1) (Kagey et al., 2010). We also identified about 16,300 DHS bearing promoter-like histone modifications.

We first analyzed the distribution of remodelers at each category of distal regulatory elements. Our analysis shows that the studied remodelers all bind to canonical enhancers, though with distinct intensities (Figure 1A). The highest binding signals were observed for Chd4 and Brg1. Ep400, Chd6, Chd8 and Smarca3 were also found enriched at most canonical enhancers. Relatively lower signals were detected for Chd1, Chd9, Smarca5, Smarcad1 and ATRX. Finally, a weak signal was observed for Alc1, which could be detected at low levels at only a subset of enhancers.

Enhancers can be distinguished by their level of H3K27ac, which identifies active (high H3K27ac) and poised (low H3K27ac) enhancers. The latter category is believed to be inactive

in ES cells and corresponds to enhancers that become activated during differentiation and development (Creyghton et al., 2010). We observed that remodelers bind to both active and poised enhancers, with an about two-fold higher enrichment at active enhancers (Figure 1B). In addition to canonical enhancer elements, the ES cell genome includes a large number (39,462) of DNA elements that have most characteristics of enhancer elements, such as DNase I hypersensitivity, presence of mediator and cohesin, and pTFs. We refer to these elements as non-canonical enhancer-like, because they differ from their canonical counterparts by being bound either at low levels or by various combinations of the three OSN pTFs (see below). Interestingly, analysis of these non-canonical enhancer-like elements, revealed also binding of remodelers at these sites (Supplementary Figure 1). This is an indication that these non-canonical, enhancer-like elements might also be part of the repertoire of ES cell distal regulatory elements.

Next, we checked whether the different remodelers are present at the 20,182 identified CTCF binding sites. CTCF is an architectural protein that plays an important role in creating boundaries between topologically associating domains in chromosomes, further facilitating interactions between transcription regulatory sequences such as enhancers and promoters (Ong and Corces, 2014). Almost all remodelers were bound (though with distinct intensities) at CTCF sites, with the exception of Chd1 and Chd9 (Figure 2A). Smarca3 and Chd4 were the remodelers that presented the highest binding signal intensity. Brg1, Ep400, Chd6, Chd8, Smarca5, Smarcad1 and ATRX were also detected on CTCF-bound elements, but only a very low signal could be detected for Alc1. To detect potential difference in the binding pattern of the different remodelers, we performed a clustering analysis of the 20,182 CTCF, which revealed two broad groups of elements with specific remodeler combinations (Figure 2B). The first group (A) is characterized by a widespread enrichment for most remodelers. In contrast, the second group (B) was bound by a more restricted set of remodelers, including

Smarca3, Smarca5, Smarcad1 and ATRX. Although further investigations will be required to test the importance of remodelers in the function of CTCF binding sites, these data suggests the existence of at least two distinct functional categories of CTCF-binding sites, based on remodeler distribution pattern.

Remodelers bind super enhancer elements and regulate the transcription of super enhancer-associated genes

A small subgroup of enhancer elements can be distinguished from typical enhancer elements by an especially high level of Mediator enrichment, and by their unusual size, spanning large genomic regions (Whyte et al. 2013). Super enhancers were shown to control the expression of key genes for the control of cell phenotype, and are thus of special importance in the control of ES cell fate. We tested whether some remodelers might have a specific function in regulating the activity of these cis-acting regulatory elements.

We first analyzed the ChIP-seq binding profile of each remodeler at the level of the previously identified 231 super enhancers (Whyte et al., 2013a), in comparison with their distribution at typical enhancers (TE) (Figure 3A). We observed that the remodelers that are bound to typical enhancers generally show a similar binding intensity at super-enhancers, with some important exceptions: we found both Chd1 and Chd9 present at higher levels at super-enhancers, compared to typical enhancers. In contrast, Chd4 and Brg1, which are extensively bound to enhancer elements (Figure 1A), show a similar binding intensity at typical and super enhancer (Figure 3A). These specificities of remodeler distribution are further illustrated at the Nanog locus, which contains several super enhancers (Figure 3B).

To examine the function of remodelers in the transcriptional control of the genes associated with super enhancers, we analyzed whether the expression of these genes was effected upon

remodeler depletion by knockdown (Figure 5A). Depletion of Smarcad1, Brg1, Chd4 or Ep400 resulted in high number of deregulated genes in this category, revealing that these remodelers are important actors in the control of super enhancer activity. Chd1, which is particularly enriched at super enhancers, was also required for the proper level of expression of a significant number of genes associated with super enhancer, revealing a new function for this factor. In contrast, Chd9 depletion caused the deregulation of only one gene associated with super enhancer, despite its higher enrichment at these elements. Moderate deregulation of SE-associated genes was observed for Chd8 and Chd6. Finally, Smarca3 and Atrx knockdown had almost no effect on the expression of SE-associated genes.

To confirm the transcriptomic results regarding SE-associated genes deregulation obtained upon remodeler depletion, we performed qPCR validation experiments on a selection of SE-associated genes (Figure 5B). Downregulation of F2rl1, a SE-associated gene was observed upon Chd1 KD. Remarkably, depletion of Chd4 caused the downregulation of the two pTFs Oct4 (*Pou5f1*) and Nanog, revealing a function of Chd4 in controlling the expression of major pTFs. Nanog expression was also downregulated upon depletion of Chd8.

Altogether, this analysis suggests that remodelers play an essential function in the maintenance of ES cell fate through the regulation of super enhancer function.

Differential regulation of promoter types by remodelers

In our previous study (article 1), we analyzed the binding pattern of Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400 and Brg1 in the promoter regions. We concluded that each of these remodeler bind specific nucleosome positions at promoters, except Chd2 which bind nucleosomes in gene bodies, suggesting a function in transcription elongation.

The additional remodelers Atrx, Smarcad1, Smarca5, Smarca3 and Alc1 were generally present at lower intensities at promoters compared to remodelers in the first series, with Smarca3 and Smarca5 having the highest binding signal (Figure 4A). Alc1 seems to be present at particularly low levels at promoters.

To understand the link between remodeler binding at promoters and functional control of gene expression, we analyzed the transcriptome of ES cells depleted of each remodeler.

The transcriptome of ES cells depleted of Chd1, Chd4, Chd6, Chd8, Chd9, Ep400 and Brg1 is described in article 1. As we observed in article 1, remodelers differentially affect the regulation of genes having H3K4me3-only versus bivalent promoters, we performed a similar analysis to analyze the the consequence of depleting Atrx, Smarca3, Smarca5 and Smarcad1. Among these four remodelers, Smarcad1 was the most involved in transcriptional control (Figure 4B). We observed that Smarcad1 acts both as an activator and a repressor of H3K4me3-only genes, but that it is mostly a repressor of bivalent genes. Smarca3 and Smarca5 had a comparatively lower impact on the transcriptome, with a number of genes deregulated in the same order of magnitude as Chd1. Finally, Atrx depletion caused the deregulation of very few genes, showing that this remodeler has only a minor function in ES cells.

Analysis of remodeler function in the control of ES cell transcriptional network

In order to understand how remodelers contribute to the regulation of ES cell transcriptional circuitry, we analyzed how each remodeler positively or negatively regulates the target genes of Oct4, Sox2 and Nanog (OSN), using a dataset published by Ivanova et al in 2006 (Ivanova et al., 2006b). This analysis revealed that remodelers have each a distinct contribution to the ES cell transcriptional circuitry. Ep400 and Chd4 positively regulate the expression of 17%

and 10% of OSN-activated genes, respectively, and thus contribute to a large extend to the regulation of the ES cell transcription circuitry. To a lesser extent Smarcad1 and Chd1 (Figure 6) also acted as co-activators of OSN at their target genes, as depletion of these remodelers caused the downregulation of about 4% of OSN-activated target genes. Similarly, Brg1 seems to play mostly a co-activator function at OSN-target genes (8% of these genes are activated by both Brg1 and OSN), but has also a co-repressive function at (4% of OSN-target genes are repressed by both OSN and Brg1). Chd8, Chd6, Smarca5, Smarca3, regulate only few genes controlled by OSN, and thus are expected to play a minor function in the regulation of ES cell transcriptional circuitry. Finally, Atrx and Chd9 do not contribute to the ES cell pTFs-controlled transcriptional circuitry.

Conclusion

In this study we conducted a comparative analysis of how various remodelers belonging to the Snf2 family of ATP-dependent chromatin remodeling factors contribute to the transcriptional control in ES cells. We show that a considerable number of remodelers bind proximal and distal regulatory elements at variable intensities. Moreover, we show that a number of remodelers interfere directly in the transcriptional regulation the ES self-renewal and pluripotency network and are essential for ES cell phenotype conservation.

The implication of chromatin remodeling factors in the control of the ES cell state has been studied before where remodelers were shown to be a part of the pluripotency core network (van den Berg et al., 2010; Liang et al., 2008; Pardo et al., 2010) . Other studies demonstrated the importance of certain remodelers such as Ep400, Chd1 and Brg1 in the conservation of the ES cell state (Fazzio et al., 2008; Gaspar-Maia et al., 2009; Ho et al., 2011). In our study we try to attribute to a global understanding of the role of large set of Snf2 remodelers in the

control of the ES cell state. Previously we have demonstrated how remodelers integrate promoter nucleosomal architecture to regulate ES cell transcription programs.

We have shown that remodelers bind canonical enhancer elements characterized by the abundant presence of the core transcription regulatory network; in addition to non-canoncial enhancer-like elements that might increase ES cell regulatory elements repertoire. Ep400, Brg1, Chd4, Chd1 and Smarcad1 were shown to be of special importance in the regulation of SE-associated genes, genes typically important in the ES cell state maintenance. Moreover, we demonstrate that remodelers, in particular Smarca3 present significant binding patterns along with architectural proteins such as CTCF (Ong and Corces, 2014) conferring a probable role of remodelers in the definition of chromatin boundaries and in the assuring of long-range promoter-enhancer interactions. Indeed, remodelers were also detected on promoter elements of a large number of genes, where they regulate in various ways the expression of active and bivalent promoters of genes. For instance, Ep400, Chd4 and Smarcad1 seem to exert a rather positive function on active genes and negatively control the expression of bivalent genes. On the contrary, Brg1 rather activates the expression of bivalent promoter-associated genes. Furthermore, we show that remodelers especially Ep400, Brg1, Chd4, Chd1 and Smarcad1 contribute to the ES pluripotency and self-renewal network.

To sum up, in the current study we conducted a genome-wide analysis of the distribution and function in ES cells of the different chromatin remodeling factors belonging to the Snf2 family of remodelers. More concentration should be shed on how such various yet related remodelers contribute synergistically or antagonistically to the control of the ES cell state.

Materials and Methods

Cell line and ES cell culture

Mouse 46C ES cells have been described previously (Ying et al., 2003b) . 46C ES cells and their tagged derivatives were cultured at 37°C, 5% CO2, on mitomycin C-inactivated mouse embryonic fibroblasts, in DMEM (Sigma) with 15% foetal bovine serum (Invitrogen), L-Glutamine (Invitrogen), MEM non-essential amino acids (Invitrogen), pen/strep (Invitrogen), 2-mercaptoethanol (Sigma), and a saturating amount of leukemia inhibitory factor (LIF), as described in reference (Southon and Tessarollo, 2009).

Knock-in of a TAP-tag in the genes encoding the remodelers through homologous recombination in ES cells

The recombineering technique (Liu et al., 2003) was adapted to construct all targeting vectors for homologous recombination in ES cells previously described in article 1 (De Dieuleveult et al., 2015). Briefly, a TAP-tag was inserted into the subcloned DNA, immediately 3' to the coding sequence. The TAP-tag was (FLAG)3-TEV-HA for Chd1, Chd2, Chd4, Chd6, Chd8, Ep400, Brg1, Smarca3, Smarcad1, Smarca5, Atrx and Alc1 and 6His-FLAG-HA for Chd9. Cell line verification and recombineering success are also described in article 1.

Antibodies

Antibodies used for western blot were as follows:

Monoclonal antibodies anti-HA (H7, Sigma H3663) and anti-FLAG (M2, Sigma F1804), anti-GAPDH (Abcam ab9485).

Tandem affinity purification of the remodeler-nucleosome complexes

As described previously (De Dieuleveult et al., 2015) ES cells were fixed either with formaldehyde, or with a combination of disuccinimidyl glutarate (DSG) and formaldehyde, then permeabilized with IGEPAL, and incubated with micrococcal nuclease (MNase) in order to fragment the genome into mononucleosomes. This nucleosome preparation was next incubated with agarose beads coupled with an antibody anti-HA or anti-FLAG. Anti-HA-agarose (ref. A2095) and anti-FLAG-agarose (ref. A2220) beads were purchased from Sigma. After a series of washes, tagged-remodeler-nucleosome complexes were eluted, either by TEV protease cleavage or by peptide competition. The eluted complexes were then subjected to a second immunopurification step, using beads coupled to the antibody specific of the second HA or FLAG epitope. After elution, DNA was extracted from the highly purified mononucleosome fraction, and processed for high-throughput sequencing. As a negative control, chromatin from untagged ES cells was subjected to the same protocol to define background signal. A detailed version of this protocol is available on the protocol exchange website: http://dx.doi.org/10.1038/protex.2014.040

High-throughput sequencing of tandem ChIP samples

After crosslink reversion, phenol-chloroform extraction and ethanol precipitation, the DNA from remodeller-nucleosome complexes was quantified using the picogreen method (Invitrogen) or by running 1/20e of the ChIP material on a High sensitivity DNA chip on a 2100 Bioanalyzer (Agilent, USA). 5 to 10 ng of ChIP DNA were used for library preparation according to the Illumina ChIP-seq protocol (ChIP-seq sample preparation kit). Following end-repair and adapter ligation, fragments were size-selected on an agarose gel in order to purify genomic DNA fragments between 140 and 180 bp. Purified fragments were next amplified (18 cycles) and verified on a 2100 Bioanalyzer before clustering and single-read sequencing on a Genome Analyzer (GA) or GA II (Illumina), according to manufacturer's instructions.

RNA preparation from ES cells depleted of each remodeler by shRNA

We used the pHYPER shRNA vector for remodeller depletion in ES cells, as previously described (Berlivet et al., 2010). shRNA design was performed using DESIR software (http://biodev.extra.cea.fr/DSIR/DSIR.html). Below are listed the shRNA selected for each remodeller.

Smarca3 shRNA 1: CGCACGAGAGAACAGTAAA

Smarca3 shRNA 2: AGCAGGATCTCCTACGATA

Smarca5 shRNA 1: GGATTCGATAGTAATTCAA

Smarca5 shRNA 2: CCTCCTTCGTCGAATTAAA

Smarcad1 shRNA 1: GGACTATAGCAGTTGTGAA

Smarcad1 shRNA 2: GACGTAGTTATAAGACTTA

Atrx shRNA 1: CGATGTATTGACAAAGCAA

Atrx shRNA 2: GATGCTAGATCATCAGTAA

Alc1 shRNA 1: GAAGGTAGAGACTATTCTA

Alc1 shRNA 2: CGAATTGGACATGCTACAA

Each shRNA was transfected in its corresponding tagged ES cell line, in order to follow remodeler depletion by Western blotting using antibodies against Flag or HA epitopes (Supplementary Figure 3).

The pHYPER shRNA vectors were transfected in ES cell by electroporation, using an Amaxa nucleofector (Lonza). 24 h after transfection, puromycin (2 μg/ml) selection was applied for an additional 48 h period, before cell collection and RNA preparation, except for Smarcad1

and Alc1, for which cells were collected after 32 h of selection. Total RNA was extracted using an RNeasy Kit (Qiagen). Total RNA yield was determined using a NanoDrop ND-100 (Labtech). Total RNA profiles were recorded using a Bioanalyzer 2100 (Agilent). For each remodeler, RNA was prepared from three independent transfection experiments, and processed for transcriptome analysis.

Transcriptome analysis in remodeler depleted ES cells

cRNA was synthesized, amplified, and purified using the Illumina TotalPrep RNA Amplification Kit (Life Technologies) following Manufacturer's instructions. Briefly, 200 ng of RNA were used to prepare double-stranded cDNA using a T7 oligo (dT) primer. Second strand synthesis was followed by *in vitro* transcription in the presence of biotinylated nucleotides. cRNA samples were hybridized to the Illumina BeadChips Mouse WG-6v2.0 arrays. These BeadChips contain 45,281 unique 50-mer oligonucleotides in total, with hybridization to each probe assessed at 30 different beads on average. 26,822 probes (59%) are targeted at RefSeq transcripts and the remaining 18,459 (41 %) are for other transcripts. BeadChips were scanned on the Illumina iScan scanner using Illumina BeadScan image data acquisition software (version 2.3). Data were then normalized using the 'normalize quantiles' function in the GenomeStudio Software. Following analyses were done using Genespring software.

To identify genes with a log2-ratios significantly different between the mutant and wild- type strain, p-values were calculated for each gene using a moderated t-test. The moderated t-test applied here was based on an empirical Bayes analysis and was equivalent to shrinkage (or expansion) of the estimated sample variances towards a pooled estimate, resulting in a more stable inference. Finally, adjusted p-values were calculated using the Benjamini & Hochberg method and filtered using the thresholds of 0.05 for the p-value and 1.5 for the fold-change.

Quantitative RT-PCR

Random-primed reverse transcription was performed at 52°C in 20μl using Maxima First strand cDNA synthesis kit (Thermo Scientific) with 10 μg of total RNA isolated from ES cells (Qiagen), quantified with NanoDrop instrument (Thermo Scientific). Reverse transcription products were diluted 40-fold before use. Composition of quantitative PCR assay included 2.5 μl of the diluted RT reaction, 0.2 to 0.5 mM forward and reverse primers, and 1X Maxima SYBR Green qPCR Master Mix (Thermo Scientific). Reactions were performed in a 10μl total volume. Amplification was performed as follows: 2 min at 95°C, 40 cycles at 95°C for 15 sec and 60°C for 60 sec in the ABI/Prism 7900HT real-time PCR machine (Applied Biosystems). The real-time fluorescent data from quantitative PCR were analyzed with the Sequence Detection System 2.3 (Applied Biosystems). Each quantitative real-time PCR was performed using the set of primer pairs validated for their specificity and efficiency of amplification. All reactions were performed in triplicates, using RNA prepared from three independent cell transfection experiments. Control reactions without enzyme were verified to be negative. Relative expression was calculated after normalization with three reference genes (*Actb*, *Nmt1* and *Ddb1*), validated for this study.

Analysis of ChIP-seq datasets

Lists of genes

The list of 14,623 genes used in Fig. 1 and 2 was obtained by filtering all mm9 RefSeq genes. We removed redundancies (that is, genes having the same start and end sites), unmappable genes and those with high ChIP-seq background, as well as genes shorter than 2 kb. The purpose of this last filtering step was to unambiguously distinguish the promoter region from the end of the genes in heat-maps. Transcription start sites (TSS) correspond to the location of the 5' RNA end.

Lists of genes with H3K4me3 and bivalent promoters: We first defined, among the 14,623 RefSeq genes, those with a promoter positive for H3K4me3 (accession number: GSM590111). Operating with the seqMINER platform, tag densities from this dataset were collected in a -500/+1,000 bp window around the TSS, and subjected to three successive rounds of k-means clustering, in order to remove all genes with a promoter negative for H3K4me3. We next conducted on this series of H3K4me3-positive promoters three successive rounds of k-means clustering, using several published datasets for H3K27me3. The genes with a promoter positive for H3K27me3 in four distinct H3K27me3 datasets (accession numbers: GSM590115, GSM590116, GSM307619 and GSM392046/GSM392047) were considered as bivalent. We eventually obtained a list of 6,481 genes with H3K4me3-only promoters, and a list of 3,411 bivalent genes.

List of enhancer-like elements and CTCF-binding sites from DHS data

We used DNaseI-Seq data from the Mouse ENCODE Consortium (GSM1004653) for the identification of DNase hypersensitive (DHS) regions in the mouse ES cell genome. DHS regions were defined using MACS 2.0 (Feng et al., 2012) (default setting), which resulted in the identification of 139,454 DHS regions. Each of these DHS regions was represented as a 500bp window (-250bp / +250bp) centred on the midpoint of the DHS peak. DHS regions overlapping with the blacklisted (high background signal) genomic areas (mm9) were removed, resulting in a final list of 138,582 DHS regions.

Enhancer-like DHS regions

After having removed promoter-like DHS regions from the list of 138,582 DHS regions, we next retrieved those highly positive for Med1, Oct4, Sox2 and Nanog, following the criteria described for enhancer function in ES cells (Chen et al., 2008b; Kagey et al., 2010) . We obtained a list of 24,547 canonical putative enhancer elements, further repartitioned according

to H3K27ac to active and poised enhancer-like elements. Further clustering analysis allowed us to identify an additional number of potential enhancer-like elements (non-canonical: various presence of Oct4/Sox2/Nanog) that we divided into six categories (Supplementary figure 1).

CTCF-bound DHS regions

After having removed promoter-like DHS regions and the enhancer-like DHS regions from the list of 138,582 DHS regions, we identified CTCF-binding sites also bound by the cohesin Smc1 (Kagey et al., 2010). Clustering analysis helped us in identifying two groups of CTC-binding site with distinct remodeler binding composition.

Transcriptomic data correlation with pluripotency gene networks and SE-associated genes

In order to evaluate the degree of involvement of each remodeler in the core pluripotency network in ES cells, data from (Ivanova et al., 2006) pattern 2 of the 237 genes downregulated upon core transcription factor depletion were compared with transcriptomic data of deregulated genes upon remodelling factor depletion.

The same analysis was done to reveal SE-associated genes that are deregulated upon remodeler factor depletion by using the data of the 231 identified SE-associated genes identified in (Whyte et al., 2013b).

Figure legends

Figure 1

Remodelers' distribution at ES cell enhancer elements

A) Heatmap representing ChIP-seq binding of remodelers at 24,357 DHS canonical enhancer-like elements ranked from highest to lowest H3K27ac presence. Color intensity represents sequencing tag counts. For each enhancer-like element, remodeler occupancy is indicated within a 6kb window centered on the DHS. **B)** ChIP-seq binding profiles (mean density) of the remodelers (series 1 and series 2) at active and poised enhancers based on H3K27ac presence within a 6kb window centered on DHS.

Figure 2

Remodelers' distribution at ES cell CTCF-binding sites

A) Heatmap representing ChIP-seq binding of remodelers at 20,182 DHS CTCF-binding sites based on CTCF and Smc1 presence. Color intensity represents sequencing tag counts. For each CTCF-binding element, remodeler occupancy is indicated within a 3kb window centered on the DHS. Clustering analysis reveals the presence of two distict groups of CTCF-binding sites with different remodeler-binding profiles. **B)** ChIP-seq binding profiles (mean density) of the remodelers (series1 and series 2) at group A and B at CTCF binding sites within a 3kb window centered on DHS.

Figure 3

Differential remodeling factor binding on typical enhancer (TE) and super enhancer (SE)

A) Comparative ChIP-seq binding profiles (mean density) of each remodeler at TE within a 6kb window and at SE within a 30kb window. ChIP-seq binding profiles of pluripotency transcription factors (Oct4/Sox2/Nanog) and Mediator subunit (Med1) are also illustrated. Signals at TE appear more precise and centered, on the contrary to signals at SE which disperse on a larger genomic scale. **B**) Binding profiles at a representative locus. Scores indicate reads per 10 million. Grey rectangles represent super-enhancers.

Figure 4

Remodelers bind promoter elements and regulate differentially H3K4me3-only and bivalent promoters

A) Heatmap representing ChIP-seq binding of remodelers at 14,623 RefSeq promoters. Color intensity represents sequencing tag counts. For each promoter, remodeler occupancy is indicated within a 10kb window centered on the TSS. **B)** The transcriptome of ES cells depleted for each remodeler was analyzed by microarray hybridization. The number of upregulated (light grey) and down regulated (dark grey) genes was defined for gene with H3K4me3-only promoters (upper) and bivalent promoters (lower). A 1.5 fold-change threshold was used. Statistical analysis was performed using a 2-sample test for equality of proportions with continuity correction. *$p<0.05$, **$p<0.01$, ***$p<0.001$

Figure 5

Transcriptional regulation of SE-associated genes by remodelers

A) The transcriptome of ES cells depleted for each remodeler was compared to a set of SE-associated genes. The numbers of upregulated (light grey) and down regulated (dark grey) SE-associated genes are shown. **B**) qPCR validation analysis of the downregulation SE-associated genes upon the depletion of remodelers.

Figure 6

Remodelers contribute to the ES cell transcription network

The transcriptome of ES cells depleted for each remodeler was compared to Ivanonva et al. 2006 published transcriptome data in ES cells depleted for the core transcription factors (Oct4/Sox2/Nanog). The numbers of upregulated (light grey) and down regulated (dark grey) genes upon each remodeler depletion that are downregulated by Oct4/Sox2/Nanog depletion are shown.

Supplementary Figure legends

Figure 1S

Binding of remodelers at non-canonical enhancer-like elements

Table showing the different non-canonical enhancer-like categories. ChIP-seq binding profiles (mean density) of each remodeler (from series 1 and 2) at the different categories of non-canonical enhancer-like elements within a 6kb window. Remodelers are presented by different colors for each series.

Figure 2S

Snf2-dependent chromatin remodeling factors

A) A tree illustration representing the different subfamilies that compose the Snf2 family of chromatin remodeling enzymes. **B**) The studied Snf2 ATP-dependent chromatin remodeling factors and their composing domains.

Figure 3S

Western blot analysis of depleted ES cell for each remodeler

Western blot validation analysis of the efficiency of the used shRNA in the depletion of the series 2 of remodelers. Anti-flag antibody was used to detect each remodeler (tagged ES cells used) and anti-GAPDH was used for normalization.

Figure 1

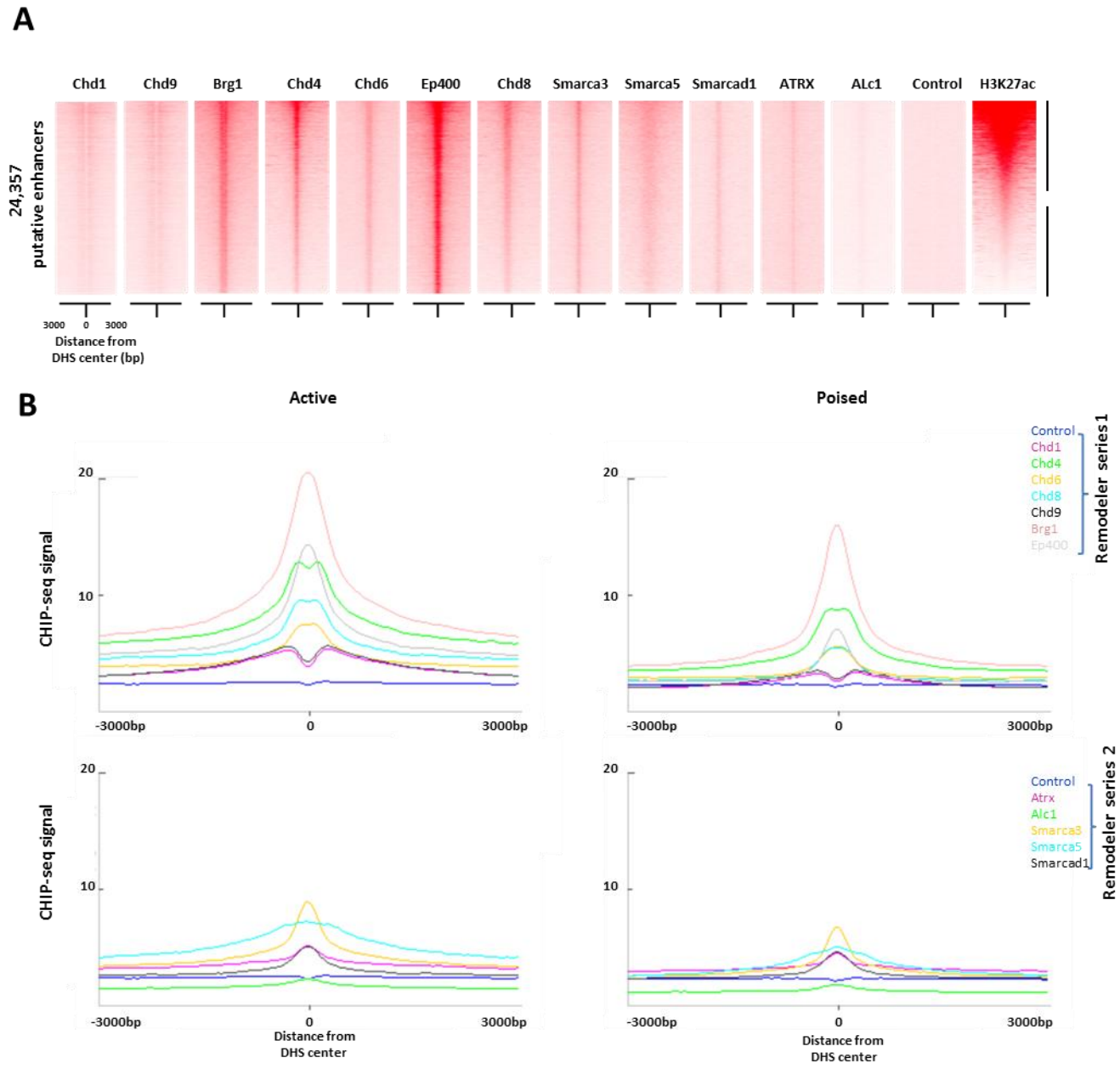

A
Chd1
Chd9
Brg1
Chd4
Chd6
Ep400
Chd8
Smarca3
Smarca5
Smarcad1
ATRX
ALc1
Control
H3K27ac
24,357 putative enhancers
3000 0 3000
Distance from DHS center (bp)
B
Active
Poised
CHIP-seq signal
20
10
-3000bp
0
3000bp
Control
Chd1
Chd4
Chd6
Chd8
Chd9
Brg1
Ep400
Remodeler series 1
Control
Atrx
Alc1
Smarca3
Smarca5
Smarcad1
Remodeler series 2
Distance from DHS center

Figure 2

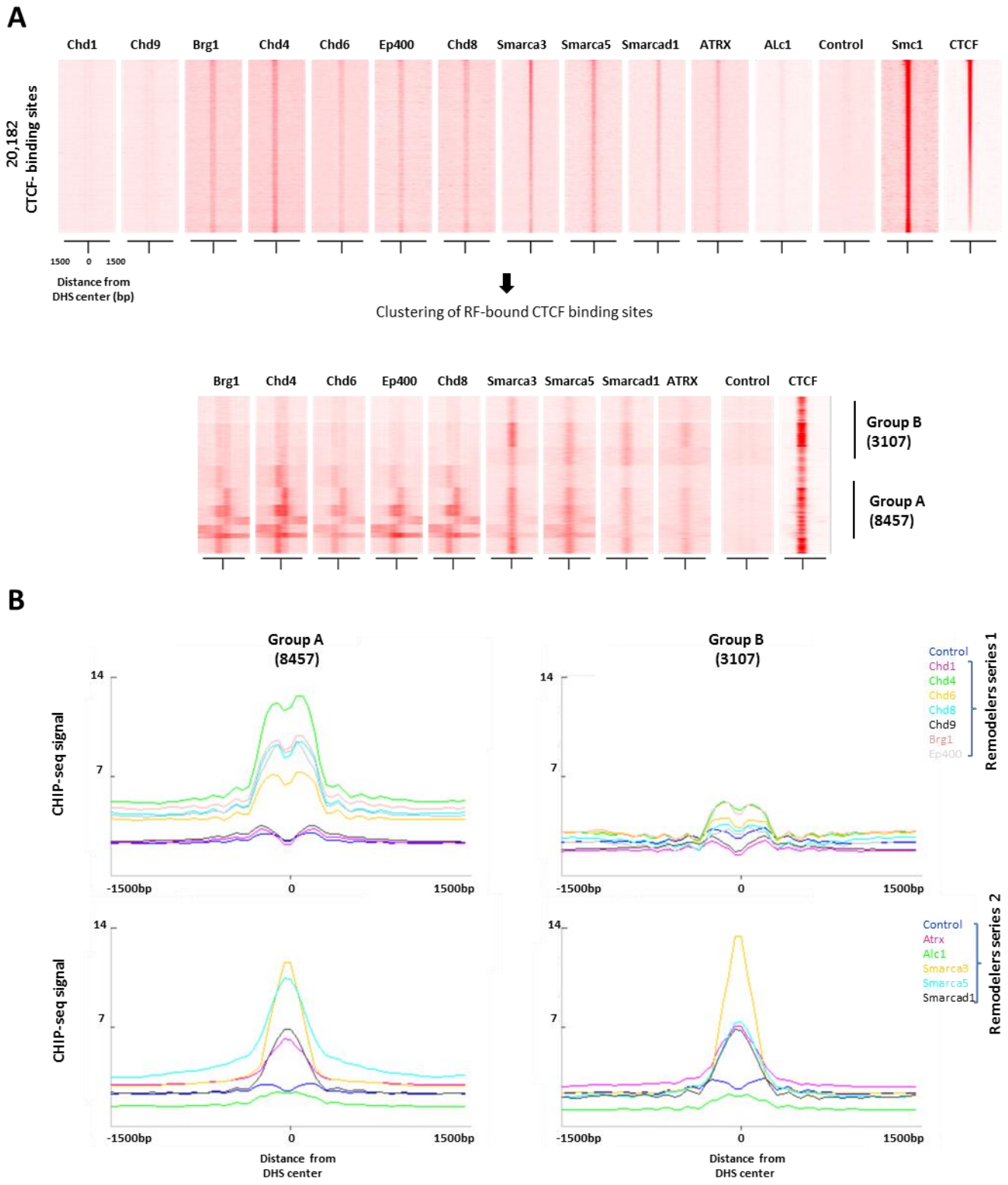

A
Chd1
Chd9
Brg1
Chd4
Chd6
Ep400
Chd8
Smarca3
Smarca5
Smarcad1
ATRX
ALc1
Control
Smc1
CTCF
20,182
CTCF- binding sites
1500
0
1500
Distance from
DHS center (bp)
Clustering of RF-bound CTCF binding sites
Brg1
Chd4
Chd6
Ep400
Chd8
Smarca3
Smarca5
Smarcad1
ATRX
Control
CTCF
Group B
(3107)
Group A
(8457)
B
Group A
(8457)
Group B
(3107)
CHIP-seq signal
14
7
-1500bp
0
1500bp
Control
Chd1
Chd4
Chd6
Chd8
Chd9
Brg1
Ep400
Remodelers series 1
Control
Atrx
Alc1
Smarca3
Smarca5
Smarcad1
Remodelers series 2
Distance from
DHS center

Figure 3

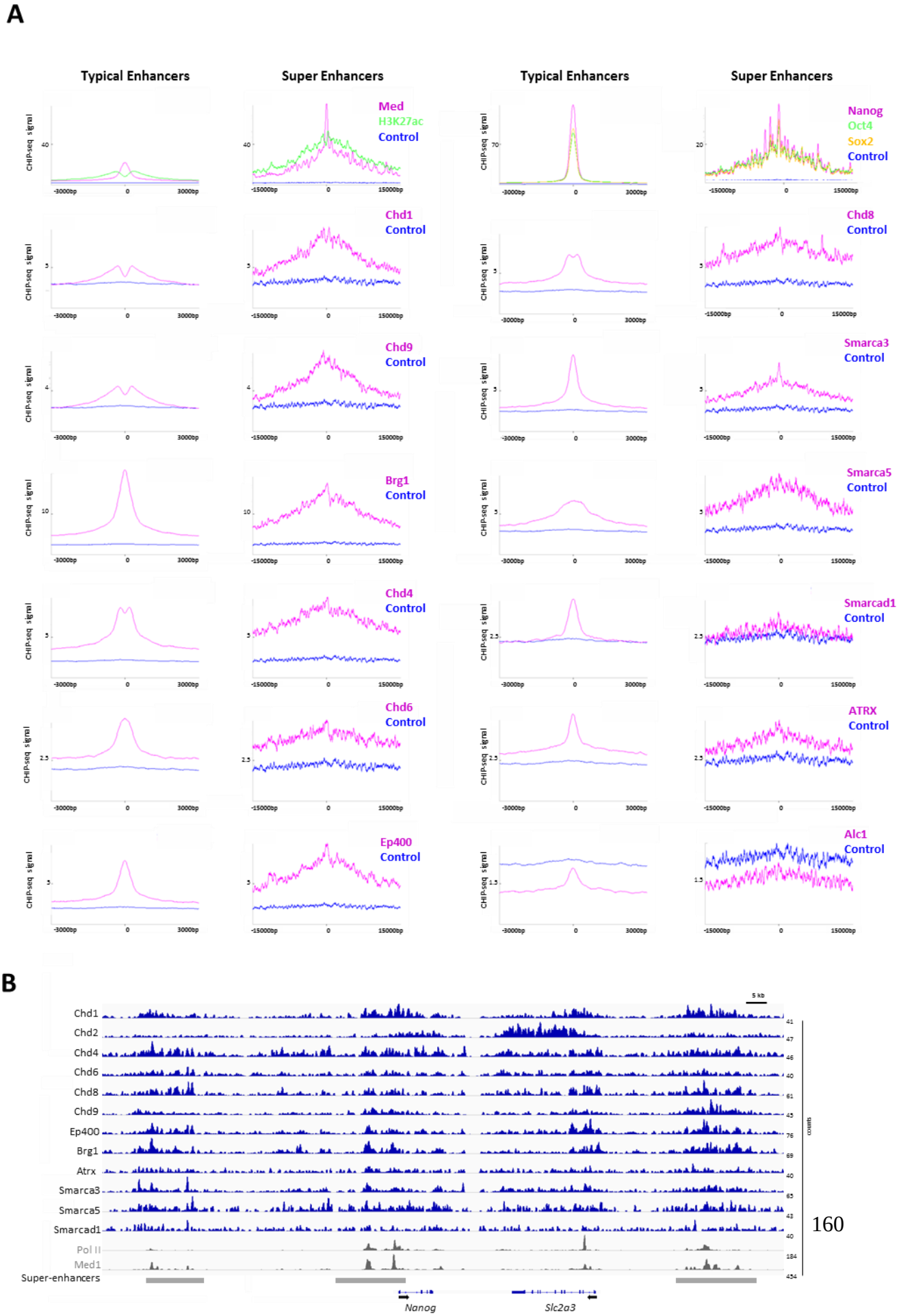

A
Typical Enhancers
Super Enhancers
Typical Enhancers
Super Enhancers
CHIP-seq signal
Med
H3K27ac
Control
Nanog
Oct4
Sox2
Control
Chd1
Control
Chd8
Control
Chd9
Control
Smarca3
Control
Brg1
Control
Smarca5
Control
Chd4
Control
Smarcad1
Control
Chd6
Control
ATRX
Control
Ep400
Control
Alc1
Control
-3000bp
0
3000bp
-15000bp
15000bp
B
5 kb
Chd1
Chd2
Chd4
Chd6
Chd8
Chd9
Ep400
Brg1
Atrx
Smarca3
Smarca5
Smarcad1
Pol II
Med1
Super-enhancers
counts
Nanog
Slc2a3

Figure 4

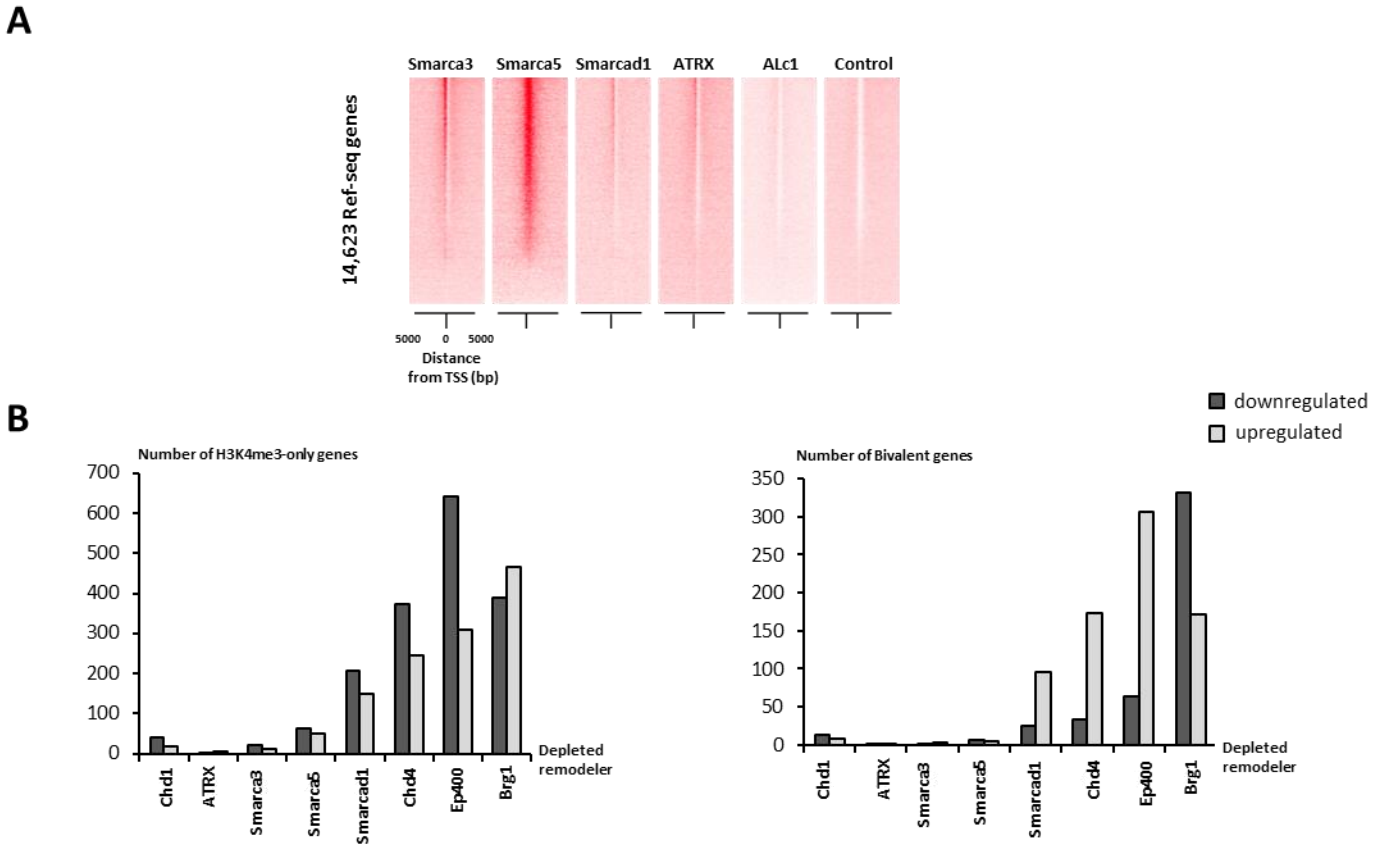

A
Smarca3
Smarca5
Smarcad1
ATRX
ALc1
Control
14,623 Ref-seq genes
5000
0
5000
Distance
from TSS (bp)
B
downregulated
upregulated
Number of H3K4me3-only genes
700
600
500
400
300
200
100
0
Chd1
ATRX
Smarca3
Smarca5
Smarcad1
Chd4
Ep400
Brg1
Depleted
remodeler
Number of Bivalent genes
350
300
250
200
150
100
50
0
Chd1
ATRX
Smarca3
Smarca5
Smarcad1
Chd4
Ep400
Brg1
Depleted
remodeler

Figure 5

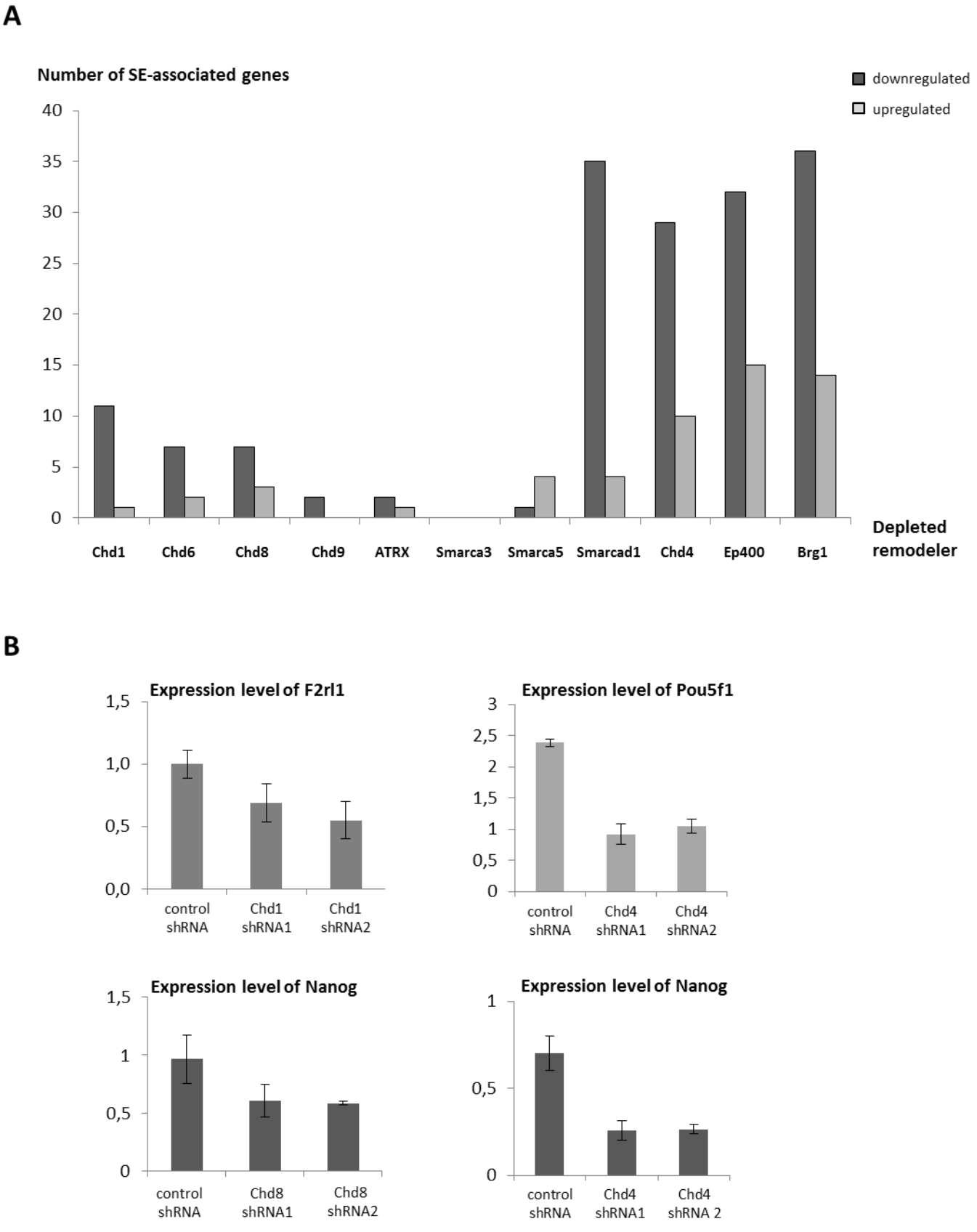

A
Number of SE-associated genes
downregulated
upregulated
40
35
30
25
20
15
10
5
0
Chd1
Chd6
Chd8
Chd9
ATRX
Smarca3
Smarca5
Smarcad1
Chd4
Ep400
Brg1
Depleted remodeler
B
Expression level of F2rl1
1,5
1,0
0,5
0,0
control shRNA
Chd1 shRNA1
Chd1 shRNA2
Expression level of Pou5f1
3
2,5
2
1,5
1
0,5
0
control shRNA
Chd4 shRNA1
Chd4 shRNA2
Expression level of Nanog
1,5
1
0,5
0
control shRNA
Chd8 shRNA1
Chd8 shRNA2
Expression level of Nanog
1
0,5
0
control shRNA
Chd4 shRNA1
Chd4 shRNA 2

Figure 6

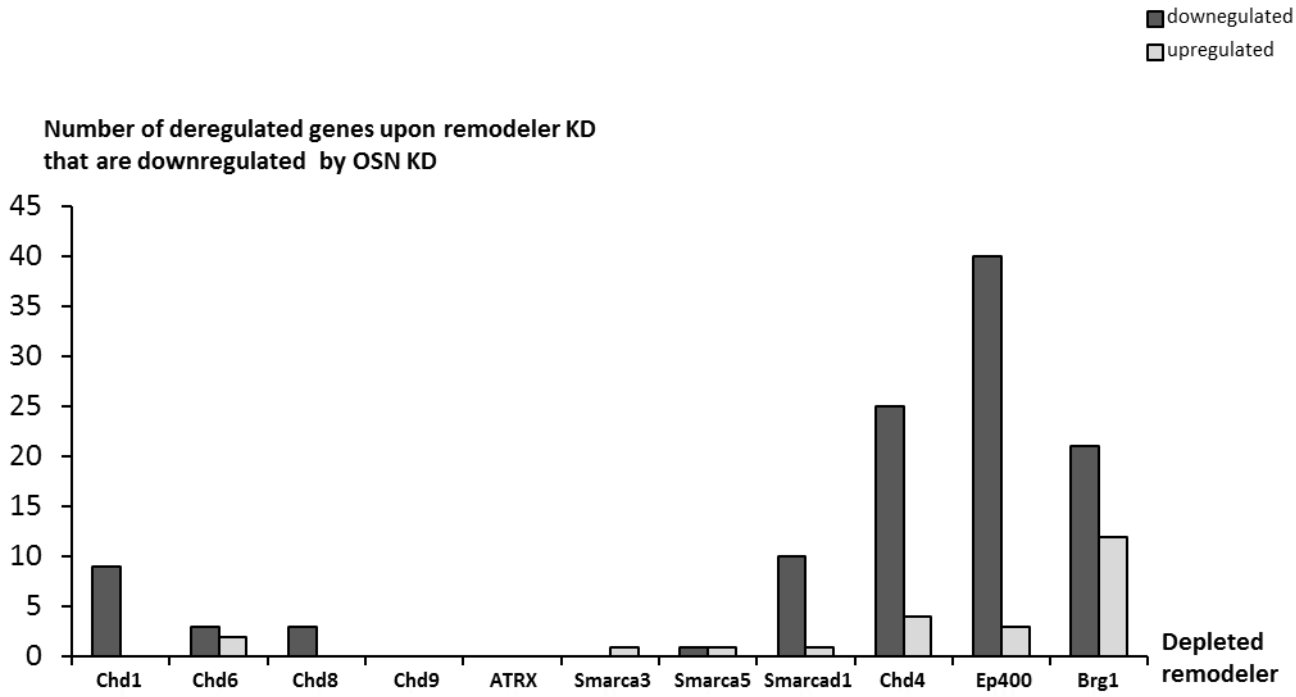
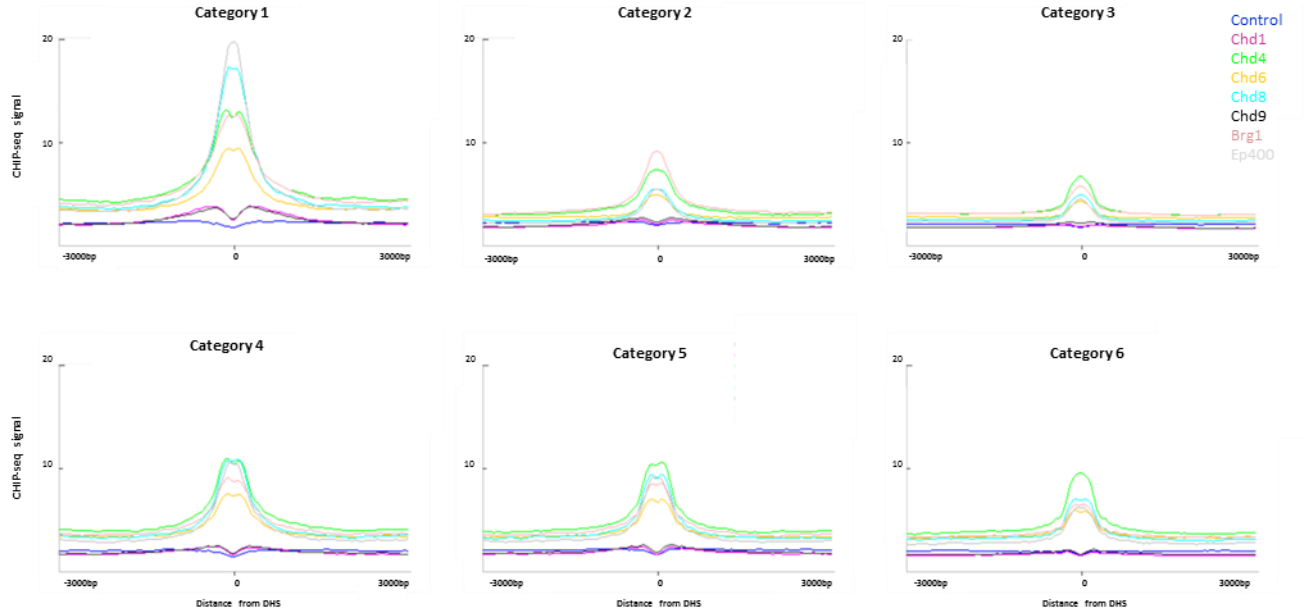

downegulated
upregulated
Number of deregulated genes upon remodeler KD
that are downregulated by OSN KD
45
40
35
30
25
20
15
10
5
0
Chd1
Chd6
Chd8
Chd9
ATRX
Smarca3
Smarca5
Smarcad1
Chd4
Ep400
Brg1
Depleted
remodeler

Figure 1S

Non-canonical enhancer-like categories	Number	H3K27ac
1- OSN moderate	2158	+++
2- OSN moderate/ Nanog dominant	5150	+/-
3- OSN weak	21744	+/-
4- ON moderate/ Sox2 negative	2530	++
5- OS moderate/ Nanong negative	2751	++
6- Oct4 weak / SN negative	4276	+

RF series 1



RF series 2

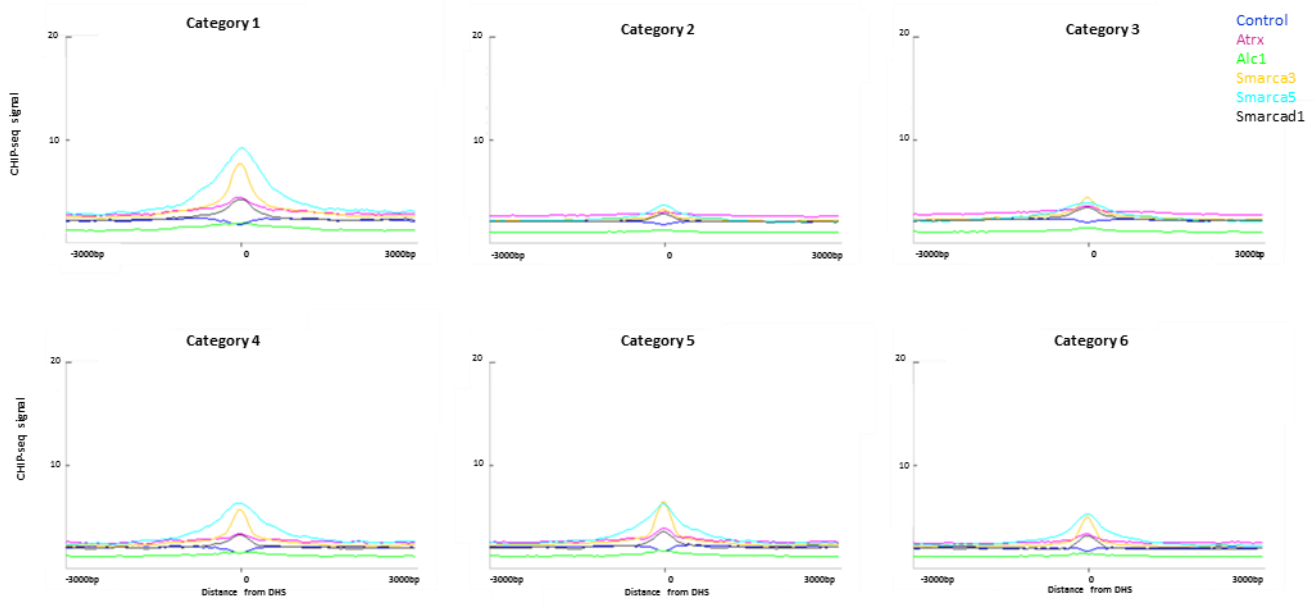


Figure 2S

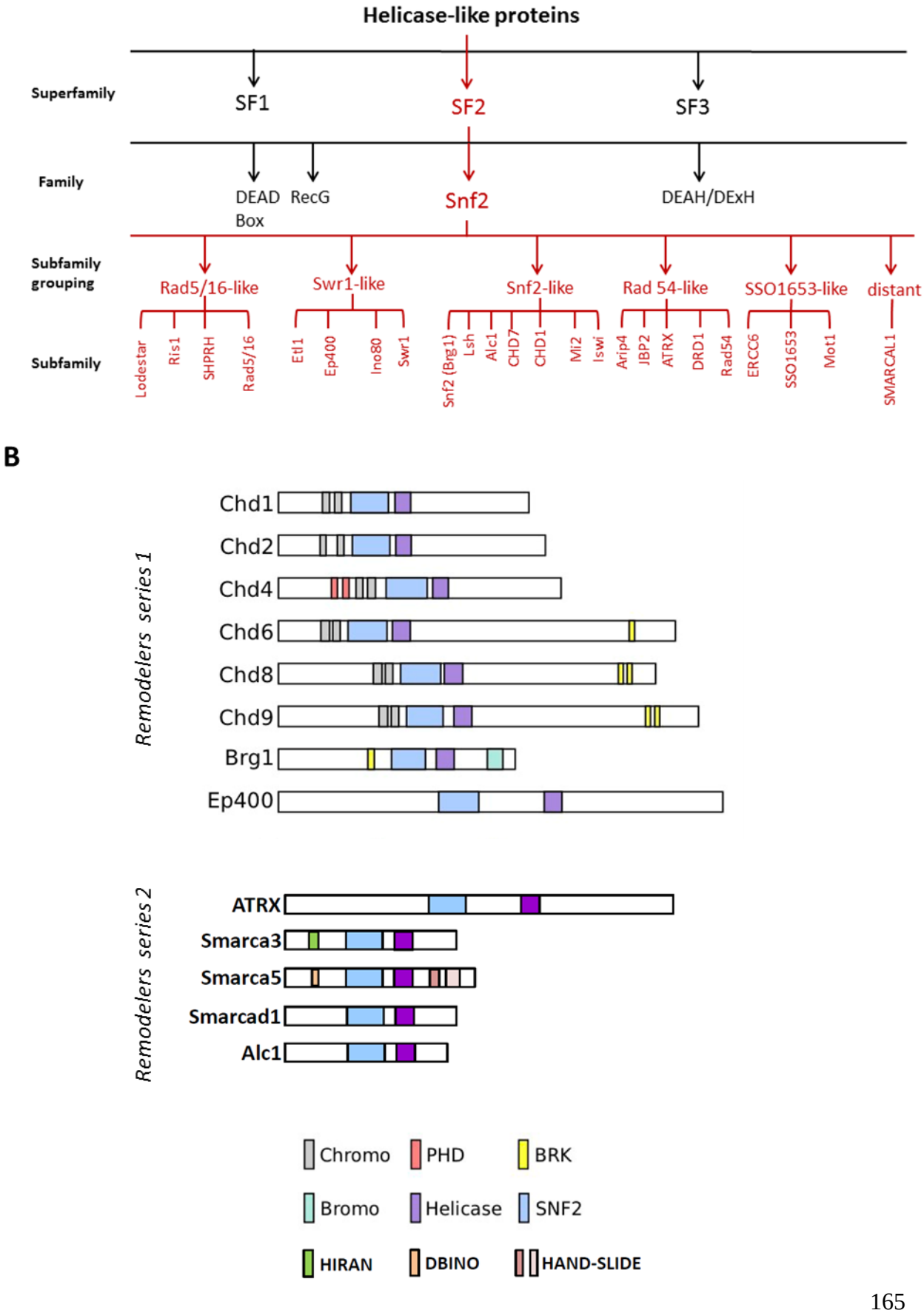
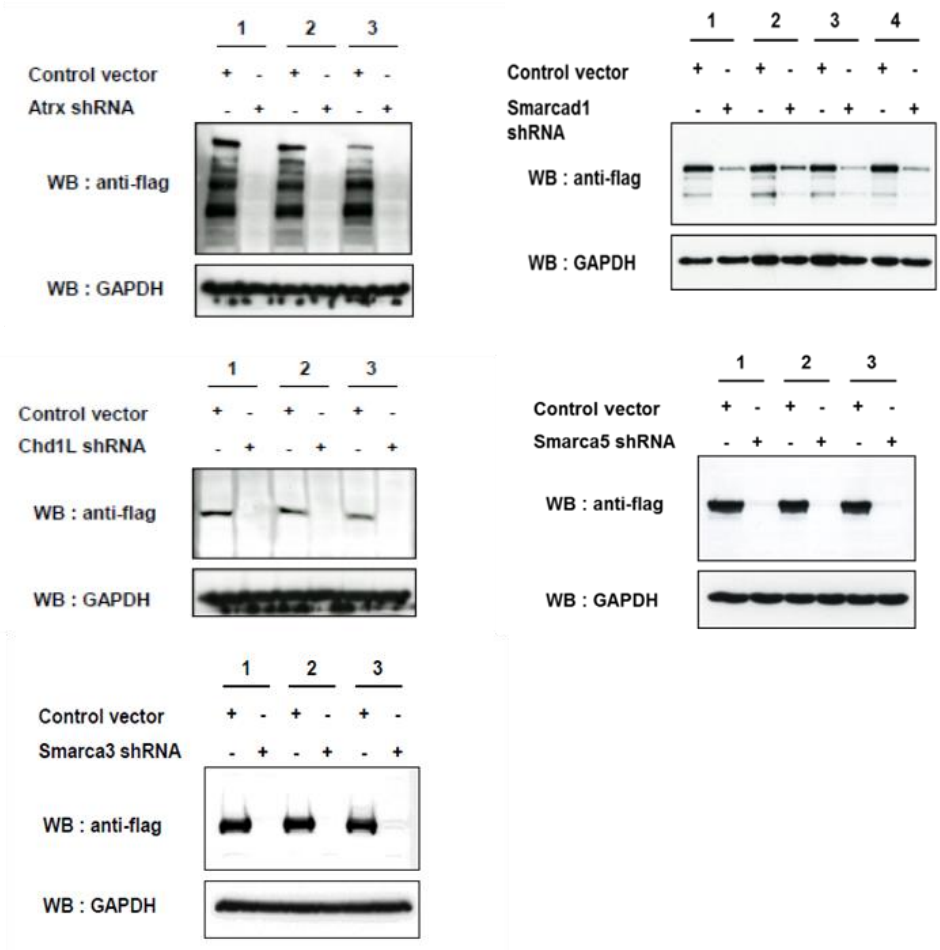

A
Helicase-like proteins
Superfamily
SF1
SF2
SF3
Family
DEAD Box
RecG
Snf2
DEAH/DExH
Subfamily grouping
Rad5/16-like
Swr1-like
Snf2-like
Rad 54-like
SSO1653-like
distant
Subfamily
Lodestar
Ris1
SHPRH
Rad5/16
Etl1
Ep400
Ino80
Swr1
Snf2 (Brg1)
Lsh
Alc1
CHD7
CHD1
Mi2
Iswi
Arip4
JBP2
ATRX
DRD1
Rad54
ERCC6
SSO1653
Mot1
SMARCAL1
B
Remodelers series 1
Chd1
Chd2
Chd4
Chd6
Chd8
Chd9
Brg1
Ep400
Remodelers series 2
ATRX
Smarca3
Smarca5
Smarcad1
Alc1
Chromo
PHD
BRK
Bromo
Helicase
SNF2
HIRAN
DBINO
HAND-SLIDE

Figure 3S


1 2 3
Control vector + - + - + -
Atrx shRNA - + - + - +
WB : anti-flag
WB : GAPDH
1 2 3 4
Control vector + - + - + - + -
Smarcad1 shRNA - + - + - + - +
WB : anti-flag
WB : GAPDH
1 2 3
Control vector + - + - + -
Chd1L shRNA - + - + - +
WB : anti-flag
WB : GAPDH
1 2 3
Control vector + - + - + -
Smarca5 shRNA - + - + - +
WB : anti-flag
WB : GAPDH
1 2 3
Control vector + - + - + -
Smarca3 shRNA - + - + - +
WB : anti-flag
WB : GAPDH

References

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010a). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010b). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

Berlivet, S., Houlard, M., and Gérard, M. (2010). Loss-of-function studies in mouse embryonic stem cells using the pHYPER shRNA plasmid vector. Methods Mol. Biol. Clifton NJ *650*, 85–100.

Chambers, I., and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene *23*, 7150–7160.

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

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

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

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

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.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Flaus, A., Martin, D.M.A., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res. *34*, 2887–2905.

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

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res. *21*, 396–420.

Ho, L., Miller, E.L., Ronan, J.L., Ho, W., Jothi, R., and Crabtree, G.R. (2011). esBAF Facilitates Pluripotency by Conditioning the Genome for LIF/STAT3Signalingand by Regulating Polycomb Function. Nat. Cell Biol. *13*, 903–913.

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.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. Curr. Opin. Cell Biol. *7*, 862–869.

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.

Liu, P., Jenkins, N.A., and Copeland, N.G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res. *13*, 476–484.

Niwa, H. (2007). How is pluripotency determined and maintained? Dev. Camb. Engl. *134*, 635–646.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

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

Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532–536.

Southon, E., and Tessarollo, L. (2009). Manipulating mouse embryonic stem cells. Methods Mol. Biol. Clifton NJ *530*, 165–185.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y., Zhou, B., Han, L., Fargo, D.C., et al. (2014). INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell *14*, 575–591.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013a). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013b). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003a). 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., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003b). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. *21*, 183–186.

Zhang, K., Li, L., Huang, C., Shen, C., Tan, F., Xia, C., Liu, P., Rossant, J., and Jing, N. (2010). Distinct functions of BMP4 during different stages of mouse ES cell neural commitment. Dev. Camb. Engl. *137*, 2095–2105.

Discussion & Conclusion

Discussion and Conclusion

The main objective of this work was to understand how ATP-dependent chromatin remodeling factors belonging to the Snf2 family interfere in the transcriptional control of mouse embryonic stem cells (ES). ES cells were shown to present a more 'open' chromatin (Meshorer et al., 2006). This chromatin state would work in parallel with a specific gene expression program that favors the expression of self-renewal genes and poises the differentiation genes. In order to maintain this epigenetic state in ES cells, several actors participate in the conservation of the genomic ES cell state. ATP-dependent Snf2 chromatin remodeling factors were shown to play important roles in the regulation of the particular chromatin landscape in ES cells. In this project, we conducted a genome-wide analysis of the distribution and function of ATP-dependent chromatin remodeling factors in mouse ES cells. To achieve this goal, we followed an adapted double experimental strategy consisting of initial ChIP-seq experiments in order to reveal the distribution of remodelers on ES cell regulatory elements followed by transcriptomic analysis on depleted ES cells for each remodeler.

We chose to study the role of a selection of thirteen chromatin remodeling factors among which several were shown to have an important role in ES cells and during differentiation. Those factors include Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400, Brg1, Smarca3, Smarca5, Smarcad1, Alc1 and Atrx.

Previous studies have concentrated on a small number of individual remodelers with no global comparative view on how those factors are recruited on the mammalian genome (Gaspar-Maia et al., 2011; Ho et al., 2009; Reynolds et al., 2012). In our study, we revealed the binding profiles of thirteen chromatin remodelers at ES promoter elements. We showed that

almost all the studied remodelers bind at various intensities promoter regions in ES cells. Some remodelers like Brg1, Chd4, Chd6, Smarca3 and Smarca5 bound similarly to all active genes, regardless of their H3K4me3/transcription level, while others, such as Chd1, Chd2, Chd9 and Ep400, were tightly linked to H3K4me3/transcription levels.

Promoters in ES cells can be active (H3K4me3-only) or poised (bivalent, H3K4me3 and H3K27me3) (Azuara et al., 2006; Bernstein et al., 2006). Closer analysis of the binding of remodelers at the two different types of promoters revealed differential binding of the various remodelers. Chd4, Chd6, Smarca5 and Brg1 bind bivalent and active promoters with similar intensities; Ep400 and Chd8 highly bind active promoters and present substantially lower enrichments at bivalent promoters, but were nevertheless bound to most (> 95 %); Chd1 and Chd9 mainly bind active promoters and at low levels to fewer than half of the total number of bivalent promoters; Smarca3 moderately binds only active promoters while Smarcad1 and Atrx present low binding intensities at only a few active promoters.

Our transcriptional analysis of ES cells depleted for the various remodelers revealed a prominent role for Ep400, Chd4, Brg1 and Smarcad1 and to a lesser extent Chd1 in the control of ES cells gene expression. Previously, Ep400 was shown to be essential in ES cell state conservation (Fazzio et al., 2008). In our study we have validated this observation; in addition we provided a better vision on the way Ep400 controls genes expression in ES cells. We showed that Ep400 acts as a gene activator at active genes (H3K4me3-only) and as a repressor at bivalent genes along with Chd4 and Smarcad1. On the other hand, Brg1 seemed to rather have a double repressing and activating role at active genes and an activating role at bivalent genes, an observation already demonstrated before in ES cells (Ho et al., 2009). Brg1 seems to counteract the inhibitory effect of the other remodeling factors at bivalent genes, seemingly to allow the low expression levels of bivalent genes that become essential during differentiation. Strikingly, and in contrary to what was published before, Chd4 known mostly

to be necessary during differentiation of ES cells as a part of the NuRD complex (Kaji et al., 2006; Reynolds et al., 2012; Sparmann et al., 2013) seems to have an important role in the conservation of the ES cell state. As Chd4-depleted ES cells die shortly after its depletion, this observation might be due to a better depletion of Chd4 in our experiment.

Several studies have demonstrated the involvement of remodelers in the core pluripotency transcriptional network of ES cells (van den Berg et al., 2010; Liang et al., 2008; Pardo et al., 2010). The further analysis of the genes deregulated upon the depletion of the factors Ep400, Chd4, Brg1, Smarcad1 and Chd1 revealed their implication in the core pTF regulatory network of ES cells, where a significant number of genes are common targets with Oct4/Sox2/Nanog. We showed that Ep400 and Chd4 positively regulate the expression of OSN-activated genes. To a lesser extent Smarcad1 and Chd1 act as co-activators of pluripotency genes along with Oct4/Sox2/Nanog. On the other hand, Brg1 seems to play a dual co-repressive and co-activator role at those genes. This further demonstrates the variable control mechanisms exerted by remodelers to control the ES cell state.

To further understand how remodelers control the ES cell genomic landscape, we analyzed the distribution of a selection of remodelers on nucleosomes. This analysis revealed variant nucleosome binding profiles depending on the nucleosome free region (NFR) size of genes. We have demonstrated that at active genes with short NFR, Ep400 and Chd4 bind equally the -1 and +1 nucleosomes and positively regulate their transcription; while Brg1 seems to mainly bind the +1 nucleosome and negatively control the transcription of such genes. At active genes with long NFR, we observe a prominent role for Brg1 bound at the -1 nucleosome and acting as a gene activator. On the other hand, at bivalent genes that usually possess long NFR; Ep400, Chd4 and Brg1 all bind the -1 nucleosome, and as mentioned previously, Ep400 and Chd4 negatively regulate this class of genes while Brg1 exerts a positive regulatory effect. Moreover, we observed that for both promoter types transcription initiates on either side

(sense and antisense) of the remodeler-bound nucleosomes where the TSS is present at the -1 nucleosome for short NFR and downstream the -1 nucleosome for long NFR. The -1 nucleosome position (presenting the TSS) of genes with short NFR is enriched with remodelers needed to rapidly evict the histone octamer in order to allow the formation of the preinitiation complex; in contrast the TSS of genes with long NFR located in a non-canonical histone region contains less remodeler-enrichment probably due to the fact that these CpG island rich sites are readily unstable (Deaton and Bird, 2011; Krinner et al., 2014) and need less but necessary remodeler involvement in order to stabilize (case of Ep400 and chd4) or reposition (case of Brg1) such labile nucleosomes.

Next, we analyzed the binding of the different remodelers at distal enhancer elements. CHIP-seq data revealed also a large remodeler presence on such elements in the ES cell genome. We showed that the studied remodelers all bind to canonical enhancers (characterized by high core pTFs binding), though with distinct intensities. The highest binding signals were observed for Chd4, Brg1 and Ep400 as expected with regard to their important role previously described. Interestingly, we have shown that the ES cell genome contains an additional large number of enhancer-like regulatory sequences bound by remodeling factors that we call non-canonical and differ from canonical enhancers by the variant presence/combination of the pTFs. This observation allows speculating that the regulatory sequence repertoire is much bigger than expected. Interestingly, remodelers showed also binding at super-enhancers of ES cells. Super-enhancers were shown to be particularly present for important pluripotency and self-renewal genes in ES cells (Whyte et al., 2013). Among the 231 super-enhancer associated genes, Brg1, Ep400, Chd4 and Smarcad1 regulate positively a considerable number of such genes revealing that these remodelers are important actors in the control of super enhancer activity, confirming their importance in the control of the ES cell state. Interestingly, Chd1 was also required for the proper level of expression of a significant number of genes

associated with super enhancer, revealing a new function for this factor. Remarkably, depletion of Chd4 caused the downregulation of the two pTFs Oct4 (Pou5f1) and Nanog, revealing a function of Chd4 in controlling the expression of major pTFs. Nanog expression was also downregulated upon depletion of Chd8. Altogether; this analysis suggests that remodelers play an essential function in the maintenance of ES cell fate through the regulation of super enhancer function.

Furthermore, we showed that the set of thirteen remodeling factors also bind architectural elements represented by CTCT-binding sites. CTCF is an architectural protein that plays an important role in creating boundaries between topologically associating domains in chromosomes further facilitating interactions between transcription regulatory sequences such as enhancers and promoters (Ong and Corces, 2014). Almost all remodelers were found bound at CTCF sites at the exception of Chd1 and Chd9 conferring a probable role of remodelers in the definition of chromatin boundaries and in assuring of long-range promoter-enhancer interactions. Smarca3 was the remodeler that presented the highest binding signal intensity. Interestingly, clustering analysis CTCF-binding sites revealed two dominant groups with various remodelers binding. Group A showed binding patterns for all remodelers. On the other hand, group B showed specific binding for each of Smarca3, Smarca5, Smarcad1 and ATRX. Although further investigations will be required to test the importance of remodelers in the function of CTCF binding sites, these data suggests the existence of at least two distinct functional categories of CTCF-binding sites, based on remodeler distribution pattern.

To conclude, our results create a better understanding of how the various ATP-dependent Snf2 chromatin remodeling factors contribute to the transcriptional control of the ES cell state. Our future goal is to analyze more profoundly the involvement of the chromatin remodelers in the transcriptional regulation of the ES cell genome. To achieve this goal,

several experimental approaches are to be considered. First, we would like to concentrate more on the effect of the depletion of chromatin remodelers in particular Ep400, Chd4, Brg1 and Smarcad1 on several transcriptional aspects. CHIP-Exo (Chromatin immunoprecipitation followed by exonuclease digestion) will be performed on ES cells depleted for each of the factors mentioned above, this will allow the detection of any Polymerase II aberrant binding. Furthermore, we will conduct Mnase-seq experiments for the detection of various modifications in nucleosome positioning upon remodeler depletion. Second, it will be interesting to focus on the role of the less studied chromatin remodelers with potential functions in ES cells such as Smarcad1 and Smarca5 using similar approaches as for the first series of remodelers.

In order to understand better how remodelers interfere in the transcriptional regulation during development, it will be essential to analyze the distribution and function of the various remodelers during ES cell differentiation to create a comparative genomic profile between the two cell states and to reveal the new set of remodelers required during differentiation. Moreover, analysis of remodeler functions in various types of tissues will be essential to understand how such a heterogeneous yet functionally related family of remodelers controls the transcription of the mammalian genome.

References

References

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.

Abranches, E., Bekman, E., and Henrique, D. (2013). Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression. PLoS ONE *8*, e59928.

Abranches, E., Guedes, A.M.V., Moravec, M., Maamar, H., Svoboda, P., Raj, A., and Henrique, D. (2014). Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. Development *141*, 2770–2779.

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572–576.

Ambrosetti, D.C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol. Cell. Biol. *17*, 6321–6329.

van Amerongen, R., and Nusse, R. (2009). Towards an integrated view of Wnt signaling in. Development *136*, 3205–3214.

Andreu-Vieyra, C.V., Chen, R., Agno, J.E., Glaser, S., Anastassiadis, K., Stewart, A.F., and Matzuk, M.M. (2010). MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. PLoS Biol. *8*.

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

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. *30*, 2817–2828.

Aubert, J., Dunstan, H., Chambers, I., and Smith, A. (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. Nat. Biotechnol. *20*, 1240–1245.

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

Bach, C., Mueller, D., Buhl, S., Garcia-Cuellar, M.P., and Slany, R.K. (2009). Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. Oncogene *28*, 815–823.

Baniahmad, A., Steiner, C., Köhne, A.C., and Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell *61*, 505–514.

Al-Baradie, R., Yamada, K., St Hilaire, C., Chan, W.-M., Andrews, C., McIntosh, N., Nakano, M., Martonyi, E.J., Raymond, W.R., Okumura, S., et al. (2002). Duane radial ray syndrome (Okihiro syndrome) maps to 20q13 and results from mutations in SALL4, a new member of the SAL family. Am. J. Hum. Genet. *71*, 1195–1199.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

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

Behrens, J., von Kries, J.P., Kühl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. Nature *382*, 638–642.

van den Berg, D.L.C., Zhang, W., Yates, A., Engelen, E., Takacs, K., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2008). Estrogen-Related Receptor Beta Interacts with Oct4 To Positively Regulate Nanog Gene Expression. Mol. Cell. Biol. *28*, 5986–5995.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010a). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010b). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

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.

Bertani, S., Sauer, S., Bolotin, E., and Sauer, F. (2011). RETRACTED: The Noncoding RNA Mistral Activates Hoxa6 and Hoxa7 Expression and Stem Cell Differentiation by Recruiting MLL1 to Chromatin. Mol. Cell *43*, 1040–1046.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. *23*, 2484–2489.

Birke, M., Schreiner, S., García-Cuéllar, M.-P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res. *30*, 958–965.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell *157*, 1445–1459.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the Distance: A Current View of Enhancer Action. Science *281*, 60–63.

Boeuf, H., Hauss, C., Graeve, F.D., Baran, N., and Kedinger, C. (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. J. Cell Biol. *138*, 1207–1217.

Boffelli, D., Nobrega, M.A., and Rubin, E.M. (2004). Comparative genomics at the vertebrate extremes. Nat. Rev. Genet. *5*, 456–465.

Boyer (2009). The chromatin signature of pluripotent cells. StemBook.

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.

Brazil, D.P., Yang, Z.-Z., and Hemmings, B.A. (2004). Advances in protein kinase B signalling: AKTion on multiple fronts. Trends Biochem. Sci. *29*, 233–242.

Briggs, R., and King, T.J. (1952). Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. Proc. Natl. Acad. Sci. *38*, 455–463.

Brock, H.W., and Fisher, C.L. (2005). Maintenance of gene expression patterns. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *232*, 633–655.

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.

Brook, F.A., and Gardner, R.L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. Proc. Natl. Acad. Sci. *94*, 5709–5712.

Buehr, M., and Smith, A. (2003). Genesis of embryonic stem cells. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 1397–1402.

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev. *13*, 2465–2477.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., et al. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Mol. Cell *6*, 1287–1295.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev. Biol. *210*, 30–43.

Cadigan, K.M. (2002). Wnt signaling – 20 years and counting. Trends Genet. *18*, 340–342.

Cao, R., and Zhang, Y. (2004). SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. Mol. Cell *15*, 57–67.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. Science *298*, 1039–1043.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. Mol. Cell *20*, 845–854.

Carlone, D.L., Lee, J.-H., Young, S.R.L., Dobrota, E., Butler, J.S., Ruiz, J., and Skalnik, D.G. (2005). Reduced Genomic Cytosine Methylation and Defective Cellular Differentiation in Embryonic Stem Cells Lacking CpG Binding Protein. Mol. Cell. Biol. *25*, 4881–4891.

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.

Chambers, I., and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene *23*, 7150–7160.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell *113*, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chapman, D.L., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S.I., Gibson-Brown, J.J., Cebra-Thomas, J., Bollag, R.J., Silver, L.M., and Papaioannou, V.E. (1996). Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *206*, 379–390.

Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PloS One *7*, e41374.

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

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

Chi, T.H., Wan, M., Lee, P.P., Akashi, K., Metzger, D., Chambon, P., Wilson, C.B., and Crabtree, G.R. (2003). Sequential roles of Brg, the ATPase subunit of BAF chromatin remodeling complexes, in thymocyte development. Immunity *19*, 169–182.

Clapier, C.R., and Cairns, B.R. (2009a). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

Clapier, C.R., and Cairns, B.R. (2009b). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

Clevers, H. (2006). Wnt/β-Catenin Signaling in Development and Disease. Cell *127*, 469–480.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.-H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev. *26*, 1714–1728.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev. *22*, 746–755.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. Cell Rep. *7*, 1456–1470.

Côté, J., Quinn, J., Workman, J.L., and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53–60.

Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. Science *309*, 1369–1373.

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

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. *19*, 24–32.

Cui, K., Zang, C., Roh, T.-Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80–93.

Curtis, C.D., and Griffin, C.T. (2012). The chromatin-remodeling enzymes BRG1 and CHD4 antagonistically regulate vascular Wnt signaling. Mol. Cell. Biol. *32*, 1312–1320.

Dang, W., and Bartholomew, B. (2007). Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. Mol. Cell. Biol. *27*, 8306–8317.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Davis, S., Miura, S., Hill, C., Mishina, Y., and Klingensmith, J. (2004). BMP receptor IA is required in the mammalian embryo for endodermal morphogenesis and ectodermal patterning. Dev. Biol. *270*, 47–63.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.-F., Songyang, Z., Thomson, J.A., and Zwaka, T.P. (2008). Ronin Is Essential for Embryogenesis and the Pluripotency of Mouse ES Cells. Cell *133*, 1162–1174.

Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H.G., and Stewart, A.F. (2014). Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. Dev. Camb. Engl. *141*, 526–537.

Di-Gregorio, A., Sancho, M., Stuckey, D.W., Crompton, L.A., Godwin, J., Mishina, Y., and Rodriguez, T.A. (2007). BMP signalling inhibits premature neural differentiation in the mouse embryo. Development *134*, 3359–3369.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Do, J.T., and Schöler, H.R. (2004). Nuclei of embryonic stem cells reprogram somatic cells. Stem Cells Dayt. Ohio *22*, 941–949.

Doble, B.W., Patel, S., Wood, G.A., Kockeritz, L.K., and Woodgett, J.R. (2007). Functional Redundancy of GSK-3α and GSK-3β in Wnt/β-Catenin Signaling Shown by Using an Allelic Series of Embryonic Stem Cell Lines. Dev. Cell *12*, 957–971.

Dottori, M., Gross, M.K., Labosky, P., and Goulding, M. (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. Development *128*, 4127–4138.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

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.

Egli, D., Rosains, J., Birkhoff, G., and Eggan, K. (2007). Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. Nature *447*, 679–685.

Eisen, J.A., Sweder, K.S., and Hanawalt, P.C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. Nucleic Acids Res. *23*, 2715–2723.

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

Faddah, D.A., Wang, H., Cheng, A.W., Katz, Y., Buganim, Y., and Jaenisch, R. (2013). Single-Cell Analysis Reveals that Expression of Nanog Is Biallelic and Equally Variable as that of Other Pluripotency Factors in Mouse ESCs. Cell Stem Cell *13*, 23–29.

Fang, X., Huang, Z., Zhou, W., Wu, Q., Sloan, A.E., Ouyang, G., McLendon, R.E., Yu, J.S., Rich, J.N., and Bao, S. (2014). The zinc finger transcription factor ZFX is required for

maintaining the tumorigenic potential of glioblastoma stem cells. Stem Cells Dayt. Ohio *32*, 2033–2047.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

Farh, K.K.-H., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817–1821.

Fazzio, T.G., and Panning, B. (2010). Condensin complexes regulate mitotic progression and interphase chromatin structure in embryonic stem cells. J. Cell Biol. *188*, 491–503.

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.

Felsenfeld, G., and Groudine, M. (2003a). Controlling the double helix. Nature *421*, 448–453.

Felsenfeld, G., and Groudine, M. (2003b). Controlling the double helix. Nature *421*, 448–453.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. Cell Stem Cell *11*, 477–490.

Flaus, A., and Owen-Hughes, T. (2011). Mechanisms for ATP-dependent chromatin remodelling: the means to the end. FEBS J. *278*, 3579–3595.

Flaus, A., Martin, D.M.A., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res. *34*, 2887–2905.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science *306*, 1574–1577.

Galan-Caridad, J.M., Harel, S., Arenzana, T.L., Hou, Z.E., Doetsch, F.K., Mirny, L.A., and Reizis, B. (2007). Zfx Controls the Self-Renewal of Embryonic and Hematopoietic Stem Cells. Cell *129*, 345–357.

Gao, H., Lukin, K., Ramírez, J., Fields, S., Lopez, D., and Hagman, J. (2009). Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. Proc. Natl. Acad. Sci. U. S. A. *106*, 11258–11263.

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

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. *12*, 36–47.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.-L., et al. (2010). Identification and characterization of enhancers

controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317–328.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.-L., and Stewart, A.F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse. Development *133*, 1423–1432.

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.

Graf, U., Casanova, E.A., and Cinelli, P. (2011). The Role of the Leukemia Inhibitory Factor (LIF) — Pathway in Derivation and Maintenance of Murine Pluripotent Stem Cells. Genes *2*, 280–297.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

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

Gupta, S., Dennis, J., Thurman, R.E., Kingston, R., Stamatoyannopoulos, J.A., and Noble, W.S. (2008). Predicting Human Nucleosome Occupancy from Primary Sequence. PLoS Comput Biol *4*, e1000134.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J. Embryol. Exp. Morphol. *10*, 622–640.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B.S., Li, P., Lim, S.L., Cao, S., Tay, J., Orlov, Y.L., et al. (2010). Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature *463*, 1096–1100.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.

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 Dev. *16*, 2650–2661.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res. *21*, 396–420.

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.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat. Cell Biol. *15*, 373–384.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

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.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934–947.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474–484.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. pnas.0812888106.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. pnas.0812888106.

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

Ho, L., Miller, E.L., Ronan, J.L., Ho, W., Jothi, R., and Crabtree, G.R. (2011b). esBAF Facilitates Pluripotency by Conditioning the Genome for LIF/STAT3Signalingand by Regulating Polycomb Function. Nat. Cell Biol. *13*, 903–913.

Hochedlinger, K., and Jaenisch, R. (2002). Nuclear transplantation: lessons from frogs and mice. Curr. Opin. Cell Biol. *14*, 741–748.

Hochedlinger, K., and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. Development *136*, 509–523.

Hong, F., Fang, F., He, X., Cao, X., Chipperfield, H., Xie, D., Wong, W.H., Ng, H.H., and Zhong, S. (2009). Dissecting Early Differentially Expressed Genes in a Mixture of Differentiating Embryonic Stem Cells. PLoS Comput Biol *5*, e1000607.

Hromas, R., Ye, H., Spinella, M., Dmitrovsky, E., Xu, D., and Costa, R.H. (1999). Genesis, a Winged Helix transcriptional repressor, has embryonic expression limited to the neural crest,

and stimulates proliferation in vitro in a neural development model. Cell Tissue Res. *297*, 371–382.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009a). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. *23*, 837–848.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009b). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. *23*, 837–848.

Ioshikhes, I.P., Albert, I., Zanton, S.J., and Pugh, B.F. (2006). Nucleosome positions predicted through comparative genomics. Nat. Genet. *38*, 1210–1215.

Ishov, A.M., Vladimirova, O.V., and Maul, G.G. (2004). Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. J. Cell Sci. *117*, 3807–3820.

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, 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. Mol. Cell. Biol. *24*, 8862–8871.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Jiang, B.-H., Chen, W.-Y., Li, H.-Y., Chien, Y., Chang, W.-C., Hsieh, P.-C., Wu, P., Chen, C.-Y., Song, H.-Y., Chien, C.-S., et al. (2015). CHD1L regulated PARP1-driven Pluripotency and Chromatin Remodeling during the Early-stage Cell Reprogramming. STEM CELLS n/a – n/a.

Jiang, H., Shukla, A., Wang, X., Chen, W., Bernstein, B.E., and Roeder, R.G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell *144*, 513–525.

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

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

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.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat. Struct. Mol. Biol. *21*, 569–571.

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 Dev. *19*, 489–501.

Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. Curr. Opin. Cell Biol. *7*, 862–869.

Kelly, V.R., Xu, B., Kuick, R., Koenig, R.J., and Hammer, G.D. (2010). Dax1 up-regulates Oct4 expression in mouse embryonic stem cells via LRH-1 and SRA. Mol. Endocrinol. Baltim. Md *24*, 2281–2291.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. *106*, 11667–11672.

Kidder, B.L., Yang, J., and Palmer, S. (2008). Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. PloS One *3*, e3932.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2004). Histone Code Modifications on Pluripotential Nuclei of Reprogrammed Somatic Cells. Mol. Cell. Biol. *24*, 5710–5720.

Kirmizis, A., Bartley, S.M., Kuzmichev, A., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2004). Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev. *18*, 1592–1605.

Knoepfler, P.S., Cheng, P.F., and Eisenman, R.N. (2002). N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. Genes Dev. *16*, 2699–2712.

Kohlhase, J., Heinrich, M., Schubert, L., Liebers, M., Kispert, A., Laccone, F., Turnpenny, P., Winter, R.M., and Reardon, W. (2002). Okihiro syndrome is caused by SALL4 mutations. Hum. Mol. Genet. *11*, 2979–2987.

Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O., and Clevers, H. (1998). Two Members of the Tcf Family Implicated in Wnt/β-Catenin Signaling during Embryogenesis in the Mouse. Mol. Cell. Biol. *18*, 1248–1256.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868–871.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet *4*, e1000242.

Ku, M., Jaffe, J.D., Koche, R.P., Rheinbay, E., Endoh, M., Koseki, H., Carr, S.A., and Bernstein, B.E. (2012). H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. Genome Biol. *13*, R85.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. *16*, 2893–2905.

Labosky, P.A., and Kaestner, K.H. (1998). The winged helix transcription factor Hfh2 is expressed in neural crest and spinal cord during mouse development. Mech. Dev. *76*, 185–190.

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

Laiosa, C.V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L., and Graf, T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity *25*, 731–744.

Laurent, B.C., Yang, X., and Carlson, M. (1992). An essential Saccharomyces cerevisiae gene homologous to SNF2 encodes a helicase-related protein in a new family. Mol. Cell. Biol. *12*, 1893–1902.

Lavelle, C., Praly, E., Bensimon, D., Le Cam, E., and Croquette, V. (2011). Nucleosome-remodelling machines and other molecular motors observed at the single-molecule level. FEBS J. *278*, 3596–3607.

Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P., and Hogan, B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev. *13*, 424–436.

Lee, J.-H., and Skalnik, D.G. (2005). CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. J. Biol. Chem. *280*, 41725–41731.

Lee, J.-H., Voo, K.S., and Skalnik, D.G. (2001). Identification and Characterization of the DNA Binding Domain of CpG-binding Protein. J. Biol. Chem. *276*, 44669–44676.

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

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron *55*, 201–215.

Li, Y., McClintick, J., Zhong, L., Edenberg, H.J., Yoder, M.C., and Chan, R.J. (2005). Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. Blood *105*, 635–637.

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

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

Liu, L., Luo, G.-Z., Yang, W., Zhao, X., Zheng, Q., Lv, Z., Li, W., Wu, H.-J., Wang, L., Wang, X.-J., et al. (2010). Activation of the imprinted Dlk1-Dio3 region correlates with pluripotency levels of mouse stem cells. J. Biol. Chem. *285*, 19483–19490.

Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell *146*, 720–731.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene *5*, 1743–1753.

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.

Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2010). Mechanism of chromatin remodeling. Proc. Natl. Acad. Sci. *107*, 3458–3462.

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. *105*, 2883–2888.

MacArthur, B.D., Sevilla, A., Lenz, M., Müller, F.-J., Schuldt, B.M., Schuppert, A.A., Ridden, S.J., Stumpf, P.S., Fidalgo, M., Ma'ayan, A., et al. (2012). Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. Nat. Cell Biol. *14*, 1139–1147.

MacDonald, B.T., Tamai, K., and He, X. (2009). Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. Dev. Cell *17*, 9–26.

MacLean-Hunter, S., Mäkelä, T.P., Grzeschiczek, A., Alitalo, K., and Möröy, T. (1994). Expression of a rlf/L-myc minigene inhibits differentiation of embryonic stem cells and embroid body formation. Oncogene *9*, 3509–3517.

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.

Malaguti, M., Nistor, P.A., Blin, G., Pegg, A., Zhou, X., and Lowell, S. (2013). Bone morphogenic protein signalling suppresses differentiation of pluripotent cells by maintaining expression of E-Cadherin. eLife *2*, e01197.

Margueron, R., and Reinberg, D. (2011a). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., and Reinberg, D. (2011b). The Polycomb Complex PRC2 and its Mark in Life. Nature *469*, 343–349.

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. (2008a). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

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. (2008b). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Massagué, J. (1998). TGF-β SIGNAL TRANSDUCTION. Annu. Rev. Biochem. *67*, 753–791.

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

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

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

McConnell, B.B., Ghaleb, A.M., Nandan, M.O., and Yang, V.W. (2007). The diverse functions of Krüppel-like factors 4 and 5 in epithelial biology and pathobiology. BioEssays News Rev. Mol. Cell. Dev. Biol. *29*, 549–557.

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

McKnight, J.N., Jenkins, K.R., Nodelman, I.M., Escobar, T., and Bowman, G.D. (2011). Extranucleosomal DNA binding directs nucleosome sliding by Chd1. Mol. Cell. Biol. *31*, 4746–4759.

Meissner, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. Nat. Biotechnol. *28*, 1079–1088.

Merrill, B.J., Gat, U., DasGupta, R., and Fuchs, E. (2001). Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. Genes Dev. *15*, 1688–1705.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006a). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell *10*, 105–116.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006b). Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. Dev. Cell *10*, 105–116.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Miller, S.A., Huang, A.C., Miazgowicz, M.M., Brassil, M.M., and Weinmann, A.S. (2008). Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. Genes Dev. *22*, 2980–2993.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S.J., Roeder, R.G., et al. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Mol. Cell *38*, 853–863.

Mishina, Y., Suzuki, A., Ueno, N., and Behringer, R.R. (1995). Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. Genes Dev. *9*, 3027–3037.

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.

Mohrmann, L., and Verrijzer, C.P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim. Biophys. Acta BBA - Gene Struct. Expr. *1681*, 59–73.

Mohrs, M., Blankespoor, C.M., Wang, Z.-E., Loots, G.G., Afzal, V., Hadeiba, H., Shinkai, K., Rubin, E.M., and Locksley, R.M. (2001). Deletion of a coordinate regulator of type 2 cytokine expression in mice. Nat. Immunol. *2*, 842–847.

Morgani, S.M., Canham, M.A., Nichols, J., Sharov, A.A., Migueles, R.P., Ko, M.S.H., and Brickman, J.M. (2013). Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. Cell Rep. *3*, 1945–1957.

Morris, S.A., Baek, S., Sung, M.-H., John, S., Wiench, M., Johnson, T.A., Schiltz, R.L., and Hager, G.L. (2014). Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. Nat. Struct. Mol. Biol. *21*, 73–81.

Mueller-Planitz, F., Klinker, H., and Becker, P.B. (2013). Nucleosome sliding mechanisms: new twists in a looped history. Nat. Struct. Mol. Biol. *20*, 1026–1032.

Müller, J., and Verrijzer, P. (2009). Biochemical mechanisms of gene regulation by polycomb group protein complexes. Curr. Opin. Genet. Dev. *19*, 150–158.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. Cell *111*, 197–208.

Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *102*, 12135–12140.

Nakashima, K., Colamarino, S., and Gage, F.H. (2004). Embryonic stem cells: staying plastic on plastic. Nat. Med. *10*, 23–24.

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J.-I., Matoba, R., Ko, M.S.H., et al. (2006). Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. Mol. Cell. Biol. *26*, 7772–7782.

Navarro, P., and Avner, P. (2009). When X-inactivation meets pluripotency: an intimate rendezvous. FEBS Lett. *583*, 1721–1727.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell *11*, 709–719.

Ng, S.-Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J. *31*, 522–533.

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

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

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

Nichols, J., 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.

Niwa, H. (2007). How is pluripotency determined and maintained? Dev. Camb. Engl. *134*, 635–646.

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

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

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

Norton, J.D. (2000). ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. J. Cell Sci. *113 (Pt 22)*, 3897–3905.

Nusse, R., Fuerer, C., Ching, W., Harnish, K., Logan, C., Zeng, A., Berge, D. ten, and Kalani, Y. (2008). Wnt Signaling and Stem Cell Control. Cold Spring Harb. Symp. Quant. Biol. *73*, 59–66.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556–562.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. TIG *17*, 520–527.

Okazaki, N., Ikeda, S., Ohara, R., Shimada, K., Yanagawa, T., Nagase, T., Ohara, O., and Koga, H. (2008). The novel protein complex with SMARCAD1/KIAA1122 binds to the vicinity of TSS. J. Mol. Biol. *382*, 257–265.

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

Olave, I., Wang, W., Xue, Y., Kuo, A., and Crabtree, G.R. (2002). Identification of a polymorphic, neuron-specific chromatin remodeling complex. Genes Dev. *16*, 2509–2517.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Paling, N.R.D., Wheadon, H., Bone, H.K., and Welham, M.J. (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. J. Biol. Chem. *279*, 48063–48070.

Pan, G.J., Chang, Z.Y., Schöler, H.R., and Pei, D. (2002). Stem cell pluripotency and transcription factor Oct4. Cell Res. *12*, 321–329.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200–213.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. Mol. Cell. Biol. *27*, 3769–3779.

Pasini, D., Bracken, A.P., Agger, K., Christensen, J., Hansen, K., Cloos, P. a. C., and Helin, K. (2008). Regulation of Stem Cell Differentiation by Histone Methyltransferases and Demethylases. Cold Spring Harb. Symp. Quant. Biol. *73*, 253–263.

Pelengaris, S., Littlewood, T., Khan, M., Elia, G., and Evan, G. (1999). Reversible activation of c-Myc in skin: induction of a complex neoplastic phenotype by a single oncogenic lesion. Mol. Cell *3*, 565–577.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. *14*, 288–295.

Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal. Mol. Cell. Biol. *26*, 7479–7491.

Pesce, M., and Scholer, H.R. (2000). Oct-4: Control of totipotency and germline determination. Mol. Reprod. Dev. *55*, 452–457.

Pesce, M., and Schöler, H.R. (2001). Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells Dayt. Ohio *19*, 271–278.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science *300*, 131–135.

Polach, K.J., and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. J. Mol. Biol. *254*, 130–149.

Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. J. Mol. Biol. *258*, 800–812.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Ramirez-Carrozzi, V.R., Nazarian, A.A., Li, C.C., Gore, S.L., Sridharan, R., Imbalzano, A.N., and Smale, S.T. (2006). Selective and antagonistic functions of SWI/SNF and Mi-2β nucleosome remodeling complexes during an inflammatory response. Genes Dev. *20*, 282–296.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114–128.

Rando, O.J., and Ahmad, K. (2007). Rules and regulation in the primary structure of chromatin. Curr. Opin. Cell Biol. *19*, 250–256.

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

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012b). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. Annu. Rev. Genet. *38*, 413–443.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W.M.P., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J. *20*, 7137–7148.

Rossant, J. (2008). Stem Cells and Early Lineage Development. Cell *132*, 527–531.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: Intricacy of writing and reading a single epigenetic mark. Mol. Cell *25*, 15–30.

Ryan, D.P., and Owen-Hughes, T. (2011). Snf2-family proteins: chromatin remodellers for any occasion. Curr. Opin. Chem. Biol. *15*, 649–656.

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

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

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. Cell *159*, 869–883.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L.G., Mardon, G., Luoh, S.-W., and Page, D.C. (1989). Putative transcription activator with alternative isoforms encoded by human ZFX gene. Nature *342*, 708–711.

Schreiner, S., Birke, M., García-Cuéllar, M.-P., Zilles, O., Greil, J., and Slany, R.K. (2001). MLL-ENL Causes a Reversible and myc-dependent Block of Myelomonocytic Cell Differentiation. Cancer Res. *61*, 6480–6486.

Schwartz, Y.B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. Nat. Rev. Genet. *14*, 853–864.

de la Serna, I.L., Carlson, K.A., and Imbalzano, A.N. (2001). Mammalian SWI/SNF complexes promote MyoD-mediated muscle differentiation. Nat. Genet. *27*, 187–190.

Sheik Mohamed, J., Gaughwin, P.M., Lim, B., Robson, P., and Lipovich, L. (2010). Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. RNA N. Y. N *16*, 324–337.

Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. Nature *406*, 541–544.

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

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532–536.

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

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.

Simon, J.A., and Kingston, R.E. (2013). Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. Mol. Cell *49*, 808–824.

Simon, J.A., and Tamkun, J.W. (2002). Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. Curr. Opin. Genet. Dev. *12*, 210–218.

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. Mol. Cell *28*, 665–676.

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

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

Sokol, S.Y. (2011). Maintaining embryonic stem cell pluripotency with Wnt signaling. Dev. Camb. Engl. *138*, 4341–4350.

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

Sparmann, A., Xie, Y., Verhoeven, E., Vermeulen, M., Lancini, C., Gargiulo, G., Hulsman, D., Mann, M., Knoblich, J.A., and van Lohuizen, M. (2013). The chromodomain helicase Chd4 is required for Polycomb-mediated inhibition of astroglial differentiation. EMBO J. *32*, 1598–1612.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

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. U. S. A. *93*, 7137–7142.

Stopka, T., and Skoultchi, A.I. (2003). The ISWI ATPase Snf2h is required for early mouse development. Proc. Natl. Acad. Sci. U. S. A. *100*, 14097–14102.

Sun, C., Nakatake, Y., Akagi, T., Ura, H., Matsuda, T., Nishiyama, A., Koide, H., Ko, M.S.H., Niwa, H., and Yokota, T. (2009). Dax1 Binds to Oct3/4 and Inhibits Its Transcriptional Activity in Embryonic Stem Cells. Mol. Cell. Biol. *29*, 4574–4583.

Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, D.A., Fletcher, C.F., Jenkins, N.A., Copeland, N.G., Klemsz, M., et al. (1996). Genesis, a Winged Helix Transcriptional Repressor with Expression Restricted to Embryonic Stem Cells. J. Biol. Chem. *271*, 23126–23133.

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. CB *11*, 1553–1558.

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.

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.

Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.-P., Marth, C., Höfer, T., and Rippe, K. (2012). Genome-wide nucleosome positioning during embryonic stem cell development. Nat. Struct. Mol. Biol. *19*, 1185–1192.

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.

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.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Thomson, M., Liu, S.J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875–889.

Tosi, A., Haas, C., Herzog, F., Gilmozzi, A., Berninghausen, O., Ungewickell, C., Gerhold, C.B., Lakomek, K., Aebersold, R., Beckmann, R., et al. (2013). Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex. Cell *154*, 1207–1219.

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

Tropepe, V., Hitoshi, S., Sirard, C., Mak, T.W., Rossant, J., and van der Kooy, D. (2001). Direct Neural Fate Specification from Embryonic Stem Cells: A Primitive Mammalian Neural Stem Cell Stage Acquired through a Default Mechanism. Neuron *30*, 65–78.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat. Genet. *36*, 1296–1300.

Uranishi, K., Akagi, T., Sun, C., Koide, H., and Yokota, T. (2013). Dax1 associates with Esrrb and regulates its function in embryonic stem cells. Mol. Cell. Biol. *33*, 2056–2066.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W.B., Cho, C.H.-H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339–1349.

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. Nature *461*, 199–205.

Voon, H.P.J., Hughes, J.R., Rode, C., De La Rosa-Velázquez, I.A., Jenuwein, T., Feil, R., Higgs, D.R., and Gibbons, R.J. (2015). ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. Cell Rep. *11*, 405–418.

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

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

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y., Zhou, B., Han, L., Fargo, D.C., et al. (2014). INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell *14*, 575–591.

Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R., and Crabtree, G.R. (1996). Diversity and specialization of mammalian SWI/SNF complexes. Genes Dev. *10*, 2117–2130.

Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L., and Blelloch, R. (2008). Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nat. Genet. *40*, 1478–1483.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell *10*, 440–454.

Watanabe, S., Umehara, H., Murayama, K., Okabe, M., Kimura, T., and Nakano, T. (2006). Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. Oncogene *25*, 2697–2707.

Wegner, M. (2010). All purpose Sox: The many roles of Sox proteins in gene expression. Int. J. Biochem. Cell Biol. *42*, 381–390.

Wegner, M., and Stolt, C.C. (2005). From stem cells to neurons and glia: a Soxist's view of neural development. Trends Neurosci. *28*, 583–588.

Wend, P., Holland, J.D., Ziebold, U., and Birchmeier, W. (2010). Wnt signaling in stem and cancer stem cells. Semin. Cell Dev. Biol. *21*, 855–863.

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.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

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.

Won, K.-J., Agarwal, S., Shen, L., Shoemaker, R., Ren, B., and Wang, W. (2009). An Integrated Approach to Identifying Cis-Regulatory Modules in the Human Genome. PLoS ONE *4*, e5501.

Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R., and Choo, K.H.A. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res. *20*, 351–360.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

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.

Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F., and Richmond, T.J. (2011). Structure and mechanism of the chromatin remodelling factor ISW1a. Nature *472*, 448–453.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., and Pugh, B.F. (2012). Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers. Cell *149*, 1461–1473.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003a). 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., Nichols, J., Chambers, I., and Smith, A. (2003b). 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.

Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin To Regulate Hox Gene Expression. Mol. Cell. Biol. *24*, 5639–5649.

Yoshida-Koide, U., Matsuda, T., Saikawa, K., Nakanuma, Y., Yokota, T., Asashima, M., and Koide, H. (2004). Involvement of Ras in extraembryonic endoderm differentiation of embryonic stem cells. Biochem. Biophys. Res. Commun. *313*, 475–481.

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

Young, S.R.L., and Skalnik, D.G. (2007). CXXC finger protein 1 is required for normal proliferation and differentiation of the PLB-985 myeloid cell line. DNA Cell Biol. *26*, 80–90.

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.

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

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. *21*, 1273–1283.

Zhang, J., Tam, W.-L., Tong, G.Q., Wu, Q., Chan, H.-Y., Soh, B.-S., Lou, Y., Yang, J., Ma, Y., Chai, L., et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early

embryonic development by the transcriptional regulation of Pou5f1. Nat. Cell Biol. *8*, 1114–1123.

Zhang, K., Li, L., Huang, C., Shen, C., Tan, F., Xia, C., Liu, P., Rossant, J., and Jing, N. (2010). Distinct functions of BMP4 during different stages of mouse ES cell neural commitment. Dev. Camb. Engl. *137*, 2095–2105.

Zhang, X., Zhang, J., Wang, T., Esteban, M.A., and Pei, D. (2008). Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. J. Biol. Chem. *283*, 35825–35833.

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

Zhao, S., Nichols, J., Smith, A.G., and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol. Cell. Neurosci. *27*, 332–342.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol. Cell *29*, 69–80.

(2014). Use of Genome-Wide RNAi Screens to Identify Regulators of Embryonic Stem Cell Pluripotency and Self-Renewal - Springer. B.L. Kidder, ed. (Springer New York),.

Rapport détaillé en français

Analyse de la fonction des facteurs de remodelage de la chromatine ATP-dépendants dans le contrôle de l'expression du génome des cellules souches embryonnaires

Résumé:

Les cellules souches embryonnaires (cellules ES) constituent un excellent système modèle pour étudier les mécanismes épigénétiques contrôlant la transcription du génome mammifère. Un nombre important de membres de la famille des facteurs de remodelage de la chromatine ATP-dépendants a une fonction essentielle dans l'auto-renouvèlement des cellules ES, ou au cours de leur différentiation. On pense que ces facteurs exercent ces rôles essentiels en régulant l'accessibilité à la chromatine au niveau des éléments régulateurs de la transcription, en modulant la stabilité et le positionnement des nucléosomes. Dans ce projet, nous avons conduit une étude génomique à grande échelle du rôle d'une dizaine des remodeleurs (Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400, Brg1, Smarca3, Smarcad1, Smarca5, ATRX et Chd1l) dans les cellules ES. Pour ce faire, une double stratégie expérimentale a été utilisée. D'une part, nous avons mené des expériences d'immunoprécipitation de la chromatine suivies par un séquençage à haut-débit (ChIP-seq) sur des cellules ES étiquetées pour les différents remodeleurs afin étudier leur distribution sur le génome. D'autre part, nous avons eu recours une approche transcriptomique qui implique l'utilisation des cellules déplétées de chaque remodeleur par traitement avec des vecteurs shRNA (knockdown). Nous avons établi les profils de liaison des remodeleurs sur des éléments régulateurs (promoteurs, enhancers et sites CTCF) sur le génome, et montré que ces facteurs occupent toutes les catégories d'éléments

régulateurs du génome. La corrélation entre les données ChIP-seq et les données transcriptomiques nous a permis d'analyser le rôle des remodeleurs dans les réseaux de transcription essentiels des cellules ES. Nous avons notamment démontré l'importance particulière de certains remodeleurs comme Brg1, Chd4, Ep400 et Smarcad1 dans la régulation de la transcription dans les cellules ES.

Introduction

Les cellules souches embryonnaires (cellules ES) de mammifères ont la capacité dauto-renouveler indéfiniment et de se différencier en tous types cellulaires du corps, une propriété connue comme la pluripotence (Keller, 1995). L'auto-renouvèlement et la pluripotence sont principalement contrôlés par une série de facteurs de transcription, qui établissent ensemble un réseau de régulation de transcription auto-entretenu. Ce réseau de transcription est contrôlé par une série de facteurs de transcription, y compris Oct4 (POU5F1), Sox2 et Nanog (Chambers and Smith, 2004; Niwa, 2007; Silva and Smith, 2008) qui constituent le réseau central responsable de la conservation de l'état des cellules ES. En outre, d'autres facteurs de transcription importants intègrent le réseau de transcription ES tels que ESRRB, Tcfcp2l1, Klf4, Klf2, STAT3 et Smad1 (van den Berg et al., 2010a; Pardo et al., 2010a). Les facteurs de transcription STAT3 et Smad1 sont les effecteurs terminaux qui intègrent au réseau de transcription les stimuli externes de deux voies majeures de signalisation dans les cellules ES, la voie LIF (Leukemia Inhibitory Factor) et la voie TGF (Transformation Growth Factor) (Chen et al., 2008; Ying et al., 2003; Zhang et al., 2010).

Les facteurs de transcription sont des protéines qui se lient à l'ADN. Dans le génome des cellules ES, ils sont présents au niveau des promoteurs et des éléments régulateurs distaux (les

enhancers et les elements liés au CTCF). Ils activent la transcription des gènes nécessaires à la conservation du phénotype des cellules ES et répriment les gènes de différenciation. En outre, ces facteurs de transcription régulent leurs propres promoteurs formant une boucle autorégulatrice qui assure un rétrocontrôle positif nécessaire pour conserver la pluripotence des cellules ES. Plusieurs études ont montré que ces facteurs de transcription interagissent avec un grand nombre de facteurs de remodelage de la chromatine, y compris plusieurs membres de la famille SNF2 de facteurs de remodelage de la chromatine ATP-dépendents (aussi appelés les remodeleurs)) (van den Berg et al., 2010b; Liang et al., 2008a; Pardo et al., 2010a).

Plusieurs membres de la famille SNF2 des remodeleurs se sont avérés importants pour l'auto-renouvèlement et la différenciation des cellules ES (Fazzio et al., 2008; Gaspar-Maia et al., 2009; Ho et al., 2011; Reynolds et al., 2012a; Wang et al., 2014).

La famille SNF2 est composée de 24 sous-familles qui partagent un domaine ATPase catalytique (Flaus et al., 2006). Ces facteurs sont soupçonnés de jouer un rôle essentiel dans la modification du paysage de la chromatine par leur capacité à positionner les nucléosomes et à déterminer leur occupation dans l'ensemble du génome ce qui rend la chromatine plus ou moins accessible (Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011; Jiang and Pugh, 2009). Bien qu'il a été demontré que plusieurs membres de la famille des remodeleurs SNF2 jouent une fonction essentielle dans le contrôle du phénotype des cellules ES, il n'existe jusqu'ici aucune vision globale, comparative de la façon dont ils travaillent ensemble pour contrôler le destin des cellules ES.

Contexte du projet

Plusieurs études ont montré l'importance d'un certain nombre de facteurs de remodelage de la chromatine ATP-dépendants dans le contrôle de l'état des cellules ES et de leur différenciation. Cependant, aucune étude globale de la façon dont ces facteurs sont recrutés sur le génome des mammifères n'a été menée.

Dans ce travail, nous essayons de comprendre le rôle de la famille SNF2 de facteurs de remodelage de la chromatine dans la régulation de la transcription du génome des cellules ES chez la souris. Nous étudions la distribution des remodeleurs dans le génome, plus précisément, sur les éléments proximaux (promoteurs) et distaux (enhancers et les sites CTCF-liés). Ensuite, nous essayons de comprendre la manière dont les remodeleurs sont recrutés sur des gènes cibles et comment ils sont impliqués dans l'architecture de nucléosomes dans le but de contrôler la transcription des cellules ES. Plus loin, nous analysons comment les facteurs de remodelage de la chromatine contribuent aux réseaux de transcription qui contrôlent la pluripotence et l'auto-renouvèlement des cellules ES.

. Dans ce projet, nous avons conduit une étude génomique à grande échelle du rôle d'une dizaine des remodeleurs (Chd1, Chd2, Chd4, Chd6, Chd8, Chd9, Ep400, Brg1, Smarca3, Smarcad1, Smarca5, ATRX et Chd1l) dans le contrôle de la transcription dans les cellules ES. Une double stratégie expérimentale a été utilisée. D'une part, nous avons mené des expériences d'immunoprécipitation de la chromatine suivies par un séquençage à haut-débit (ChIP-seq) sur des cellules ES étiquetées pour les différents remodeleurs afin étudier leur distribution sur le génome. D'autre part, nous avons eu recours une approche transcriptomique qui implique l'utilisation des cellules déplétées de chaque remodeleur par traitement avec des vecteurs shRNA (knockdown).

Résultats et Discussion

Dans notre étude, nous avons révélé les profils de liaison de treize remodeleurs de la chromatine à des promoteurs des cellules ES. Nous avons montré que presque tous les remodeleurs étudiés se lient à différentes intensités aux régions promotrices dans les cellules ES. Certains remodeleurs comme Brg1, Chd4, Chd6, Smarca3 et Smarca5 se lient avec un niveau similaire à tous les gènes actifs, indépendamment de leur niveau de H3K4me3/ transcription, tandis que la liaison d'autres, tels que Chd1, Chd2, Chd9 et EP400, est liée aux niveaux H3K4me3 / transcription.

Les promoteurs dans les cellules ES peuvent être actifs (H3K4me3 seule) ou bivalents (H3K4me3 et H3K27me3) (Azuara et al., 2006; Bernstein et al., 2006). Une analyse plus approfondie de la liaison de remodeleurs à ces deux types de promoteurs révèle une liaison différente entre les remodeleurs. Chd4, Chd6, Smarca5 et Brg1 se lient aux promoteurs bivalents et actifs avec des intensités semblables. Néanmoins, EP400 et Chd8 se lient principalement aux promoteurs actifs et présentent des enrichissements sensiblement inférieurs aux promoteurs bivalents, mais sont liés à la plupart (> 95%). Chd1 et Chd9 se lient principalement aux promoteurs actifs et à des niveaux plus faibles à la moitié des promoteurs bivalents. Smarca3 se lie modérément seulement aux promoteurs actifs tandis que Smarcad1 et ATRX présentent des faibles intensités de liaison à seulement quelques promoteurs actifs.

L'analyse du transcriptome des cellules ES deplétées pour les différents remodeleurs a révélé un rôle important pour EP400, Chd4, Brg1 et Smarcad1 dans le contrôle de l'expression des gènes des cellules ES. Auparavant, EP400 a été montré comme essentiel dans la conservation de l'état des cellules ES (Fazzio et al., 2008). Dans notre étude, nous avons validé cette observation. En outre, nous avons fourni une meilleure vision sur la façon dont EP400

contrôle l'expression des gènes dans les cellules ES. Nous avons montré que EP400 agit comme un activateur des gènes actifs (H3K4me3 seule) et comme un répresseur des gènes bivalents de la même façon comme Chd4 et Smarcad1. D'autre part, Brg1 semble avoir plutôt un double rôle de répression et d'activation des gènes actifs et un rôle d'activation au niveau des gènes bivalents, ce qui est en concordance avec une observation démontrée auparavant dans des cellules ES (Ho et al., 2009). Brg1 semble contrecarrer l'effet inhibiteur des autres facteurs de remodelage sur les gènes bivalents, apparemment pour assurer les faibles niveaux d'expression de ces gènes qui deviennent indispensables lors de la différenciation. Contrairement à ce qui a été publié avant, Chd4 surtout connu pour être nécessaire au cours de la différenciation des cellules ES en tant que membre du complexe NuRD (Kaji et al., 2006; Reynolds et al., 2012b; Sparmann) semble avoir un rôle important aussi dans la conservation de l'état des cellules ES.

Plusieurs études ont démontré l'implication des remodeleurs dans le réseau transcriptionnel central de la pluripotence des cellules ES (van den Berg et al., 2010a; Liang et al., 2008b; Pardo et al., 2010b). L'analyse des gènes dérégulés dans les cellules ES déplétées pour différents remodeleurs a révélé l'implication de Ep400, Chd4, Brg1, Smarcad1 et Chd1 dans le réseau transcriptionnel central des cellules ES, où un nombre important de gènes sont des cibles communes avec Oct4/Sox2/Nanog (OSN). Nous avons montré que Ep400 et Chd4 régulent positivement l'expression des gènes activés par OSN. Smarcad1 et Chd1 agissent aussi, mais à un plus faible niveau, comme des co-activateurs de gènes de pluripotence activés par OSN. D'autre part, Brg1 semble exercer un double rôle de co-répresseur et co-activateur sur ces gènes. Cela démontre que les mécanismes de contrôle de l'état des cellules ES exercés par les remodeleurs de la chromatine sont variables.

Ensuite, nous avons analysé la liaison des différents remodelers à des éléments regulateurs distaux (enhancers et site liés au CTCF). Les données de ChIP-seq ont révélé également une

grande présence des remodeleurs sur ces éléments dans le génome des cellules ES. Nous avons montré que les remodeleurs étudiés se lient aux enhancers canoniques (caractérisés par des niveaux d'expression élevés des facteurs OSN) avec des intensités différentes. Les signaux de liaison les plus élevés ont été observés pour Chd4, Brg1 et Ep400. Nous avons démontré que le génome des cellules ES contient un grand nombre supplémentaire d'enhancers liés par des facteurs de remodelage de la chromatine que nous appelons les enhancers non-canoniques. Les enhancers non-canoniques diffèrent des enhancers canoniques par une plus faible présence d'OSN. Cette observation permet de spéculer que le répertoire des séquences régulatrices dans le génome des cellules ES est beaucoup plus grand que prévu. Pour aller plus loin, nous avons démontré que les remodeleurs étaient présentes sur les super-enhancers des cellules ES. Dans les cellules ES, les super-enhancers sont particulièrement présentes sur les gènes importants pour la pluripotence et l'auto-renouvèlement (Whyte et al., 2013). Parmi les 231 gènes associés auxsuper-enhancers, Brg1, Ep400, Chd4 et Smarcad1 régulent positivement un nombre considérable de ces gènes. CHD1 a été également nécessaire pour le bon niveau d'expression d'un nombre important de gènes associés à des super-enhacers, révélant une nouvelle fonction de ce facteur. Remarquablement, l'épuisement des Chd4 dérégule négativement l'expression des facteurs Oct4 et Nanog, ce qui révèle une fonction importante de Chd4 dans le contrôle de l'expression des facteurs de transcription majeurs des cellules ES. L'expression de Nanog a également été dérégulée lors de la déplétion de Chd8. En conclusion, cette analyse suggère que les remodeleurs jouent un rôle essentiel dans le maintien des cellules ES par la régulation des super-enhancers.

En plus, nous avons démontré que l'ensemble des treize facteurs de remodelage se lie également à des éléments architecturaux représentés par des sites CTCF-liés. CTCF est une protéine architecturale qui joue un rôle important dans la création des limites entre les TAD (Topologically Associated Domains) et dans la facilitation des interactions entre les séquences

régulatrices telles que les enhancers et les promoteurs (Ong and Corces, 2014). Presque tous les remodeleurs ont été trouvés liés sur les sites CTCF à l'exception de CHD1 et Chd9, ce qui confère aux remodeleurs un rôle probable dans la définition des TAD et dans l'assurance des interactions entre les promoteurs et les enhancers. Smarca3 était le remodeler qui a présenté la plus forte intensité de liaison. L'analyse des sites de liaison CTCF analyse ont révélé deux groupes dominants liés différemment par les remodeleurs. D'une part, le groupe A est lié par tous les remodeleurs de la même manière. D'autre part, le groupe B est lié spécifiquement par les remodelerus Smarca3, Smarca5, Smarcad1 et Atrx. Bien que d'autres investigations soient nécessaires pour valider l'importance des remodeleurs dans la fonction de sites de liaison CTCF, ces données basées sur la distribution des remodeleurs suggèrent l'existence d'au moins deux catégories fonctionnelles distinctes de sites CTCF-liés.

Pour conclure, nos résultats créent une meilleure compréhension de la façon dont les différents facteurs SNF2 de remodelage de la chromatine ATP-dépendants contribuent au contrôle transcriptionnel associé auu maintien de l'état des cellules ES. Notre objectif pour l'avenir est d'analyser plus profondément l'implication des remodeleurs de la chromatine dans la régulation de la transcription du génome des cellules ES. Pour atteindre cet objectif, plusieurs approches expérimentales sont à considérer. Tout d'abord, nous tenons à nous concentrer davantage sur l'effet de la déplétion des remodeleurs en particulier Ep400, Chd4, Brg1 et Smarcad1 sur plusieurs aspects de la transcription. Des expériences ChIP-Exo (immunoprécipitation de la chromatine suivie par une digestion exonucléase) seront réalisées sur des cellules ES déplétées pour chacun des facteurs mentionnés ci-dessus, ce qui permettra la détection de toute présence aberrante de polymérase II. En outre, nous allons mener des expériences MNase-seq pour la détection de diverses modifications dans le positionnement des nucléosomes lors de la déplétion des remodeleurs. Deuxièmement, il serait intéressant de

se concentrer sur le rôle des remodeleurs moins étudiés avec des fonctions potentielles dans les cellules ES tels que Smarcad1 et Smarca5 en utilisant des approches similaires.

Afin de mieux comprendre comment les remodeleurs interfèrent dans la régulation de la transcription au cours du développement, il serait essentiel d'analyser la distribution et la fonction des divers remodeleurs au cours de la différenciation des cellules ES. Cela aiderait à créer un profil génomique comparatif entre les deux états cellulaires et de révéler la nouvelle série des remodeleurs nécessaires lors de la différenciation. En outre, l'analyse des fonctions des remodeleurs dans divers types de tissus serait essentielle pour mieux comprendre comment une telle famille contrôle la transcription du génome des mammifères.

Bibliographie

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.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010a). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

van den Berg, D.L.C., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010b). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

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.

Chambers, I., and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene *23*, 7150–7160.

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.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

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.

Flaus, A., Martin, D.M.A., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res. *34*, 2887–2905.

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

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res. *21*, 396–420.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. pnas.0812888106.

Ho, L., Miller, E.L., Ronan, J.L., Ho, W., Jothi, R., and Crabtree, G.R. (2011). esBAF Facilitates Pluripotency by Conditioning the Genome for LIF/STAT3Signalingand by Regulating Polycomb Function. Nat. Cell Biol. *13*, 903–913.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

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.

Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. Curr. Opin. Cell Biol. *7*, 862–869.

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

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

Niwa, H. (2007). How is pluripotency determined and maintained? Dev. Camb. Engl. *134*, 635–646.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010a). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010b). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

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

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012b). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532–536.

Sparmann, A. The chromodomain helicase Chd4 is required for Polyco... [EMBO J. 2013] - PubMed - NCBI.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y., Zhou, B., Han, L., Fargo, D.C., et al. (2014). INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell *14*, 575–591.

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.

Zhang, K., Li, L., Huang, C., Shen, C., Tan, F., Xia, C., Liu, P., Rossant, J., and Jing, N. (2010). Distinct functions of BMP4 during different stages of mouse ES cell neural commitment. Dev. Camb. Engl. *137*, 2095–2105.