Transcription Factor Networks in Embryonic and Neural Stem Cells

Studies on transcription factor interactions and localization

Erik Engelen

Colofon

ISBN: 978-94-6169-496-6

The studies presented in this thesis were conducted at the department of Cell Biology, Erasmus MC, Rotterdam and financially supported by NWO ALW open grant; 819.02.008 and VIDI grant; 91756346.

Cover design by Erik Engelen. Colloidal Coomassie stained Bis-Tris protein gel loaded with a FLAG-Sox2 purification sample.

Printed by: Optima Grafische Communicatie B.V.

Transcription Factor Networks in Embryonic and Neural Stem Cells

Studies on transcription factor interactions and localization

Transcriptiefactor netwerken in embryonale en neurale stamcellen

Studies over transcriptiefactor interacties en lokalisatie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 12 maart 2014 om 15.30 uur

door

Erik Marinus Petrus Engelen

geboren te Roosendaal en Nispen

Promotiecommissie

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Prof.dr. J.N.J. Philipsen Prof.dr. A.B. Houtsmuller Prof.dr. D. Huylebroeck
Copromotor:	Dr. R.A. Poot

Voor mijn vader en moeder

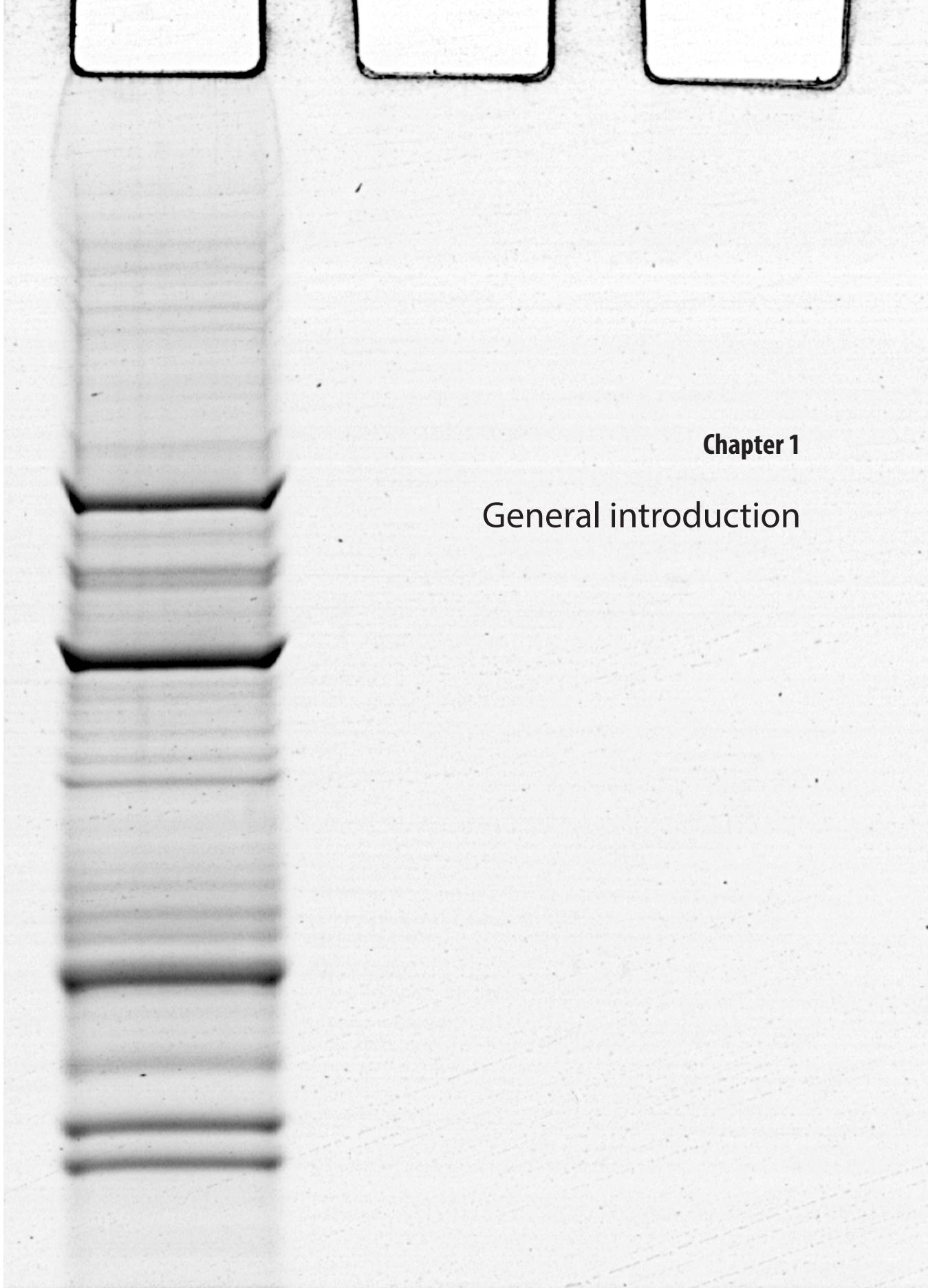
Be a simple kind of man

(Lynyrd Skynyrd, Simple Man 1973)

Table of contents

Abbreviations

bp	base pair
BRD	bromodomain
CHD	chromodomain helicase DNA-binding
ChIP	chromatin immunoprecipitation
DBD	DNA binding domain
DNA	deoxyribonucleic acid
EGF	epidermal growth factor
ENCODE	encyclopedia of DNA elements
ESC	embryonic stem cell
GTF	general transcription factor
FGF	fibroblast growth factor
HDAC	histone deacetylase
HMG	high-mobility-group
hPTM	histone post-translational modification
ICM	inner cell mass
iPSC	induced pluripotent stem cell
kDa	kilo Dalton
KDM	lysine demethylase
KMT	lysine methyltransferase
LIF	leukemia inhibitory factor
MS	mass spectrometry
ncRNA	non-coding RNA
NSC	neural stem cell
NXT	nuclear extract
Oct	octamer-binding protein
PcG	Polycomb group
PIC	pre-initiation complex
Pol II	RNA polymerase 2
PTM	post translational modifications
RNA	ribonucleic acid
RNAi	RNA interference
Sox	Sry-related HMG box
TF	transcription factor
TFIID	transcription factor IID complex
TSS	transcription start site



Chapter 1

General introduction

Introduction

The genetic material of any organism is also referred to as the genome and is passed on from generation to generation. The genome contains the hereditary information that is needed to construct the organism and to ensure its survival. This information is encoded in the form of deoxyribonucleic acid (DNA). DNA consists of two complementary strands of consecutively arranged nucleotides, composed of the nucleobase adenine (A), thymine (T), guanine (G) or cytosine (C). Due to selective pairing (A pairs to T, and G to C) they form a double helical structure which was first revealed by Watson and Crick in 1953 [1]. This discovery led to a revolution in genomic research that over 50 years later resulted in the first sequenced draft of the human genome [2]. The diploid human genome comprises roughly 2 x 3 billion nucleotides which are divided over 22 paired chromosomes and the two sex chromosomes. In the genome are units, referred to as genes, that code for proteins or non-coding RNA molecules. Expression of genes varies between cell types and during the various stages of development of an organism. Therefore gene expression is temporally and spatially controlled by a myriad of regulatory mechanisms. This chapter will introduce the basic concepts of gene expression and review the current knowledge on how transcription can be regulated. It will highlight recent advances made on the identification of regulatory elements throughout the genome with the aid of post-translational modifications (PTMs) on chromatin. The second part of the introduction will focus on embryonic stem cells (ESCs) and neural stem cells (NSCs). Their origin and function will be discussed, including the potential these cells have for application in hypothesis-driven fundamental research. The role of the transcription factor Sox2 in both of these cell types will get specific attention.

Transcriptional regulation

Basic concepts of transcription

The human genome contains approximately 22,000 protein-coding genes; the regions that actually contain instructions for making proteins are encoded in only 1.5% of the whole genome [3]. The expression of only a fraction of these genes is required depending on the cell type or differentiation state during development. Gene expression programs need to be properly tuned to ensure that appropriate genes are transcribed at the right time during cell-fate and/or activity/behavior decisions in response to signals from the environment. Transcription of genes is directed by sequence elements residing at the start of the gene in a region called the promoter. There are two types of promoter, focused core promoters and dispersed promoters. They can be distinguished by the presence of

a defined transcription start site (TSS) in focused core promoters, as opposed to multiple TSSs present in dispersed promoters. Dispersed promoters are usually observed with constitutively expressed genes such as housekeeping genes. Genes that are cell type specific and need to be tightly regulated are characterized by focused core promoters [4]. The focused core promoter encompasses -40 to +40 base pairs (bps) relative to position +1 from the TSS [5]. The first step in the process of transcription is the binding of one or multiple sequence specific DNA binding transcription factors to the promoter region that, with the support of other factors (e.g. chromatin remodelers), facilitate the accessibility of the promoter DNA. After this initial phase, defined sequence elements in the exposed promoter region can be recognized by the general transcription factors (GTFs). The first GTF to interact with the promoter is the transcription factor IID complex (TFIID). TFIID binds to several of these promoter elements via different subunits, which leads to a cascade of other general transcription factors (e.g. TFIIA, TFIIB) to bind the promoter region. The presence of the general transcription factor complexes at the core promoter is followed by the recruitment of the RNA polymerase II complex (Pol II). Pol II is the enzyme that synthesizes RNA molecules according to the DNA template. The joint complex of GTFs and Pol II located on the promoter is referred to as the pre-initiation complex (PIC) and is responsible for transcription of all protein-coding genes and non-coding RNAs in the eukaryotic genome, except for rRNAs and tRNAs which are transcribed by RNA polymerase I or III. After assembly of the PIC, the serine 5 residue (Ser5) in the carboxy-terminal domain (CTD) of Pol II is phosphorylated by general transcription factor TFIIH. The CTD is part of the largest subunit of Pol II and consists of 52 repeats of a heptapeptide. Within each of those repeats three specific residues can be phosphorylated throughout the whole CTD, contributing to the regulation of Pol II activity. The phosphorylation of Ser5 results in transcription initiation of Pol II. Following initiation, Pol II enters into an early elongation phase until it is blocked at approximately +20 to +60 nucleotides. This proximal-promoter pausing of Pol II is caused by binding of DRB-sensitivity-inducing factor (DSIF) and negative elongation factor (NELF), that inhibit further elongation of Pol II [6]. When positive transcription elongation factor b (P-TEFb) is recruited to the paused Pol II complex, either directly or via additional factors, the kinase activity of P-TEFb phosphorylates the DSIF-NELF complex, which results in its disassociation. P-TEFb subsequently phosphorylates the Ser2 residue within the CTD of Pol II. The phosphorylated CTD serves as a platform for binding of RNA processing factors that couple RNA processing to transcription. Following Ser2 phosphorylation, Pol II is released into productive elongation and synthesized RNA molecules will be further processed to mRNA, transported to the cytoplasm and translated into proteins (**Fig. 1**) [7].

Recent studies demonstrated that within the serial steps of the transcriptional process, according to the standard model described above, transcription is tightly

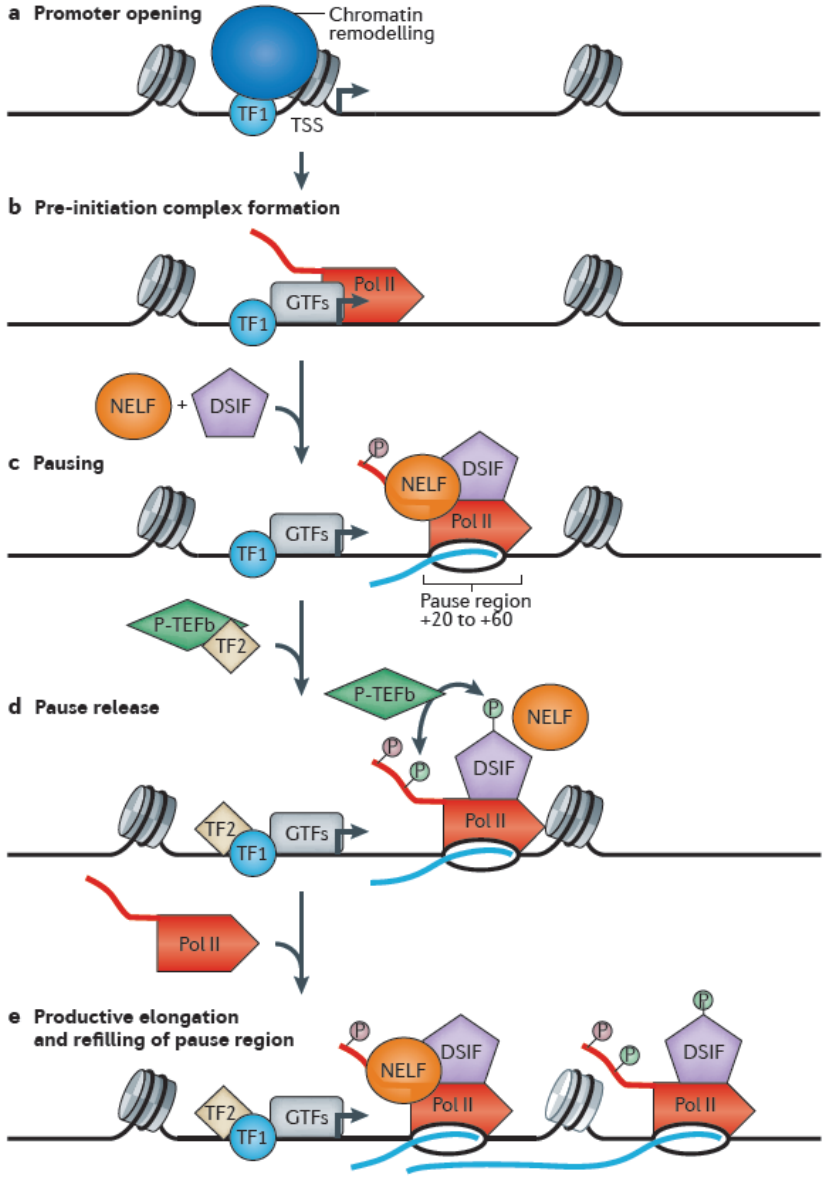


Figure 1 | Process of transcription.

Schematic representation of the process of transcription. a) Promoter opening; Sequence-specific TFs localize to a transcriptional start site (TSS) and recruit chromatin remodeling complexes (blue) to open the chromatin structure to allow for additional factors to enter the site b) Pre-initiation complex formation; General transcription factors (grey) bind to target sequences in the promoter region and recruit the Pol II complex (red) c) Pausing; Pol II proceeds +20 to +60 bp into the gene and pauses due to the binding of NELF (orange) and DSIF (grey) d) Pause release; Additional TFs recruit P-TEFb which phosphorylates the CTD of Pol II and DSIF which causes NELF to disassociate and Pol II to elongate e) Productive elongation; In the presence of both TF1 and TF2, transcription is maintained by binding of a new Pol II complex to achieve efficient RNA production. Adapted from [7].

regulated at the stage of paused Pol II. Nearly 30% of genes in human embryonic stem cells (hESCs) exhibit initiation of Pol II but no productive elongation, indicating that there are regulatory steps in transcription that occur after the formation of the PIC and initiation of Pol II [8]. There are several reasons why promoter-proximal pausing of Pol II is suggested as a general mechanism in transcriptional control. One important role of Pol II pausing is to serve as a post-initiation mechanism required to prevent uncontrolled divergent transcription at bidirectional promoters, where initiated Pol II can transcribe in both the sense and antisense direction[9,10]. Second, pausing of Pol II could facilitate the assembly of RNA processing factors onto the CTD domain of Pol II, which is proposed to couple transcription with mRNA processing events [11,12]. Recently it has also been demonstrated that certain signaling pathways can influence promoter-proximal pausing and therefore the rate of transcription by Pol II [13]. Also, most genes that exhibit paused Pol II are involved in stimulus-response pathways and it is therefore proposed that pausing could be a mechanism for synchronizing rapid gene activation [14,15].

There are reports that describe the role of sequence-specific DNA binding transcription factors in the release of paused Pol II. It has been demonstrated that the transcription factor c-Myc plays an important role in this process by recruiting P-TEFb to the paused Pol II complex [16]. Recently a new technique (permanganate-ChIP-seq) enabled the determination of the exact location of promoter-proximal pausing of Pol II throughout the whole *Drosophila melanogaster* (fruit fly) genome [17]. These accurate pinpoints were used to identify an associated DNA sequence motif, called motif 1. Motif 1 binding protein (M1bp) was subsequently identified and demonstrated to interact with these sites and shown to function as a novel pause release factor [18]. In addition to providing an extra level of gene expression control, the presence of the paused Pol II maintains an accessibility chromatin structure by preventing redisposition of nucleosomes within the transcription site. Consequently, this allows for additional regulators to enter the transcription site [19].

Chromatin

In eukaryotic cells, the genome is packaged into chromatin that is compartmentalized in the nucleus. Chromatin is a dynamic complex of DNA and proteins that can be condensed to form a “closed” state named heterochromatin, or be more loosely packed into an “open” state named euchromatin (**Fig. 2**). The compaction of DNA into chromatin renders the DNA inaccessible for transcription and other genomic processes such as replication and repair. This chromatin barrier has to be actively overcome before the basal transcription machinery is able to interact with the DNA and transcribe genes. Therefore chromatin can be assumed as a general mechanism to prevent random transcription events occurring all over the genome and organization of chromatin is key in transcriptional regulation.

In the early 70's, electron microscope studies first revealed the basic structure of euchromatin, the 10nm fiber, which appears as "beads on a string" [20]. The beads observed on the DNA are actually the smallest subunit of chromatin, the nucleosomes. The nucleosome consists of a ≈147 bp DNA strand wrapped around an octamer core of two copies of each of the histone proteins H2A, H2B, H3 and H4 [21]. Consecutive nucleosomes are connected by a small DNA linker of 20 to 60 bp, which can be bound by histone protein 1 (H1) (**Fig. 3a**). The regulation of transcription via chromatin structure is most prominent at the nucleosomal level. The nucleosomes can be moved and/or modified by various epigenetic mechanisms that aid in the control of gene expression. This adds another layer to the regulation of transcription without changing the actual DNA coding sequence.

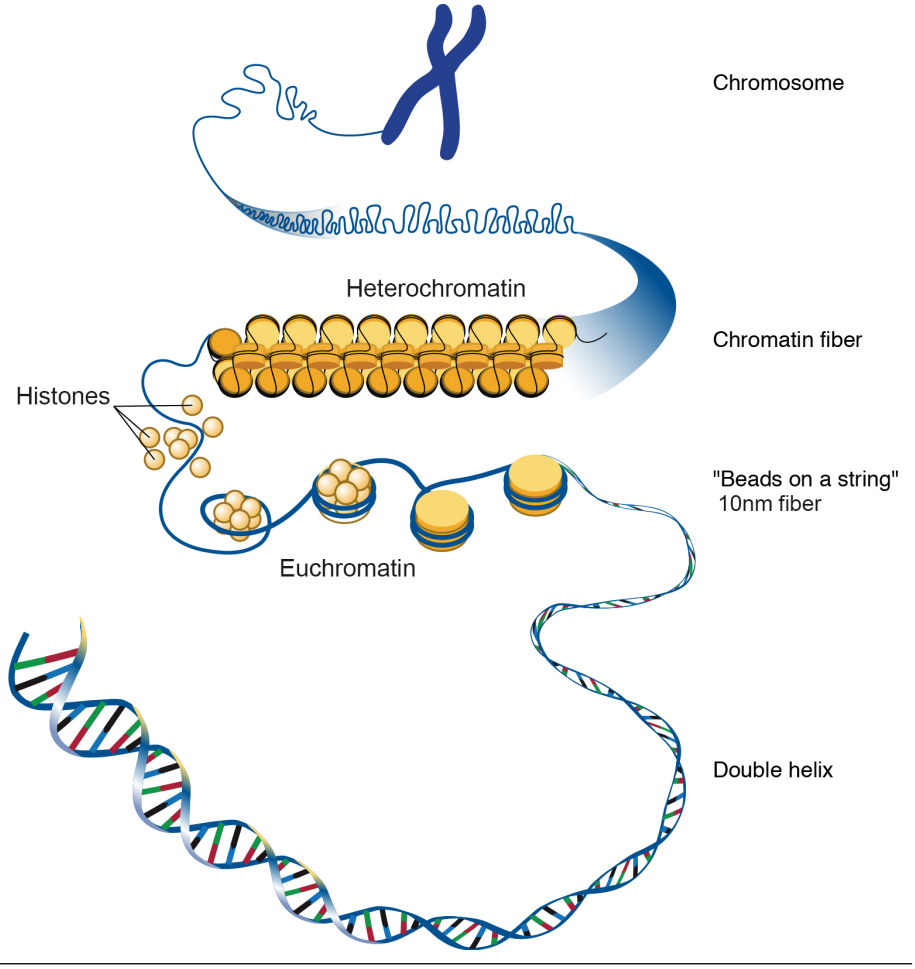


Figure 2 | Chromatin structure.
The various shapes of chromatin from chromosome to various chromatin fibers. DNA in the densely condensated fibers (heterochromatin) is difficult to access while decondensed chromatin (euchromatin) is more accessible for the basal transcription machinery. Adapted from Darryl Leja, NHGRI.

Chromatin remodeling

For the basal transcription machinery to gain access to the DNA and transcribe genes, the chromatin structure has to be remodeled. Chromatin remodeling is performed by large multi-subunit complexes that utilize ATP hydrolysis to disrupt nucleosome-DNA contacts by the repositioning, reconstitution or ejection of nucleosomes. These ATP-dependent chromatin remodeling complexes are all characterized by the presence of a motor subunit that belongs to the SNF2-like superfamily of ATPases [22]. From the shared homology of the catalytic ATPase domain and other characteristic protein motifs, 27 of these active subunits can be identified in human [23]. Despite the similarity between the domains within these proteins, they are genetically non-redundant and perform specific functions within the complex they reside in. The active subunits can be subdivided into four main families of chromatin remodelers; SWI/SNF family, ISWI family, INO80/SWR1 family and the CHD family.

The best studied family is that of SWI/SNF chromatin remodelers, which was first discovered in yeast and shown to be able to move or eject nucleosomes to provide access to DNA [24,25]. These actions are suggested to maintain a nucleosome depleted region (NDR) at the TSS in active promoter regions. In mammals the active subunit of the complex is Brg1 (or Brm in fly and human) and therefore the mammalian complex is named Brg/Brm associated factors, or BAF complex. Whereas the yeast SWI/SNF complex is monomorphic, the BAF complexes can be assembled in different variants depending on particular subunit combinations. The exchange of these variable subunits during development gives rise to tissue-specific functions that expand the ways in which the BAF complexes can contribute to transcription regulation. For instance a distinctive BAF complex in mouse ESCs (esBAF) has been shown to be important for the maintenance of gene regulatory networks that govern ESC pluripotency, while in mouse NSCs the replacement of subunits from a neural progenitor (np)BAF to neuron (n)BAF complex configuration acts as a switch in neuronal differentiation [26,27].

Similar to the SWI/SNF family of chromatin remodelers, the ISWI family comprises several distinct complexes that are involved in various processes, from DNA repair to replication and gene transcription [28]. In contrast to the SWI/SNF family, the ISWI family is often implicated in the repression of transcription. The active subunits in mammalian ISWI complexes are SNF2H or SNF2L and have the ability to generate regularly spaced nucleosome arrays [29]. In yeast it has been reported that an ISWI complex can overcome unfavorable sequence conditions for nucleosome placement which is suggested to aid silencing of transcription [30]. The capacity to generate regularly spaced nucleosome arrays is also proposed to be essential for the initiation and maintenance of heterochromatin formation supporting the suggestion ISWI complexes function in negative regulation of

transcription [31].

The INO80/SWR1 family of chromatin remodelers is unique for the reason that the catalytic subunits contain a split ATPase domain. This allegedly accounts for the specific functions they perform separate from the other families. In addition to transcription regulation, the INO80/SWR1 family functions in other diverse cellular processes, such as DNA repair, cell cycle checkpoint mechanisms and control of telomere stability [23]. The ability of the INO80 complex to evict nucleosomes is suggested to be necessary in transcriptional regulation of YY1 target genes in human cells [32]. However most studies describe a role for INO80/SWR1 complexes in the replacement of canonical histones with variant ones. Histone variants are non-allelic isoforms of the canonical histones that differ in their amino acid sequence [33]. These differences supply each histone variant with distinct physical and structural properties and consequently alter the function of the nucleosome into which they are incorporated. Therefore the incorporation of histone variants leads to profound changes in chromatin structure and DNA accessibility that is important in the regulation of many biological processes including transcription. The best studied are two histone variants for canonical histones H2A and H3, namely H2A.Z and H3.3. Indeed when H2A.Z or H3.3 are incorporated into nucleosomes, these become differently positioned onto the DNA [34]. The yeast SWR1-related Tip60/Ep400 chromatin remodeling complex in humans has been implicated to specifically replace H2A histones with the histone variant H2A.Z in nucleosomes at promoter regions [35].

The CHD family of remodelers is characterized by a double chromatin organization modifier domain (chromodomain) that is located in the N-terminal region and a highly conserved central ATPase domain. Proteins containing a chromodomain are generally considered to be regulators of chromatin structure since the domain has the ability to bind to methylated histone proteins thereby facilitating chromatin interactions which will be discussed below. The CHD family consists of nine members that can be further divided into three subcategories by the presence of additional structural motifs. The CHD family members are non-redundant and have very diverse functions that extend from regulation of different steps in the transcription cycle to functions in RNA processing [36]. The first subfamily is made up by Chd1 and Chd2 that are characterized by an AT-hook domain, an unique DNA binding domain that binds specifically to AT-rich sequences [37]. They act mainly as monomers and are reported to be localized together with Pol II at sites of active transcription. After depletion of Chd1 in mouse ESCs heterochromatin formation was observed indicating Chd1 is required for the maintenance of an open chromatin state [38]. The second subfamily contains Chd3/4, also referred to as Mi-2alpha and beta. Chd3/4 are the central components of the nucleosome-remodeling and histone deacetylase (NuRD) complex. The NuRD complex is implicated in repression of transcription via the combinatorial action of nucleosome remodeling with histone deacetylation by

histone deacetylases 1 and 2 (HDAC1/2) [36]. The third CHD subfamily is the most diverse with family members Chd5 to Chd9 that each contain a variety of additional functional motifs. Chromatin remodeling activity linked to transcriptional regulation has only been demonstrated for family members Chd7 and Chd8 [39,40].

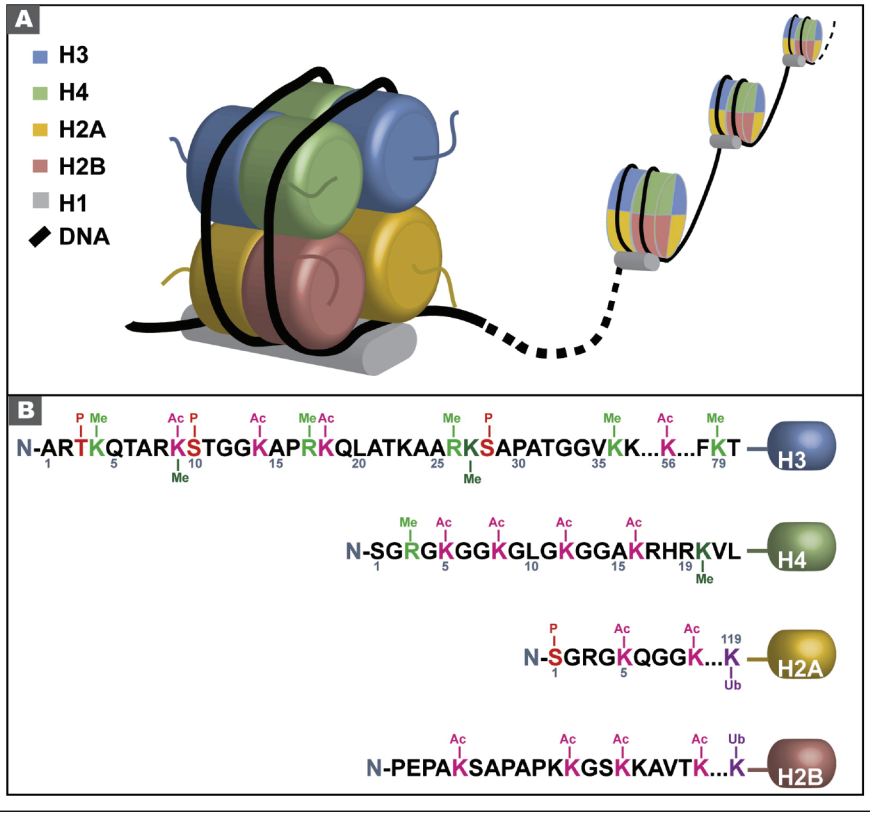


Figure 3 | The nucleosome as the smallest subunit of chromatin.
a) Schematic representation of the nucleosome. DNA (black line) wrapped around an octamer (colored cylinders) of histones b) N-terminal histone tails depicting the residues and their specific modification. Adapted from [230].

Chromatin modification

Besides chromatin remodeling the chromatin can also be altered by the placement of covalent modifications onto the DNA or histone proteins within the nucleosome. These modifications can directly alter chromatin structure or indirectly by the attraction or repulsion of chromatin remodeling complexes. Often various modifications coincide with each other and act together to regulate chromatin structure.

In mammals, the most stable and best studied epigenetic mechanism of chromatin modification is the methylation of cytosine in DNA. This usually occurs at symmetric CpG dinucleotides that are scarcely spread over the genome. Cytosine methylation is deposited and maintained by a conserved family of DNA methylases (Dnmt1, Dnmt3a and b) [41]. A

small percentage of CpGs is located in CG-enriched stretches of DNA called CpG islands that are typically unmethylated. These CpG islands are prevalent at the 5 prime regions of housekeeping genes and genes involved in developmental processes [42]. DNA methylation is highly linked to repression of transcription and promotes heterochromatin formation [43]. The positioning of methyl groups on the DNA can prevent interaction of sequence specific transcription factors or serve as binding platforms for proteins containing a methyl binding domain (MBD) such as MeCP2 and Mbd1, 2 and 4 [44]. These proteins are often associated with other known complexes (e.g. Mbd2 is a component of the NuRD complex) to repress transcription [45].

Besides DNA, histone proteins are substrates for modification as well. The histones can be chemically modified from within their globular domains to the N-terminal tails that extend outside the nucleosomes. Over 100 distinct histone post-translational modifications (PTMs) have been described ranging from the well-known forms of lysine acetylation, lysine methylation and serine/threonine phosphorylation (**Fig. 3b**), to the more unfamiliar forms such as SUMO-ylation and the more recently discovered crotonylation [46,47]. Histone PTMs are involved in a plethora of genomic processes where the placement of the modifications is highly linked to nucleosome dynamics via chromatin remodeling.

The first discovered modification on histones was acetylation, that was observed at highly transcribed genes [48]. These findings suggested that histone acetylation is important for facilitating transcription, possibly through a direct effect on chromatin structure. Acetylation neutralizes positively charged lysine residues on histones. The charge neutralization results in weakened charge-dependent interactions between the histones and nucleosomal DNA, linker DNA or adjacent histones, which increases the accessibility of DNA to other factors such as the basal transcription machinery. Histone acetylation at promoter regions and in transcribed genes has a high turnover by the constant placement and removal of the modification by so called “writers” and “erasers”. Acetyl-groups are actively laid down on lysine residues by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs) that are both associated with regions of active transcription [49]. The acetylation of nucleosomes at promoters and within gene bodies allows for efficient Pol II transition by the weakened interaction between DNA and histones, while deacetylation strengthens this interaction and therefore promotes chromatin reassembly after Pol II transcription [50].

Two chromatin remodeling complexes have been described that both have acetyltransferase or deacetylase capabilities in addition to remodeling activity. In humans, the Tip60/Ep400 can reconstitute nucleosomes by incorporating H2A.Z via its Ep400 subunit, while the Tip60 subunit subsequently can acetylate histones, thereby tightly linking remodeling activity with histone modifications for activation of transcription [51]. In contrast, the NuRD complex combines nucleosome placement with deacetylase activity of

Hdac1/2 subunits to repress transcription at regulatory regions in the genome [52].

Lysine acetylation on histone proteins can also serve as a recognition site for so called "reader" proteins that usually bring in additive activities. The acetyl modification can be recognized by proteins containing a bromodomain (BRD) that are often present in chromatin remodeling and/or modifier complexes and other transcriptional regulators [53]. An example of BRD containing proteins is the CREB binding protein or its paralog Ep300 (referred to as CBP/Ep300). Besides their ability to interact with acetylated histone lysines, they contain an enzymatic HAT domain by which they can acetylate many different proteins, including histones [54]. Their main substrates are histone 3 lysine 18 (H3K18) and H3K27 which are both associated with active transcription [55].

Acetyl modification on histones is assumed to be a general mechanism for generating weaker bonds between nucleosomes and the DNA which subsequently leads to a higher nucleosome turn over. These dynamics consequently improve DNA accessibility allowing for processes such as transcription to occur. This notion is supported by the observation that novel modifications, like acetylation, neutralize the charge of specific histone residues [46].

Another important form of modifying histone proteins is by mono-, di- or trimethylation (me1, me2 or me3) on lysine residues. There is an additional form of mono- or di-methylation on arginine residues, but its effects on nucleosome dynamics are less well understood. Unlike acetylation, methylation does not influence the charge of histone residues and therefore the effect is not as direct. Histone lysine methylation is considered to serve as a mechanism to modulate nucleosome stability. An example is the recognition of H3K9me3 by the fission yeast ortholog of HP1, Swi6 [56]. The dimerization of Swi6 on pairs of adjacent H3K9me3 modified nucleosomes stabilizes the nucleosomes which leads to heterochromatin formation.

The methylated residues can increase or decrease the affinity for certain reader proteins to interact with the modified nucleosome. The readers of the methylated nucleosomes subsequently determine a functional outcome, such as activation or repression of transcription. Certain methylation marks are associated with sites of active transcription and others with repressed regions. The main lysine methylation marks associated with active transcription are H3K4me2/3 and H3K36me2/3 and can be found on active promoters and in transcribed gene bodies, respectively [57]. H3K9me2/3 and H3K27me3 are involved in heterochromatin formation and Polycomb-mediated gene silencing and therefore marks that repress transcription. The cross-talk between these variably methylated nucleosomes contributes to the modular regulation of transcription. For instance, the H3K4me3 mark, deposited on active promoter regions by Trithorax group proteins, can be bound by the Taf3 subunit of the GTF complex thereby facilitating transcription [58]. The presence of H3K4me3 prevents binding of the Polycomb repressive

complex 2 (PRC2) [59]. PRC2 is responsible for establishing H3K27me3 modification that eventually leads to compaction of chromatin and therefore repression of transcription. H3K4me3 recruits the GTFs to the promoter and antagonizes the PRC2-mediated repression of transcription by placing H3K27me3. The cross-talk between H3K4me3 and H3K27me3 serves as a modulator in transcriptional regulation.

Lysine residues on histones are methylated by lysine methyltransferases (KMTs) that come in two classes. The largest class is characterized by a conserved catalytic SET domain, responsible for methylation of lysines [60]. In humans there are 48 SET domain-containing KMTs, and only one member from the second class, Dot1l, which does not contain a SET domain [61]. An interesting aspect of methylation is that methyl groups can be added to single lysine residues in distinct states. The residue can be unmodified or acquire mono-, di-or tri-methyl groups that are added in a progressive fashion. The addition of multiple methyl groups on histone lysines can be catalyzed by different KMTs or one single KMT. In the latter case the addition of the methyl groups is highly dependent on the accessibility and residence time of KMTs, which is regulated by surrounding histone modifications, underlying DNA sequence or interaction partners [57].

Every distinct state of histone lysine methylation creates a recognition site for particular readers to interact with the modified nucleosomes, although more methyl groups can also increase the affinity for one specific reader such as, heterochromatin protein 1 (HP1). HP1 binds H3K9me1 but interacts with an increased affinity with H3K9me2 and me3 [62]. Readers of lysine methylation are characterized by the presence of a methylation interaction domain such as Tudor, chromo, PWWP, MBT and PHD [63]. Like the BRD that recognizes acetylated lysines, these domains are often present in proteins with chromatin-remodeling capabilities that can either stabilize or destabilize nucleosomes. For example in yeast, placement of H3K36me3 by Set2 in gene bodies behind the transcription elongating Pol II complex is implemented to stabilize nucleosome occupancy. The H3K36me3 demarcation activates HDAC complex Rpd3S and enhances binding of the repressive chromatin remodeler Isw1b, which prevents histone turnover. Deacetylation and low nucleosome turn over increases nucleosome stability which in turn prevents cryptic transcription from within gene bodies by Pol II [64,65].

Methylated lysines can be actively removed by lysine demethylases (KDMs). There are several classes of KDMs that each have specific abilities in removing methyl groups from mono-, di-, or trimethylated lysines [66]. KDMs are involved in both activation and repression of transcription by removing active or repressive associated methylation marks, respectively. In some cases the same KDM can do both. For instance, the first discovered KDM, Kdm1a (Lsd1), was demonstrated to be a corepressor of transcription by demethylation of the active H3K4me2 mark [67]. Another study demonstrated that the interaction of androgen receptor (AR) with Kdm1a promotes transcription by Kdm1a

mediated removal of the repressive H3K9me1/2 mark [68]. These findings indicate that the function of methyltransferase and demethylases is complex and context dependent.

Important to note here is the interpretation on activation and repression of transcription as a consequence of chromatin remodeling and histone modifications. The classical terms, activation and repression, are perceived as static processes that once initiated, determine the functional outcome of gene expression. However, more evidence suggests that these processes are dynamic and interlaced with each other, leading to well-balanced mechanisms in which subtle changes can influence the functional outcome. In addition, a recent study demonstrated that key transcription factors recruit the repressive NuRD complex in parallel with de-repressing complexes, containing Utx and Wdr5, to target genes. This dual recruitment restricts the reactivation of genes that are necessary for somatic cell reprogramming [69]. These findings suggest there is a "gas and brake" model, in which activating and repressing complexes compete with each other and thereby fine-tune transcriptional output.

Another example is the colocalization of HATs and HDACs on actively transcribed regions, indicating that histone acetylation marks are constantly placed and removed at the same promoters [49]. Further, it was demonstrated that inactive genes, which are primed for activation by the presence of H3K4me3, showed low levels of both HATs and HDACs. These results indicate that the placement and removal of histone acetylation is an ongoing process in which the chromatin state is opened and reset at gene promoters.

Methylation and demethylation by KMTs and KDMs is a more complex process and is found to be context dependent. This is due to the large variety of KMTs and KDMs and their presence in multiple subunit complexes that can contain writer, eraser and reader capabilities [70]. For instance, it was shown that equilibrium between the activating Mll2/Utx complex and repressive PRC2/Rbp2 complex is important for regulation of developmental genes [71,72]. These mechanisms are further complicated by combinations of various histone modifications (called multi-valency) which creates an almost infinite number of possibilities for protein complexes to regulate a multitude of different processes, including transcription [73].

Transcription factors and enhancers

Regulation of gene transcription in a temporal and spatial manner is essential for cell fate decisions and the development of an organism. ATP-dependent chromatin remodelers, histone modifiers and the basal transcription machinery alone, lack the sequence specificity necessary to orchestrate these defined patterns of gene expression. Sequence-specific DNA-binding transcription factors (TFs) have evolved to direct these complexes to regulatory regions throughout the genome to facilitate the transcription or repression of genes. TFs are characterized by a DNA-binding domain (DBD) that recognizes a specific

DNA sequence (motif), that is approximately between 6 to 12 bps in eukaryotes [74]. In humans there are approximately 1400 validated TFs containing a DBD and more are still being identified [75]. The binding of TFs is not limited to motifs within promoter regions. Besides the core promoter, gene expression is controlled by many other *cis*-regulatory regions that reside in the genome and operate at greater distances from the TSS such as enhancers [76], silencers [77], insulators [78] and tethering elements [79]. Enhancers are DNA segments of a few hundred base pairs long that contain DNA motifs that serve as binding sites for TFs. TFs on enhancers recruit coactivators that facilitate chromatin remodeling/modification or communication between enhancers and promoters. The joint action of TFs and coactivators on enhancers results in the activation of transcription at specific promoters and therefore plays a key role in the regulation of gene expression.

Most of the TF binding sites in the genome cannot be readily bound by TFs due to 1) low affinity of the TF (monomer) or 2) packaging of the DNA into chromatin. In that case, cooperative binding of multiple TFs could overcome this barrier. Some TFs have the ability to interact with motifs that reside in nucleosomal DNA. These are called pioneer transcription factors that play an important role in the initiation of transcriptional activation. The best studied examples of these factors are forkhead box A proteins (Foxa). Crystal structure analysis has demonstrated that Foxa proteins have similar structures to that of the linker histone protein. This similarity in structure allows for the simultaneous binding of Foxa TFs and core histone proteins to nucleosomal DNA. In contrast to linker histones, the DBD in Foxa proteins is responsible for the sequence specific location of Foxa proteins [80]. Pioneer factors can prime specific regulatory regions and therefore act to initiate a gene regulatory network (GRN).

The collaborative action of multiple TFs or the priming by “pioneer” TFs provide the sequence specificity that facilitates recruitment of general chromatin remodeling/modifying complexes to regulatory regions, such as promoters and enhancers, in the genome. Consequently, nucleosome displacement facilitates the further binding of other TFs, coactivators and the basal transcription machinery to initiate gene expression. It is important to note that TFs themselves can also modulate nucleosome positioning. For example, upon removal of Myb proteins in yeast the size of NDRs was significantly reduced and it was demonstrated that binding of Gal4 disrupts the formation of a nucleosome at a specific location [81]. These are indications that the interactions between histones and TFs are dynamic.

To create spatio-temporal gene expression patterns, TFs not only work in a combinatorial fashion but can also function in a sequential manner to progress through different levels of GRNs during development. TFs with an identical DNA binding domain, usually within the same family, have the ability to interact with the same motifs and hand-over the activation of a gene during a developmental process. An example are the Sox

factors in neurogenesis [82]. In ESCs, Sox2 is bound to multiple enhancers in the genome. Most enhancers are associated with actively transcribed regions that are important for maintenance of ESC pluripotency, but Sox2 is also bound to enhancers that only become active when the ESCs are specified into the neuronal lineage. Upon differentiation into neurons, Sox2 is replaced by Sox3 and later Sox11 [83]. These observations imply that TF binding to enhancers is not always indicative of activity in the traditional sense, but may function to maintain enhancers poised for subsequent use.

Nucleosomes surrounding poised and/or active enhancer regions within the genome can distinguished by a H3K4me1 mark in the absence of H3K4me3 [84]. The mechanism for the deposition of H3K4me1 around enhancers is still unknown. Often H3K4me1 precedes binding of TFs that are associated with activation of transcription and it is therefore likely that H3K4me1 is linked to the priming of enhancers by pioneer factors. Indeed, some studies demonstrate that deposition of H3K4me1 coincides with binding of pioneer TFs, such as FoxA and PU.1 [85,86]. However, other evidence suggests that H3K4me1 is necessary for FoxA1 to bind [87]. H3K4me1 deposition at enhancer regions functions as a module to promote or repulse the binding of reader proteins that facilitate or maintain an accessible chromatin structure. For instance the presence of methylated H3K4 interferes with the interaction of Dnmt3l, a co-factor of Dnmt3a/b complexes. This blockage is suggested to prevent silencing of enhancers by DNA methylation [76]. However, in Drosophila H3K4me1 is present at enhancer regions despite the absence of DNMTs and therefore other mechanisms must play a role as well. These mechanisms most likely involve direct interactions with H3K4me1. Although most readers for H3K4 methylation can recognize most forms of methylation, Tip60, a subunit of the Tip60/Ep400 chromatin remodeling complex preferentially recognizes H3K4me1 over higher forms of H3K4 methylation [88]. The incorporation of H2A.Z into the neighboring nucleosomes by the Tip60/Ep400 complex results in increased DNA accessibility that facilitates the interaction of additional factors to the enhancer region.

Subsequent binding of lineage-specific TFs to poised enhancers can either stimulate or inhibit the activation of the enhancer. This is achieved by the interplay of TFs with coactivators or corepressors of which most do not have sequence-specific DNA binding capacities. Coactivators are near-ubiquitously expressed proteins or protein assemblies that facilitate in the activation of transcription. Coactivators can function as chromatin remodelers (e.g. BAF complex), modifiers (e.g. Ep300/CBP) or mediators of crosstalk with the basal transcription machinery at promoters (e.g. Mediator complex). Two essential transcriptional coactivators that are mainly associated with enhancers, are the acetyltransferase Ep300 and its paralog CREB binding protein (CBP). Apart from many other proteins, including several TFs, the major substrates for acetylation by Ep300/CBP are H3K18 and H3K27 [54]. The acetylation of the enhancer region is followed by the recruitment

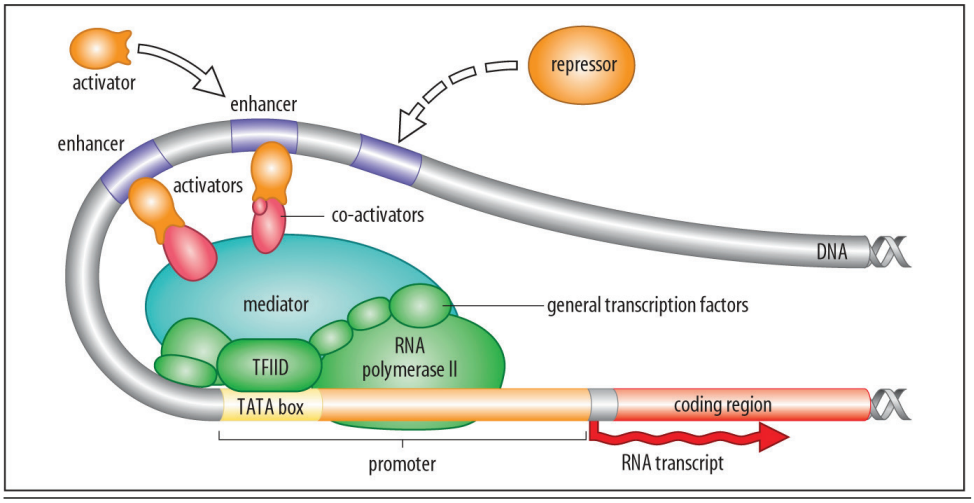


Figure 4 | Enhancer mediated regulation of transcription.
Enhancers regulate transcription by binding of activators or repressors, which are delivered to associated promoters via DNA looping. Adapted from [231].

of Pol II. The presence of Pol II leads to the synthesis of small enhancer-templated non-coding RNA (eRNA) molecules [89]. The combination of H3K27Ac, Pol II recruitment and eRNA production are considered to be the hallmarks for active enhancers [90].
Important to note is that Ep300/CBP localization to poised enhancers can occur without the display of acetyltransferase activity. The regulation of Ep300/CBP mediated acetylation is still under investigation and certain mechanisms, involving substrate availability or signaling events, have been proposed on how Ep300/CBP activation could be controlled [54].

Active enhancers are thought to regulate transcription by delivering supplementary factors to the promoter region that are necessary for transcription activation or repression. These factors have the ability to either facilitate or block PIC assembly and regulate the transition from initiated Pol II into productive elongation. The delivery of factors from enhancers to associated promoters occurs through looping of one or multiple enhancer regions onto promoter regions in a three dimensional chromatin architecture (**Fig. 4**). Long-range interactions between enhancers and promoters can be visualized by technologies that demonstrate chromatin interactions such as chromatin conformation capture, 3C for short [91]. For instance, extensive research on the Beta-globin locus revealed highly integrated loops being formed between promoters and multiple enhancers located in a region termed the locus control region. The 3D architectural structure regulates the expression of the various globin genes during different phases of mammalian development [92].

The Mediator complex plays an important role in bridging enhancers regions to the promoter of the regulated gene. Mediator consists of multiple subunits that can

interact with various TFs, other coactivators and the basal transcription machinery at the promoter. For instance, the human subunit Med26 was shown to recruit the super elongation complex that contains P-TEFb which is necessary for activation of Pol II towards productive elongation [93]. The Mediator complex can also change its conformational state which subsequently influences accessibility and facilitates recruitment of additional components, such as the Cohesin complex. The Cohesin complex is a ring-like structure that has the ability to connect two DNA segments. The direct interaction between Mediator and Cohesin suggests they cooperate to arrange three-dimensional chromatin architecture that enables communication between enhancer and promoters [94].

Recently, a study investigated the localization of Mediator and compared it to that of key ESC transcription factors, Oct4, Sox2 and Nanog in mouse ESCs [95]. They identified two types of enhancers that could be distinguished by low and high levels of Mediator that was spread over a few hundred kb or large clusters of around 50kb. These large Mediator-rich enhancers domains were termed super-enhancers. Super-enhancers were found to be located near genes that have high expression levels and are important for ESC identity. Upon knock-down of Mediator or Oct4, the expression of these genes was highly affected compared to genes associated with regular enhancers and led to a loss of ESC pluripotency and differentiation. Super-enhancers were shown to exist also in other cell types, including certain cancer cells, where they were found to be associated with genes involved in cancer cell biology [96]. Therefore, cell type-specific super-enhancers are relevant for maintenance of cell identity and can be used to determine cell type-specific gene regulatory networks.

The interplay of enhancers and promoters is essential in the regulation of gene expression. Therefore identification of these *cis*-regulatory elements within the genome is crucial for the understanding of the regulation of GRNs.

Promoter and enhancer annotation

Chromatin immunoprecipitation (ChIP) experiments were developed to investigate the binding of a protein of interest to specific DNA regions. ChIP followed by deep-sequencing (ChIP-seq) is now routinely used to investigate the genome-wide localization of a protein[97]. This technique has greatly benefited from the completion of most sequenced eukaryotic genomes and recent developments in high-throughput sequencing technologies, which make it possible to map DNA fragments to the genome. Localization of proteins, such as TFs or specific PTM histones, by ChIP-seq can be utilized to predict regulatory elements and therefore stands at the base of the functional annotation of the genome.

Early genome-wide studies on histone PTMs demonstrated that they were not randomly distributed throughout the genome but localized at specific regions, for example at promoters. Transcriptionally active promoters are marked by the accumulation of histone acetylation on various H3 and H4 residues and the specific demarcation of H3K4me3 near

the TSS [84]. These regions colocalized with Pol II, had a high nucleosome turn-over or were even depleted from nucleosomes, demonstrated by a saddle distribution of H3K4me3 around the TSS. Most strikingly, it was described that putative enhancers were marked by H3K4me1 and devoid of H3K4me3, thereby establishing the realization that the specific methylation state of H3K4 could be used to distinguish between promoters and enhancers. This led to a revolution in epigenomic research, where the localization of histone PTMs can be used as a tool to annotate regulatory regions within the genome.

Histone PTMs can not only be utilized to determine the location but also the current state of a certain regulatory region. Although acetylation was already known to be generally associated with transcriptional activity, several studies demonstrated that specifically H3K27Ac can be a mark for active enhancers [90,98]. Genome-wide profiles for H3K27Ac revealed that a portion of all H3K4me1 marked enhancers were also associated with H3K27Ac. These double-positive (H3K4me1+/H3K27Ac+) enhancers correlated with higher gene expression of nearby genes compared to genes located in the vicinity of enhancers that were negative for H3K27Ac. Occurrence of short RNA reads originating from the H3K4me1+/H3K27Ac+ enhancers was also observed, as well as colocalization of cell identity TFs and the coactivator Ep300, the main HAT enzyme for establishing H3K27Ac. Together these findings demonstrate that the location of histone PTMs, particularly H3K27Ac, could distinguish between two enhancer states, namely the H3K4me1+/H3K27Ac+ enhancers that are engaged in activation of gene transcription and H3K4me1+/H3K27Ac- enhancers that are in an inactive or poised state (Table 1).

Further investigation showed that a portion of H3K4me1+/H3K27Ac- poised enhancers that are associated with developmentally regulated genes, were also marked by H3K27me3 in ESCs (Table1) [99]. These double H3K4me1+/H3K27me3+ labeled enhancers also showed colocalization of Ep300 [98]. H3K27me3 is a mark associated with Polycomb-mediated silencing and is established by the PRC2 complex [100]. In ESCs, H3K27me3 was also found to be localized together with H3K4me3 at promoters of developmentally regulated genes [101]. These bivalent promoters were demonstrated to be in a poised state, which enables them to quickly act on differentiation cues and engage in transcriptional activation [102]. Detection of H3K27me3 at a subset of enhancers marked by H3K4me1 suggests that apart from poised promoters there are also poised enhancers.

The knowledge that histone PTMs can distinguish between different *cis*-regulatory elements, such as promoters and enhancers, allowed for the categorization of the genome in different chromatin states. In one study, histone PTM profiles and their spatial location in the genome were used to characterize six broad classes of different chromatin states; promoter, enhancer, insulator, transcribed, repressed and inactive regions in nine different human cell lines [103]. These six chromatin states showed distinct correlation with TSSs, transcript levels, DNase I hypersensitivity and transcription factor binding, which indicated

Table 1 | Overview of histone marks and associated localization.

Histone PTM	Associated factor	Annotation
H3K4me3+ H3K27Ac+	Pol II	Active promoter
H3K4me1+ H3K27Ac+	Pol II, Ep300	Active enhancer
H3K4me1+ H3K27Ac-	-	Intermediate enhancer
H3K4me1+ H3K27me3+	Ep300	Poised enhancer

the validity of the characterization based on histone PTM profiles. When transcript levels were compared with promoter and enhancer states, these could be divided into active, weak and poised subclasses, which allows for an accurate prediction of chromatin state.

The culmination of genome annotation is the establishment of the Encyclopedia of DNA Elements (ENCODE) project, carried out by a consortium of laboratories that aimed to map all regulatory regions throughout the genome [104]. This resulted in the ENCODE database that contains a multitude of datasets such as genome-wide histone PTM profiles and binding profiles for over 100 different factors such as chromatin remodelers/modifiers, histone variants and TFs. These profiles were correlated with datasets for DNase I hyper sensitivity (DNase-seq), gene expression (RNA-seq) and other chromatin state determining techniques, which led to interesting findings. The major discovery of the human ENCODE project is that a large part of the genome (~80%), which was believed to be mostly "junk DNA", displayed at least one biochemical function in one of the datasets. However, it is still under debate if this is an overestimation, since a full test for direct functionality has not been performed [104].

Genome-wide profiling of regulatory regions determines a specific epigenetic landscape that is distinct in different cell types. Therefore, an epigenetic landscape can be used to identify a specific cell and the state it is in. The ENCODE database currently estimates the presence of approximately 400.000 enhancers in various states within the human genome. This large set of potential enhancers suggests there are many different possibilities to regulate a relatively small number of genes. Several studies demonstrated that active promoters and enhancers in a specific cell type are associated with genes that are involved in cell type specific functions [98,99,103]. For instance, in skeletal muscle cells (HSMM) active promoters were highly associated with extracellular structure genes, while those in lymphoblastoid cells (GM12878) were associated with immune response genes [103]. Interestingly, it appeared that the state of enhancers differs more than that of promoters between cell types and enhancer clusters are significantly more cell type specific [103]. Therefore, genome-wide enhancer activity patterns (active versus poised) are a better predictor for cell identity as opposed to active promoters.

The next phase in epigenomics is to use enhancer annotation to analyze the

progression of the activity of enhancers during the development of an organism. A recent study has addressed this issue by profiling various histone PTMs in different cell stages from the developing Drosophila embryo [105]. The results in this study support a model for enhancer activation, where intermediate enhancers, marked by H3K4me1, can progress to poised or active states during development, by subsequent marking of H3K27me3 or H3K27Ac, respectively.

Epigenomic profiling can predict and annotate *cis*-regulatory regions, such as enhancers and promoters, which can be correlated to their associated genes via proximity and gene expression analysis. Nevertheless, it still remains unclear how an enhancers selects a promoter, sometimes across a considerable distance spreading several kilobases. For example, an enhancer that regulates the *Sonic hedgehog* gene (*Shh*) important for limb development was identified 1Mb away from its target gene [106]. Direct associations between enhancers and promoters can be measured by 3C experiments on a single locus scale. To perform these experiments on a genome-wide scale, 3C-based technologies, such as 4C, 5C and Hi-C were created [107-109]. Technologies, such as 4C, can identify interactions between a region of interest (e.g. promoter) and the whole genome, thereby identifying chromatin interactions (e.g. putative enhancers) [107]. Results from experiments that combine 4C with epigenomic profiling can be used to further investigate how the genome is used within the three dimensional space of the nucleus.

The functional annotation of the genome is not only of importance in understanding how the genome is organized and operates, but it can also aid in investigating human disease. Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) that are associated with phenotypes of human diseases [110]. It was demonstrated that these SNPs were enriched in functional elements, such as TSSs and enhancers, annotated by the ENCODE consortium [111]. Furthermore, SNPs were found to be concentrated in DNase I hypersensitive sites (DHSs) in many different cell and tissue types. DNase I hypersensitivity is a hallmark for functional regulatory activity. SNPs can disturb the binding of TFs in these regulatory regions and alter regulation of nearby genes. Therefore, mapping of SNP containing DHSs can predict disruption of regulatory pathways and provide new insights in the causes of human disorders [112]. This greatly benefited the identification of the functional SNPs and allowed for the formulation of hypotheses that explain the biological mechanism behind the SNP.

Genome-wide annotation of regulatory elements in various types of tumor cells could also be of interest in cancer research. Current epigenomic profiling of various types of cancer has mainly focused on DNA methylation patterns combined with gene expression analysis. Previous studies have demonstrated that global histone modification patterns between healthy and tumorigenic cells vary greatly [113]. Also there are numerous reports on chromatin remodelers/modifiers that are involved in various types of cancer (reviewed

in [114]). Therefore to employ epigenomic profiling data of histone PTMs in tumor cells will increase the understanding of tumor formation and could give rise to new epigenetic biomarkers that aid in the characterization, diagnosis and improve therapies.

Stem cells

The adult human body consists of approximately 300 different cell types including various stem cells. Stem cells are unique from any other cell within the organism and are distinguished by two main characteristics. First, they are unspecialized cells that have the ability to self-renew thereby maintaining their own population. Second, they have the potential to differentiate into other cell types depending on certain physiological conditions and environmental cues. Stem cells play an important role during early growth and development of the embryo. However, some stem cell populations also persist within the adult organism, where they function to replenish tissue and organ specific cells. This part will introduce two main types of stem cells; 1) embryonic stem cells (ESCs) and, 2) neural stem cells (NSCs) as an example for adult stem cells. The transcription factor Sox2 is discussed at the end of this chapter.

Embryonic stem cells

The starting point in mammalian development is the fertilized oocyte, called zygote. During the pre-implantation phase, the totipotent zygote is programmed to undergo a series of cell divisions and several lineage specifications to form the blastocyst (**Fig. 5a**) [115]. This early structure consists of different cell layers, including the trophectoderm that will form the placenta and the epiblast that will form the embryo (**Fig. 5b**). The blastocyst contains the inner cell mass (ICM) which is comprised of unspecialized cells that have the capacity to generate all cell lineages that form the embryo. In the early eighties, these cells were extracted from the ICM of the mouse blastocyst, plated and propagated *in vitro* [116,117]. These embryonic stem cells (ESCs) were found to proliferate indefinitely while remaining in a naïve state similar to the cells in the ICM. ESCs self-renew and are pluripotent, meaning they are able to differentiate into a cell lineage of any of the three main germ layers, ectoderm, endoderm and mesoderm, depending on the culture conditions. This ability can be demonstrated when ESCs are grafted to the adult mouse which results in the formation of teratocarcinomas [117]. The full potential of mouse ESCs was shown when they were injected into blastocysts. In these experiments, *in vitro* cultured ESCs fully reintegrate in the developing embryo resulting in chimeras that show contribution to all tissues, including colonization to the germline [118]. Therefore, mouse ESCs have been used extensively to conceive genetically engineered mouse lines in various fields of research.

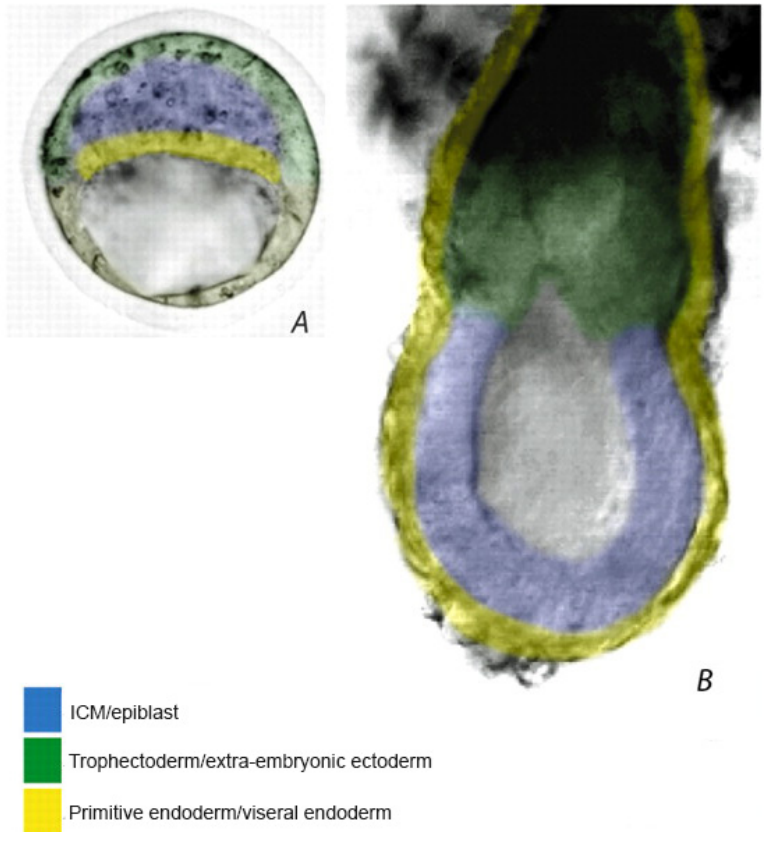


Figure 5 | Mouse blastocyst and implanted embryo.
a) Mouse blastocyst around developmental day 4.5. Trophectoderm (green) Inner cell mass (blue) b) Implanted mouse embryo around developmental day 5.5. Epiblast (blue) Adapted from [232].

For example, by the use of homologous recombination it is possible to engineer clonal genome-edited (e.g. knock-in or knock-out) ESCs lines for a gene of interest that can be used to generate mice for studying gain- or loss-of-function *in vivo* [119].

ESCs are also a valuable model system to study early embryonic development and cell lineage specification. The unspecialized state of ESCs allows for the determination of factors and environmental cues that are required to differentiate into a certain somatic cell lineage [120]. Moreover, this also allows for the investigation of a certain cell type that is difficult to obtain otherwise. For instance, mouse ESCs were used to generate neurons following a controlled schedule of cytokines and growth factors [121]. ESCs are also widely used in a range of studies from large genomic protein-DNA interaction studies to focused mechanistic studies of individual regulators. The ENCODE consortium has included both human and mouse ESCs as model systems in their investigation for the functional annotation of the genome [104,122].

Culture conditions of ESCs are critical for self-renewal and maintenance of the pluripotent state. Mouse ESCs were originally established on a layer of supporting cells in the presence of serum [117,118]. As supporting cells, mouse skin fibroblasts were used, which were irradiated to avoid their proliferation in the ESC culture. A few years later it was discovered that the feeder cells and the serum can be replaced by the addition of the cytokine leukemia inhibitory factor (LIF) and low concentrations of bone morphogenetic protein 4 (BMP4) [123,124]. The presence of LIF in the culture medium was found to stimulate several signaling pathways that all result in the phosphorylation of the effector protein Stat3. The activation of Stat3 has been shown to maintain ESCs and prevent differentiation by inducing Kruppel-like factor 4 (Klf4) [37,125]. BMPs, in particular BMP4, appeared to be the beneficial ingredient in serum to maintain ESC self-renewal. The binding of BMP4 to its receptor activates the expression of genes encoding basic helix-loop-helix type inhibitors of differentiation (Ids) such as Id1. Overexpression of Id1 in mouse ESCs results in self-renewal in the absence of BMP4 or serum [126]. However, serum is still commonly used in ESC culture due to the relatively high costs of purified BMP4 proteins.

Feeder cells and serum were also used to isolate the first ESCs from human blastocysts [127]. However, it became clear that in order to maintain human ESCs, different conditions are required. Instead of LIF/BMP4 stimulated pathways, it was discovered that hESCs depend on FGF2 and Activin signaling pathways to maintain self-renewal [128]. This difference was first attributed to species divergence. However, later it was found that mouse epiblast stem cells (EpiSC) could be derived under these same conditions [129]. Compared to mouse ESCs that are generated from pre-implantation blastocysts, mouse EpiSC are isolated from a post-implantation stage blastocyst and therefore represent a more “primed” state towards differentiation [130]. Human ESCs are more similar to mouse EpiSCs with respect to global gene expression and growth characteristics [131]. The EpiSC-like state of human ESCs is accompanied by low single cell viability and karyotype instability [132,133]. However, several studies demonstrated the generation of naïve human ESCs using cellular reprogramming techniques, although these cells were still difficult to maintain [134,135]. A recent study defined the optimal conditions necessary to derive naïve human ESC lines from primed ESC lines and human blastocysts by the use of a cocktail of selected cytokines and small molecule inhibitors [136]. These naïve ESCs were stable and displayed an epigenetic landscape resembling that of naïve mouse ESCs. The establishment of these conditions provides for the derivation of stable naïve human ESCs that are a valuable tool for genetic modification and future research.

Recently, it was discovered that mouse ESCs can be cultured and maintained in a medium without LIF and serum. Instead, a medium is used that contains two small molecules that block the FGF receptor pathway and activate Wnt signaling, by specifically inhibiting the kinase activity of MEK and GSK3. This culture condition is referred to as

2i or the ground state (see below), and demonstrates that LIF/Stat3 signaling is not essential for ESC maintenance [137]. However, ESCs remain sensitive to LIF signaling and 2i conditions supplemented with LIF showed increased ESC derivation and cloning efficiency. Therefore, 2i plus LIF conditions are preferred for the generation of new ESC lines [138]. Several studies investigated the differences in ESC characteristics between both culture conditions (reviewed in [139]). ESCs cultured in 2i conditions appear more morphologically homogenous compared to conventional ESCs cultured in LIF/Serum containing medium. On the molecular level there are several differences, such as the near absence of global DNA methylation and fewer bivalent chromatin marks [140,141]. These findings led to the suggestion that 2i-cultured ESCs represented a more naïve state, or "ground state" compared to conventional cultured ESCs [142]. The existence of mouse EpiSCs that are more primed towards differentiation, suggests that conventional ESCs grown on LIF and serum are in an intermediate state between that of ESCs cultured on 2i and the EpiSCs. However, ESCs cultured on 2i or LIF and serum are shown to have the same differentiation potential [140]. Future investigations will be needed to gain more insight in the relevance of the changes between the states of various ESC types.

Transcriptional regulation of pluripotency

ESCs receive extrinsic signals that stimulate self-renewal or cause cell lineage specification. However, there is also an intrinsic network of transcription factors at play to maintain the pluripotent state. This network has a core regulatory circuitry that consists of three transcription factors; Oct4, Sox2 and Nanog [143]. In the center of this network is Oct4, which is encoded by the *Pou5f1* gene and contains a POU domain by which it interacts with the DNA [144]. Oct4 is essential in early development and maintenance of ESCs *in vitro*, as loss of function experiments result in differentiation into trophoblasts [145,146]. Interestingly, overexpression of Oct4 in ESCs also results in differentiation, which is an indication that besides pluripotency genes it also regulates lineage specification genes [146].

Oct4 does not act alone in maintaining pluripotency and physically interacts with a myriad of factors as shown by several interaction studies [147-149]. The best-characterized partner of Oct4 is the high-mobility-group (HMG) box containing transcription factor Sox2. Their cooperation was first demonstrated on an enhancer of *Fgf4* [150]. Oct4 and Sox2 interact with each other via their DNA binding domains and together bind the DNA on a joined motif [151,152]. Sox2 knock-out mice are post-implantation lethal and knock-down in ESCs leads to a similar phenotype as seen in the Oct4 knock-down ESCs, namely differentiation to trophoblasts. In Sox2 knock-out ESCs, pluripotency is rescued by overexpression of Oct4. This suggest that the unique role of Sox2 is to activate Oct4 expression and Sox2 can be replaced by other close Sox family members at different Oct4/Sox2 targets genes [153].

Another essential factor within the core circuitry maintaining pluripotency in

ESCs is the homeodomain-containing protein, Nanog. As is the case for Oct4 and Sox2, knock-out mice for Nanog are early embryonic lethal due to the absence of the epiblast [154]. However, when Nanog *null* ESCs were generated they could be maintained and were able to fully contribute to chimeras [155]. Therefore, Nanog is essential for the formation of ESCs but can be omitted once they are formed. Overexpression of Nanog enables ESCs to autonomously self-renew, even in the absence of LIF, BMP4 or 2i. Nanog is therefore considered a moderator for pluripotency and part of the core circuitry [155].

Aside from the core regulatory circuitry, many other factors act in concert with Oct4, Sox2 and Nanog to maintain ESC pluripotency. They are part of an outer network that sustains self-renewal and controls lineage specification. Nuclear receptor Esrrb, T-cell leukemia oncogene 1 and T-box transcription factor 3 were identified in an RNA interference (RNAi) screen and shown to be implicated in maintenance of self-renewal [156]. Follow up studies demonstrated that Esrrb interacts and colocalizes with Oct4 and Nanog on target genes [157,158]. A search for downstream operators of LIF signaling identified Kruppel-like factor Klf4 [159]. Later it was demonstrated that Klf4 is a direct target of Stat3 thereby underlining its role in ESC maintenance [160]. Many other regulators of ESC maintenance have been identified through genetic or proteomic screens [148,149,161,162]. The co-factors Mediator and Cohesin are also essential for maintaining pluripotency of ESCs [94]. Although the function of some factors is not fully understood, they ultimately regulate or facilitate various steps within the transcription process.

The core regulatory circuitry controls the maintenance of ESCs, but can also respond to signals of cell lineage specification. This is achieved through the transcription factors Oct4, Sox2 and Nanog that maintain the expression of one another through an interconnected auto-regulatory loop. Secondly, they activate expression of a large variety of genes needed in maintaining the ESC state while they repress cell lineage specification genes preventing differentiation. Appropriate expression levels of each the core transcription factors are crucial to sustain the balance between self-renewal and differentiation [163]. This allows for a quick reaction to environmental cues to engage cell lineage expression programs that lead to differentiation. The core transcription factors co-occupy the same binding sites in the genome that were shown to have enhancer activity [164]. Additional factors, part of the outer circuitry, often colocalize to these binding sites where they facilitate either in the activation, poising or repression of the genes associated with enhancer.

The ability of ESCs to self-renew and maintain pluripotency requires genes that stimulate lineage specification to be repressed and susceptible for activation at the same time. Apart from the core regulatory circuitry, transcriptional activation or repression in ESCs is also tightly regulated by chromatin structure. ESCs and other stem cells (e.g. hematopoietic stem cells [165]) have a relatively “open” chromatin structure compared to

somatic cells [166]. This phenomenon is observed in electron microscopy examinations as well as multiple biochemical assays [167,168]. In one study it was demonstrated that chromatin compaction increased upon *in vitro* stimulated differentiation with retinoic acid of mouse ESCs, which suggested that chromatin becomes compacted during specification [169]. Another line of evidence is the relatively low level of total repressive chromatin marks such as H3K9me3 that was detected when compared to differentiated cells. Consistent with the increase of chromatin compaction upon ESCs differentiation an increase of heterochromatic proteins such as HP1 was observed as well [170]. One chromatin regulator that is essential for maintaining the "open" chromatin state is the ATP-dependent chromatin remodeler Chd1 [38]. Knock-down of Chd1 by RNAi showed an increase in heterochromatin formation and resulted in a loss of pluripotency in mouse ESCs.

A direct consequence of the open chromatin state in ESCs is hyperactive transcription that occurs throughout the genome [171]. To avoid transcriptional noise, transcriptional activation and repression of genes is regulated on a local chromatin level at individual genes. This is achieved by several mechanisms in which the core transcription factors cooperate with general chromatin regulators that alter chromatin structure through DNA methylation, remodeling, modification and also use non-coding RNAs. Therefore, many chromatin regulators and GTFs are highly expressed in ESCs and are essential for viability or controlled differentiation. For example, Oct4 was shown to interact and colocalize with SetDB1 at gene promoters. Here, SetDB1 establishes H3K9me3 to locally repress promoters of cell lineage specification genes. Knockdown of SetDB1 led to a derepression of these developmental regulator genes and therefore differentiation [172].

Another mechanism to locally repress gene transcription involves the Polycomb group (PcG) class chromatin modifier complexes. The PcG repressor complexes PRC1 and PRC2 co-occupy promoters of key developmental regulator genes together with the core transcription factors [173]. The enzymatic activity of the PRC2 complex specifically establishes H3K27me3, which is recognized by the PRC1 complex. PRC1 specifically mono-ubiquitinates H2AK119 which leads to the inhibition of Pol II activity [102]. However, the promoters of these cell lineage specification regulators need to be rapidly activated upon developmental cues and therefore silencing of these regions is counteracted by the placement of active transcription marks such as H3K4me3 by mammalian homologues of the Trithorax group proteins [174]. These regions marked by active and repressive histone modifications are referred to as bivalent domains and were first discovered in mouse ESCs [101]. The associated genes are poised for future expression and were shown to be rapidly induced upon environmental cues that initiate differentiation.

During cell lineage specification, bivalent domains found in ESCs lose one of the modifications over time which results in gene activation or repression. However, new bivalent domains were observed to form upon differentiation and therefore these domains

are not strictly limited to ESCs and were also found in other stem cells [175,176]. Moreover, later studies demonstrated that bivalent domains were not only observed at promoters but are also present within poised enhancers that regulate cell specification genes [177]. Although the PcG group proteins are required for the establishment of bivalent domains and gene silencing they were found not to be essential for the maintenance of ESCs [178]. Loss-of-function experiments for components of the PRC complexes demonstrated that PcG deficient ESCs are found to have cell lineage specification defects. For example, mouse ESCs that were deficient for PRC2 failed to form teratocarcinomas indicating that PRC components are essential in executing differentiation programs appropriately [179].

In addition to the repression by SetDB1 and PcG complexes, several other chromatin remodeling/modifying complexes have been identified to be essential for maintenance of the pluripotent state of ESCs. A variant of the mammalian SWI/SNF complex was identified in mouse ESCs [26]. This esBAF complex contains both Brg1 and Arid1a that have been both shown to be essential for ESC viability [180]. The complex facilitates activation of transcription and was found to be localized on promoters that were also bound by the core transcription factors Oct4, Sox2 and Nanog [181]. Interactions between the core transcription factors and subunits of the NuRD complex have also been described [182]. The core subunit of the NuRD complex is the methyl-CpG binding domain protein 3 (Mbd3). The core pluripotency TFs recruit Mbd3/NuRD to important pluripotency genes (including themselves), where it modulates transcriptional output. The balance between activation by the core TFs and repressive activity of NuRD results in a heterogenic transcriptional state. This allows ESCs to have a homogeneous appearance but remain responsive to cues that cause differentiation [183]. Another chromatin regulator complex that is implicated in ESC pluripotency is Tip60/Ep400 [184]. The subunits Tip60 and Ep400 have acetyltransferase and remodeling activity. The complex was found to be targeted to promoters by Nanog but they also recognize H3K4me3 via the Ing3 subunit [185]. At these promoters, they can facilitate gene activation. However, Tip60/Ep400 was also shown to regulate repression at bivalent domains [185].

The importance of key transcription factors in establishing pluripotency in ESCs was demonstrated in 2006, when the pioneering work of Yamanaka and colleagues demonstrated that ectopic expression of only four transcription factors, Oct4, Sox2, Klf4 and c-Myc (OSKM) in mouse embryonic and adult skin fibroblasts was able to reprogram these cells into an ESC-like state [186]. These induced pluripotent stem cells (iPSCs) were similar to ESCs and shown to be able to differentiate into all cells from the three germ layers in teratoma assays. Similar studies demonstrated that when iPSCs were injected into blastocysts they contribute to all tissues of the developing embryo and were able to form viable animals [187,188]. This discovery led to a revolution in stem cell research and the reprogramming capability of TFs had become a new standard of investigating their transcriptional network. Different variations on the original experiment were used and

demonstrated that many other factors and different somatic cells from various species, including human, could be used to generate iPSCs (reviewed in [189]). The core factor Oct4 was irreplaceable in all these reprogramming experiments. However, a recent study demonstrated that it was possible to reprogram even without Oct4 by using somatic lineage factors [190]. The idea behind this experiment is that pluripotent stem cells balance between two cell states (mesoderm and ectoderm). Induction of somatic cells with factors for these two states at the same time will induce a pluripotent equilibrium state therefore omitting the need for the core factor of pluripotency, Oct4. Cell reprogramming is not limited for induction to the pluripotent state and can also be used to generate other types of cells. For example, direct reprogramming of somatic cells into NSCs was demonstrated by overexpression of a different set of factors where Oct4 was replaced with other neural specific Oct factors, Brn2/4 [191].

Reprogramming of somatic cells to iPSC and other cells has tremendous potential for medical applications. Patient cells could be reprogrammed and genetically treated *in vitro* to alter a genetic defect. Cured cells could then be transplanted back into diseased patients. This ability to generate cells for transplantation therapies that originate from the host will eliminate rejection and the need for immune suppressive drugs. However, reprogramming techniques need to be improved due to the tumorigenic potential of reprogrammed cells [192]. Reprogrammed cells are also an invaluable tool to study disease. Somatic cells from patients can be used as model systems to test compounds and treatments that could aid in curing disease or alleviate symptoms.

Neural stem cells

During early stages of development, the pluripotent stem cells that make up the ICM in the blastocyst differentiate into cells of one of the three germ layers. These cell layers will eventually generate all the tissues within the embryo. The cells transit through several developmental stages and become more specialized, losing their pluripotency. The vertebrate nervous system is conceived from the ectoderm, which is further induced to specify into neuroectoderm, which forms the neural plate [193]. In the last phase of gastrulation, the neural plate folds to form the neural tube from which all cells of the nervous system are derived and eventually gives rise to the brain and spinal cord. During this developmental process different neural progenitor cell populations arise in the nervous system, which contribute to generating all the cell types necessary for the formation of the central and peripheral nervous system. Some of these NSC populations are maintained during adult life where they function to regenerate and replenish the tissue during life and upon injury [194]. These adult NSCs reside in specialized stem cell niches in the brain, such as the subventricular zone (SVZ) and the hippocampus [195]. Embryonic and adult NSCs are characterized by being capable to self-renew and are multipotent. They can give rise to the

three neural cell lineages; neurons, astrocytes and oligodendrocytes, whilst maintaining their own population [193].

The ability of NSCs to self-renew allows for long-term *in vitro* cell culture under conditions that mimic the cellular microenvironment or stem cell niche. Mouse NSCs can be obtained from three main sources. First, they can be isolated from the brain and spinal cord of developing embryos [195]. Secondly, adult NSCs can be isolated from the adult mouse brain sections that contain germinal centers, such as the SVZ [196]. The third source for obtaining NSCs in culture is the derivation of NSCs from pluripotent cells such as ESCs or iPSCs.

There are many different types of NSCs that vary in their molecular characteristics and differentiation potential [197]. The derivation of a specific NSC population depends on the isolation procedure and/or conditions they are cultured in (reviewed in [193]). The NSCs used in this thesis were derived from 46C ESCs according to a serum-free monolayer protocol designed in the laboratory of Austin Smith [198]. The process of neural differentiation can be followed with these 46C ESCs, by the expression of a Sox1-GFP reporter [199]. Sox1 is an early marker of neural differentiation that is expressed during neuroectoderm specification [200]. The protocol stimulates ESCs to form heterogeneous cellular aggregates containing neural progenitors. After formation of floating neurospheres, they will settle and attach to culture flasks. This allows for outgrowth of neural stem cells. These NSCs (**Fig. 6**) are maintained and propagated in the presence of the growth factors fibroblast growth factor 2 (FGF2) and epidermal growth factor (EGF). The adherent NSCs grow as monolayers on gelatin coated culture and have a bipolar morphology. They are characterized by the expression of NSC markers such as Pax6 and Nestin, but do not show expression of markers that are associated with neuronal or astrocyte differentiation such as Beta-tubulin and GFAP. These NSCs self-renew and are multipotent as shown by their differentiation into neurons, astrocytes and oligodendrocytes [201]. They have a high clonogenicity and can be grown in sufficient amounts to perform biochemical studies as described in this thesis.

NSCs have become of special interest for stem cell transplantation therapies, in which the regenerative potential of NSCs can be utilized to repair damage after injury in the nervous system. Also NSCs serve as an excellent model system to study the molecular background of neurological disorders and could aid the development of drug treatments that prevent their upcoming or alleviate symptoms.

Sox2 as a transcriptional regulator in NSCs

As described above, there is a thorough understanding of transcriptional control of pluripotency in ESCs by the core circuitry factors Oct4, Sox2 and Nanog. However, little is known about transcriptional control of self-renewal and multipotency in NSCs. Surprisingly, it was found that the transcription factor Sox2 is not only essential for early

development and maintenance of ESCs, but is also necessary for maintaining the balance between self-renewal and differentiation in NSCs. Before the early functions of Sox2 were discovered, conserved expression of Sox2 was observed in the developing nervous system of multiple species [202]. However, experiments to investigate the functions of Sox2 later in development were severely hampered due to the early embryonic lethality of Sox2 knock-out mice [203]. Independent investigations using tissue specific disruption of Sox2 in the nervous system demonstrated that Sox2 is essential to maintain NSC populations and subsequent neurogenesis in adult brains of mice [204,205]. Further, it was demonstrated that overexpression of Sox1-3 prevents NSCs to differentiate by counteracting the activity of pro-neural factors in chick embryos. These same pro-neural factors can induce neurogenesis by suppressing the function of Sox1-3 [206]. Sox2 is not only important for ESCs and NSCs, but

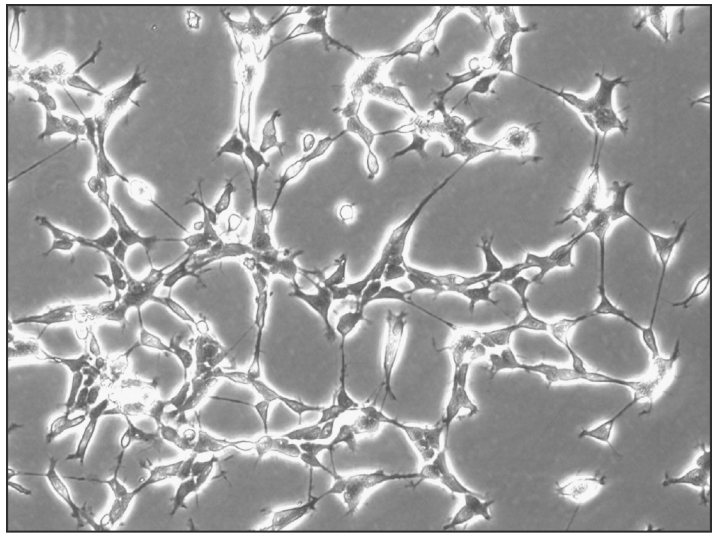


Figure 6 | Adherent NSC culture.
Bright-field image of a 46C ESC derived adherent NSC culture. Magnification 20x.

has been shown to be essential in several other adult stem cell and progenitor populations in the gastrointestinal and respiratory tract as well as in the sensory organs [207]. Therefore, Sox2 is a transcription factor that functions at many different levels of various cell types by directing the balance between self-renewal and differentiation.

Sox2 and the Sox family

Sox2 is a member of the family of Sox transcription factors with diverse roles during development. The Sox family is classified under the superfamily of HMG domain containing proteins. The HMG domain is composed of three alpha helices that form an L-shaped structure that allows for binding to the minor groove of the DNA helix. This interaction leads

to widening of the minor groove and compression of the major groove, that causes bending of the DNA [208]. Therefore, proteins containing an HMG domain can alter the conformation of the DNA which increases the accessibility for other proteins and the plasticity of the DNA. This ability is demonstrated to be essential in the formation of enhanceosomes that are complexes of transcription factors associated on enhancer elements [209].

The HMG superfamily can be divided into two groups of proteins, based on the number of HMG domains present in the protein and the ability of the HMG domain to bind DNA in a sequence-specific manner. The first group consists of proteins that contain multiple canonical HMG domains that bind to DNA with little or no sequence specificity [210]. These proteins, such as Ubf or HmgB, are ubiquitously expressed and involved in general chromatin-related processes where they serve as architectural facilitators [211,212]. The Sox family belongs to the second group within the HMG superfamily together with the related Tcf/Lef family. This group is characterized by the presence of a single non-canonical HMG domain, which is also referred to as the HMG box. This HMG box has the ability to bind DNA in a sequence-specific manner and to bend DNA at variable angles, as opposed to fixed angle [213].

The first identified member of the *Sox* gene family was identified in mouse. The gene was found to be responsible for male sex determination and is located in a region on the Y chromosome, hence the name *Sry* [214]. Further on, similar genes were identified based on sequence homology of the *Sry* HMG box and were named *Sry*-related HMG box containing genes which gave the family its name; Sox. For this reason, Sox proteins share approximately 50% or higher homology within the amino acid sequence of their HMG boxes. As new *Sox* genes were identified throughout the animal kingdom they were given a number upon their discovery [215]. In mouse and human, there are 20 Sox genes known, which are divided into 8 subgroups (designated A to H) based on homology in and outside their HMG boxes (Table 2)[209].

Members of each subgroup share at least 80% homology between their HMG boxes and have other conserved regions within their coding sequence. For instance, Sox2 that belongs to the SoxB1 group shares an identical amino acid sequence on the C-terminal side of its HMG box with the other members Sox1 and Sox3 [216]. Consequently, they have similar biochemical properties and members from the same group are often redundant when co-expressed [217]. Members from various groups share very little sequence identity outside their HMG boxes. These regions can contain transactivation or repression domains and also regions that facilitate protein-protein interactions are described [209]. The additional domains outside the HMG box account for the different functionality between subgroups. This is well demonstrated in the SoxB group proteins. Sox2 and the other SoxB1 group members contain a transcriptional activation domain while SoxB2 members such as Sox21 contain a repression domain. SoxB1 members are involved in maintaining neural progenitor

Table 2 | Sox family subgroups in mouse and human.
Overview of subgroups and the corresponding Sox proteins. Human orthologs of mouse Sox12 and Sox15 are named SOX22 and SOX20.

Subgroup	Protein
A	SRY
B1	Sox1, **Sox2,** Sox3
B2	Sox14, Sox21
C	Sox4, Sox11, Sox12 (SOX22)
D	Sox5, Sox6, Sox13
E	Sox8, Sox9, Sox10
F	Sox7, Sox17, Sox18
G	Sox15 (SOX20)
H	Sox30

populations by inhibiting neurogenesis. Expression of Sox21 has an antagonizing effect on the SoxB1 members and stimulates neurogenesis by repressing the SoxB1 members and their target genes [218].

All Sox proteins recognize and bind the DNA through a common consensus motif, 5'-(A/T)(A/T)CAA(A/T)G-3' [219]. This motif is so short and erratic that it is abundant throughout the whole genome. Furthermore, the affinity for Sox proteins to bind to DNA is relatively low compared to most other transcription factors [209]. Since there is no discrimination between Sox proteins to bind this motif, binding specificity must be influenced by other factors. Indeed, there are several ways by which differential binding by Sox factors is regulated, such as variability of the sequences flanking the consensus motif, dimerization of Sox proteins or most consequential; hetero-dimerization with other transcriptional regulators. This is nicely demonstrated by the action of Sox2 in ESCs compared to NSCs. In ESCs, Sox2 interacts with Oct4 to regulate pluripotency. Upon differentiation towards NSCs, Sox2 switches its interaction partner with another Oct factor to specify a NSC state [220]. The interaction with Brn2 causes the localization of Sox2 on the genome to change to other enhancer elements to regulate a set of genes important in NSCs. Overexpression of Brn2 in ESCs leads to a functional recruitment of Sox2 to a subset of NSC specific target genes and results in differentiation into the neural fate. Therefore, Sox2 heterodimerization is an important determinant of Sox2 function and explains how Sox2 can fulfill its multiple roles in various cell types.

The conserved motif recognition of Sox proteins is utilized in their role as pioneer factors. For example, sequential action of Sox2, Sox3 and Sox11 is important for the differentiation of ESCs to maturing neurons [101]. In ESCs, Sox2 localizes to neural-specific genes that become marked by bivalent chromatin marks. These same genes are bound by

Sox3 after differentiation into NSCs. Upon further specification, Sox3 bound genes in NSCs are subsequently bound and activated by Sox11 in mature neurons.

As mentioned above, Sox2 is expressed in various stem cell populations and is involved in the specification of different cell types. A common theme of cell fate specification by Sox2 is the antagonizing effect it has on transcription factors of alternative cell lineages [221]. For instance, Sox2 is involved in foregut development, where it specifies the formation of the esophagus and stomach [222]. During this process, it antagonizes the transcription factors Nkx2.1 and Cdx2 to establish the borders between the esophagus and trachea and stomach and intestine, respectively [222,223]. However, it is important to recognize that these antagonisms are developmental stage and cell type specific. In some cases Sox2 can antagonize one factor in one cell type and cooperate with it in a different cellular setting. A good example is the cooperation of Sox2 and Pax6 in lens development, where Sox2 and Pax6 colocalize on enhancers of the *Crystalline* gene [224]. While in optic cup progenitors Sox2 antagonizes Pax6 to specify a neurogenic fate opposed to non-neurogenic epithelium fate specification by Pax6 [225].

Sox2 functions are diverse and range from early to late development, during which it regulates the maintenance of various stem cell populations and is involved in different tissue-specification events. In humans, heterozygous mutations in *SOX2* cause SOX2 anophthalmia-esophageal-genital (AEG) syndrome that displays symptoms in the various organs SOX2 is expressed [226,227]. This syndrome is very rare (prevalence is approximately 1:250,000) and is characterized by the appearance of microphthalmia or anophthalmia (meaning patients have small or absent eyes). Severe neurological defects, such as brain malformations and mental retardation, are also associated with AEG syndrome. Surprisingly, mice that are heterozygous for *Sox2* show a mild phenotype with only minor pituitary defects resulting in reduced hormone secretion [228]. However, similar symptoms were demonstrated between patients with AEG and mice that are tissue-specific hypomorphic for *Sox2* [222,229]. Sox2 is also increasingly implemented as an oncogene in various types of cancer, in which misregulation of Sox2 often results in proliferation and anti-differentiation phenotypes [221]. Therefore, molecular investigation of Sox2 function is essential for understanding its role in transcriptional regulatory pathways in stem cells.

References

1. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
2. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
3. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
4. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**, 233-45 (2012).
5. Juven-Gershon, T. & Kadonaga, J.T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* **339**, 225-9 (2010).
6. Yamaguchi, Y. et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41-51 (1999).
7. Adelman, K. & Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-31 (2012).
8. 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 (2007).
9. Seila, A.C. et al. Divergent transcription from active promoters. *Science* **322**, 1849-51 (2008).
10. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-8 (2008).
11. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D.L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* **15**, 71-8 (2008).
12. Moore, M.J. & Proudfoot, N.J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700 (2009).
13. Danko, C.G. et al. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* **50**, 212-22 (2013).
14. Gilchrist, D.A. et al. Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. *Genes Dev* **26**, 933-44 (2012).
15. Boettiger, A.N. & Levine, M. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science* **325**, 471-3 (2009).
16. Rahl, P.B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432-45 (2010).
17. Li, J. et al. Kinetic Competition between Elongation Rate and Binding of NELF Controls Promoter-Proximal Pausing. *Mol Cell* **50**, 711-22 (2013).
18. Li, J. & Gilmour, D.S. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J* (2013).
19. Gilchrist, D.A. et al. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev* **22**, 1921-33 (2008).
20. Kornberg, R.D. & Thomas, J.O. Chromatin structure; oligomers of the histones. *Science* **184**, 865-8 (1974).
21. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60 (1997).
22. Lusser, A. & Kadonaga, J.T. Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* **25**, 1192-200 (2003).
23. Hargreaves, D.C. & Crabtree, G.R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* **21**, 396-420 (2011).
24. Whitehouse, I. et al. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature* **400**, 784-7 (1999).
25. Lorch, Y., Zhang, M. & Kornberg, R.D. Histone octamer transfer by a chromatin-remodeling complex. *Cell* **96**, 389-92 (1999).
26. Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**, 5181-6 (2009).
27. Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during

neural development. *Neuron* **55**, 201-15 (2007).
28. Erdel, F. & Rippe, K. Chromatin remodelling in mammalian cells by ISWI-type complexes--where, when and why? *FEBS J* **278**, 3608-18 (2011).
29. Cairns, B.R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193-8 (2009).
30. Whitehouse, I. & Tsukiyama, T. Antagonistic forces that position nucleosomes in vivo. *Nat Struct Mol Biol* **13**, 633-40 (2006).
31. Kagalwala, M.N., Glaus, B.J., Dang, W., Zofall, M. & Bartholomew, B. Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *EMBO J* **23**, 2092-104 (2004).
32. Conaway, R.C. & Conaway, J.W. The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem Sci* **34**, 71-7 (2009).
33. Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J. & Maltais, L.J. The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-98 (2002).
34. Thakar, A. et al. H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry* **48**, 10852-7 (2009).
35. Venters, B.J. & Pugh, B.F. How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol* **44**, 117-41 (2009).
36. Murawska, M. & Brehm, A. CHD chromatin remodelers and the transcription cycle. *Transcription* **2**, 244-53 (2011).
37. Hall, J.A. & Georgel, P.T. CHD proteins: a diverse family with strong ties. *Biochem Cell Biol* **85**, 463-76 (2007).
38. Gaspar-Maia, A. et al. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863-8 (2009).
39. Bouazoune, K. & Kingston, R.E. Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proc Natl Acad Sci U S A* **109**, 19238-43 (2012).
40. Thompson, B.A., Tremblay, V., Lin, G. & Bochar, D.A. CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol* **28**, 3894-904 (2008).
41. Hermann, A., Gowher, H. & Jeltsch, A. Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci* **61**, 2571-87 (2004).
42. Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010-22 (2011).
43. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-20 (2013).
44. Roloff, T.C., Ropers, H.H. & Nuber, U.A. Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics* **4**, 1 (2003).
45. Attwood, J.T., Yung, R.L. & Richardson, B.C. DNA methylation and the regulation of gene transcription. *Cell Mol Life Sci* **59**, 241-57 (2002).
46. Tan, M. et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-28 (2011).
47. Shiio, Y. & Eisenman, R.N. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225-30 (2003).
48. Pogo, B.G., Allfrey, V.G. & Mirsky, A.E. RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proc Natl Acad Sci U S A* **55**, 805-12 (1966).
49. Wang, Z. et al. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019-31 (2009).
50. Waterborg, J.H. Dynamics of histone acetylation in vivo. A function for acetylation turnover? *Biochem Cell Biol* **80**, 363-78 (2002).
51. Clapier, C.R. & Cairns, B.R. The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**, 273-304 (2009).
52. Kunert, N. & Brehm, A. Novel Mi-2 related ATP-dependent chromatin remodelers. *Epigenetics* **4**, 209-11 (2009).

53. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett* **586**, 2692-704 (2012).
54. Holmqvist, P.H. & Mannervik, M. Genomic occupancy of the transcriptional co-activators p300 and CBP. *Transcription* **4**, 18-23 (2013).
55. Wang, Z. et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903 (2008).
56. Canzio, D. et al. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol Cell* **41**, 67-81 (2011).
57. Zentner, G.E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20**, 259-66 (2013).
58. Lauberth, S.M. et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021-36 (2013).
59. Schmitges, F.W. et al. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-41 (2011).
60. Jenuwein, T. The epigenetic magic of histone lysine methylation. *FEBS J* **273**, 3121-35 (2006).
61. Albert, M. & Helin, K. Histone methyltransferases in cancer. *Semin Cell Dev Biol* **21**, 209-20 (2010).
62. Fischle, W. et al. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**, 1870-81 (2003).
63. Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. & Patel, D.J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**, 1025-40 (2007).
64. Venkatesh, S. et al. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452-5 (2012).
65. Maltby, V.E. et al. Histone H3 lysine 36 methylation targets the Isw1b remodeling complex to chromatin. *Mol Cell Biol* **32**, 3479-85 (2012).
66. Black, J.C., Van Rechem, C. & Whetstine, J.R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**, 491-507 (2012).
67. Shi, Y. et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-53 (2004).
68. Metzger, E. et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436-9 (2005).
69. Rais, Y. et al. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65-70 (2013).
70. Cloos, P.A., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev* **22**, 1115-40 (2008).
71. Agger, K. et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731-4 (2007).
72. Pasini, D. et al. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* **22**, 1345-55 (2008).
73. Ruthenburg, A.J., Li, H., Patel, D.J. & Allis, C.D. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* **8**, 983-94 (2007).
74. Spitz, F. & Furlong, E.E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613-26 (2012).
75. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-63 (2009).
76. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-37 (2013).
77. Li, L.M. & Arnosti, D.N. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. *Curr Biol* **21**, 406-12 (2011).
78. Phillips-Cremins, J.E. & Corces, V.G. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**, 461-74 (2013).
79. Akbari, O.S. et al. A novel promoter-tethering element regulates enhancer-driven gene expression at

the bithorax complex in the Drosophila embryo. *Development* **135**, 123-31 (2008).
80. Zaret, K.S. & Carroll, J.S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-41 (2011).
81. Hartley, P.D. & Madhani, H.D. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445-58 (2009).
82. Pevny, L. & Placzek, M. SOX genes and neural progenitor identity. *Curr Opin Neurobiol* **15**, 7-13 (2005).
83. Bergsland, M. et al. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* **25**, 2453-64 (2011).
84. Heintzman, N.D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).
85. Serandour, A.A. et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**, 555-65 (2011).
86. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
87. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-70 (2008).
88. Jeong, K.W. et al. Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. *Nat Struct Mol Biol* **18**, 1358-65 (2011).
89. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-4 (2011).
90. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
91. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
92. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-65 (2002).
93. Takahashi, H. et al. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**, 92-104 (2011).
94. Kagey, M.H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-5 (2010).
95. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
96. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).
97. Collas, P. The current state of chromatin immunoprecipitation. *Mol Biotechnol* **45**, 87-100 (2010).
98. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83 (2011).
99. Zentner, G.E., Tesar, P.J. & Scacheri, P.C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**, 1273-83 (2011).
100. Kerppola, T.K. Polycomb group complexes--many combinations, many functions. *Trends Cell Biol* **19**, 692-704 (2009).
101. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).
102. Stock, J.K. et al. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**, 1428-35 (2007).
103. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
104. Consortium, E.P. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
105. Bonn, S. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**, 148-56 (2012).

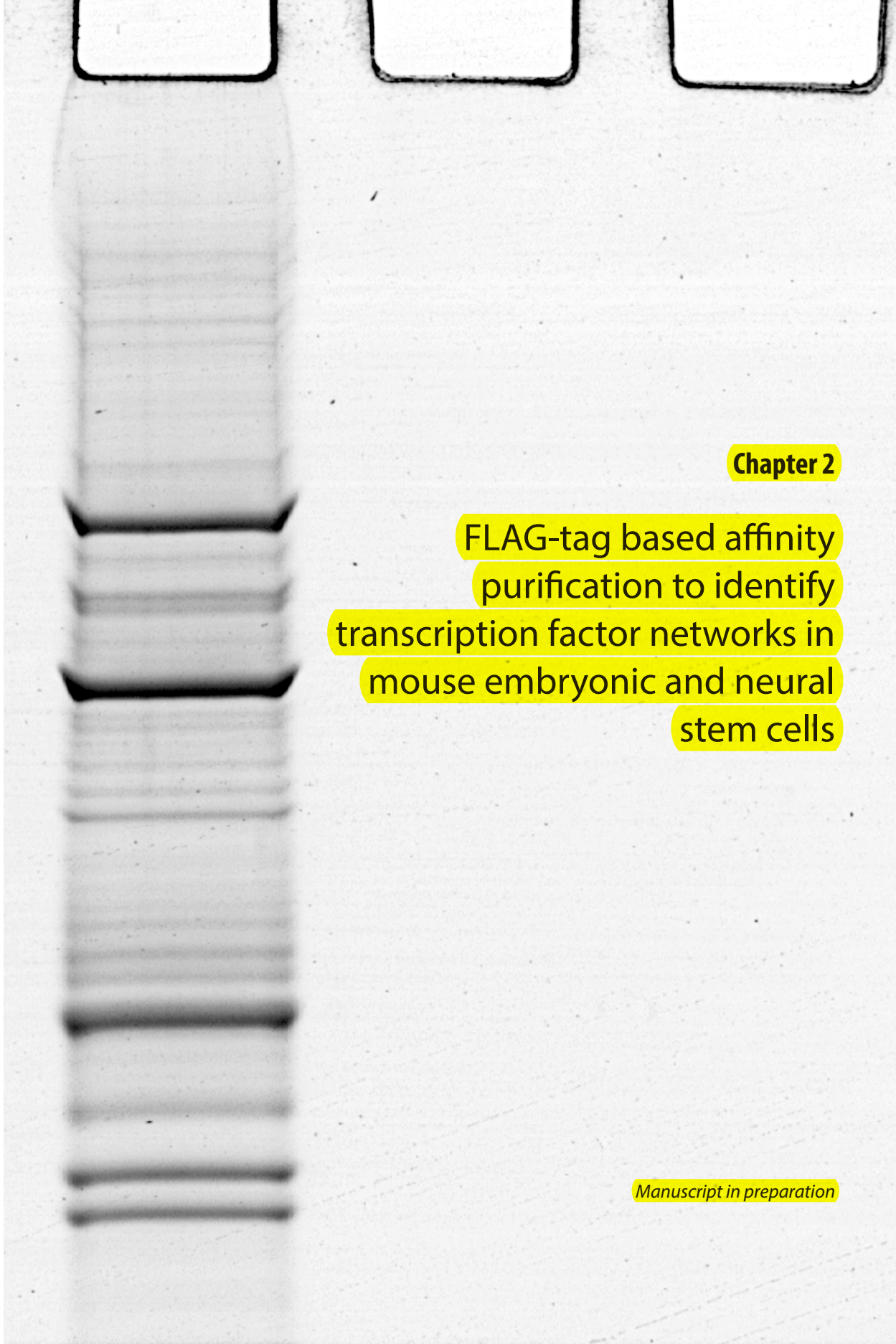
106. Lettice, L.A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-35 (2003).
107. Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-54 (2006).
108. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
109. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc* **2**, 988-1002 (2007).
110. Manolio, T.A. Bringing genome-wide association findings into clinical use. *Nat Rev Genet* **14**, 549-58 (2013).
111. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res* **22**, 1748-59 (2012).
112. Maurano, M.T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
113. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-34 (2011).
114. Dawson, M.A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).
115. Nichols, J. & Smith, A. Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect Biol* **4**, a008128 (2012).
116. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8 (1981).
117. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6 (1981).
118. Bradley, A., Evans, M., Kaufman, M.H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-6 (1984).
119. Tolosa de Talamoni, N. et al. Comparative immunolocalization of the plasma membrane calcium pump and calbindin D28K in chicken retina during embryonic development. *Eur J Histochem* **46**, 333-40 (2002).
120. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-55 (2005).
121. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* **2**, 1034-43 (2007).
122. Mouse, E.C. et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* **13**, 418 (2012).
123. Smith, A.G. et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90 (1988).
124. Ying, Q.L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-92 (2003).
125. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-60 (1998).
126. Romero-Lanman, E.E., Pavlovic, S., Amlani, B., Chin, Y. & Benezra, R. Id1 maintains embryonic stem cell self-renewal by up-regulation of Nanog and repression of Brachyury expression. *Stem Cells Dev* **21**, 384-93 (2012).
127. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-7 (1998).
128. Vallier, L., Alexander, M. & Pedersen, R.A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**, 4495-509 (2005).
129. Brons, I.G. et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-5 (2007).
130. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487-92 (2009).
131. Tesar, P.J. et al. New cell lines from mouse epiblast share defining features with human embryonic

stem cells. *Nature* **448**, 196-9 (2007).
132. Watanabe, K. et al. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat Biotechnol* **25**, 681-6 (2007).
133. Draper, J.S. et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol* **22**, 53-4 (2004).
134. Hanna, J. et al. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A* **107**, 9222-7 (2010).
135. Buecker, C. et al. A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell* **6**, 535-46 (2010).
136. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* (2013).
137. Ying, Q.L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-23 (2008).
138. Blair, K., Wray, J. & Smith, A. The liberation of embryonic stem cells. *PLoS Genet* **7**, e1002019 (2011).
139. Marks, H. & Stunnenberg, H.G. Transcription regulation and chromatin structure in the pluripotent ground state. *Biochim Biophys Acta* (2013).
140. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604 (2012).
141. Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360-9 (2013).
142. Wray, J., Kalkan, T. & Smith, A.G. The ground state of pluripotency. *Biochem Soc Trans* **38**, 1027-32 (2010).
143. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940-54 (2011).
144. Scholer, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. & Gruss, P. New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**, 435-9 (1990).
145. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-91 (1998).
146. Niwa, H., Miyazaki, J. & Smith, A.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-6 (2000).
147. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-8 (2006).
148. van den Berg, D.L. et al. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369-81 (2010).
149. Pardo, M. et al. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **6**, 382-95 (2010).
150. Ambrosetti, D.C., Basilico, C. & Dailey, L. 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**, 6321-9 (1997).
151. Remenyi, A. et al. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**, 2048-59 (2003).
152. Chambers, I. & Tomlinson, S.R. The transcriptional foundation of pluripotency. *Development* **136**, 2311-22 (2009).
153. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-35 (2007).
154. Silva, J. et al. Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-37 (2009).
155. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-4 (2007).
156. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-8 (2006).
157. van den Berg, D.L. et al. Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol Cell Biol* **28**, 5986-95 (2008).
158. Zhang, X., Zhang, J., Wang, T., Esteban, M.A. & Pei, D. Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* **283**, 35825-33 (2008).

159. Li, Y. et al. Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood* **105**, 635-7 (2005).
160. Hall, J. et al. Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. *Cell Stem Cell* **5**, 597-609 (2009).
161. Chia, N.Y. et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316-20 (2010).
162. Hu, G. et al. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* **23**, 837-48 (2009).
163. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).
164. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-17 (2008).
165. Spangrude, G.J., Heimfeld, S. & Weissman, I.L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58-62 (1988).
166. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* **12**, 36-47 (2011).
167. Park, S.H. et al. Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. *Ultrastruct Pathol* **28**, 229-38 (2004).
168. Mattout, A. & Meshorer, E. Chromatin plasticity and genome organization in pluripotent embryonic stem cells. *Curr Opin Cell Biol* **22**, 334-41 (2010).
169. Schaniel, C. et al. Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells* **27**, 2979-91 (2009).
170. Meshorer, E. et al. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-16 (2006).
171. Efroni, S. et al. Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437-47 (2008).
172. Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B. & Young, R.A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* **23**, 2484-9 (2009).
173. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53 (2006).
174. Schuettengruber, B., Martinez, A.M., Iovino, N. & Cavalli, G. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**, 799-814 (2011).
175. Golebiewska, A., Atkinson, S.P., Lako, M. & Armstrong, L. Epigenetic landscaping during hESC differentiation to neural cells. *Stem Cells* **27**, 1298-308 (2009).
176. Oguro, H. et al. Poised lineage specification in multipotential hematopoietic stem and progenitor cells by the polycomb protein Bmi1. *Cell Stem Cell* **6**, 279-86 (2010).
177. Rada-Iglesias, A. et al. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633-48 (2012).
178. Vastenhouw, N.L. & Schier, A.F. Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol* **24**, 374-86 (2012).
179. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**, 3769-79 (2007).
180. Ho, L. & Crabtree, G.R. Chromatin remodelling during development. *Nature* **463**, 474-84 (2010).
181. Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* **106**, 5187-91 (2009).
182. Liang, J. et al. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* **10**, 731-9 (2008).
183. Reynolds, N. et al. NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10**, 583-94 (2012).
184. 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**, 162-74 (2008).
185. Fazzio, T.G., Huff, J.T. & Panning, B. Chromatin regulation Tip(60)s the balance in embryonic stem cell self-renewal. *Cell Cycle* **7**, 3302-6 (2008).
186. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-76 (2006).
187. Wernig, M. et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-24 (2007).
188. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7 (2007).
189. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev* **24**, 2239-63 (2010).
190. Shu, J. et al. Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell* **153**, 963-75 (2013).
191. Han, D.W. et al. Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell Stem Cell* **10**, 465-72 (2012).
192. Lee, A.S., Tang, C., Rao, M.S., Weissman, I.L. & Wu, J.C. Tumorigenicity as a clinical hurdle for pluripotent stem cell therapies. *Nat Med* **19**, 998-1004 (2013).
193. Conti, L. & Cattaneo, E. Neural stem cell systems: physiological players or in vitro entities? *Nat Rev Neurosci* **11**, 176-87 (2010).
194. Gage, F.H. Mammalian neural stem cells. *Science* **287**, 1433-8 (2000).
195. Temple, S. The development of neural stem cells. *Nature* **414**, 112-7 (2001).
196. Pollard, S.M., Conti, L., Sun, Y., Goffredo, D. & Smith, A. Adherent neural stem (NS) cells from fetal and adult forebrain. *Cereb Cortex* **16 Suppl 1**, i112-20 (2006).
197. Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu Rev Neurosci* **32**, 149-84 (2009).
198. Conti, L. et al. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol* **3**, e283 (2005).
199. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-6 (2003).
200. Pevny, L.H., Sockanathan, S., Placzek, M. & Lovell-Badge, R. A role for SOX1 in neural determination. *Development* **125**, 1967-78 (1998).
201. Glaser, T., Pollard, S.M., Smith, A. & Brustle, O. Tripotential differentiation of adherently expandable neural stem (NS) cells. *PLoS One* **2**, e298 (2007).
202. Collignon, J. et al. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* **122**, 509-20 (1996).
203. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-40 (2003).
204. Ferri, A.L. et al. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**, 3805-19 (2004).
205. Cavallaro, M. et al. Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. *Development* **135**, 541-57 (2008).
206. Bylund, M., Andersson, E., Novitch, B.G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* **6**, 1162-8 (2003).
207. Arnold, K. et al. Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* **9**, 317-29 (2011).
208. Zhu, D., Fang, J., Li, Y. & Zhang, J. Mbd3, a component of NuRD/Mi-2 complex, helps maintain pluripotency of mouse embryonic stem cells by repressing trophectoderm differentiation. *PLoS One* **4**, e7684 (2009).
209. Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y. & Pallavi, B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int J Biochem Cell Biol* **39**, 2195-214 (2007).

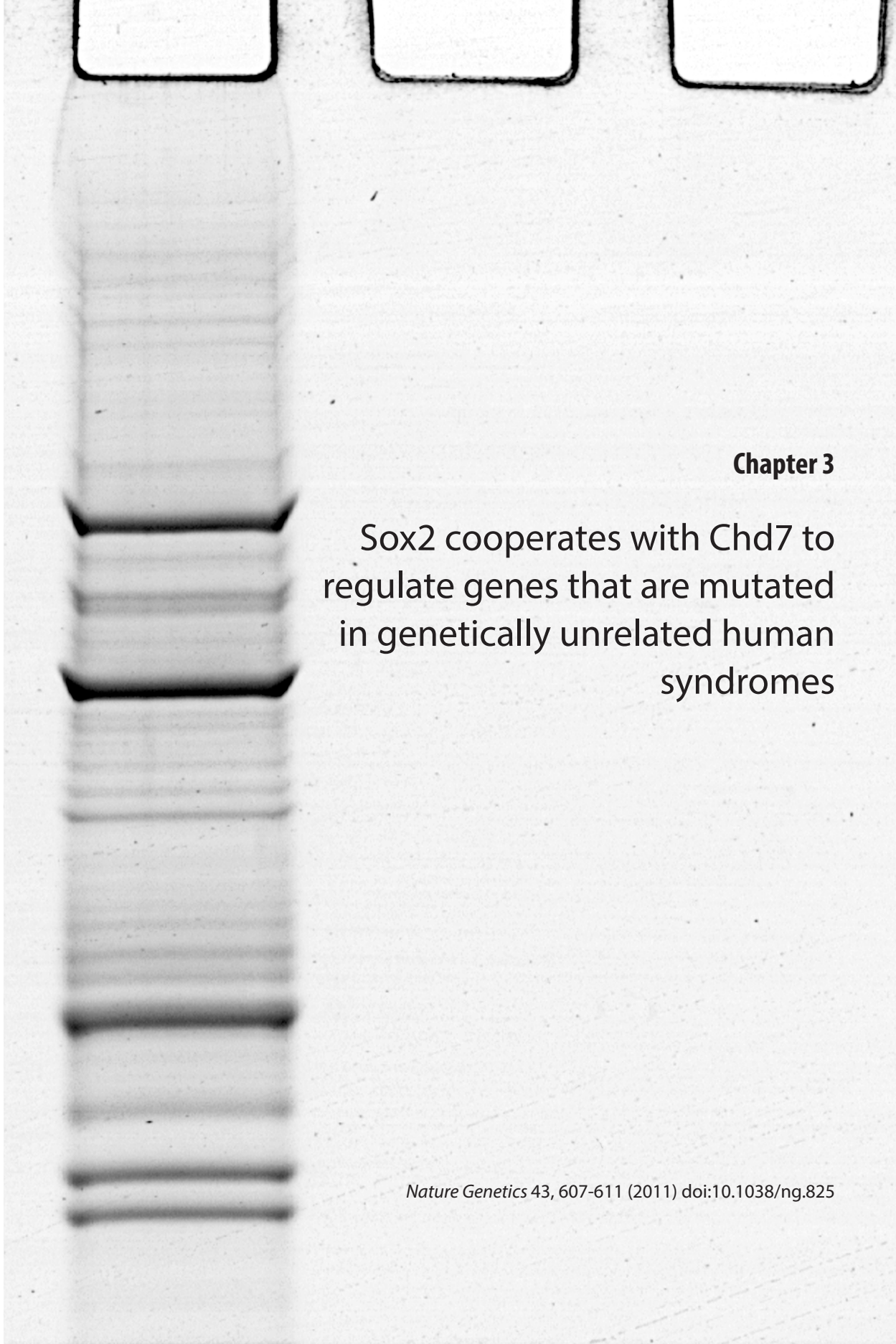
210. Laudet, V., Stehelin, D. & Clevers, H. Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res* **21**, 2493-501 (1993).
211. Jantzen, H.M., Admon, A., Bell, S.P. & Tjian, R. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. *Nature* **344**, 830-6 (1990).
212. Stros, M. HMGB proteins: interactions with DNA and chromatin. *Biochim Biophys Acta* **1799**, 101-13 (2010).
213. Kiefer, J.C. Back to basics: Sox genes. *Dev Dyn* **236**, 2356-66 (2007).
214. Gubbay, J. et al. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245-50 (1990).
215. Wilson, M. & Koopman, P. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr Opin Genet Dev* **12**, 441-6 (2002).
216. Kamachi, Y., Uchikawa, M. & Kondoh, H. Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet* **16**, 182-7 (2000).
217. Wegner, M. All purpose Sox: The many roles of Sox proteins in gene expression. *Int J Biochem Cell Biol* **42**, 381-90 (2010).
218. Sandberg, M., Kallstrom, M. & Muhr, J. Sox21 promotes the progression of vertebrate neurogenesis. *Nat Neurosci* **8**, 995-1001 (2005).
219. Harley, V.R., Lovell-Badge, R. & Goodfellow, P.N. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res* **22**, 1500-1 (1994).
220. Lodato, M.A. et al. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet* **9**, e1003288 (2013).
221. Sarkar, A. & Hochedlinger, K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15-30 (2013).
222. Que, J. et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* **134**, 2521-31 (2007).
223. Zorn, A.M. & Wells, J.M. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* **25**, 221-51 (2009).
224. Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev* **15**, 1272-86 (2001).
225. Matsushima, D., Heavner, W. & Pevny, L.H. Combinatorial regulation of optic cup progenitor cell fate by SOX2 and PAX6. *Development* **138**, 443-54 (2011).
226. Fantes, J. et al. Mutations in SOX2 cause anophthalmia. *Nat Genet* **33**, 461-3 (2003).
227. Williamson, K.A. et al. Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. *Hum Mol Genet* **15**, 1413-22 (2006).
228. Kelberman, D. et al. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest* **116**, 2442-55 (2006).
229. Kiernan, A.E. et al. Sox2 is required for sensory organ development in the mammalian inner ear. *Nature* **434**, 1031-5 (2005).
230. Hamon, M.A. & Cossart, P. Histone modifications and chromatin remodeling during bacterial infections. *Cell Host Microbe* **4**, 100-9 (2008).
231. Wolpert, L. *Principles of development*, 551 S. (Oxford University Press, Oxford, 2006).
232. Zernicka-Goetz, M. Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development* **129**, 815-29 (2002).



Chapter 2

FLAG-tag based affinity purification to identify transcription factor networks in mouse embryonic and neural stem cells

Manuscript in preparation



Chapter 3

Sox2 cooperates with Chd7 to regulate genes that are mutated in genetically unrelated human syndromes

Nature Genetics 43, 607-611 (2011) doi:10.1038/ng.825

Sox2 cooperates with Chd7 to regulate genes that are mutated in genetically unrelated human syndromes

Erik Engelen[1*], Umut Akinci[1*], Jan Christian Bryne[2], Jun Hou[1], Cristina Gontan[3†], Maaike Moen[1], Dorota Szumska[4], Christel Kockx[5], Wilfred van IJcken[5], Dick H. W. Dekkers[6], Jeroen Demmers[6], Erik-Jan Rijkers[7], Shoumo Bhattacharya[4], Sjaak Philipsen[1], Larysa H. Pevny[8], Frank G. Grosveld[1], Robbert J. Rottier[3], Boris Lenhard[2], Raymond A. Poot[1]

[1] Department of Cell Biology, Erasmus MC, Dr. Molewaterplein 50, 3015 GE, Rotterdam, The Netherlands
[2] Department of Biology, University of Bergen, Thormøhlensgate 53A, N-5008 Bergen, Norway
[3] Department of Pediatric Surgery, Erasmus MC, The Netherlands
[4] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX1 0RD, United Kingdom
[5] Center for Biomics, Erasmus MC, The Netherlands
[6] Proteomics Center, Erasmus MC, The Netherlands
[7] Department of Biochemistry, Erasmus MC, The Netherlands
[8] Department of Genetics, University of North Carolina Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
† Current Address, Department of Reproduction and Development, Erasmus MC, The Netherlands

* These authors contributed equally to this work

ABSTRACT

The HMG-box transcription factor Sox2 plays a role throughout neurogenesis[1] and also acts at other stages of development[2], as illustrated by the multiple organs affected in the anophthalmia syndrome caused by *SOX2* mutations[3-5]. Here we combined proteomic and genomic approaches to characterize gene regulation by Sox2 in neural stem cells (NSCs). Chd7, a chromatin remodeling ATPase associated with CHARGE syndrome[6,7], was identified as a Sox2 transcriptional co-factor. Sox2 and Chd7 physically interact, have overlapping genome-wide binding sites and regulate a set of common target genes, including *Jag1*, *Gli3* and *Mycn*, genes mutated in the syndromes of Alagille, Pallister-Hall and Feingold, which show malformations also associated with SOX2- or CHARGE syndrome[8-10]. Regulation of disease-associated genes by a Sox2-Chd7 complex provides a plausible explanation for several malformations associated with SOX2- or CHARGE syndrome. Indeed, we found that *Chd7*-haploinsufficient embryos displayed severely reduced expression of *Jag1* in the developing inner ear.

RESULTS

As a first step to gain more insight in the transcriptional network in which Sox2 operates, we identified Sox2-interacting proteins in neural stem cells (NSCs). Sox2 is essential for the *in vivo* maintenance of mouse embryonic and adult NSCs and subsequent neurogenesis[1]. NSCs are therefore an appropriate cell type to study gene regulation by Sox2. NSCs that stably express FLAG-Sox2[11] (F-Sox2) have a normal morphology and expressed NSC markers such as Nestin, RC2[11] and Pax6 (**Supplementary fig. 1a**). To identify interaction partners, F-Sox2 was purified from NSC nuclear extract by a FLAG-affinity based protocol[12], proteins separated by PAA gel (**Supplementary fig. 1b**) and analyzed by mass spectrometry. We identified 50 Sox2-interacting factors that were specifically present in two F-Sox2 purifications (**table 1** and **Supplementary table 1**). Many of these interactions were validated by an independent method, immunoprecipitation (IP) of endogenous Sox2 (**Supplementary table 1**). NuRD complex subunits Mi2-beta and MTA2 were confirmed to interact with Sox2 by IP-western blot (**Fig. 1a,b**). Interestingly, many of the identified Sox2-interacting factors, such as transcription factors Nfi-beta and Twist1 and chromatin modifying complexes SWI-SNF and SMRT, are involved in neural development (**Supplementary table 2**). We conclude that Sox2 interacts with multiple factors with importance for neurogenesis.

table1 | Interaction partners of Sox2 in neural stem cells.

Protein	Average mascot[a]	Average peptides[a]
Sox2	479	5
Spalt (3)	888	12
NuRD complex (9)	727	11
Trrap complex (3)	421	6
SMRT/NcoR complex (5)	307	6
SWI/SNF complex	271	4
Transcription factors		
Chd7	1,253	20
Cutl1	1,192	17
Dbc1	679	10
Zeb1	312	4
Ctbp1	242	3
Supt16h	187	3
Rfx3	167	2
Sox8	167	3
Hoxa5	166	2
Mef2	159	3
Nfi-β	139	3
Ctbp2	109	2
Nac1	104	2
Tead1	87	1
Snf2h	87	2
Sox5	69	1
Twist1	63	1
Tcf3	46	1
Zfp191	41	1
Other		
Exp4	1,667	24
Skiv2l2	179	3
Dock7	135	2
Dnaja2	93	1

[a]Averages are from three experiments (Supplementary table 1). Parentheses indicate the number of identified subunits.

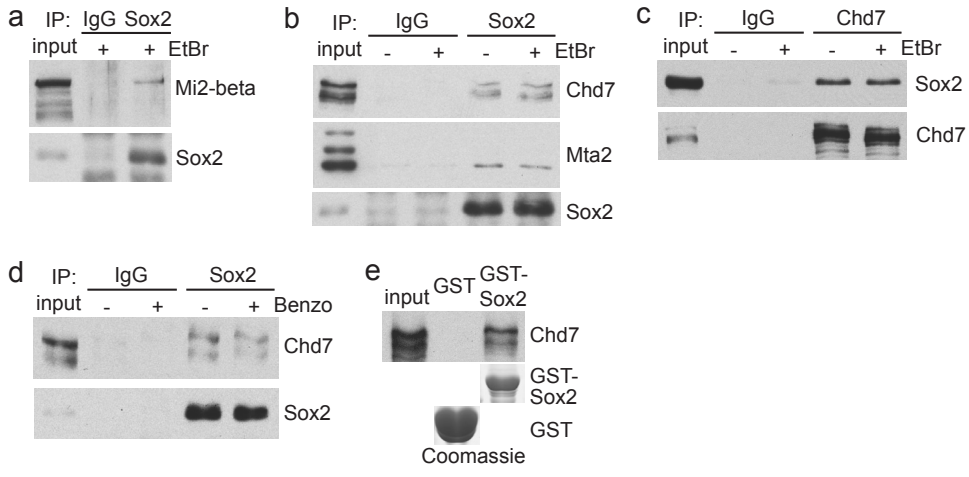


Figure 1 | Interaction partners of Sox2 and Chd7 in neural stem cells.
a-b,d, co-precipitation of Mi2-beta (a), Mta2 and Chd7 (b) and Chd7 (d) in Sox2 immunoprecipitations. c, Co-precipitation of Sox2 in Chd7 immunoprecipitations. Proteins were detected by western blot. Ethidium bromide (EtBr) or nuclease benzonase (Benzo) was added where indicated. e, Co-precipitation of Chd7 with GST-Sox2 pull down. Chd7 was detected by western blot, GST and GST-Sox2 stained by Coomassie blue in left and right bottom panels, respectively. a-e, Fraction of input loaded is 5%.

Prominent on the list of Sox2 interactors with many peptides identified (**table 1** and **Supplementary table 1**) is Chd7, a member of the family of CHD chromatin remodeling ATPases. Chd7$^{-/-}$ mouse embryos have neural and other defects and die at embryonic day (E) 10.5[13,14]. In humans, *CHD7* haploinsufficiency causes CHARGE syndrome[6,7] (incidence ~ 1:10,000), a clustering of coloboma, heart malformation, atresia of the choanae, retarded growth and development, genital anomalies and ear anomalies/ deafness. *Chd7* and *Sox2* have a similar expression pattern in E14.5 mouse embryos with high expression in the ventricular zones of the brain, the pituitary gland, the olfactory bulbs, the eyes and inner ears (Genepaint[15], **Supplementary fig. 2** and references therein). CHARGE syndrome overlaps in several features reported for the anophthalmia syndrome caused by *SOX2* mutations, such as malformations of the esophagus and trachea, genital abnormalities and pituitary defects[3-5,7], which occasionally leads to a misdiagnosis[3]. We confirmed that Sox2 interacts with Chd7 by Sox2 IP-western, Chd7 IP-western and GST-pull down (**Fig. 1b-e**). The Sox2-Chd7 interaction was insensitive to Ethidium Bromide (**Fig. 1b,c**) and the nuclease Benzonase (**Fig. 1d**) and therefore unlikely to be mediated by DNA. We subsequently immunoprecipitated Chd7 and analyzed the spectrum of Chd7 binding partners by mass spectrometry. Although the Chd7 IPs were efficient and depleted Chd7 from nuclear extract, only three transcription factors were consistently identified; Sox2, Olig1 and Zbtb20 (**table 2** and **Supplementary table 3**). This suggests that Chd7 is a

specialized co-factor for Sox2 and a limited number of other transcription factors in NSCs.

table2 | Interaction partners of Chd7 in neural stem cells.

Protein	Average mascot[a]	Average peptides[a]
Chd7	7,084	99
Spalt (1)	684	12
SWI/SNF complex (4)	677	10
NuRD complex (3)	559	9
Transcription factors		
Zbtb20	337	6
Sox2	198	3
Olig1	85	1
Other		
Vars	1,230	21

[a]Averages are from three experiments (Supplementary table 1). Parentheses indicate the number of identified subunits.

To investigate gene regulation by Sox2 and Chd7, we performed shRNA-mediated knock down (kd) of Sox2 or Chd7 in NSCs **(Supplementary fig. 3**) followed by microarray analysis after 48 hours. Strikingly, 43% of the misregulated genes in Chd7-kd NSCs were also misregulated in Sox2-kd NSCs (**Fig. 2a** and **Supplementary table 4**), and often in the same direction (**Fig. 2b** and **Supplementary table 4**), showing that gene regulation by Sox2 and Chd7 are correlated (**Fig. 2a**). Sox2+Chd7-activated genes include the Sox2-target gene *Egfr*[16], and genes from the Notch and Sonic Hedgehog (Shh) pathways (**Fig. 2b** and **Supplementary table 4**), supporting the reported role of Sox2 in activating Shh and Notch genes to facilitate development of the brain[17,18] and eyes[19]. Subsequently, the binding sites of Sox2 on a genome-wide scale were determined by Sox2 chromatin immunoprecipitation, and sequencing of the bound DNA (ChIP-seq). The resulting set of Sox2-peaks (~7400 peaks, **Supplementary table 5,** including ~6300 peaks near genes, **Supplementary table 6**) showed high enrichment of the Sox2 consensus motif (**Fig. 2c**) suggesting they represent genuine Sox2 binding sites. We also observed enrichment of the Helix-Loop-Helix (HLH) motif (**Fig. 2d)** and a G-rich motif (**Supplementary fig. 4a**). Enrichment of the HLH motif may indicate a role for Sox2-interacting HLH-factors Zeb1 and Twist1 in the regulation of Sox2-targets. Mutations in *TWIST1* cause Saethre-Chotzen syndrome, which shares some phenotypic overlap with CHARGE syndrome[20], possibly due to the role of *TWIST1* as a target gene of CHD7 in neural crest formation[21].

Sox2 binding sites were found to be often located near transcription start sites (TSS, **Fig. 2e**) and were especially enriched near the TSS of CpG-island promoters (**Supplementary fig. 4b**). Interestingly, Sox2-activated genes, and in particular Sox2+Chd7-activated genes, were enriched for Sox2 binding sites (**Fig. 2f**). This positive correlation of Sox2-binding and Sox2-dependent gene activation suggests that these Sox2-bound and –activated genes (**Supplementary table 7)** are direct Sox2 target genes.

Chd7 genome localisation was recently determined in NSCs and ES cells[22,23]. To analyze the genome-wide co-localization of Sox2 and Chd7 on the NSC genome, we

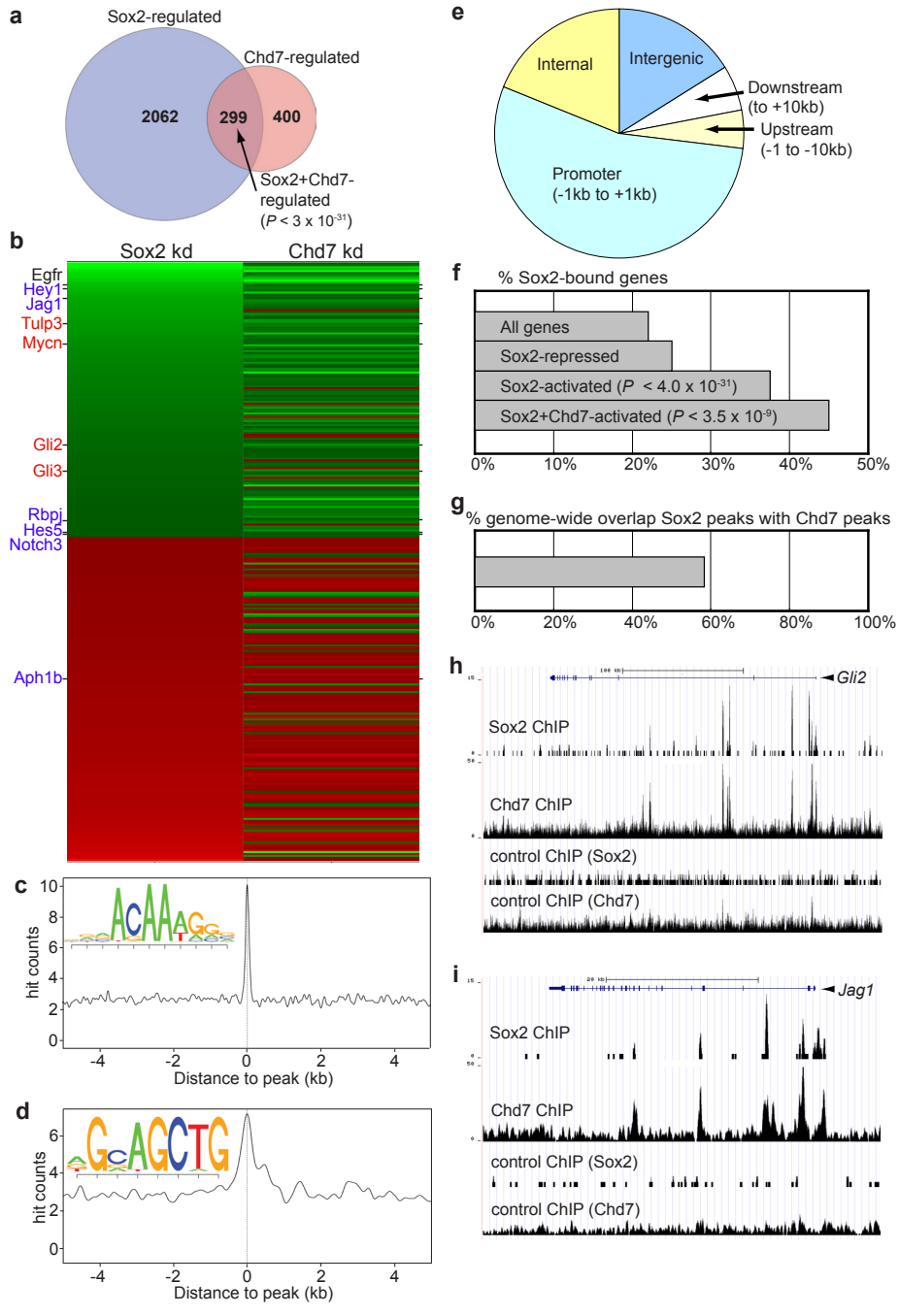


Figure 2 | Target genes of Sox2 and Chd7 in neural stem cells.

a, Venn diagram of number of genes mis-regulated in Sox2 kd NSCs and/or Chd7 kd NSCs, as determined by microarray. Fisher's exact test P-value for the correlation of gene-misregulation by Sox2 or Chd7 is indicated. b, Heatmap of genes regulated by Sox2 and Chd7. Shades of green and red indicate down-regulation and up-regulation, respectively. Note that genes are often mis-regulated in the same direction in Sox2 kd NSCs and

Chd7 kd NSCs. Blue and red labeling indicates members of the Notch pathway or Sonic Hedgehog pathway, respectively. c, Enrichment of Sox2 consensus DNA motif in Sox2 ChIP-seq peaks. d, Enrichment of HLH transcription factor DNA motif in Sox2 peaks. e, Pie-chart of distribution of Sox2 peaks. f, Percentage of genes with Sox2 peaks within the different indicated categories of genes on the microarray. Fisher's exact test P-values for correlation with Sox2-bound genes are indicated. g, Genome-wide overlap of Sox2 peaks with Chd7 peaks. Percentage of the genome-wide Sox2 peaks (Supplementary table 6) that overlaps with Chd7 peaks (Supplementary table 8) is indicated. h,i, Localization of Sox2 and Chd7 on Sox2-Chd7 target genes. Sequence reads from the indicated ChIP experiments were plotted relative to chromosomal position. Sox2 ChIP, Chd7 ChIP and their corresponding control (IgG) ChIP experiments are shown. Genome locations of the Gli2 locus (h) and Jag1 locus (i) are shown. Arrowhead indicates direction of the locus, scale bar represents 100 kb (h) or 20 kb (i).

determined the binding sites of Chd7 by ChIP-seq (**Supplementary table 8**). Strikingly, the majority (58%) of genome-wide Sox2 binding sites overlap with Chd7 binding sites (**Fig. 2g**) suggesting that Chd7 is an important co-factor for Sox2. Examples of Sox2 and Chd7 genomic co-localization are shown for Sox2-Chd7 target genes *Jag1* and *Gli2* (**Fig. 2h,i**) and *Tulp3* and *Hes5* (**Supplementary fig. 5**). Chd7 has more binding sites on the genome (~23000, **Supplementary table 8**, of which ~16000 near or in genes, **Supplementary table 9**) than Sox2, suggesting that Chd7 is also involved in gene regulation with transcription factors other than Sox2. Interestingly, the HLH motif is more enriched in Sox2 peaks that overlap with Chd7 peaks (**Supplementary fig. 6a**), suggesting that HLH factors may be involved in the regulation of Sox2 targets, especially in the context of Chd7.

The physical interaction of Sox2 and Chd7 and the overlap in regulated genes and genomic localization indicated that Sox2 and Chd7 may act synergistically in gene activation. To investigate this further, we focused on seven genes out of the set of 48 identified Sox2-Chd7 target genes (**Supplementary table 10**). These genes are part of the Notch pathway (*Jag1*, *Rbpj*, *Hes5*) or Shh pathway (*Gli2*, *Gli3*, *Mycn*, *Tulp3*). Mutations in human *GLI2*, *GLI3* and *MYCN* cause pituitary hypoplesia[10,24] and esophagal atresia[9], respectively. We confirmed down-regulation of the selected genes upon Sox2 or Chd7 kd (**Fig. 3a**). We subsequently analyzed their expression in NSCs from E12.5 embryos of *Sox2*$^{+/-}$ mice and *Chd7*$^{+/-}$ mice[13,25]. All selected genes were again found to be down-regulated, except *Hes5* in *Sox2*$^{+/-}$ NSCs (**Fig. 3b)**. Thus, the expression of these genes was also affected in the context of Sox2 or Chd7 haploinsufficiency. We confirmed by ChIP analysis that Chd7 binds at Sox2 binding sites in all tested genes (**Fig. 3c**). Knock-down of Sox2 not only reduced Sox2 binding, as expected (**Fig. 3d**), but also reduced binding of Chd7 (**Fig. 3e**), suggesting that Sox2 facilitates the recruitment of Chd7. Chd7 knock-down did not affect Sox2 binding (**Supplementary fig. 6b**). Sox2 or Chd7 knock-down also did not affect the H3K4me3 histone mark at the TSS of their target genes (**Supplementary fig. 6c**). We conclude that Sox2 and Chd7 cooperate to activate a set of common target genes with potential relevance for SOX2 anophtalmia syndrome and CHARGE syndrome. Analysis of human NSCs that were positive for markers SOX2 and NESTIN (**Supplementary fig. 7a-c**), using ChIP and knock down (**Supplementary fig. 7d,e**), showed that SOX2 and CHD7 bind

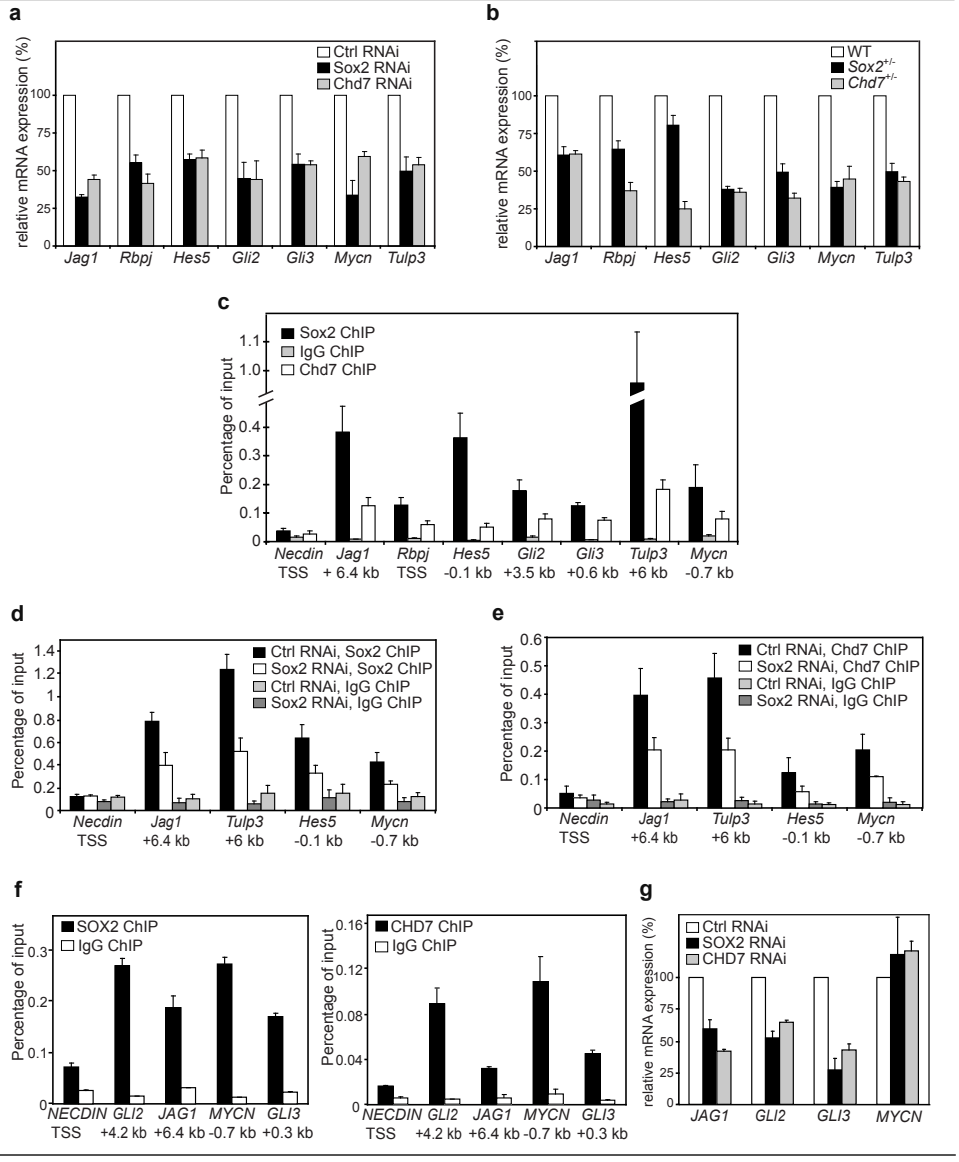


Figure 3 | Sox2 and Chd7 regulate genes of the Notch and Sonic Hedgehog signaling pathways.

a, Analysis by RT-PCR of mRNA levels of indicated genes in Sox2 knock-down (kd) NSCs, Chd7 kd NSCs or control (ctrl) kd NSCs . b, Analysis of mRNA levels in NSCs isolated from E12.5 embryos of wild type (WT) mice, Sox2+/- mice or Chd7+/- mice. c, Binding of Sox2 and Chd7 to the indicated genes, detected by chromatin immunoprecipitation (ChIP) with the indicated antibodies. Distance of Sox2 binding site to gene transcription start site (TSS) is indicated if ≥ 0.1 kb. Precipitated DNA for the indicated genes is shown as percentage of input DNA, Necdin TSS is used as negative control region. d, e, Sox2-dependence of Sox2 binding (d) or Chd7 binding (e) to Sox2 binding sites in the indicated genes, using Sox2 kd NSCs or control kd NSCs, ChIP is with the indicated antibodies. f, Binding of SOX2 (left panel) and CHD7 (right panel) to the indicated genes in human NSCs, detected by ChIP with indicated antibodies. Distance of SOX2 and CHD7 binding site to gene transcription start site (TSS) is indicated if ≥ 0.1 kb. Precipitated DNA for indicated genes is shown as a percentage of input DNA, NECDIN TSS is used as negative control region. S.e.m. is indicated of two independent experiments. g, Analysis by RT-PCR of mRNA levels of indicated genes in human NSCs

with SOX2 knock-down (kd), CHD7 kd or control kd. a-e, g, S.e.m. is indicated of three independent experiments.

disease-relevant genes *JAG1*, *GLI2*, *GLI3* and *MYCN* (**Fig. 3f**) and activate the expression of *JAG1*, *GLI2* and *GLI3* (**Fig. 3g**), indicating that the regulation of these genes by SOX2 and CHD7 has a high level of conservation in human cells.

Most individual features of CHARGE syndrome or SOX2 anophthalmia syndrome only occur in a subset of patients[3,4,7] and mutant mice[25,26], suggesting variability in the underlying gene regulation, which complicates its molecular analysis. An exception is the malformation of the semicircular canals in the inner ear and the accompanying vestibular defects in CHARGE syndrome, which is nearly fully penetrant, both in patients[7] and in *Chd7* haploinsufficient mice[13,14,25]. Interestingly, heterozygosity for *Jag1* causes similar semi-circular canal malformations in mice[27] and humans[8], where *JAG1* mutations cause Alagille syndrome[28]. *Jag1* is expressed in the otocyst, at the start of semicircular canal development (E10.5)[29]. We found that *Jag1* expression is dramatically reduced in *Chd7*$^{+/-}$ E10.5 otocysts, compared to wt otocysts (**Fig. 4a,b**). *Jag1* expression was not significantly affected in *Sox2*$^{+/-}$ otocysts (**Fig. 4c**), in line with the normal development of the semi-circular canals in these mice[30]. We conclude that Chd7 regulates *Jag1* expression in the developing inner ear. Reduced *Jag1* expression due to lower Chd7 levels provides a rationale for the CHARGE-associated defects of the vestibular apparatus.

In summary, we characterized here three aspects of Sox2 in NSCs in a combined and unbiased approach; its associated proteins, its binding sites and its regulated genes. We identified putative Sox2 co-factors and target genes, many of which are themselves involved in NSC identity and maintenance. We focused on gene-regulation by Sox2 and Chd7, two proteins found to physically interact. Sox2 and Chd7 are expressed in a similar pattern during development, including in many organs that can be affected in SOX2 anophthalmia syndrome or CHARGE syndrome. We show here that Sox2 and Chd7 cooperatively activate target genes. Heterozygosity for *SOX2*, *CHD7* or *MYCN* (which causes Feingold syndrome[9]) are the only known genetic causes of trachea-esophagal malformations in humans (incidence ~ 1:3500)[31]. Pituitary and genital anomalies are part of syndromes caused by mutations in *SOX2*, *CHD7* and *GLI3* (Pallister-Hall syndrome[10]), whereas CHARGE syndrome and Alagille syndrome share similar vestibular defects. An extrapolation of our results may suggest that the common features of these syndromes, which were not previously linked, have a similar cause at the molecular level (**Fig. 4d**). In addition to *Mycn*, *Gli2*, *Gli3* and *Jag1*, our list of Sox2-Chd7 targets may contain more genes which, when mutated, cause malformations normally associated with SOX2 anophthalmia syndrome or CHARGE syndrome.

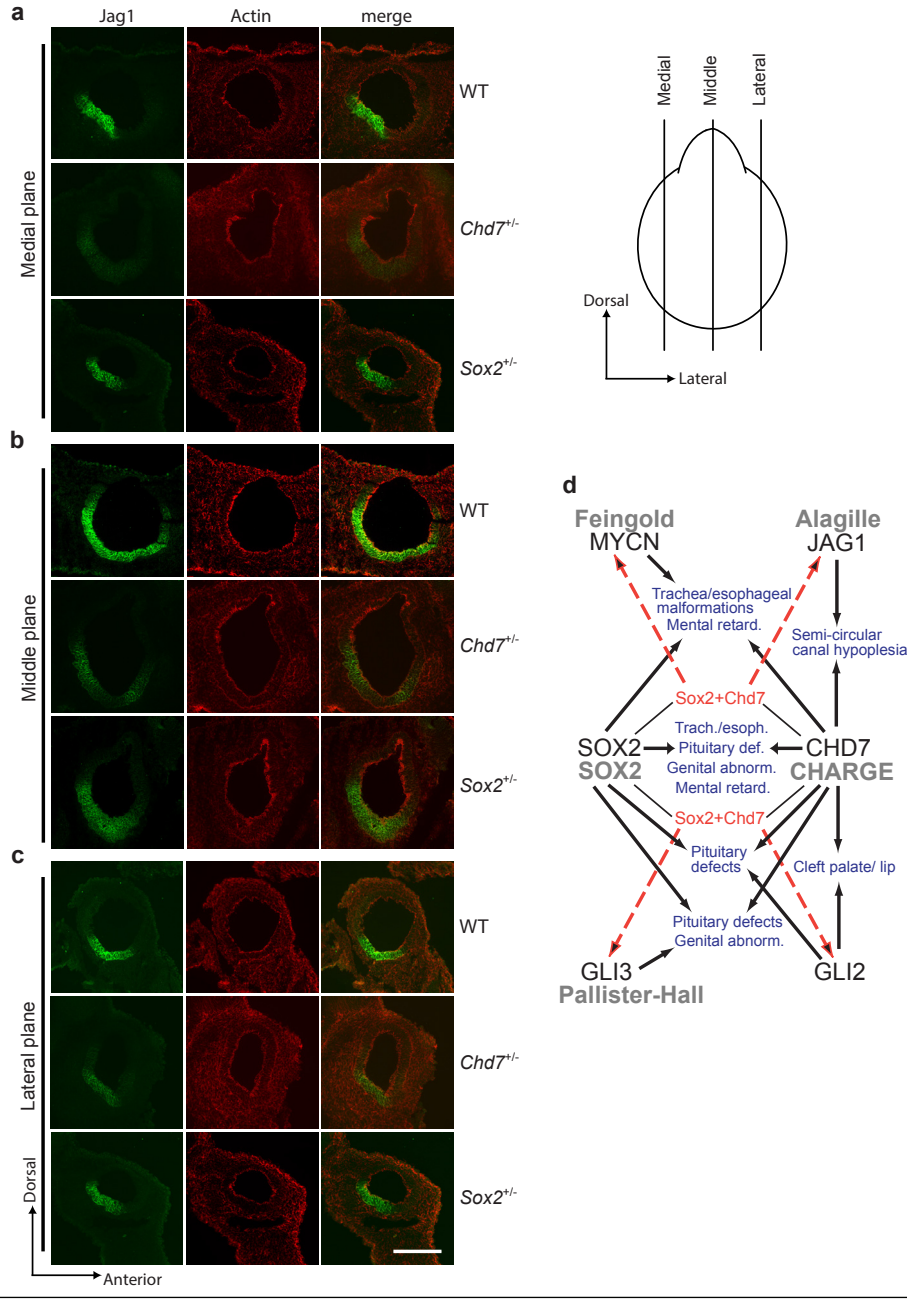


Figure 4 | Expression of Jag1 is strongly reduced in otocysts of Chd7 heterozygous embryos.
a,b,c, Right panel, schematic drawing of E10.5 otocyst with the three investigated planes indicated. Left panels, sagittal cryosections of medial planes (a), middle planes (b) and lateral planes (c) of otocysts from wild type (WT), Chd7+/- and Sox2+/- E10.5 embryos, stained with anti-Jag1 antibodies (green, left images), stained for actin using Phalloidin (red, central images), or the merged images (right images). Representative images are shown from otocyst stainings of 8 wild-type embryos, 8 Chd7+/- embryos and 5 Sox2+/- embryos. Orientation is indicated, bar is 200µm. d, Hypothetical model for mechanistic links between shared malformations in

different human syndromes. Syndromes in grey lettering, haploinsufficiency for gene associated with syndrome in black lettering, associated malformations/defects indicated by black arrow, transcriptional regulation of genes by Sox2+Chd7 in NSCs indicated by dotted red arrows, shared malformations/defects in blue. T.e.f./ E.a.; Trachea-esophagal fistula/ Esophagal atresia, Ment. ret.; Mental retardation, S.c.c. hyp.; Semi-circular canal hypoplesia, Pit. hyp.; Pituitary hypoplasia, Genit. ab.; Genital abnormalities, Cl. pal./lip; Cleft palate/lip.

Acknowledgements

We thank Gina Abelo for advice on the otocyst stainings, Austin Smith for 46C ES cells, Steven Pollard for advice on deriving neural stem cells, Zeliha Ozgür for micro-array hybridizations, Mirjam van den Hout-van Vroonhoven for Illumina GAP analyses, and Paul Wade for Mi2-beta antibody. R.A.P., E.E. and U.A. were supported by a Vidi grant, ALW-open program grant and a Chemical Sciences ECHO grant, respectively, all from the Netherlands Organisation for Scientific Research (NWO). J.C.B. was supported by EuTRACC, B.L. was supported by grants from the Norwegian Research Council (YFF) and Bergen Research Foundation. C.G. and R.J.R. were supported in part by the Sophia Foundation for Medical Research.

Author contributions

E.E. and U.A. performed nearly all experiments and analyzed the data. J.C.B. and B.L. normalized the ChIP-seq data and performed all bioinformatic analyses. J.H. and S.P. normalized and formatted the microarray gene-expression data. C.G., R.A.P. and R.J.R. created the F-Sox2 ES cells. D.S and S.B assisted in the mouse work, M.M. performed the GST-pull down experiment, C.K. and W.v.IJ. performed the microarray analyses and Illumina-sequencing of the ChIP material. D.H.W.D. and J.D. performed the mass spectrometry analyses. E.-J.R. provided bioinformatic assistance in the early stages of this work. L.H.P. provided *Sox2* COND mice. F.G.G. set up the ChIP-sequencing facility and bioinformatics infrastructure. R.J.R. created *Sox2*$^{+/-}$ mice from *Sox2* COND mice. R.A.P. designed the study, analyzed the data and wrote the manuscript with support from co-authors.

Database accession codes

ChIP sequencing data are available through the Sequence Read Archive, accession code ERP000239. Microarray data are available through the EBI ArrayExpress database, accession code E-MEXP-2743.

METHODS

Neural stem cell culturing and derivation.

F-Sox2 NSCs[11] and wild-type NSCs, used for biochemical and RNAi studies, were cultured, as described[32], using N2B27 medium (Stem Cell Sciences) supplemented with EGF and FGF (both from Peprotech). *Sox2*$^{+/-}$, *Chd7*$^{+/-}$ and wild-type littermate NSCs, used for gene expression studies, were derived from forebrains of E12.5 mouse embryos, filtered through a 70 µm cell strainer (Falcon) and cultured in N2B27 with EGF + FGF, as described[32]. Human NSCs (ES cell-derived) were purchased from Invitrogen (N7800-100) and cultured as described[33], on laminin (Roche) coated dishes in Euromed-N (Euroclone) supplemented with N2/B27 (Invitrogen), EGF and FGF (Peprotech) and L-glutamine (Invitrogen).

Identification of interacting proteins of Sox2 and Chd7.

FLAG-Sox2 was purified from F-Sox2 NSC[11] nuclear extract prepared from 2x10^8 NSCs using a FLAG-affinity protocol, as described[12]. Control purifications were from nuclear extract from wild-type NSCs. Immunoprecipitation of Sox2 with a Sox2 antibody (Y-17, sc-17320) or immunoprecipitation of Chd7 with a Chd7 antibody (ab-31824, Abcam) from NSC nuclear extract was as described[12], EtBr (25µg/ml) or Benzonase (150U/µl, Novagen) was added, where indicated. GST-pull downs were as described[12]. Identification of interacting proteins by mass spectrometry was as described[12], proteins were included if specifically identified in both purifications of F-Sox2, or both immunoprecipitations of Chd7.

Transfection with shRNA constructs.

shRNA sequence for mouse Sox2 was described[34]. shRNA sequences for mouse Chd7 and human SOX2 and CHD7 (**Supplementary table 11**) were designed with help of Whitehead siRNA selection program (http://jura.wi.mit.edu/bioc/siRNAext/) and cloned into pSuper-puro (Oligoengine). pSuper-control-shRNA (Dharmacon) was used as a control. 3x10^6 NSCs were transfected with pSuper constructs by electroporation using an Amaxa Nucleofector and Nucleofector Kit V (Lonza). Puromycin (1µg/ml^{-1}) was added after 24 h and NSCs were harvested 48h after electroporation.

Expression analyses.

For expression analysis by microarray, total RNA was isolated in experimental triplicates from NSCs electroporated with the different shRNA constructs and converted into biotin-labeled ssDNA, as described[35], and hybridized to GeneChIP Mouse Genome 430 2.0 arrays (Affymetrix) according to manufacturer's recommendations. Array data quality control,

normalization and statistical analysis were as described[36]. Quantitative real time PCR analyses on cDNA transcribed from total RNA with Superscript™ II Reverse Transcriptase, was performed on a DNA Engine Opticon2/ CFX96 (Biorad) and normalized on *CalR* or *Hprt* expression. Primer sequences are listed in **Supplementary table 12**.

Mice.

Chd7$^{+/edy}$ mice[25,37], here called Chd7$^{+/-}$ mice, were obtained from the EMMA consortium and maintained in an FVB background. *Sox2*$^{+/floxKO}$ mice, here called Sox2$^{+/-}$ mice, were generated by crossing *Sox2*$^{+/COND}$ mice[19] with mice expressing CRE recombinase from a CAG promoter[38] and maintained on a C57Bl/6 background. All animal studies were in conducted under the guidelines for animal experimentation approved by the Erasmus University Animal Welfare Committee.

Immunostainings.

For immunostaining of the otocysts, sagittal cryostat sections of 10 µm thickness of heads of E10.5 embryos were fixed in PBS-2% PFA for 15 minutes at room temperature, permeabilized with PBS-0.1%Triton (PBS-0.1T) for 2 x 10 min. and incubated in blocking buffer (PBS-0.5%BSA, 0.15% glycin) for 15 minutes. Slides were incubated overnight at 4°C in blocking buffer with anti-Jag1 (1:100; Santa Cruz Biotechnology; H-114). After washing in PBS-0.1T, slides were incubated for 1 hour in blocking buffer with Alexa 488 goat anti-rabbit (1:200; Invitrogen; A11008) and Alexa 594-conjugated phalloidin (1:100; Invitrogen; A12381), washed and mounted in Vectashield Mounting Medium with DAPI (Vector laboratories). Mouse NSCs were grown on poly-D-lysine coated cover slips, fixed in PBS-4% PFA, permeabilized with PBS-0.4T in PBS, blocked in PBS-10% fetal calf serum, incubated in anti-Pax6 (1:10 dilution, Developmental Studies Hybridoma Bank) for 2 hours at room temperature, washed, incubated in Alexa 594 goat anti-mouse IgG (1:200; Invitrogen; A11032), washed and mounted in Vectashield Mounting Medium with DAPI (Vector laboratories). Human NSCs were grown on laminin-coated coverslips and stained with antibodies against SOX2 (AF2018, R&D systems) and NESTIN (mAb1259, R&D systems) as described above. Digital images were captured on a Zeiss microscope Axio Imager Z1.

Chromatin immunoprecipitations (ChIP) and high-throughput sequencing.

For large scale chromatin preparation, 10^8 NSCs were crosslinked with 2mM disuccinimidyl glutarate (Thermo Fisher Scientific) and 1% formaldehyde, as described[39], nuclei were isolated in 50mM Tris-Cl pH 8.0, 1mM EDTA, 1%SDS, lysed in pre-IP buffer (10mM Tris, 10mM NaCl, 3mM $MgCl_2$, 1mM $CaCl_2$). Chromatin was prepared and ChIP performed

according to the Millipore on-line protocol using 15µg of antibodies against Sox2 (Y17, sc-17320) and Chd7 (ab-31824) or goat IgG (sc-2028) as control. ChIP DNA library preparation, ChIP-sequencing on an Illumina Genome Analyzer or HiSeq2000, processing of the raw data and mapping the peaks to the mouse genome (NCBI build 37.1), was as described[35]. ChIP sequencing data have been submitted to the Sequence Read Archive, accession nr. pending. For small scale ChIP, 10^7 NSCs, electroporated with the different shRNA constructs, were directly lysed in pre-IP buffer and ChIP performed as above, using 5µg of antibodies. Small scale ChIP in human NSCs was performed as described above, using antibodies against SOX2 (AF2018, R&D systems) and CHD7 (Bethyl laboratories A301-223A). Primers for amplification of genomic regions by qPCR are listed in **Supplementary table 13**.

Bioinformatic analyses.

Significance estimations of Sox2 ChIP peaks and Chd7 peaks was calculated using the Poisson distribution, as described[40], followed by multiple testing correction by controlling the false discovery rate (fdr). A threshold false discovery rate (fdr) of $2x10^{-10}$ was applied for Sox2 peaks and $1x10^{-13}$ for Chd7 peaks. Derivation of motifs was performed using MEME[41] on genomic sequences of 400 bps centered on Sox2 peaks. Mapping Sox2 peaks and Chd7 peaks to different regions of genes was performed with R/Bioconductor (http://www.bioconductor.org).

REFERENCES

1. Pevny, L.H. & Nicolis, S.K. Sox2 roles in neural stem cells. *Int J Biochem Cell Biol* **42**, 421-424 (2010).
2. Guth, S.I. & Wegner, M. Having it both ways: Sox protein function between conservation and innovation. *Cell Mol Life Sci* **65**, 3000-18 (2008).
3. Williamson, K.A. et al. Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. *Hum Mol Genet* **15**, 1413-22 (2006).
4. Fantes, J. et al. Mutations in SOX2 cause anophthalmia. *Nat Genet* **33**, 461-3 (2003).
5. Kelberman, D. et al. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest* **116**, 2442-55 (2006).
6. Vissers, L.E. et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* **36**, 955-7 (2004).
7. Zentner, G.E., Layman, W.S., Martin, D.M. & Scacheri, P.C. Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *Am J Med Genet A* **152A**, 674-86 (2010).
8. Okuno, T., Takahashi, H., Shibahara, Y., Hashida, Y. & Sando, I. Temporal bone histopathologic findings in Alagille's syndrome. *Arch Otolaryngol Head Neck Surg* **116**, 217-20 (1990).
9. van Bokhoven, H. et al. MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome. *Nat Genet* **37**, 465-7 (2005).
10. Kang, S., Graham, J.M., Jr., Olney, A.H. & Biesecker, L.G. GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet* **15**, 266-8 (1997).
11. Gontan, C. et al. Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors. *J Cell Biol* **185**, 27-34 (2009).
12. van den Berg, D.L. et al. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369-81 (2010).
13. Hurd, E.A. et al. Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mamm Genome* **18**, 94-104 (2007).
14. Alavizadeh, A. et al. The Wheels mutation in the mouse causes vascular, hindbrain, and inner ear defects. *Dev Biol* **234**, 244-60 (2001).
15. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res* **32**, D552-6 (2004).
16. Hu, Q. et al. The EGF receptor-sox2-EGF receptor feedback loop positively regulates the self-renewal of neural precursor cells. *Stem Cells* **28**, 279-86 (2010).
17. Bani-Yaghoub, M. et al. Role of Sox2 in the development of the mouse neocortex. *Dev Biol* **295**, 52-66 (2006).
18. Favaro, R. et al. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat Neurosci* **12**, 1248-56 (2009).
19. Taranova, O.V. et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev* **20**, 1187-202 (2006).
20. Howard, T.D. et al. Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. *Nat Genet* **15**, 36-41 (1997).
21. Bajpai, R. et al. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958-62 (2010).
22. Schnetz, M.P. et al. Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Res* **19**, 590-601 (2009).
23. Schnetz, M.P. et al. CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet* **6**, e1001023.
24. Roessler, E. et al. Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proc Natl Acad Sci U S A* **100**, 13424-9 (2003).
25. Bosman, E.A. et al. Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum Mol Genet* **14**, 3463-76 (2005).
26. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function.

Genes Dev **17**, 126-40 (2003).
27. Kiernan, A.E. et al. The Notch ligand Jagged1 is required for inner ear sensory development. *Proc Natl Acad Sci U S A* **98**, 3873-8 (2001).
28. Li, L. et al. Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet* **16**, 243-51 (1997).
29. Brooker, R., Hozumi, K. & Lewis, J. Notch ligands with contrasting functions: Jagged1 and Delta1 in the mouse inner ear. *Development* **133**, 1277-86 (2006).
30. Kiernan, A.E. et al. Sox2 is required for sensory organ development in the mammalian inner ear. *Nature* **434**, 1031-5 (2005).
31. Shaw-Smith, C. Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology. *J Med Genet* **43**, 545-54 (2006).

Methods references

32. Conti, L. et al. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol* **3**, e283 (2005).
33. Sun, Y. et al. Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. *Mol Cell Neurosci* **38**, 245-58 (2008).
34. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-8 (2006).
35. Soler, E. et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* **24**, 277-89 (2010).
36. Hou, J. et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS One* **5**, e10312 (2010).
37. Kiernan, A.E. et al. ENU mutagenesis reveals a highly mutable locus on mouse Chromosome 4 that affects ear morphogenesis. *Mamm Genome* **13**, 142-8 (2002).
38. Sakai, K. & Miyazaki, J. A transgenic mouse line that retains Cre recombinase activity in mature oocytes irrespective of the cre transgene transmission. *Biochem Biophys Res Commun* **237**, 318-24 (1997).
39. Nowak, D.E., Tian, B. & Brasier, A.R. Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715-25 (2005).
40. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* **36**, 5221-31 (2008).
41. Bailey, T.L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* **3**, 21-9 (1995).

Supplementary Information

Supplementary figures

Supplementary figure 1: Information on mouse neural stem cells and FLAG-Sox2 purification

Supplementary figure 2: Expression patterns of *Sox2* and *Chd7* in E14.5 mouse embryos

Supplementary figure 3: Sox2 and Chd7 knock-down verification

Supplementary figure 4: Additional information on genome-binding by Sox2

Supplementary figure 5: Localization of Sox2 and Chd7 on target genes *Tulp3* and *Hes5*

Supplementary figure 6: Additional information on gene regulation by Sox2 and Chd7

Supplementary figure 7: Verification of markers and knock down of Sox2 and Chd7 in human neural stem cells

Supplementary tables

Supplementary table 1: Sox2 interacting proteins, identified by mass spectrometry

Supplementary table 2: Mouse phenotype of Sox2 interacting proteins

Supplementary table 3: Chd7 interacting proteins, identified by mass spectrometry

Supplementary table 11: shRNA sequences

Supplementary table 12: Primers for amplification cDNA from NSC RNA samples

Supplementary table 13: Primers for amplification for genomic regions in ChIP material

Supplementary tables 4 – 10 can be found online at:

http://www.nature.com/ng/journal/v43/n6/full/ng.825.html#supplementary-information

Supplementary table 4: List of genes regulated by Sox2 and Chd7, identified by microarrays

Supplementary table 5: Genome-wide Sox2 peaks, identified by ChIP-sequencing

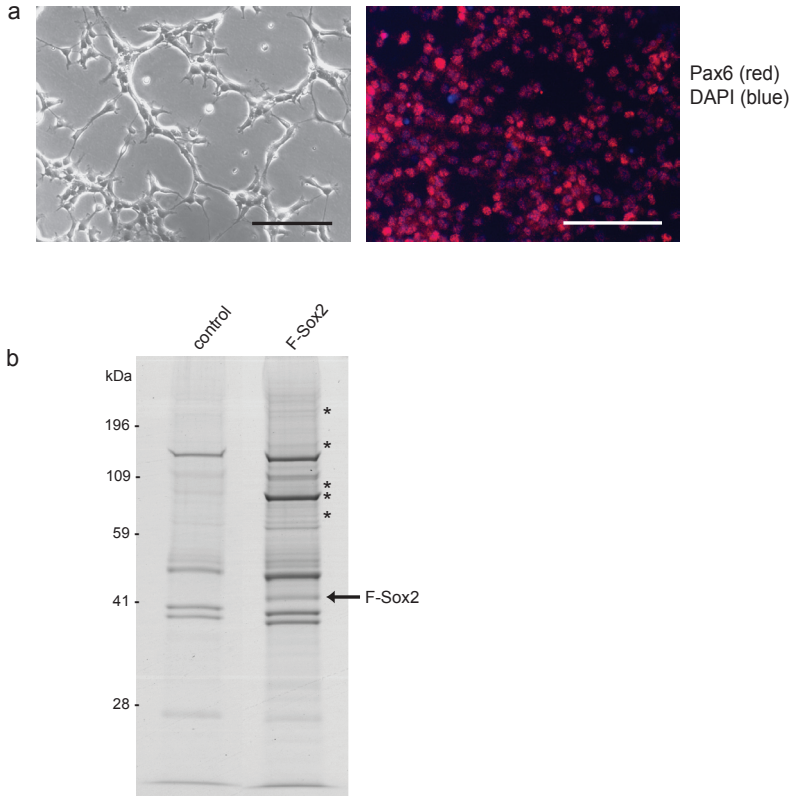
Supplementary table 6: Sox2 peaks within 10 kb of genes.

Supplementary table 7: List of genes that are activated and bound by Sox2.

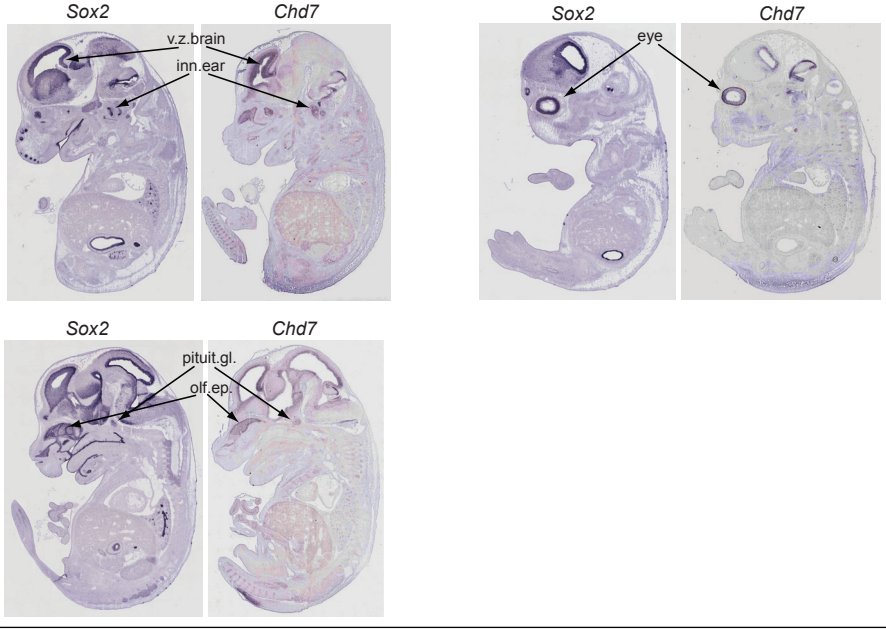
Supplementary table 8: Genome-wide Chd7 peaks, identified by ChIP-sequencing.

Supplementary table 9: Chd7 peaks within 10 kb of genes.

Supplementary table 10: List of genes that are activated and bound by Sox2 and Chd7.

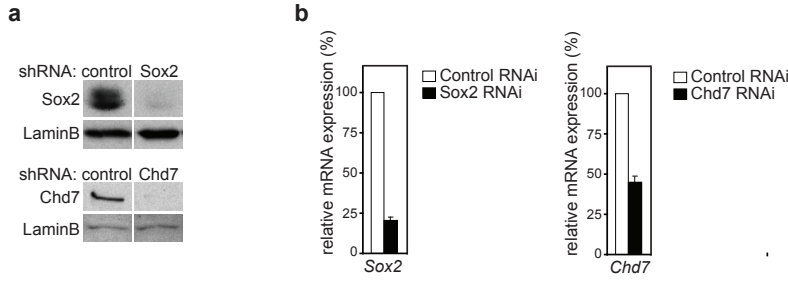


Supplementary figure 1 | Additional information on neural stem cells and F-Sox2 purification.
(a) F-Sox2 expressing neural stem cells (NSCs) have normal morphology and express marker Pax6. Left panel shows a phase-contrast image, right panel shows staining by Pax6-antibody. Bar, 100 mm. (b) Representative PAA gel of an F-Sox2 purification from F-Sox2 expressing NSCs and control purification from NSCs. F-Sox2 protein is indicated by arrow. Asterisks indicate bands present in the F-Sox2 purification and not in the control purification and likely represent Sox2-interacting proteins.

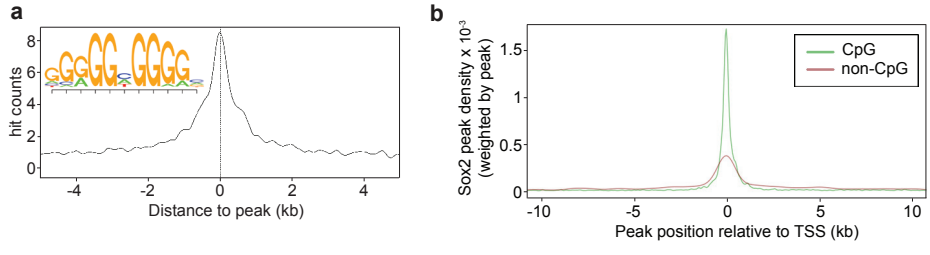


Supplementary figure 2 | Gene expression patterns of Sox2 or Chd7 overlap in E14.5 mouse embryos.
Images were taken from the Genepaint database(1). Gene expression patterns were determined on cryosections of E14.5 embryos by RNA in situ hybridization with the indicated gene-specific probes. Sox2 and Chd7 were observed to be expressed in the ventricular zone of the brain (v.z.brain), the inner ear (inn. ear), the eye, the olfactory epithelium (olf. ep.) and the pituitary gland (pituit. gl.). Expression pattern data in the embryo that support and/or extend the above images were reported for Sox2(2-6) and Chd7 (7-11).

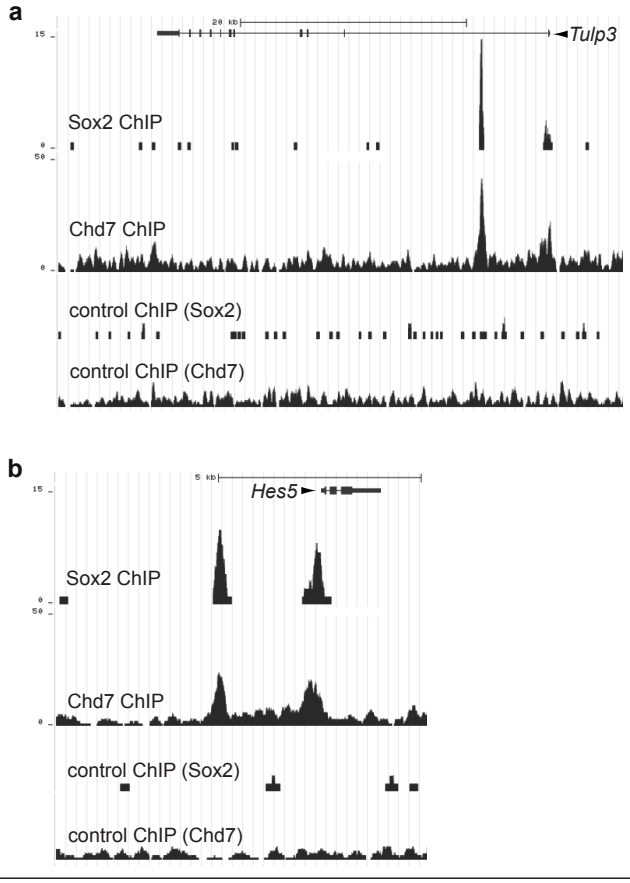
(1) Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res 32, D552-6 (2004).
(2) Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17, 126-40 (2003).
(3) Ferri, A.L. et al. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development 131, 3805-19 (2004).
(4) Kiernan, A.E. et al. Sox2 is required for sensory organ development in the mammalian inner ear. Nature 434, 1031-5 (2005).
(5) Taranova, O.V. et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev 20, 1187-202 (2006).
(6) Kelberman, D. et al. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. J Clin Invest 116, 2442-55 (2006).
(7) Que, J. et al. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development 134, 2521-31 (2007).
(8) Hurd, E.A. et al. Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. Mamm Genome 18, 94-104 (2007).
(9) Bosman, E.A. et al. Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. Hum Mol Genet 14, 3463-76 (2005).
(10) Layman, W.S. et al. Defects in neural stem cell proliferation and olfaction in Chd7 deficient mice indicate a mechanism for hyposmia in human CHARGE syndrome. Hum Mol Genet 18, 1909-23 (2009).
(11) Bergman, J.E., Bosman, E.A., van Ravenswaaij-Arts, C.M. & Steel, K.P. Study of smell and reproductive organs in a mouse model for CHARGE syndrome. Eur J Hum Genet 18, 171-7 (2010).



Supplementary figure 3 | Knock-down of Sox2 and Chd7 by RNA interference.
(a) Proteins levels in extracts from mouse NSCs transfected with the indicated shRNA-expressing plasmids were detected by western blot. LaminB serves as an equal loading control. (b) mRNA levels as detected by RT-PCR for Sox2 (left panel) and Chd7 (right panel) in mouse NSCs transfected with the indicated shRNA-expressing plasmids. S.e.m. is indicated of three independent experiments.

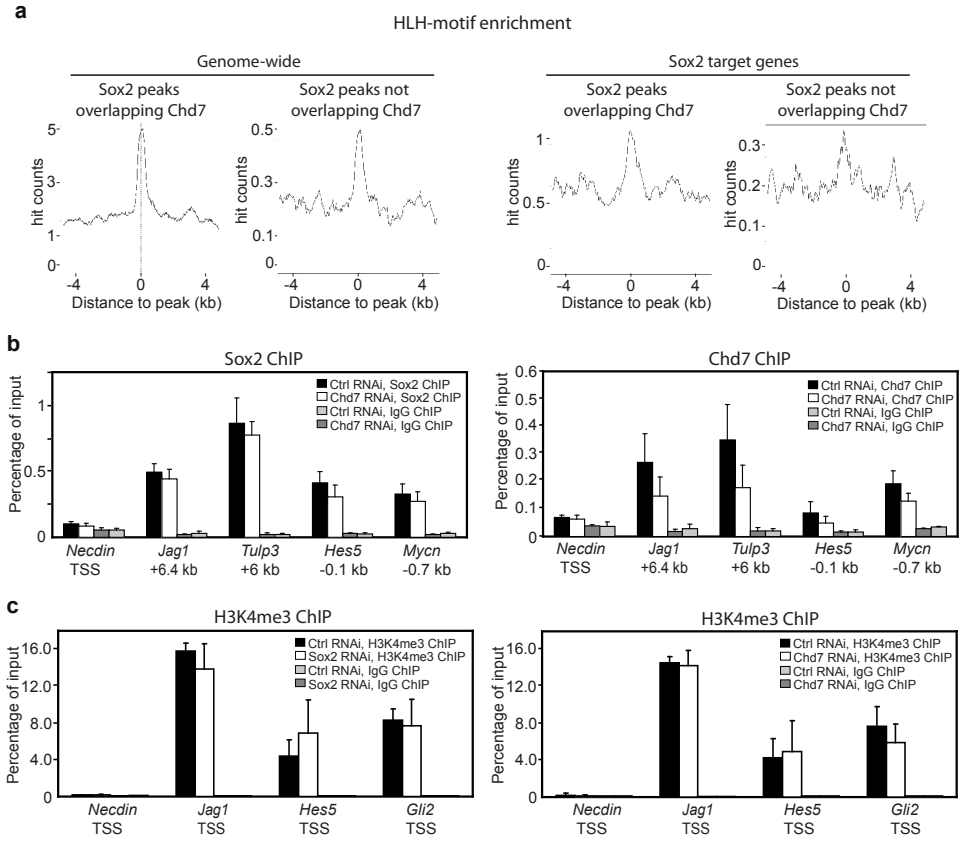


Supplementary figure 4 | Additional information on genome binding by Sox2.
(a) Enrichment of the G-rich motif in Sox2 peaks. (b) Density of Sox2 peaks relative to transcriptional start site (TSS) of CpG island promoters and non-CpG island promoters.



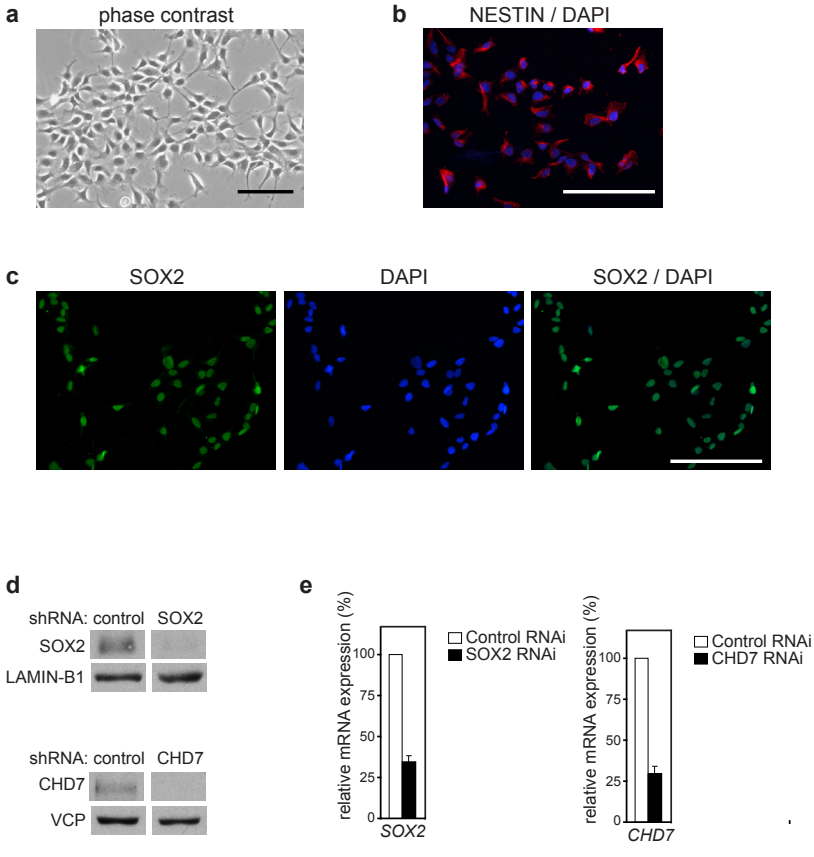
Supplementary figure 5 | Localization of Sox2 and Chd7 on Sox2-Chd7 target genes.

The number of overlapping sequence reads from the indicated ChIP experiments were plotted relative to chromosomal position on the UCSC genome browser. Sox2 ChIP, Chd7 ChIP and their corresponding control (IgG) ChIP experiments are shown. Genome locations of the Tulp3 locus (a) and Hes5 locus (b) are shown. Arrowhead indicates direction of the locus, scale bar represents 20 kb (a) or 5 kb (b).



Supplementary figure 6 | Characterization of gene regulation by Sox2 and Chd7.

(a) The HLH motif is more enriched in Sox2 peaks that overlap with Chd7 peaks. HLH motif enrichment is shown for genome-wide Sox2 peaks or Sox2 peaks in Sox2 target genes (Supplementary table 7). (b) Chd7 does not affect Sox2 recruitment. Chd7-dependence of Sox2 binding (left panel) or Chd7 binding (right panel) to binding sites in the indicated genes, using Chd7 kd NSCs or control kd NSCs, ChIP is with the indicated antibodies, S.e.m. is indicated of two independent experiments. (c) No change in H3K4me3 histone modification levels upon Sox2 kd (left panel) or Chd7 kd (right panel) on the transcription start site (TSS) of the indicated genes, as assessed by H3K-4me3 ChIP. S.e.m. is indicated of two independent experiments, Necdin TSS is a negative control genomic region.



Supplementary figure 7 | Human neural stem cell characterization and knock-down.
Human neural stem cells have normal morphology and express markers NESTIN and SOX2. (a) phase-contrast image. (b) Immunostaining with anti-NESTIN, DAPI counterstain. (c) Immunostaining with anti-SOX2 (left panel), counterstain with DAPI (middle panel) and merge (right panel). Scale bars, 100 mm. (d) Knock-down of SOX2 and CHD7 by RNA interference. Proteins levels in extracts from human NSCs transfected with the indicated shRNA-expressing plasmids were detected by western blot. LAMIN-B1 serves as an equal loading control for SOX2 knock-down (upper panel) and VCP (Valosin Containing Protein) for the CHD7 knock-down (lower panel). (e) mRNA levels as detected by RT-PCR for SOX2 (left panel) and CHD7 (right panel) in human NSCs transfected with the indicated shRNA-expressing plasmids. S.e.m. is indicated of three independent experiments.

Supplementary table 1 | Sox2-interacting proteins, as identified by mass spectrometry analysis of purified Sox2 samples.

		F-Sox2 #1		F-Sox2 #2		Sox2-IP		
Protein	**Accession**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Average mascot**
Sox2	gi|127140986	315	5	508	5	614	6	479
NuRD complex								
Mi2-beta (Chd4)	gi|39204553	1362	24	1910	32	967(164)	16(3)	1413
Mta2	gi|51491880	1021	15	1140	16	479	8	880
Mta1	gi|15077051	882	13	1262	18	314	4	819
Gatad2b	gi|120577529	745	13	1271	15	315	4	777
Hdac1	gi|2347180	589	10	1005	14	473	9	689
Hdac2	gi|3023934	525	9	797(105)	12(1)	394(187)	8(3)	572
Rbbp7	gi|2494892	594	9	770(75)	13(1)	264(259)	5(4)	543
Gatad2a	gi|148696823	406	8	920	13	193	2	506
Mbd3	gi|7305261	345	5	487	7	199	4	344
Spalt proteins (NuRD associated)[c]								
Sall3	gi|49257163	778	13	2490	33	1061	15	1443
Sall2	gi|49087134	414	6	1587	18	-	-	667
Sall1	gi|11496251	349	5	1316	18	-	-	555
SMRT/ NcoR complex								
SMRT (NcoR2)	gi|119226235	703	14	1678	31	173	3(1)	851
Ira1 (Tbl1xr1)	gi|12006108	259	5	406	4	230	4	298
Sin3a	gi|91980275	68	2	308	6	61	2	146
Hdac3	gi|6840851	162	3	252	4	-	-	138
NcoR1	gi|119624899	133	3	168	3	-	-	100
SWI/ SNF complex								
Baf170 (Smarcc2)	gi|38565930	326	5	1054	13	358(102)	6(2)	579
RbAp48 (Rbbp4)	gi|1016275	473(147)	7(4)	656(104)	9(1)	376(255)	6(4)	502
Baf53a (Actl6a)	gi|23396474	463	6	472	6	187	2	374
Baf155 (Smarcc1)	gi|30851572	320	5	449	8	284(102)	5(2)	351
Baf60a (Smarcd1)	gi|27502706	121	3	198	3	-	-	106
Baf250 (Smarcf1)	gi|14150461	176	4	63	2	58(54)	1(1)	99
Ini1 (Smarcb1)	gi|6755578	67	1	72	1	-	-	46
Trrap complex								
Ruvbl2	gi|9790083	265	4	857	13	693(255)	10(5)	605
Ruvbl1	gi|6755382	217	3	969	13	223(98)	4(1)	470
Trapp	gi|38605208	254	7	193	1	113(63)	4(2)	187
Transcription factors								
Chd7	gi|124487249	1141	22	2179	31	438	8	1253
Cutl1	gi|60360228	619	11	1936	26	1021	14	1192
Dbc1	gi|94397239	504	7	1533	22	-	-	679
Zeb1	gi|141796995	203	3	532	7	200	3	312
Ctbp1	gi|3452507	296	4	430	6	-	-	242
Supt16h	gi|110287968	183	3	295	5	83	1	187
Rfx3	gi|34328189	133	2	368	5	-	-	167
Sox8	gi|33563276	238	5	263	5	-	-	167
Hoxa5	gi|6754232	221	3	200	2	78	1	166
Myef2	gi|27819594	266	4	162	3	50	1	159
Nfi-beta	gi|1524157	164	3	200	4	52	1	139
Ctbp2	gi|15426462	128	2	199	3	-	-	109
Nac1	gi|31543309	229	3	84	2	-	-	104
Tead1	gi|3041733	65	1	195	3	-	-	87
Snf2h	gi|14028669	184	4	82	3	-	-	87
Sox5	gi|83404978	90	2	117	2	-	-	69
Twist1	gi|6755907	118	2	72	1	-	-	63
Tcf3	gi|4151036	73	2	64	1	-	-	46
Zfp191	gi|33636732	61	2	63	1	-	-	41
Other								
Exp4	gi|10048438	1330	22	2042	27	1629	23	1667
Skiv2l2	gi|21312352	164	4	373	5	-	-	179
Dock7	gi|122065171	304	5	102	1	-	-	135
Dnaja2	gi|9789937	121	2	158	2	-	-	93

[a] Mascot score for the specified protein in the Sox2 sample. Mascot score for the specified protein in the corresponding control sample, if present, is between brackets.

[b] Number of identified unique, non-redundant peptides for the specified protein in the Sox2 sample. Number of identified unique peptides in the control purification is between brackets.

[c] van den Berg, D.L. et al. An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 6, 369-81 (2010) and references therein.

Supplementary table 2 | Sox2-interacting proteins with a neural phenotype.

Supplementary Table 2. Sox2-interacting proteins with a neural pher

Protein	Knock-out or heterozygous neural phenotype in mice
NuRD complex	
Gatad2a	KO dies at E9.5 with abnormal neural fold formation[1]
Hdac2	KO has increased synaptic plasticity and memory formation[2]
Spalt proteins	
Sall1	KO has defects in olfactory neurogenesis[3]
Sall2	KO shows neural tube closure defects[4]
Sall3	KO has defect in the formation of cranial nerve and olfactory interneurons[5]
SWI/SNF complex	
Baf155	+/- has abnormal neuroepithelial differentiation and brain development[6]
SMRT/ NcoR complex	
SMRT	KO has reduced NSC maintenance and abnormal brain development, dies at E16.5[7]
NcoR1	KO has reduced NSC maintenance and increased differentiation to astrocytes[8]
Transcription factors	
Chd7	KO has neuroepithelium hypoplesia, dies at E10.5[9,10], +/- has defects in olfactory neurogenesis[11,12]
Cutl1	Cutl1-Cutl2 DKO have no Reelin-expressing cortical interneurons[13]
Nfi-beta	KO has severe defects in development of forebrain, hippocampus and dentate gyrus[14]
Ctbp2	KO has abnormal forebrain and midbrain development[15]
Twist1	KO shows neural tube closure defects, dies at E11.5[16]
Zfp191	KO disrupts oligodendrocyte function and has post-natal myelination defects[17]

(1) Marino, S. & Nusse, R. Mutants in the mouse NuRD/Mi2 component P66alpha are embryonic lethal. PLoS ONE 2, e519 (2007).
(2) Guan, J.S. et al. HDAC2 negatively regulates memory formation and synaptic plasticity. Nature 459, 55-60 (2009).
(3) Harrison, S.J., Nishinakamura, R. & Monaghan, A.P. Sall1 regulates mitral cell development and olfactory nerve extension in the developing olfactory bulb. Cereb Cortex 18, 1604-17 (2008).
(4) Bohm, J. et al. Sall1, sall2, and sall4 are required for neural tube closure in mice. Am J Pathol 173, 1455-63 (2008).
(5) Parrish, M. et al. Loss of the Sall3 gene leads to palate deficiency, abnormalities in cranial nerves, and perinatal lethality. Mol Cell Biol 24, 7102-12 (2004).
(6) Kim, J.K. et al. Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. Mol Cell Biol 21, 7787-95 (2001).
(7) Jepsen, K. et al. SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. Nature 450, 415-9 (2007).
(8) Hermanson, O., Jepsen, K. & Rosenfeld, M.G. N-CoR controls differentiation of neural stem cells into astrocytes. Nature 419, 934-9 (2002).
(9) Hurd, E.A. et al. Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. Mamm Genome 18, 94-104 (2007).
(10) Alavizadeh, A. et al. The Wheels mutation in the mouse causes vascular, hindbrain, and inner ear defects. Dev Biol 234, 244-60 (2001).
(11) Layman, W.S. et al. Defects in neural stem cell proliferation and olfaction in Chd7 deficient mice indicate a mechanism for hyposmia in human CHARGE syndrome. Hum Mol Genet 18, 1909-23 (2009).
(12) Bergman, J.E., Bosman, E.A., van Ravenswaaij-Arts, C.M. & Steel, K.P. Study of smell and reproductive organs in a mouse model for CHARGE syndrome. Eur J Hum Genet 18, 171-7 (2010).
(13) Nieto, M. et al. Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. J Comp Neurol 479, 168-80 (2004).
(14) Barry, G. et al. Specific glial populations regulate hippocampal morphogenesis. J Neurosci 28, 12328-40 (2008).
(15) Hildebrand, J.D. & Soriano, P. Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. Mol Cell Biol 22, 5296-307 (2002).
(16) Chen, Z.F. & Behringer, R.R. twist is required in head mesenchyme for cranial neural tube morphogenesis. Genes Dev 9, 686-99 (1995).
(17) Howng, S.Y. et al. ZFP191 is required by oligodendrocytes for CNS myelination. Genes Dev 24, 301-11.

Supplementary table3 | Chd7-interacting proteins as identified by mass spectrometry analysis of Chd7 immunoprecipitates.

		Chd7-IP 1		Chd7-IP 2		
Protein	**Accession**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Average mascot**
Chd7	gi|124487249	7066	97	7102	100	7084
NuRD complex						
Mta2	gi|51491880	616(62)	11(1)	1051	17	834
Mta1	gi|15077051	531	7	509	10	520
Gatad2a	gi|148696823	100	1	292	5	196
Spalt proteins (NuRD associated)[c]						
Sall3	gi|115528513	315	7	1053	16	684
SWI-SNF complex						
Baf170 (Smarcc2)	gi|38565930	414	6	1781(102)	24(2)	1098
Brg1 (Smarca4)	gi|76253779	104	3	1321(245)	19(5)	713
Baf250 (Smarcf1)	gi|14150461	62	1	1203(54)	18(1)	633
Baf57 (Smarce1)	gi|10181166	82	1	444	6	263
Transcription factors						
Zbtb20	gi|9790133	374	6	300	5	337
Sox2	gi|127140986	81	2	314	3	198
Olig1	gi|7385152	80	1	89	1	85
Other						
Vars	gi|12643967	666	10	1794	31	1230

[a,b,c,] Equivelant to Supplementary table 1.

Supplementary table 11 | shRNA sequences

Gene	shRNA sequence
Sox2	5'-GGTTGATATCGTTGGTAAT-3'
Chd7	5'-GCCAGCCGTCGGACCATTC-3'
SOX2	5'-AGACTAGGACTGAGAGAAA-3'
CHD7	5'-GCTGTTCTGTAACATAGTG-3'

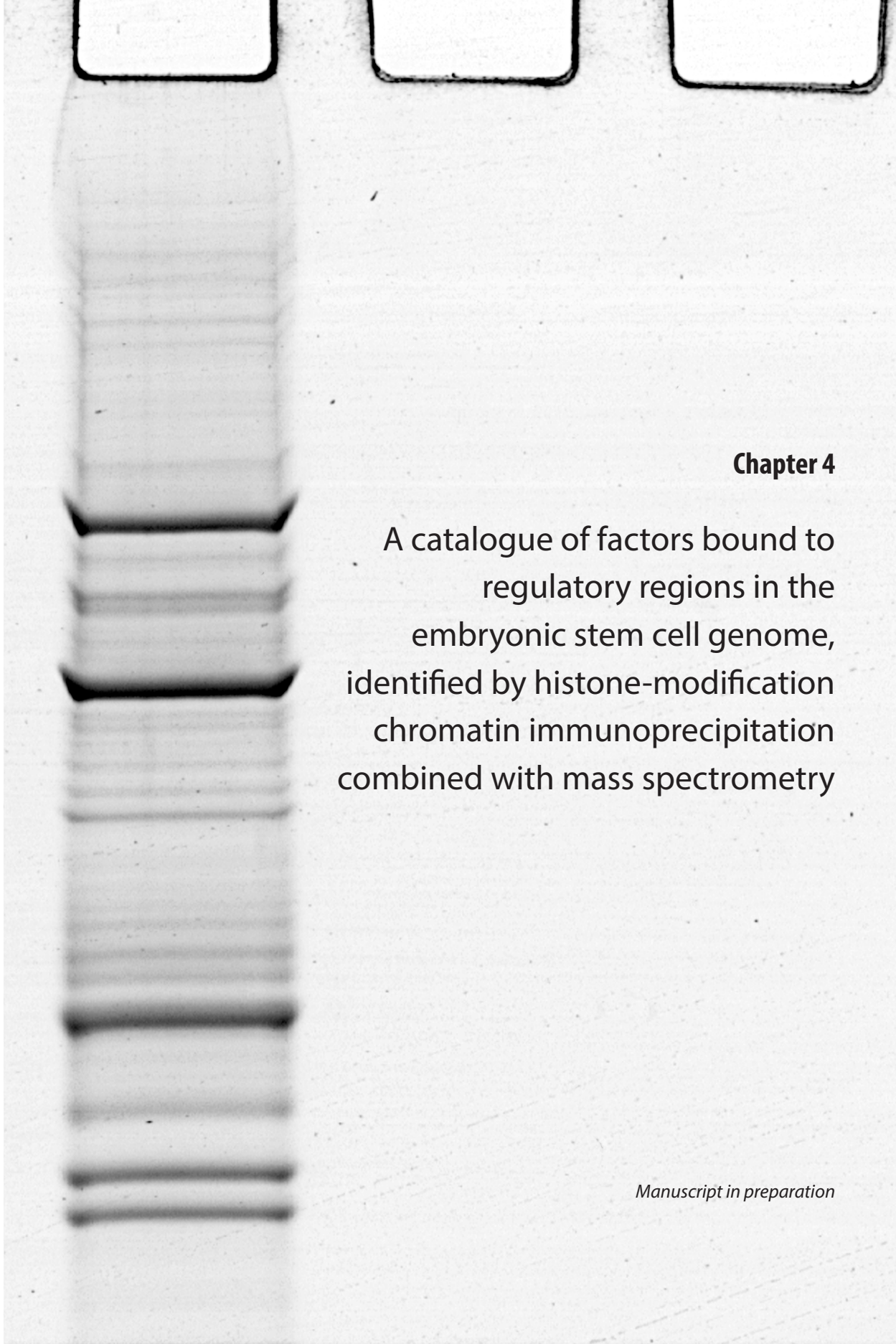
Supplementary table 12 | Primers for amplification cDNA from NSC RNA samples

Gene	Orientation	Sequence
Jag1	forward	5'-TTGGCTGCAATAAGTTCTGT -3'
	reverse	5'-TGCAGTCACCTGGAAGTTTA -3'
Hes5	forward	5'-AGCTACCTGAAACACAGCAA-3'
	reverse	5'-GCTGGAAGTGGTAAAGCAG-3'
Rbpj	forward	5'-GCAAAAGTTGCACAGAAGTC-3'
	reverse	5'-CCTATTCCAATAAACGCACA-3'
Gli2	forward	5'-CATCTGAAAGAGAGGGGACT -3'
	reverse	5'-GGTCACACGTGGACTAGAGA -3'
Gli3	forward	5'-CTTGCCCTTCATTAGGATCT-3'
	reverse	5'-CAGAGCCATCTGGTGATAGT-3'
Tulp3	forward	5'-AAGCCTCAGGTTCTCTCTGT-3'
	reverse	5'-GCTCCTCGTCATAGTTCACA-3'
Mycn	forward	5'-GTGTCTGTTCCAGCTACTGC-3'
	reverse	5'-GCTCCTCGTCATAGTTCACA-3'
JAG1	forward	5'-AGTCCTAAGCATGGGTCTTG-3'
	reverse	5'-CCAGTTGGTCTCACAGAGG-3'
GLI2	forward	5'-AAGAAAGTGATGATGCGATG-3'
	reverse	5'-ACTTTTGGCTTCTTGCTTCT-3'
GLI3	forward	5'-AATGTTCCTAGAGGGTCTGC-3'
	reverse	5'-GTTCCTCACTGACTTTGCTG-3'
MYCN	forward	5'-ACAAGGCCCTCAGTACCTC-3'
	reverse	5'-ACAGTGATGGTGAATGTGGT-3'
Sox2	forward	5'-AAACATGGCAATCAAATGTC-3'
	reverse	5'-TTGCCAGTACTTGCTCTCAT-3'
Chd7	forward	5'-AACCTGTCCTCCACTACAGC-3'
	reverse	5'-TCACTAGCTGAGCGTTCTGT-3'
Hprt	forward	5'-AGCCTAAGATGAGCGCAAGT-3'
	reverse	5'-ATGGCCACAGGACTAGAACA-3'
CalR	forward	5'-GACTTTCTGCCACCCAAG-3'
	reverse	5'-GTTCCCACTCTCCATCCA-3'
SOX2	forward	5'-GTTCATCGACGAGGCTAAG-3'
	reverse	5'-GTTCATGTGCGCGTAACT-3'
CHD7	forward	5'-CATTAGTGGGAGTGAGGACA-3'
	reverse	5'-TGTCCATTTTCTGAGACCAC-3'
HPRT	forward	5'-TGGAAAGGGTGTTTATTCCT-3'

	reverse	5’-GCTTTGATGTAATCCAGCAG-3’
CALR	forward	5’-GCTCCAGGAATACACCCAAA-3’
	reverse	5’-CAGCTCATGCTCGTCAATGT-3’

Supplementary table 13 | Primers for amplification genomic regions in ChIP material

Genomic region	Orientation	Sequence
Necdin	forward	5’-GGTCCTGCTCTGATCCGAAG-3’
	reverse	5’-GGGTCGCTCAGGTCCTTACTT-3’
Jag1	forward	5’-GAGTTGGCTGGACTGACTGA-3’
	reverse	5’-ATCCTGAGAATGTCCCGAGT-3’
Rbpj	forward	5’-TCTGCACCCACACCTACATC-3’
	reverse	5’-TGTTCACTTTGCACCCACA-3’
Hes5	forward	5’-GGGAAAAGGCAGCATATTG-3’
	reverse	5’-CACGCTAAATTGCCTGTGAA-3’
Gli2	forward	5’- TTGCCTTTTCCCAATTCTCT -3’
	reverse	5’- CCCGGGCTGATAAATTAAAA -3’
Gli3	forward	5’- GATCAGTCAGGCCATCCAC -3’
	reverse	5’- CCGCAAAACAAAGAACTTCA -3’
Tulp3	forward	5’-GTGTGAGCTGGATTCTTCAG-3’
	reverse	5’-GACAGGAAATGACTCCTGGT-3’
Mycn	forward	5’- AGGGACTGGGCTAGAAACCT -3’
	reverse	5’- TCGTTTTTCAGACTGCAAGC -3’
NECDIN	forward	5’- CTGTTTGGGCTGAGAAGAT -3’
	reverse	5’- AAGAAACTTGACCCCAACAT -3’
JAG1	forward	5’- GCAGAGCGGTAAGCACTTAAT -3’
	reverse	5’- GTTTGGATGGCGGTTTATTT -3’
GLI2	forward	5’- TGAAATTGCTCCTGCACTTC -3’
	reverse	5’- ATGTCGGATGACCCTTTCTC -3’
GLI3	forward	5’- CCTTTTGACAGCCATTTTCA -3’
	reverse	5’- GAAGTTCGGGGACTTGACAG -3’
MYCN	forward	5’- TCGGACTACCCTTCTTTCGT -3’
	reverse	5’- GGGAGACCGATGCTTCTAAC -3’
Jag1 promoter	forward	5’- AGGAAAGAAAGCCGAGAGGT -3’
	reverse	5’- GCACGACTGGAAAACAACAC -3’
Gli2 promoter	forward	5’- GTGGGGGAGAGTCTGTGTTC -3’
	reverse	5’- GCAATCCATCAGCGTCTCT -3’



Chapter 4

A catalogue of factors bound to regulatory regions in the embryonic stem cell genome, identified by histone-modification chromatin immunoprecipitation combined with mass spectrometry

Manuscript in preparation

A catalogue of factors bound to regulatory regions in the embryonic stem cell genome, identified by histone-modification chromatin immunoprecipitation combined with mass spectrometry.

Erik Engelen[1], Johannes H. Brandsma[1], Maaike Moen[1], Luca C. Signorile[1], Dick H. Dekkers[2], Jeroen Demmers[2], Christel Kockx[3], Wilfred van IJcken[3], Debbie L.C. van den Berg[1,4], Raymond A. Poot[1]

[1] Department of Cell Biology, Erasmus MC, Dr. Molewaterplein 50, 3015 GE, Rotterdam, The Netherlands
[2] Proteomics Center, Erasmus MC, The Netherlands
[3] Center for Biomics, Erasmus MC, The Netherlands
[4] Division of Molecular Neurobiology, MRC-National Institute for Medical Research, London NW7 1AA, United Kingdom

ABSTRACT

The ENCODE consortium has mapped the genome-wide locations of transcriptional enhancers and promoters in many different cell types and thereby provided a molecular signature of their cell identity. Factors that bind those regulatory regions can regulate cell identity but have not been systematically identified. Here we purified native enhancers, promoters or heterochromatin from embryonic stem cells (ESCs) and identified their associated factors, using an adapted chromatin immunoprecipitation protocol against several histone modifications followed by factor identification by mass spectrometry (ChIP-MS). We identified approximately 250 factors that are enriched in a particular chromatin fraction, suggesting their specific association with enhancers, promoters or heterochromatin. Analysis of genome-wide localization data by our ChIP-MS procedure shows the high accuracy of localization prediction. Strikingly, more than a quarter of the identified factors have been documented to be important for maintaining ES cell pluripotency and include Oct4, Esrrb, Klf5 and Dppa2, ESC-specific factors that promote reprogramming to pluripotency. We further show that Dppa2 binds the promoters of testis expressed genes in ESCs. Our ChIP-MS protocol is adaptable to other histone modifications and can be applied to any cell type in culture to identify the spectrum of factors that occupy their regulatory regions and may define their identity.

INTRODUCTION

A mammalian genome can support the generation of few hundreds of different cell types in an adult organism. These cell types differ in their gene expression profiles as a direct consequence of differences in the activation state of their gene promoters and distal *cis*-regulatory elements often collectively called enhancers. The ENCODE consortium has generated a wealth of data on the genomic signatures of many different types of mouse and human cells [1,2]. In particular, the genome-wide identification of regulatory regions such as transcriptional enhancers and promoters and their state of activity has the potential to increase our understanding of the epigenetic identity of a particular cell type. Nevertheless, the identity of cells is not determined by the location of enhancers and promoters *per se*, but to a large extent by transcription factors that bind these DNA elements, as is becoming increasingly clear from many experiments including recent ones of reprogramming of cell identity [3]. It is therefore of interest to purify native transcriptional enhancers and promoters of a particular cell type and identify the proteins that bind to these regulatory regions, which could allow for the *de novo* identification of cell identity determination factors. Here we modified the protocol for chromatin immunoprecipitation of different histone modifications and identified the proteins bound to precipitated chromatin fractions by mass spectrometry (ChIP-MS). Performing ChIP-MS on ESCs identified 245 factors that we predict to bind to promoters, enhancers or heterochromatin. These respective three sets of factors contain ubiquitously expressed factors, some of which were previously identified to bind to specific modified histone tail peptides, euchromatin or heterochromatin in HeLa cells [4,5]. However, we also identified ESC-specific transcription factors such as Oct4, Esrrb, Dppa2 and Klf5 that are not only important for maintaining ESC self-renewal but also known to facilitate reprogramming of somatic cells to pluripotency [3,6-8]. Comparison with genome-wide data sets that are available for over 29 ChIP-MS-identified factors showed a high accuracy of our prediction of localization. We determined the genome-wide localization of Dppa2, a factor that is highly expressed in ESCs and shown to be essential for normal development [9] and a facilitator of induced pluripotent stem cell (iPSC) formation. Our method is applicable to any histone or DNA modification for which a chromatin immunoprecipitation protocol is available and to any cell type that can be grown in sufficient quantities.

RESULTS & DISCUSSION

ChIP-MS rationale and procedure

Transcriptional enhancers and promoters have been defined by the chemical modification of their associated histones, mostly histone H3 [10]. Poised enhancers were found to contain

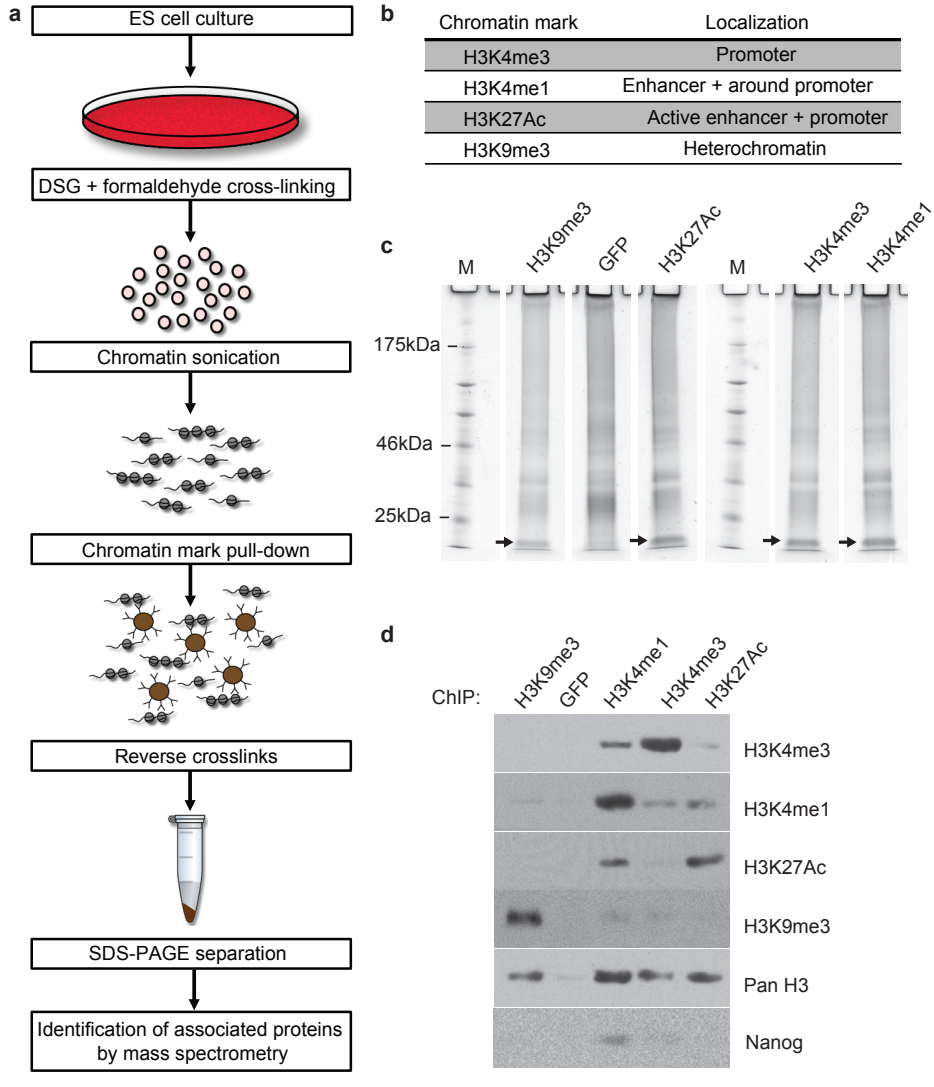


Figure 1 | ChIP-MS procedure for identification of promoter and enhancer associated factors.
a, Flowchart of the ChIP-MS procedure. **b**, Chromatin marks and their localization on the genome. **c**, Chromatin fractions and GFP control sample separated by 10% SDS-PAGE. A compressed band for the histones can be observed in all fractions and is absent in the GFP control sample (arrows). Molecular weight marker depicted by M. **d**, Precipitation of the specific chromatin marks and their detection, including that of Nanog, by Western blot.

histone H3 that is mono-methylated at lysine 4 (H3K4me1) [11], and upon subsequent activation also accumulate the H3K27 acetylation mark (H3K27Ac) [12,13]. Promoters of transcribed genes contain H3K4me3 marks and the level of their activity is reflected by the presence of histone acetylation, including the H3K27Ac mark [10]. Inactive (hetero)chromatin is marked by H3K9me3 [14]. The presence or absence of these and other chromatin marks

was used to postulate 15 different chromatin regions in the mammalian genome, including promoters and enhancers with different levels of activation [10].

We performed large scale chromatin immunopreciptations in duplicate against four histone marks, H3K4me3, H3K4me1, H3K27Ac and H3K9me3, and against GFP as a control, in mouse ESCs (**Fig. 1a, b**). Crosslinking of the chromatin was performed with a protein-protein cross-linker, DSG, followed by standard formaldehyde cross-linking, with the idea to increase cross-linking efficiency of genome-bound factors to the chromatin [15-17]. Bound protein factors were de-cross-linked and eluted by prolonged heating in protein denaturing conditions [18], loaded on an SDS-PAGE gel and analyzed by mass spectrometry. A representative protein gel showed that the (unresolved) histones precipitated with each histone modification antibody but not with the GFP control (**Fig. 1c**). Analysis by Western blot revealed that comparable amounts of chromatin were precipitated in the different histone-modification ChIPs, as indicated by the total content of histone H3 (**Fig. 1d**). ChIP against H3K4me1 efficiently precipitated H3K4me1-marked chromatin and also precipitated minor amounts of H3K4me3- and H3K27Ac-marked chromatin (**Fig. 1d**, H3K4me1 ChIP). This was to be expected as these histone marks slightly overlap at promoters and enhancers, respectively [12]. H3K4me3 ChIP precipitated also minor amounts of H3K4me1-marked chromatin and H3K27Ac-marked chromatin. H3K27Ac ChIP co-precipitated minor amounts of H3K4me1 and H3K4me3-marked chromatin due to their overlap at enhancers and promoters, respectively. H3K9me3 ChIP precipitated H3K9me3-marked (hetero)chromatin with negligible contamination of the other types of chromatin.

Prediction of genome localization of identified factors by ChIP-MS

We analyzed the different precipitated chromatin fractions and GFP-control fractions by mass spectrometry for an unbiased identification of the protein factors present in each fraction. We identified 251 factors that were 3-fold or more enriched by average emPAI score, a semi-quantitative measure for the amount of protein present [19], in one chromatin fraction over another chromatin fraction, whereas we had no or very low presence in the GFP control (**Supplementary table 1, 2**). Within this set, six factors were only present in the H3K27Ac fraction which fails to discriminate between promoters, enhancers and heterochromatin (**Supplementary table 1-3, Fig. 2**).

An early indication that of our ChIP-MS method worked came from the identification of RNA polymerase subunits pol2a,b,c,e and g and associated TFIID subunits Taf1-7 predominantly or solely in the H3K4me3 fraction (**Supplementary table 1, 2 and 3**), which was to be expected as these factors together form the Pol II holo-complex at active promoters. All four identified subunits of the NuRD complex had their highest score in the H3K4me1 fraction (**Supplementary table 1, 2 and 3**), predicting a predominant

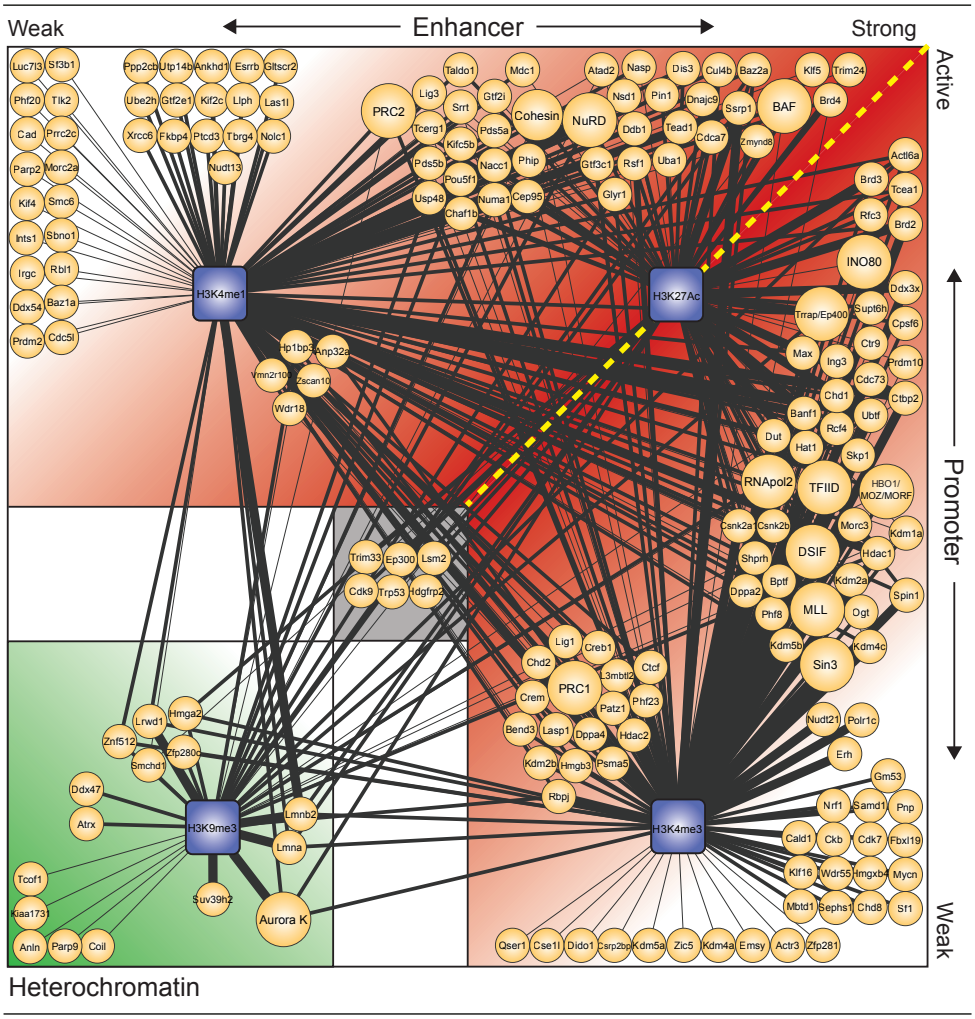


Figure 2 | Catalogue of factors identified by ChIP-MS.
Overview of factors (small orange circles) and complexes (large orange circles) identified by ChIP-MS purifications for the specific chromatin marks (purple squares). The thickness of the line indicates the enrichment of a factor/complex in the associated chromatin fraction as determined by average emPAI score. Thin lines (1 pixel) represent average emPAI scores <0.05, intermediate lines (3 pixel) represent average emPAI scores ≥0.05 and thick lines (6 pixel) represent average emPAI scores ≥0.3. Factors are divided over 4 quadrants according to their prediction by ChIP-MS. Identified factors assigned to the heterochromatin fraction are displayed in the bottom left corner (green). Factors assigned to the enhancer or promoter chromatin fractions are displayed on a horizontal axis in the top quadrant or vertical axis in the right quadrant, respectively. The factors in the enhancer and promoter quadrants are ranked according to their H3K27Ac ratios mark (red gradient). Factors of which their prediction is unclear are located in the middle quadrant (gray). Figure is generated with the aid of cytoscape [42].

localization at enhancers, as was experimentally shown for its key subunit Chd4 [20]. H3K9 methyltransferase Suvar3-9 binds pericentric heterochromatin and was indeed observed solely in the H3K9me3 fraction [21]. We assigned to identified factors the locations "promoter", "enhancer" and "heterochromatin" on the basis of the fractions in which

they have the highest emPAI value (**Supplementary tables 1, 2, and 3**, see methods). This annotation is not absolute, as factors can be present in more than one location, but it does provide clarity and facilitates a more systematic comparison with published genome-wide localization data (see below).

We identified a large number of subunits of established chromatin modifying complexes such as the BAF complex, Sin3 complex, MLL complex and NuRD complex (**Supplementary table 3, Fig. 2**). Strikingly, our prediction of localization between different subunits of the same complex is nearly 100% identical (**Supplementary table 3**), which suggests a low level of false identifications due to spurious chromatin binding.

Arguably, the most interesting factors present in a given cell type are those that determine cell identity. These have been relatively well characterized in ESCs due to many studies that identified factors that are indispensable for maintaining ESC pluripotency[22-24] or factors that reprogram somatic cells towards ESC-like induced pluripotent cells [3,6,7]. A literature search revealed a pluripotency phenotype for more than one quarter of our identified factors (68 out of 251, **Supplementary table 1**). This category includes key ESC pluripotency factors such as Oct4 (Pou5f1), Esrrb and Klf5 (**Supplementary table 1, 2**), which can be part of a combination of 3-4 factors that reprograms somatic cells to pluripotent stem cells [3,6,8]. Our ChIP-MS data predict that all three factors bind predominantly to enhancers, in agreement with published genome-wide localization data [25]. Another well-known factor important for ESCs, Nanog, is difficult to identify by mass spectrometry [26]. Western blot analysis on our ChIP-MS samples revealed that Nanog is also predominantly present in the H3K4me1 fraction (**Fig. 1d**), suggesting it binds to enhancers, in agreement with published data [25].

Verification of ChIP-MS prediction accuracy with published genome-wide localization data

Our list of 245 identified factors includes 29 factors for which the genome-wide localization has been determined in mouse ESCs (**Fig. 3**). To verify whether our localization prediction by ChIP-MS is accurate, we compared our prediction from the ChIP-MS analysis (**Fig. 3,** left panel) with the correlation of genome-wide binding sites of a factor with the different histone marks (**Fig. 3**, right panel). Of the 17 factors predicted by ChIP-MS to be predominantly promoter-associated, 14 were indeed primarily colocalized on the genome with the active promoter mark H3K4me3 (**Fig. 3**). The three "false" identifications Ctr9, Ctcf and Cbx7 were all borderline cases where the genome-wide correlation with H3K4me3 and H3K4me1 (which differentiates between "promoter" and "enhancer" prediction) was very similar (**Fig. 3**, right panel). Moreover, the correlation with any of the four tested histone marks was low for Ctcf and Cbx7. In short, the wrongly annotated factors either had no clear association with (active) promoters or enhancers (Ctcf and Cbx7) or had similar

Protein	H3K4me3	H3K27Ac	H3K4me1	H3K9me3	Ratio	Prediction	H3K4me3	H3K27Ac	H3K4me1	H3K9me3	Location
Taf3	**0.02**	**0.02**	0.00	0.00		Promoter	0.74	0.76	0.03	-0.53	
Kdm2a	**0.16**	0.03	0.03	0.00	0.19	Promoter	0.66	0.57	-0.02	-0.41	
Rbbp5	**0.14**	0.00	0.00	0.00	0.00	Promoter	0.66	0.56	0.03	-0.34	
Polr2a	**0.15**	0.13	0.06	0.00	0.87	Promoter	0.63	0.71	0.08	-0.52	
Wdr5	**0.60**	0.27	0.17	0.00	0.45	Promoter	0.61	0.53	0.03	-0.40	
Tcea1	0.12	**0.26**	0.05	0.00	2.17	Promoter	0.52	0.65	0.15	-0.50	
Mll2	**0.21**	0.00	0.02	0.00	0.00	Promoter	0.51	0.60	0.16	-0.48	
Taf1	**0.03**	0.00	0.00	0.00		Promoter	0.48	0.43	0.06	-0.27	Promoter
Supt5h	**0.22**	0.16	0.06	0.00	0.73	Promoter	0.46	0.56	0.13	-0.48	
Kdm1a	**0.06**	0.02	0.02	0.02		Promoter	0.46	0.73	0.26	-0.59	
Hdac1	**1.02**	0.30	0.38	0.11	0.29	Promoter	0.45	0.17	-0.14	0.00	
Hdac2	**0.40**	0.00	0.10	0.00	0.00	Promoter	0.44	0.48	0.12	-0.31	
Kdm5b	**0.25**	0.02	0.12	0.00	0.08	Promoter	0.41	0.22	-0.02	-0.22	
Rnf2	**0.35**	0.00	0.30	0.00	0.00	Promoter	0.39	0.24	0.10	-0.42	
Ctr9	**0.03**	**0.03**	0.00	0.00		Promoter	0.19	0.45	0.26	-0.48	
Cbx7	**0.10**	0.00	**0.10**	0.00	0.00	Promoter	0.09	-0.14	0.14	0.02	Enhancer
Ctcf	**0.07**	0.00	0.05	0.00		Promoter	0.08	0.17	0.15	-0.30	
Chd4	0.27	0.29	**0.44**	0.17	0.66	Enhancer	0.00	0.38	0.52	-0.36	
Esrrb	0.00	0.08	**0.20**	0.00	0.40	Enhancer	0.00	0.27	0.36	-0.31	
Oct4	0.16	0.10	**0.26**	0.05	0.38	Enhancer	0.12	0.35	0.34	-0.34	
Brd4	0.03	**0.21**	0.10	0.00	2.10	Enhancer	0.28	0.66	0.34	-0.51	Enhancer
Smc1a	0.08	0.12	**0.22**	0.06	0.55	Enhancer	0.13	0.35	0.27	-0.47	
Smarca4	0.14	**0.28**	0.22	0.03	1.27	Enhancer	0.19	0.28	0.25	-0.22	
Rad21	0.00	0.03	**0.03**	0.00		Enhancer	0.02	0.21	0.24	-0.36	
Suz12	0.31	0.12	**0.40**	0.07	0.30	Enhancer	0.18	-0.10	0.09	-0.08	
Jarid2	0.16	0.05	**0.21**	0.04	0.13	Enhancer	0.51	0.30	0.05	-0.43	Promoter
Mtf2	0.16	0.06	**0.22**	0.06	0.27	Enhancer	0.33	0.00	-0.09	-0.10	
Ezh2	0.07	0.04	**0.09**	0.00		Enhancer	0.25	-0.21	-0.16	0.03	
Atrx	0.00	0.00	0.00	**0.09**		Heterochrom.	-0.22	-0.59	-0.38	0.63	Heterochrom.

Figure 3 | Comparison of ChIP-MS with correlation of genome-wide localization of identified factors and histone modification profiles.

Average emPAI scores for identified factors in corresponding chromatin fraction are given in the left panel. The ChIP-MS based prediction of their localization to enhancer, promoter or heterochromatin is indicated in the middle column. Correlation values between genome-wide profiles of individual factors and histone modification profiles are given in the panel on the right. Factors are organized based on their predicted association to promoter, enhancer or heterochromatin by ChIP-MS and accordingly ranked on the correlation of localization on the genome and the respective histone mark. H3K27Ac/H3K4me3 ratios for promoter-associated factors and H3K27Ac/H3K4me1 ratios for enhancer-associated factors are given in the fifth column on the left.

associations with promoters and enhancers (Ctr9). We conclude that ChIP-MS predicts (predominant) association with promoters with a high accuracy (82%) with the few wrong identifications in the expected grey areas.

Of the 11 factors predicted by ChIP-MS to be predominantly associated with

enhancers, 7 (64%) were indeed primarily associated on the genome with enhancer mark H3K4me1 (**Fig. 3**). All misidentified factors (Jarid2, Mtf2, Ezh2, Suz12) are members of the Polycomb group (PcG) of repressor proteins, which showed their highest association with H3K4me3 marked areas on the genome, predicting promoter binding. PcG factors are abundant in ESCs and show broad binding around inactive promoters that have low but detectable H3K4me3 and H3K4me1 levels and high H3K27me3 levels, a histone mark outside the scope of this study [20,27,28]. Binding to such inactive promoters is hard to detect by our ChIP-MS set-up, due to the low H3K4me3 levels, and was not our aim. However, PcG factors are well characterized and would therefore be easy to recognize [27,29]. Setting aside PcG factors, our ChIP-MS experiments had a striking 100% success rate (7/7) in correctly predicting predominant enhancer binding by factors such as Oct4 (Pou5f1) and Esrrb, two key pluripotency and reprogramming transcription factors as well as Smarca4 (Brg1) and Chd4, catalytic subunits of the SWI-SNF and NuRD chromatin modifying complexes (**Fig. 3**). Atrx was correctly assigned by ChIP-MS to bind H3K9me3-containing heterochromatin (**Fig. 3**). We conclude that our ChIP-MS experiments predicted with a near 100% accuracy, binding to (active) promoters, enhancers and heterochromatin, as assigned by their associated histone mark.

The level of H3K27 acetylation present at a promoter or enhancer correlates with its activity [10,12,13]. We assessed whether the relative level of a factor in the H3K27Ac ChIP-MS sample correlates with its localization with H3K27Ac on the genome and therefore provides relevant information. From the factors with genome-wide information (**Fig. 3**), we selected the factors for which the highest emPAI value was 0.1 or higher, to be well above the detection limit of our ChIP-MS experiment. We calculated for each of these factors the ratio of the emPAI score in the H3K27Ac sample and H3K4me3 sample for predicted promoter binders and the ratio of the emPAI score of the H3K27Ac sample and H3K4me1 sample for predicted enhancer binders (**Fig. 3**). These ratios were compared for predicted promoter binders and enhancer binders to the correlation of genome-wide binding of this factor with H3K27Ac-marked regions (**Fig. 3**, right panel). Indeed, promoter predicted factors such as Tcea1, Pol II subunit Polr2a and transcription elongation factor Supt5h show high H3K27Ac/H3K4me3 ChIP-MS ratios and have high correlation with H3K27Ac on the genome (**Fig. 3**). Promoter-predicted factors such as Hdac1 and 2, H3K4 demethylase Kdm5b and PcG factors Rnf2 and Cbx7 have lower H3K27Ac/H3K4me3 ChIP-MS ratios and are indeed the factors with the lowest genome-wide associations with H3K27Ac (**Fig. 3**). Kdm2a and MLL complex subunits Rbbp5 and Mll2 have low H3K27Ac/H3K4me3 ratios but still high genome-associations with H3K27Ac. Here the H3K27Ac ChIP-MS score does not correlate well with H3K27Ac association on the genome. The enhancer-predicted factor with the highest H3K27Ac/H3K4me1 ratio, Brd4, also has the highest correlation with H3K27Ac on the genome among the enhancer binders (**Fig. 3**). Transcription factors

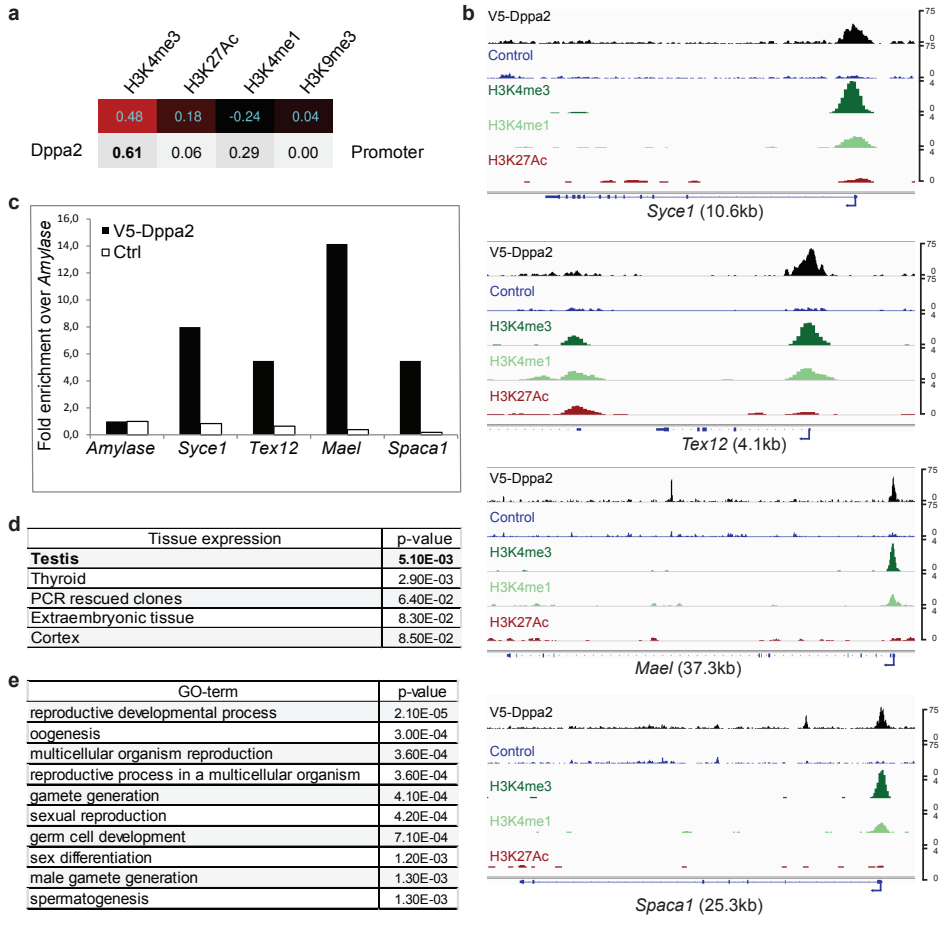
Oct4 and Esrrb, NuRD catalytic subunit Chd4 and Cohesin complex subunit Smc1a have intermediate H3K27Ac ChIP-MS ratios and intermediate genome-wide association with H3K27Ac. PcG factors Suz12, Jarid2 and Mtf2 have low H3K27Ac ChIP-MS ratios and low genome-wide association with H3K27Ac. The only predicted enhancer binder where the H3K27Ac ChIP-MS ratio provides a poor prediction is Smarca4 which has a high H3K27Ac ChIP-MS ratio and but only an intermediate genome-wide association with H3K27Ac. We also assessed the relative H3K27Ac levels of different subunits within the same biochemical complex (**Supplementary table 3**) and found highly consistent values in most cases. We conclude from our above analyses that the H3K27Ac ChIP-MS value provides a good indication of colocalization with H3K27Ac-rich regions on the genome. Accordingly, we ranked both, predicted promoter binders and enhancer binders, by their H3K27Ac ratio in Figure 2 to provide a prediction of the relative activity of the promoters or enhancers that these factors bind to.

Dppa2 localizes on testis-expressed genes in ESCs

We identified developmental pluripotency-associated 2 (Dppa2) factor to be present highest in the H3K4me3 chromatin fraction and therefore predict its association to promoters (**Fig. 4a**, **Supplementary table 1**). Dppa2 is characterized by a putative DNA binding (SAP) domain and has a close family member Dppa4, with whom it has been shown to form heterodimers [9,30]. Indeed, we also find Dppa4 highest in the H3K4me3 fraction (**Supplementary table 1**). Dppa2 is exclusively expressed in early embryonic cells of the inner cell mass, primordial germ cells and other pluripotent cells, and was shown to be implicated in maintaining the pluripotency state [30-32]. However, Dppa2-deficient ESCs can be generated albeit a reduced proliferation rate [9]. Strikingly, Dppa2 in combination with Oct4, Esrrb, Klf4 and c-Myc were sufficient to reprogram mouse embryonic fibroblasts into germline competent iPSCs [3].

As Dppa2 appears a very interesting factor and its genome-wide localization is unknown, we cloned Dppa2 and added a V5-tag to perform V5-ChIP followed by high-throughput sequencing. The genome-wide binding profile for V5-Dppa2 displayed a high correlation with that of H3K4me3 thereby confirming the prediction based on the ChIP-MS data (**Fig. 4a**). Binding of Dppa2 to the promoters of individual genes can be observed (**Fig. 4b**). Interestingly, Dppa2 appears to have a relatively wide binding pattern at these promoters, which overlaps with the localization H3K4me1 and H3K4me3. The levels of H3K27Ac are low, but can be detected at these promoters, which is also reflected by the genome-wide binding profile correlation and suggests these promoters are only weakly active.

We then examined Dppa2-associated genes with the functional annotation tool



d

Tissue expression	p-value
Testis	**5.10E-03**
Thyroid	2.90E-03
PCR rescued clones	6.40E-02
Extraembryonic tissue	8.30E-02
Cortex	8.50E-02

e

GO-term	p-value
reproductive developmental process	2.10E-05
oogenesis	3.00E-04
multicellular organism reproduction	3.60E-04
reproductive process in a multicellular organism	3.60E-04
gamete generation	4.10E-04
sexual reproduction	4.20E-04
germ cell development	7.10E-04
sex differentiation	1.20E-03
male gamete generation	1.30E-03
spermatogenesis	1.30E-03

Figure 4 | Dppa2 binds promoters of testis expressed genes.
a, ChIP-MS values vs ChIPseq values. **b**, Examples of Dppa2 binding sites on promoters; *Syce1*, *Tex12*, *Mael* and *Spaca1*. Sequence reads are plotted for V5-Dppa2 and Control (ChIP of non-V5 expressing cell line) experiments relative to chromosomal position. Size of each gene is indicated. Mouse ENCODE tracks for the indicated histone modifications are shown in log-scale. **c**, Binding of Dppa2 to the promoters of testis genes detected by ChIP. Fold enrichment over *Amylase* is shown. **d**, Genes bound by Dppa2 are significantly expressed in testis. DAVID output on tissue expression database. **e**, Genes bound by Dppa2 are significantly involved in GO-terms related to reproduction and testis development.

DAVID for expression and GO-term analysis [33,34]. Interestingly, a proportion of genes bound by Dppa2 in their promoter appeared to be expressed in testis (**Fig. 4d**). Furthermore, the top ten of GO-terms associated with Dppa2 bound genes indicates roles in reproductive developmental processes and spermatogenesis related processes (**Fig. 4e**). Since expression analysis of Dppa2-deficient ESCs is available we observed that several of Dppa2 bound genes are indeed down-regulated, suggesting Dppa2 is involved in the regulation of expression of these genes [9]. Although, a full analysis still needs to be performed to examine the full extent of genes regulated by Dppa2. It was reported however that expression of

key pluripotency genes in Dppa2-deficient ESCs was not altered. In our Dppa2 ChIP-seq data, we do not observe binding of Dppa2 to these pluripotency genes, such as Pou5f1 or Nanog (data not shown), consistent with the fact Dppa2 does not have a direct effect on their expression. Therefore, more investigation into the function of Dppa2 in regulating pluripotency is necessary.

ChIP-MS and related studies

Interest for proteomic investigation of chromatin has increased over the last few years due to the realization of the influence of epigenetic elements on transcription. Various strategies have been designed to examine protein architecture at specific chromatin regions. A number of studies analyzed the protein factors directly binding to H3K4me3- or H3K9me3-containing peptides or reconstituted nucleosomes using HeLa cell extracts [4,5]. Out of the 174 factors that we find associated with H3K4me3 marked chromatin (**Supplementary table 1**), in total 30 were also observed to be associated with H3K4me3-containing peptides in any of these studies (**Supplementary table 4**). We find H3K4me3 as the only or dominant histone mark associated with all of these overlapping factors and accordingly they were all assigned to locate at promoters (**Supplementary table 4**). Four of the 49 factors that we find binding to H3K9me3 chromatin were also observed in the above studies (**Supplementary table 5)**. Of these, for three factors (Lrdw1, Atrx and Smchd1) H3K9me3 is the dominant mark (heterochromatin). Chd4 was observed in the H3K9me3 fraction but was highest in the H3K4me1, in agreement with its ChIP-seq data (**Fig. 2**). We conclude that our data are highly consistent with other studies that use a different approach.

Soldi *et al.* used a ChIP-MS approach with a different protocol, to identify factors that bind H3K4me3-marked chromatin or H3K9me3-marked chromatin [35]. In the data, 18 factors and 10 factors were identified that overlap with factors that we find associated with H3K4me3-marked and H3K9me3-marked chromatin, respectively (**Supplementary table 4, 5**). Although, in the majority of cases, these marks are dominant for our factors, we predicted for a number of factors to be mostly associated with enhancers. For example, Chd4, Brd4, Smc1a and Smarca4 whose genome-wide binding profiles predict enhancer association confirming our prediction by ChIP-MS (**Fig. 3**). This demonstrates the additional value of the H3K4me1 chromatin fraction in our approach. Importantly, the above studies did not identify key cell identity transcription factors such as Oct4, Esrrb and Nanog, as these are not expressed in HeLa cells.

CONCLUSION

We report here a method that combines histone modification ChIP with mass spectrometry analysis (ChIP-MS) to systematically identify factors associated with enhancer, promoter or heterochromatin regions. We demonstrate the accuracy of our prediction by correlation between genome-wide histone modification profiles with the genome-wide localization of an identified factor. Precipitation of particular chromatin fractions by ChIP-MS is efficient and allows for successful identification of specific cell-identity factors. We also identified Dppa2, a factor specifically expressed in ESCs and iPSCs, and found it to be predominantly associated with promoters of testis expressed genes.

It is known that cell identity is directed by the action of specifically expressed transcription factors acting on tissue-specific enhancers and promoters. The ability to systematically identify these cell identity factors acting on regulatory regions by ChIP-MS is therefore a great tool, which can be used in other cell types in culture. For instance, it would be interesting to perform ChIP-MS for the same histone modifications in neural stem cells (NSCs) to identify novel factors that are key for NSC identity. Also our approach is applicable for other histone modifications that mark different regulatory regions. For example, H3K27me3 is a demarcation for poised enhancers and promoters in ESCs [28]. ChIP-MS could be used to identify novel factors specifically interacting with these bivalent domains in ESCs. Another example is H3K36me3 that is associated with actively transcribed regions and involved in diverse processes, such as histone turnover, repression of cryptic transcription and RNA splicing [36]. Finally, our ChIP-MS approach is applicable to any other chromatin modification (e.g. DNA methylation [37]) for which ChIP can be performed which can improve understanding of protein composition at various specific chromatin regions.

URLs

DAVID Bioinformatics resources; http://david.abcc.ncifcrf.gov, Cytoscape open source software; http://www.cytoscape.org

METHODS

Cell Lines and constructs.

Mouse embryonic stem cell line CGR8 was grown on gelatin-coated dishes without feeders as previously described [16]. In brief, cells were cultured in Glasgow minimal essential medium supplemented with leukemia inhibitory factor, 15% fetal bovine serum, 0.25% sodium bicarbonate, 1mM glutamine, 1mM sodium pyruvate, nonessential amino acids, 50 μM beta-mercaptoethanol, and penicillin-streptomycin. The coding sequence for Dppa2 was amplified from mouse ES cell cDNA. All cDNAs were cloned including a N-terminal V5-tag into a pPyCAG driven expression vector. CGR8 cells were transfected with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. Clones were selected with 1μg/ml puromycin (Sigma), and expression of stable V5-tagged proteins in selected clones was tested by Western blot analysis with anti-V5 antibody (1:2000) (Invitrogen).

Chromatin extract preparation for ChIP-MS.

For each experimental condition, $300x10^6$ cells were used. For chromatin extract preparation cells were washed on plate three times with 1x PBS and treated with 2mM Disuccinimidyl glutarate (DSG) (Thermo Scientific) in 1x PBS for 45 minutes at room temperature. Cells were washed in 1x PBS again three times and subsequently cross-linked with 1% buffered formaldehyde (50mM Hepes-KOH pH 7.5, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 11% Formaldehyde (Merck)) for 12 min at room temperature. Cells were then washed two times in cold 1x PBS and collected by centrifugation. All subsequent steps were performed on ice with pre-cooled buffers. Cells were lysed according to Boyer *et al.* [38]. In brief, cells were collected and resuspended in LB1 (50mM Hepes-KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) and sonicated on a Soniprep 150 (MSE). After 10 minutes incubation, cells were pelleted by centrifugation and resuspended in LB2 (10mM Tris-HCl, pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA). After 10 minutes incubation, cells were pelleted and resuspended in freshly prepared LB3 (10mM Tris-HCl, pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine) and sonicated appropriately. For quality control after sonication it was verified whether the DNA fragment size lay between 200 and 1000 bp, as expected.

ChIP-MS procedure.

Antibodies used for immunoprecipitation were anti-H3K4me1 (ab8895, Abcam), anti-H3K4me3 (ab8580, Abcam), anti-H3K27Ac (ab4729, Abcam), anti-H3K9me3 (ab8898,

Abcam) and anti-GFP (sc8334, Santa Cruz Biotechnology). To prevent immunoglobin contamination during the mass spectrometry analysis we cross-linked 50µg of antibodies to 500µl Protein G magnetic beads (Invitrogen) with Dimethyl Pimelimidate (DMP)(Sigma). In this procedure, all steps were performed at room temperature. Antibodies and beads were incubated together for 1 hour and washed twice in 0.2M Na-borate pH 9.0. The antibody-bead complexes were treated for 30 minutes with 20mM DMP in 0.2M Na-borate buffer and afterwards washed and incubated for two hours with 0.2M ethanolamine pH 8.0. Cross-linked antibody-bead complexes were equilibrated in LB3 buffer and subsequently blocked with 0.5mg/ml BSA (New England Biolabs) and 0.2mg/ml sonicated salmon sperm DNA (Stratagene) for one hour. Antibody- bead mixture was incubated with approximately 10mg chromatin extract overnight rotation at 4°C. The next day beads were transferred to no stick tubes (Alpha laboratories) and washed five times 5 minutes in RIPA buffer (50mM Hepes-KOH, pH 7.6, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-deoxycholate). After washing, the beads were boiled for 35 minutes at 95°C in 2x Laemmli buffer and supernatant was transferred to a fresh tube for storage at -20°C. For protein identification, ChIP-MS samples were run on 10% precast SDS-PAGE gels (NuPage Invitrogen) and stained by sensitive Colloidal Coomassie stain (Invitrogen). Gel lanes were sliced and analysis on LQT orbitrap was performed as described in van den Berg *et al.* [39]

Prediction of localization on the genome for the identified factors.

Factors were selected based on 3-fold enrichment of emPAI score over any other chromatin fraction and none to low detection in the GFP control sample. Subsequently, average emPAI scores were calculated from two individual ChIP-MS experiments. Localization of a factor was predicted according to the following rules; enhancer association when the average emPAI score of H3K4me1 was larger than H3K4me3 or H3K9me3, promoter association when the average emPAI score of H3K4me3 equaled or was larger than H3K4me1 or H3K9me3 and heterochromatin when the average emPAI score of H3K9me3 was larger than H3K4me1 or H3K4me3.

Western blotting.

Chromatin fraction samples were separated on a 15% SDS-PAGE gel and transferred to nitrocellulose membrane by semi-dry blotting for 30 minutes at 15V/270mA (Biorad). Membranes were blocked in 5% skim milk in TBS-T 0.05%. Membranes were probed over-night with primary antibodies against H3K4me1 (ab8895), H3K4me3 (ab8580), H3K27Ac (ab4729), H3K9me3 (ab8898)(Abcam) and Nanog (REC-RCAB0002PF, Cosmo Bio Ltd.). Secondary probing with anti-IgG HRP labelled antibodies was performed for 1 hour at RT. Detection was done by ECL incubation (Amersham).

Chromatin immunoprecipitation and sequencing.

For each ChIP, $100x10^6$ cells were used. As a control, cells expressing no V5-tag protein were used. Cells were double cross-linked and lysed as described above. Chromatin fragments were checked to ascertain that they were between 200-1000 bp on an agarose gel. To avoid unspecific interactions, the chromatin was pre-cleared with Protein A agarose (Millipore) for 30 minutes rotation at 4°C. For immunoprecipitation, V5 agarose beads (Sigma) were blocked with 0.5mg/ml BSA for one hour at room temperature. Chromatin was incubated with V5 beads overnight. The following day, the beads were washed 5 times with RIPA buffer as described for CHIP-MS. Precipitated protein-DNA complexes were eluted from beads with elution buffer 50mM Tris-HCl, pH 8.0, 10mM EDTA and 1% SDS for 30 minutes at 65°C. Samples were de-cross-linked overnight at 65°C and treated the following day with proteinase K for one hour at 45°C. DNA was extracted by phenol/chloroform/isoamyl extraction and precipitated with 100% ethanol. DNA samples were analysed by quantitative PCR or used for library generation followed by next-generation sequencing as described in [17].

Bioinformatical analysis.

For calculating the correlation coefficient of our own and publicly available ChIP-seq datasets (**Supplementary table 6**), sequences with low complexity that are unlikely to map uniquely to the genome were removed. This was done using prinseq-lite with options -lc_method dust and -lc_threshold 7 for each dataset [40]. The remaining sequences were mapped using Bowtie 0.12.7 with following options "-n 2 -l 36 -p 4 -e 70 -k 1 --best". Duplicated sequences were removed. Bedtools-multiBamCov was used to count the number of sequences present in bins of 1000 bp covering the entire genome for each dataset [41]. Subsequently the Reads Per Million (RPM) were calculated for each bin and input was subtracted. The Spearman correlation coefficient between transcription factor and histone modification ChIPs was calculated from an unified list, containing the top 4000 bins with the highest RPM of each histone modification ChIP and the transcription factor investigated. Each bin was only allowed to be in the list once. Correlation table was generated using the heatmap.2 function from the R gplot package (Gregory R. Warnes, gplots: Various R programming tools for plotting data, 2012).
For analysis of the Dppa2 ChIP-seq data, MACS 1.4.2 was used for peak calling using default settings. For Gene Ontology analysis we used DAVID 6.7 with the highest significant binding sites covering the TSS as input [31,32].

REFERENCES

1. Mouse, E.C. et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* **13**, 418 (2012).
2. Consortium, E.P. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
3. Buganim, Y. et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209-22 (2012).
4. Vermeulen, M. et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-80 (2010).
5. Bartke, T. et al. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-84 (2010).
6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-76 (2006).
7. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101-6 (2008).
8. Feng, B. et al. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* **11**, 197-203 (2009).
9. Nakamura, T., Nakagawa, M., Ichisaka, T., Shiota, A. & Yamanaka, S. Essential roles of ECAT15-2/Dppa2 in functional lung development. *Mol Cell Biol* **31**, 4366-78 (2011).
10. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).
11. Heintzman, N.D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).
12. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-6 (2010).
13. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83 (2011).
14. Peters, A.H. et al. Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet* **30**, 77-80 (2002).
15. Nowak, D.E., Tian, B. & Brasier, A.R. Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques* **39**, 715-25 (2005).
16. van den Berg, D.L. et al. Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol Cell Biol* **28**, 5986-95 (2008).
17. Engelen, E. et al. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet* **43**, 607-11 (2011).
18. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5 (1970).
19. Ishihama, Y. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265-72 (2005).
20. Ram, O. et al. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* **147**, 1628-39 (2011).
21. Peters, A.H. et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-37 (2001).
22. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-8 (2006).
23. Hu, G. et al. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* **23**, 837-48 (2009).
24. Chia, N.Y. et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316-20 (2010).

Supplementary information

References for pluripotency phenotype in ESCs of factors identified by ChIP-MS

Supplementary tables

Supplementary table 1: ChIP-MS identified factors with averaged emPAI scores

Supplementary table 2: ChIP-MS identified factors with individual experimental mascot and emPAI scores

Supplementary table 3: Complexes identified by ChIP-MS

Supplementary table 4: Overlap identifications for H3K4me3 with Vermeulen *et al.*[1], Bartke *et al.*[2], Nikolov *et al.*[3] and Soldi *et al.*[4]

Supplementary table 5: Overlap identifications for H3K9me3 with Vermeulen *et al.*[1], Bartke *et al.*[2], Nikolov *et al.*[3] and Soldi *et al.*[4]

Supplementary table 6: Accession numbers of studies used for genome-wide correlation

References for pluripotency phenotype in ESCs of factors identified by ChIP-MS

Ahctf1 (Chia *et al.* Nature 2010)[5], Arid1a (Gao *et al.* PNAS 2008)[6], Ash2l (Wan *et al.* J Biol Chem 2012)[7], Aurkb (Lee *et al.* Cell Stem Cell 2012)[8], Banf1 (Cox *et al.* J Cell Sci 2011)[9], Bptf (Landry *et al.* PLoS Genet 2008)[10], Cbx7 (O'Loghlen *et al.* Cell Stem Cell 2012)[11], Cdc73 (Wang *et al.* Mol Cell Biol 2008)[12], Cdk9 (Hu *et al.* Genes Dev 2009)[13], Chaf1b (Bilodeau *et al.* Genes Dev 2009)[14], Chd1 (Gaspar-Maia *et al.* Nature 2009)[15], Ctbp2 (Tarleton *et al.* Mech Dev 2010)[16], Ctr9 (Hu *et al.* Genes Dev 2009)[13](Ding *et al.* Cell Stem Cell 2009)[17], Dmap1 (Fazzio *et al.* Cell 2008)[18], Dppa2, Dppa4 (Maldonado-Saldivia *et al.* Stem Cells 2007)[19], Dpy30 (Jiang *et al.* Cell 2011)[20], Ep300 (Zhong *et al.* J Biol Chem 2009)[21], Ep400 (Fazzio *et al.* Cell 2008)[18], Esrrb (Ivanova *et al.* Nature 2006)[22], Ezh2 (Shen *et al.* Mol Cell 2008)[23], Hcfc1 (Chia *et al.* Nature 2010)[5], Hdac1 (Dovey *et al.* Proc Natl Acad Sci 2010)[24], Ing5 (Hu *et al.* Genes Dev 2009)[13], Ino80 (Chia *et al.* Nature 2010)[5], Jarid2 (Shen *et al.* Cell 2009)[25], Kdm1a (Whyte *et al.* Nature 2012)[26], Kdm2b (He *et al.* Nat Cell Biol 2013)[27], Kdm4c (Loh *et al.* Genes Dev 2007)[28], Kdm5b (Xie *et al.* EMBO J 2011)[29], Klf5 (Ema *et al.* Cell Stem Cell 2008)[30], L3mbtl2 (Qin *et al.* Cell Stem Cell 2012)[31], Max (Hishida *et al.* Cell Stem Cell 2011)[32], Mtf2 (Walker *et al.* Cell Stem Cell 2010)[33], Mycn (Chappell *et al.* Genes Dev 2013)[34], Myst2, Myst3 (Kagey *et al.* Nature 2010)[35], Nacc1 (Wang *et al.* Nature 2006)[36], Nfrkb (Chia *et al.* Nature 2010)[5], Ogt (Shafi *et al.* Proc Natl Acad Sci 2000)[37], Phf20 (Zhao *et al.* Cell 2013)[38], Phf23 (Ding *et al.* Cell Stem Cell 2009)[17], Pin1 (Nishi *et al.* J Biol Chem 2011)[39], Pou5f1 (Niwa *et al.* Nature Genetics 2000)[40], Rad21 (Nitzsche *et al.* PLoS One 2011)[41], Rbbp5 (Jiang *et al.* Cell 2011)[20], Rif1 (Loh *et al.* Nature Genetics 2006)[42], Rnf2 (van der Stoop *et al.* PLoS One 2008)[41] (Hu *et al.* Genes Dev 2009)[13] (Fazzio *et al.* Cell 2008)[18], Ruvbl1, Ruvbl2 (Fazzio *et al.* Cell 2008)[18], Sin3a (McDonel *et al.* Dev Biol 2012)[43], Smarca4 (Kidder *et al.* Stem Cells 2009)[44], Smarcb1 (You *et al.* PLoS Genet 2013)[45], Smarcc1 (Fazzio *et al.* Cell 2008)[18], Smc1a (Hu *et al.* Genes Dev 2009)[13] (Fazzio *et al.* Cell 2008)[18], Smc6 (Fazzio *et al.* Cell 2008) [18](Ding *et al.* Cell Stem Cell 2009)[17], Supt4h2 (Fazzio *et al.* Cell 2008)[18], Suv39h2 (Chia *et al.* Nature 2010)[5], Suz12 (Pasini *et al.* Mol Cell Biol 2007)[46], Taf1, Taf2, Taf4a, Taf5, Taf6 (Pijnappel *et al.* Nature 2013)[47], Taf3 (Liu *et al.* Cell 2011)[48] (Pijnappel *et al.* Nature 2013)[47] , Taf7 (Chia *et al.* Nature 2010)[5], Tpr (Chia *et al.* Nature 2010)[5] (Fazzio *et al.* Cell 2008)[18], Trrap (Fazzio *et al.* Cell 2008)[18], Wdr5 (Ang *et al.* Cell 2011)[49], Zfp281 (Wang *et al.* Nature 2006)[36] (Fidalgo *et al.* Stem Cells 2011)[50], Zscan10 (Yu *et al.* J Biol Chem 2009)[51].

1. Vermeulen, M. et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967-80 (2010).
2. Bartke, T. et al. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-84 (2010).
3. Nikolov, M. et al. Chromatin affinity purification and quantitative mass spectrometry defining the in-

teractome of histone modification patterns. *Mol Cell Proteomics* **10**, M110 005371 (2011).
4. Soldi, M. & Bonaldi, T. The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. *Mol Cell Proteomics* **12**, 764-80 (2013).
5. Chia, N.Y. et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316-20 (2010).
6. Gao, X. et al. ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* **105**, 6656-61 (2008).
7. Wan, M. et al. The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *J Biol Chem* **288**, 5039-48 (2013).
8. Lee, D.F. et al. Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. *Cell Stem Cell* **11**, 179-94 (2012).
9. Cox, J.L. et al. Banf1 is required to maintain the self-renewal of both mouse and human embryonic stem cells. *J Cell Sci* **124**, 2654-65 (2011).
10. Landry, J. et al. Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genet* **4**, e1000241 (2008).
11. O'Loghlen, A. et al. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33-46 (2012).
12. Wang, P. et al. Parafibromin, a component of the human PAF complex, regulates growth factors and is required for embryonic development and survival in adult mice. *Mol Cell Biol* **28**, 2930-40 (2008).
13. Hu, G. et al. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* **23**, 837-48 (2009).
14. Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B. & Young, R.A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev* **23**, 2484-9 (2009).
15. Gaspar-Maia, A. et al. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460**, 863-8 (2009).
16. Tarleton, H.P. & Lemischka, I.R. Delayed differentiation in embryonic stem cells and mesodermal progenitors in the absence of CtBP2. *Mech Dev* **127**, 107-19 (2010).
17. Ding, L. et al. A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4**, 403-15 (2009).
18. 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**, 162-74 (2008).
19. Maldonado-Saldivia, J. et al. Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells* **25**, 19-28 (2007).
20. Jiang, H. et al. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* **144**, 513-25 (2011).
21. Zhong, X. & Jin, Y. Critical roles of coactivator p300 in mouse embryonic stem cell differentiation and Nanog expression. *J Biol Chem* **284**, 9168-75 (2009).
22. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-8 (2006).
23. Shen, X. et al. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491-502 (2008).
24. Dovey, O.M., Foster, C.T. & Cowley, S.M. Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A* **107**, 8242-7 (2010).
25. Shen, X. et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**, 1303-14 (2009).
26. Whyte, W.A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221-5 (2012).
27. He, J. et al. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat Cell Biol* **15**, 373-84 (2013).
28. Loh, Y.H., Zhang, W., Chen, X., George, J. & Ng, H.H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases

regulate self-renewal in embryonic stem cells. *Genes Dev* **21**, 2545-57 (2007).
29. Xie, L. et al. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J* **30**, 1473-84 (2011).
30. Ema, M. et al. Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. *Cell Stem Cell* **3**, 555-67 (2008).
31. Qin, J. et al. The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell Stem Cell* **11**, 319-32 (2012).
32. Hishida, T. et al. Indefinite self-renewal of ESCs through Myc/Max transcriptional complex-independent mechanisms. *Cell Stem Cell* **9**, 37-49 (2011).
33. Walker, E. et al. Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **6**, 153-66 (2010).
34. Chappell, J., Sun, Y., Singh, A. & Dalton, S. MYC/MAX control ERK signaling and pluripotency by regulation of dual-specificity phosphatases 2 and 7. *Genes Dev* **27**, 725-33 (2013).
35. Kagey, M.H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-5 (2010).
36. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-8 (2006).
37. Shafi, R. et al. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci U S A* **97**, 5735-9 (2000).
38. Zhao, W. et al. Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination. *Cell* **152**, 1037-50 (2013).
39. Nishi, M. et al. A distinct role for Pin1 in the induction and maintenance of pluripotency. *J Biol Chem* **286**, 11593-603 (2011).
40. Niwa, H., Miyazaki, J. & Smith, A.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-6 (2000).
41. Nitzsche, A. et al. RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS One* **6**, e19470 (2011).
42. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).
43. McDonel, P., Demmers, J., Tan, D.W., Watt, F. & Hendrich, B.D. Sin3a is essential for the genome integrity and viability of pluripotent cells. *Dev Biol* **363**, 62-73 (2012).
44. Kidder, B.L., Palmer, S. & Knott, J.G. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* **27**, 317-28 (2009).
45. You, J.S. et al. SNF5 is an essential executor of epigenetic regulation during differentiation. *PLoS Genet* **9**, e1003459 (2013).
46. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* **27**, 3769-79 (2007).
47. Pijnappel, W.W. et al. A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495**, 516-9 (2013).
48. Liu, Z., Scannell, D.R., Eisen, M.B. & Tjian, R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* **146**, 720-31 (2011).
49. Ang, Y.S. et al. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183-97 (2011).
50. Fidalgo, M. et al. Zfp281 functions as a transcriptional repressor for pluripotency of mouse embryonic stem cells. *Stem Cells* **29**, 1705-16 (2011).
51. Yu, H.B., Kunarso, G., Hong, F.H. & Stanton, L.W. Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *J Biol Chem* **284**, 31327-35 (2009).

Supplementary table 1

ChIP-MS identified factors with averaged emPAI scores

Protein	Complex[a]	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[b]	ESC pheno-type[c]
Actl6a	BAF+INO80+Trrap/Ep400	0,29	0,16	0,12	0,08		Promoter	-
Actr3	Arp2/3	0,04					Promoter	-
Actr5	INO80		0,03				Unclear	-
Ankhd1				0,10			Enhancer	-
Anln					0,03		Heterochromatin	-
Anp32a	SET	0,17		0,18			Enhancer	-
Arid1a	BAF		0,07	0,06			Enhancer	yes
Arid4a	Sin3	0,12					Promoter	-
Ash2l	MLL	0,25					Promoter	yes
Atad2			0,03	0,04	0,04		Enhancer	-
Atrx					0,09		Heterochromatin	-
Aurkb	Aurora K	0,22	0,10	0,10	0,65		Heterochromatin	yes
Banf1		0,66	0,47	0,63			Promoter	yes
Bap18	MLL	0,63					Promoter	-
Baz1a	ACF			0,01			Enhancer	-
Baz2a	NoRC	0,06	0,22	0,20	0,07		Enhancer	-
Bend3		0,24		0,04			Promoter	-
Bptf	NURF	0,08	0,02	0,02			Promoter	yes
Brd1	MOZ/MORF	0,03	0,05	0,10			Enhancer	-
Brd2		0,37	0,75	0,18			Promoter	-
Brd3		0,12	0,32				Promoter	-
Brd4		0,03	0,21	0,10			Enhancer	-
Brms1	Sin3	0,14					Promoter	-
Brms1l	Sin3	0,47					Promoter	-
Brpf1	MOZ/MORF	0,16					Promoter	-
Brpf3	MOZ/MORF	0,03					Promoter	-
Cad				0,04			Enhancer	-
Cald1		0,06					Promoter	-
Cbx7	PRC1	0,10		0,10			Promoter	yes
Cdc5l	Prp19			0,04			Enhancer	-
Cdc73	PAF1	0,10	0,10				Promoter	yes
Cdca7			0,09	0,09			Enhancer	-
Cdca8	Aurora K	0,28	0,13	0,24	1,17		Heterochromatin	-
Cdk7	TFIIH	0,05					Promoter	-
Cdk9	P-TEFb		0,09				Unclear	yes
Cep95			0,02	0,04			Enhancer	-
Chaf1b	CAF1		0,23	0,52			Enhancer	yes
Chd1		0,26	0,24	0,14	0,02		Promoter	yes
Chd2		0,05		0,03			Promoter	-
Chd4	NuRD	0,27	0,29	0,44	0,17	0,02	Enhancer	-
Chd8		0,15					Promoter	-

Protein	Complex[a]	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[b]	ESC pheno-type[c]
Ckb		0,14					Promoter	-
Coil					0,03		Heterochromatin	-
Cpsf6	CFIm	0,07	0,07				Promoter	-
Creb1		0,11		0,05			Promoter	-
Crem		0,14		0,05			Promoter	-
Cse1l		0,04					Promoter	-
Csnk2a1		0,89	0,27	0,21			Promoter	-
Csnk2b		0,82	0,26	0,08			Promoter	-
Csrp2bp	ATAC	0,04					Promoter	-
Ctbp2		0,46	0,30	0,43	0,04		Promoter	yes
Ctcf		0,07		0,05			Promoter	-
Ctr9	PAF1	0,03	0,03				Promoter	yes
Cul4b	DDB1/Cul4 ubiquitin ligase	0,17	0,18	0,18			Enhancer	-
Ddb1	DDB1/Cul4 ubiquitin ligase	0,19	0,24	0,29			Enhancer	-
Ddx3x		0,36	0,68				Promoter	-
Ddx47					0,15		Heterochromatin	-
Ddx54				0,04			Enhancer	-
Dido1		0,02					Promoter	-
Dis3			0,05	0,05			Enhancer	-
Dmap1	Trrap/Ep400	0,07		0,04			Promoter	yes
Dnajc9			0,14	0,14			Enhancer	-
Dpf2	BAF		0,22	0,12			Enhancer	-
Dppa2		0,61	0,06	0,29			Promoter	yes
Dppa4		0,58		0,39			Promoter	yes
Dpy30	MLL	0,18					Promoter	yes
Dut		0,41	0,18	0,18			Promoter	-
Emsy		0,02					Promoter	-
Ep300			0,05				Unclear	yes
Ep400	Trrap/Ep400		0,02				Unclear	yes
Erh		0,37					Promoter	-
Esrrb			0,08	0,20			Enhancer	yes
Ezh2	PRC2	0,07	0,04	0,09			Enhancer	yes
Fam60a	Sin3	0,24					Promoter	-
Fbxl19	SCF ubiquitin ligase	0,05					Promoter	-
Fkbp4				0,07			Enhancer	-
Gatad2a	NuRD	0,16	0,21	0,27	0,11		Enhancer	-
Gltscr2				0,07			Enhancer	-
Glyr1		0,19	0,24	0,26	0,07		Enhancer	-
Gm53		0,19					Promoter	-
Gtf2e1	TFIIE			0,08			Enhancer	-
Gtf2i			0,04	0,07			Enhancer	-
Gtf3c1	TFIIIC	0,01	0,01	0,02			Enhancer	-

Protein	Complex[a]	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[b]	ESC pheno-type[c]
Hat1		0,11	0,04	0,04			Promoter	-
Hcfc1	MLL	0,05					Promoter	yes
Hdac1	NuRD + Sin3 + REST	1,02	0,30	0,38	0,11		Promoter	yes
Hdac2	NuRD + Sin3 + REST	0,40		0,10			Promoter	-
Hdgfrp2			0,05				Unclear	-
Hmga2		0,14	0,14		0,28		Heterochromatin	-
Hmgb3		0,11		0,1			Promoter	-
Hmgxb4		0,20					Promoter	-
Hp1bp3		0,26		0,38	0,23		Enhancer	-
Incenp	Aurora K	0,06			0,26		Heterochromatin	-
Ing2	Sin3	0,11					Promoter	-
Ing3	Trrap/Ep400	0,04	0,04				Promoter	-
Ing4	Trrap/Ep400	0,14		0,07			Promoter	-
Ing5	MOZ/MORF	0,58	0,13	0,20			Promoter	yes
Ino80	INO80	0,05	0,05				Promoter	yes
Ints1	Integrator			0,02			Enhancer	-
Irgc				0,04			Enhancer	-
Jarid2	PRC2	0,16	0,05	0,21	0,04		Enhancer	yes
Kdm1a	BHC	0,06	0,02	0,02	0,02		Promoter	yes
Kdm2a	SCF ubiquitin ligase	0,16	0,03	0,03			Promoter	-
Kdm2b	PRC1	0,09		0,01			Promoter	yes
Kdm4a		0,03					Promoter	-
Kdm4c		0,65	0,03	0,13			Promoter	yes
Kdm5a		0,02					Promoter	-
Kdm5b		0,25	0,02	0,12			Promoter	yes
Kiaa1731	Centrosome				0,01		Heterochromatin	-
Kif2c				0,05			Enhancer	-
Kif4				0,03			Enhancer	-
Kifc5b			0,07	0,15			Enhancer	-
Klf16		0,16					Promoter	-
Klf5			0,07	0,04			Enhancer	yes
L3mbtl2		0,08		0,07			Promoter	yes
Las1l	5FMC			0,06			Enhancer	-
Lasp1		0,29		0,13			Promoter	-
Lig1		0,08		0,02			Promoter	-
Lig3			0,02	0,05			Enhancer	-
Llph				0,13			Enhancer	-
Lmna		0,13	0,08	0,08	0,48		Heterochromatin	-
Lmnb2		0,12	0,24	0,30	0,98		Heterochromatin	-
Lrwd1	ORC	0,14		0,11	0,35		Heterochromatin	-
Lsm2	U6 SnRNP		0,18				Unclear	-
Luc7l3				0,04			Enhancer	-
Max		0,11	0,11	0,10			Promoter	yes

Protein	Complex[a]	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[b]	ESC pheno-type[c]
Mbtd1		0,06					Promoter	-
Mdc1			0,03	0,05			Enhancer	-
Meaf6	Trrap/Ep400	0,51	0,19	0,19			Promoter	-
Men1	MLL	0,32	0,03	0,15			Promoter	-
Mll2	MLL	0,21		0,02		0,01	Promoter	-
Morc2a				0,03			Enhancer	-
Morc3		0,70	0,24	0,24	0,02		Promoter	-
Mta2	NuRD	0,19	0,21	0,26		0,07	Enhancer	-
Mta3	NuRD	0,24	0,14	0,27			Enhancer	-
Mtf2	PRC2	0,16	0,06	0,22	0,06		Enhancer	yes
Mycn		0,08					Promoter	yes
Myst2	HBO1	0,56	0,36	0,39	0,06		Promoter	yes
Myst3	MOZ	0,08					Promoter	yes
Myst4	MORF	0,04					Promoter	-
Nacc1		0,04	0,07	0,13			Enhancer	yes
Nasp			0,02	0,02			Enhancer	-
Nfrkb	INO80		0,03				Unclear	yes
Nolc1				0,05			Enhancer	-
Nrf1		0,08					Promoter	-
Nsd1		0,01	0,04	0,05			Enhancer	-
Nudt13				0,10			Enhancer	-
Nudt21	CFIm	0,48					Promoter	-
Numa1		0,22	0,15	0,29	0,10		Enhancer	-
Oct4		0,16	0,10	0,26	0,05		Enhancer	yes
Ogt		0,15	0,02	0,05			Promoter	yes
Parp2				0,03			Enhancer	-
Parp9					0,04		Heterochromatin	-
Patz1		0,05		0,05			Promoter	-
Pbrm1	BAF	0,02	0,11	0,11			Enhancer	-
Pds5a	Wapl		0,03	0,05			Enhancer	-
Pds5b	Wapl	0,02	0,01	0,02			Enhancer	-
Phc1	PRC1			0,04			Enhancer	-
Phf16	HBO1	0,04					Promoter	-
Phf17	HBO1	0,31	0,26	0,16	0,12		Promoter	-
Phf20				0,02			Enhancer	yes
Phf23		0,53		0,09			Promoter	yes
Phf8		0,23	0,02				Promoter	-
Phip		0,04	0,07	0,13			Enhancer	-
Pin1			0,11	0,11			Enhancer	yes
Pnp		0,20					Promoter	-
Polr1c	Pol I	0,3					Promoter	-
Polr2a	Pol II	0,15	0,13	0,06			Promoter	-
Polr2b	Pol II	0,43	0,33	0,07			Promoter	-

Protein	Complex[a]	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[b]	ESC pheno-type[c]
Polr2c	Pol II	0,57	0,16	0,11			Promoter	-
Polr2e	Pol II	0,85	0,24	0,08			Promoter	-
Polr2g	Pol II	0,22					Promoter	-
Ppp2cb				0,11			Enhancer	-
Prdm10		0,03	0,03				Promoter	-
Prdm2				0,02			Enhancer	-
Prrc2c				0,01			Enhancer	-
Psma5	Proteasome	0,07		0,06			Promoter	-
Ptcd3				0,12			Enhancer	-
Qser1		0,03					Promoter	-
Rad21	Cohesin		0,03	0,03			Enhancer	yes
Rbbp5	MLL	0,14					Promoter	yes
Rbl1				0,03			Enhancer	-
Rbpj		0,30		0,20	0,14		Promoter	-
Rfc3		0,05	0,10				Promoter	-
Rfc4		0,20	0,15	0,19			Promoter	-
Rnf2	PRC1	0,35		0,30			Promoter	yes
Rsf1	RSF	0,05	0,11	0,12			Enhancer	-
Ruvbl1	Trrap/Ep400	1,38	1,31	0,68	0,61	0,07	Promoter	yes
Ruvbl2	Trrap/Ep400	1,36	1,91	0,41	0,16		Promoter	yes
Samd1		0,26					Promoter	-
Sap130	Sin3	0,18					Promoter	-
Sap30	Sin3	0,29					Promoter	-
Sap30l	Sin3	0,09					Promoter	-
Sbno1				0,01			Enhancer	-
Sephs1		0,14					Promoter	-
Sf1		0,14					Promoter	-
Sf3b1				0,04			Enhancer	-
Shprh		0,04	0,01	0,03			Promoter	-
Sin3a	Sin3	1,34	0,15	0,30			Promoter	yes
Sin3b	Sin3	0,09					Promoter	-
Skp1	SCF ubiquitin ligase	0,45	0,23	0,10			Promoter	-
Smarca4	BAF	0,14	0,28	0,22	0,03		Enhancer	yes
Smarcb1	BAF		0,14	0,05			Enhancer	yes
Smarcc1	BAF	0,10	0,46	0,31	0,07		Enhancer	yes
Smarcd1	BAF		0,50	0,35			Enhancer	-
Smc1a	Cohesin	0,08	0,12	0,22	0,06		Enhancer	yes
Smc6	Smc5/Smc6			0,03			Enhancer	yes
Smchd1			0,02	0,03	0,04		Heterochromatin	-
Spin1		0,72	0,13	0,13			Promoter	-
Srrt		0,04	0,04	0,08			Enhancer	-
Ssrp1	FACT	0,68	0,93	0,73	0,29	0,08	Enhancer	-
Suds3	Sin3	0,41					Promoter	-

Protein	Complex[a]	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[b]	ESC pheno-type[c]
Supt4h2	DSIF	0,34					Promoter	yes
Supt5h	DSIF	0,22	0,16	0,06			Promoter	-
Supt6h		0,07	0,09				Promoter	-
Suv39h2					0,31		Heterochromatin	yes
Suz12	PRC2	0,31	0,12	0,40	0,07		Enhancer	yes
Taf1	TFIID	0,03					Promoter	yes
Taf2	TFIID	0,06	0,02				Promoter	yes
Taf3	TFIID	0,02	0,02				Promoter	yes
Taf4a	TFIID	0,11					Promoter	yes
Taf5	TFIID	0,05					Promoter	yes
Taf6	TFIID	0,20	0,14				Promoter	yes
Taf7	TFIID	0,16	0,11				Promoter	yes
Taldo1			0,05	0,11			Enhancer	-
Tbrg4				0,09			Enhancer	-
Tcea1	TFIIS	0,12	0,26	0,05			Promoter	-
Tcerg1		0,02	0,03	0,02			Enhancer	-
Tcof1					0,02		Heterochromatin	-
Tead1			0,04	0,04			Enhancer	-
Tlk2				0,04			Enhancer	-
Trim24			0,25	0,04			Enhancer	-
Trim33			0,08				Unclear	-
Trp53			0,06				Unclear	-
Trrap	Trrap/Ep400	0,05	0,05	0,01			Promoter	yes
Uba1			0,03	0,03			Enhancer	-
Ube2h				0,24			Enhancer	-
Ubtf		0,79	0,44	0,43	0,07		Promoter	-
Usp48		0,15	0,08	0,21	0,05		Enhancer	-
Utp14b				0,07			Enhancer	-
Vmn2r100		0,02		0,06			Enhancer	-
Wdr18	5FMC	0,04		0,12	0,08		Enhancer	-
Wdr5	MLL	0,60	0,27	0,17			Promoter	yes
Wdr55		0,05					Promoter	-
Xrcc6				0,13			Enhancer	-
Zfp280c		0,05			0,22		Heterochromatin	-
Zfp281		0,04					Promoter	yes
Zic5		0,03					Promoter	-
Zmynd8	Integrator		0,05	0,04			Enhancer	-
Znf512		0,13		0,24	0,25		Heterochromatin	-
Zscan10		0,04		0,09			Enhancer	yes

[a] Subunit of indicated protein complex

[b] Prediction of genome localization based on our ChIP-MS criteria

[c] ESC pluripotency phenotype (references above)

Supplementary table 2

ChIP-MS identified factors with individual experimental mascot and emPAI scores

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Actl6a	305	0,57			122	0,16	81	0,16	191	0,24			91	0,16						
Actr3	56	0,08																		
Actr5							79	0,05												
Ankhd1									111	0,19										
Anln															63	0,06				
Anp32a	75	0,17	47	0,17							74	0,36								
Arid1a					170	0,05	235	0,08	202	0,07	241	0,05								
Arid4a	79	0,24																		
Ash2l	195	0,35	77	0,14																
Atad2					89	0,06			101	0,04	80	0,04			152	0,07				
Atrx													259	0,1	264	0,07				
Aurkb			238	0,44			69	0,2	70	0,19			276	0,72	252	0,57				
Banf1	107	0,93	59	0,39	78	0,93			103	0,87	71	0,39								
Bap18	166	0,79	119	0,47																
Baz1a									55			0,02								
Baz2a	57	0,04	239	0,07	363	0,21	579	0,23	526	0,21	468	0,19	191	0,09	110	0,04				
Bend3	196	0,26	206	0,21							77	0,08								
Bptf	422	0,11	173	0,04	102	0,03			114	0,02	115	0,02								
Brd1	70	0,06			57	0,03	88	0,06	156	0,08	142	0,11								
Brd2	471	0,51	273	0,23	527	0,57	750	0,93	191	0,17	231	0,18								
Brd3	119	0,15	81	0,09	278	0,44	174	0,2												
Brd4	72	0,05			313	0,23	323	0,18			132	0,19								
Brms1	81	0,28																		
Brms1l	179	0,6	177	0,33																
Brpf1	233	0,23	91	0,08																

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Brpf3			86	0,05																
Cad											204	0,08								
Cald1	61	0,12																		
Cbx7	50	0,19									56	0,19								
Cdc5l											58	0,08								
Cdc73	82	0,2			84	0,13	62	0,06												
Cdca7							64	0,18			51	0,18								
Cdca8			191	0,56			62	0,25	97	0,23	56	0,25	344	1,17	392	1,17				
Cdk7			51	0,1																
Cdk9					61	0,18														
Cep95							45	0,04	68	0,07										
Chaf1b					45	0,45			109	1,04										
Chd1	672	0,34	382	0,18	239	0,16	641	0,32	212	0,11	323	0,16			66	0,04				
Chd2	177	0,09							96	0,05										
Chd4	840	0,42	445	0,12	685	0,35	600	0,22	1124	0,5	899	0,37	568	0,24	254	0,1			101	0,03
Chd8	516	0,23	199	0,07																
Ckb	66	0,18	60	0,09																
Coil															62	0,06				
Cpsf6			88	0,13			131	0,13												
Creb1	60	0,22									54	0,1								
Crem	51	0,28									54	0,1								
Cse1l			72	0,07																
Csnk2a1	383	1,4	115	0,37	106	0,27	106	0,27	117	0,25	58	0,17								
Csnk2b	238	1,32	80	0,32			135	0,52			51	0,15								
Csrp2bp	66	0,08																		
Ctbp2	452	0,67	154	0,24	229	0,44	72	0,16	309	0,51	188	0,34			60	0,08				
Ctcf	87	0,14									89	0,09								
Ctr9	108	0,06			91	0,06														

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Cul4b	262	0,26	79	0,07	239	0,22	195	0,14	229	0,13	275	0,22								
Ddb1	349	0,22	198	0,15	339	0,29	317	0,19	414	0,31	339	0,26								
Ddx3x			462	0,72	369	0,64	527	0,72												
Ddx47													81	0,15	103	0,15				
Ddx54									68	0,07										
Dido1	120	0,04																		
Dis3							92	0,1			84	0,1								
Dmap1	100	0,14							69	0,07										
Dnajc9							89	0,27			137	0,27								
Dpf2					189	0,36	57	0,08	127	0,15	66	0,08								
Dppa2	255	0,86	157	0,36			69	0,11	85	0,21	152	0,36								
Dppa4	231	0,92	75	0,24					59	0,23	134	0,54								
Dpy30	50	0,36																		
Dut	117	0,64	58	0,18	49	0,18	48	0,18	53	0,17	69	0,18								
Emsy	56	0,03																		
Ep300					147	0,06	135	0,04												
Ep400							101	0,03												
Erh	71	0,74																		
Esrrb					45	0,08	48	0,08	139	0,23	112	0,16								
Ezh2	99	0,09	73	0,04	47	0,04	98	0,04	79	0,04	141	0,14								
Fam60a	75	0,48																		
Fbxl19			75	0,1																
Fkbp4									62	0,14										
Gatad2a	184	0,31			131	0,24	87	0,17	138	0,22	202	0,31	72	0,11	71	0,11				
Gltscr2									153	0,13										
Glyr1	136	0,19	147	0,19	181	0,34	111	0,13	292	0,32	172	0,19	97	0,13						
Gm53			59	0,37																
Gtf2e1									76	0,15										

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Gtf2i					70	0,07			68	0,06	102	0,07								
Gtf3c1	55	0,02					53	0,02			97	0,03								
Hat1	88	0,15	45	0,07			62	0,07			63	0,07								
Hcfc1	137	0,09																		
Hdac1	453	1,65	239	0,38	118	0,3	170	0,3	176	0,28	244	0,48	87	0,14	60	0,07				
Hdac2	222	0,79							114	0,2			87		60					
Hdgfrp2							67	0,1												
Hmga2			82	0,28			56	0,28					52	0,28	87	0,28				
Hmgb3	52	0,21							63	0,2										
Hmgxb4	60	0,39																		
Hp1bp3	45	0,12	104	0,39					142	0,51	116	0,24	132	0,19	172	0,27				
Incenp			152	0,11									270	0,28	317	0,24				
Ing2	57	0,11	57	0,11																
Ing3			70	0,08			74	0,08												
Ing4	60	0,28									48	0,13								
Ing5	168	0,87	71	0,29	46	0,13	63	0,13	72	0,27	73	0,13								
Ino80	147	0,06	79	0,04	140	0,09														
Ints1									69	0,03										
Irgc									68	0,08										
Jarid2	283	0,17	309	0,14	102	0,05	128	0,05	281	0,19	424	0,23	70	0,03	104	0,05				
Kdm1a	94	0,08	91	0,04			57	0,04			86	0,04			53	0,04				
Kdm2a	435	0,24	158	0,08			66	0,06	51	0,03	62	0,03								
Kdm2b	267	0,16	71	0,02							64	0,02								
Kdm4a	67	0,06																		
Kdm4c	861	0,87	647	0,43			119	0,06	73	0,06	270	0,2								
Kdm5a	61	0,04																		
Kdm5b	519	0,3	424	0,2			107	0,04	68	0,05	347	0,18								
													53	0,01						

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Kif2c									70	0,09										
Kif4									103	0,05										
Kifc5b							71	0,14	126	0,16	108	0,14								
Klf16	100	0,31																		
Klf5					53	0,07	63	0,07			72	0,07								
L3mbtl2	94	0,15							111	0,14										
Las1l									106	0,12										
Lasp1			85	0,58							57	0,26								
Lig1	147	0,15							55	0,03										
Lig3							67	0,03			115	0,1								
Llph									53	0,25										
Lmna	96	0,1	155	0,16	130	0,16			155	0,15			451	0,55	290	0,41				
Lmnb2			173	0,24	170	0,24	154	0,24	297	0,42	173	0,17	701	1,24	435	0,71				
Lrwd1			220	0,28					59	0,05	93	0,16	163	0,28	224	0,42				
Lsm2							63	0,36												
Luc7l3									55	0,07										
Max	56	0,21					45	0,21	72	0,2										
Mbtd1	108	0,11																		
Mdc1							143	0,06	100	0,06	96	0,04								
Meaf6	130	0,63	117	0,38	60	0,38					104	0,38								
Men1	333	0,45	106	0,18			50	0,06	55	0,05	232	0,25								
Mll2	575	0,23	518	0,18							146	0,04					61	0,01		
Morc2a					107	0,1			51	0,06										
Morc3	865	0,91	626	0,48	225	0,18	437	0,3	141	0,1	501	0,38			59	0,03				
Mta2	226	0,27	97	0,1	188	0,27	137	0,15	345	0,37	139	0,15							95	0,14
Mta3	141	0,28	173	0,2	171	0,28			189	0,26	164	0,28								
Mtf2	74	0,12	111	0,19	52	0,06	45	0,06	123	0,18	161	0,26			62	0,12				
Mycn	66	0,16			66	0,16														

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Myst2	327	0,75	268	0,36	201	0,36	296	0,36	170	0,27	316	0,5			89	0,11				
Myst3	293	0,1	167	0,05																
Myst4	70	0,04	81	0,04																
Nacc1			47	0,07			111	0,13	130	0,19	52	0,07								
Nasp							58	0,04			46	0,04								
Nfrkb					58	0,05														
Nolc1									57	0,1										
Nrf1	66	0,15																		
Nsd1	62	0,02			138	0,05	96	0,02	190	0,05	163	0,04								
Nudt13									62	0,19										
Nudt21	123	0,96																		
Numa1	402	0,13	339	0,3	431	0,18	457	0,11	685	0,19	666	0,39	284	0,1	105	0,09				
Oct4	94	0,32			66	0,2			170	0,42	51	0,1	51	0,1						
Ogt	208	0,2	183	0,1			46	0,03			178	0,1								
Parp2									86	0,05										
Parp9															58	0,08				
Patz1	107	0,1							74	0,09										
Pbrm1	66	0,04			220	0,11	174	0,11	146	0,08	218	0,14								
Pds5a							99	0,05	162	0,09										
Pds5b			78	0,04			56	0,02			83	0,04								
Phc1											80	0,07								
Phf16			69	0,08																
Phf17	354	0,41	281	0,21	230	0,21	380	0,3	156	0,15	183	0,16	126	0,12	137	0,12				
Phf20											60	0,03								
Phf23	148	0,4	196	0,66							56	0,18								
Phf8	345	0,33	144	0,13	51	0,04														
Phip	152	0,05	63	0,02	134	0,07	187	0,07	248	0,1	295	0,15								
Pin1							57	0,21			48	0,21								

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Pnp	129	0,39																		
Polr1c	174	0,6																		
Polr2a	460	0,18	321	0,11	333	0,12	498	0,14	129	0,05	159	0,07								
Polr2b	709	0,67	258	0,18	407	0,34	421	0,31	63	0,05	148	0,08								
Polr2c	308	0,93	147	0,21			158	0,32			100	0,21								
Polr2e	252	1,37	100	0,33	57	0,33	71	0,15			55	0,15								
Polr2g	114	0,44																		
Ppp2cb											77	0,22								
Prdm10	63	0,06					87	0,06												
Prdm2											66	0,04								
Prrc2c									65	0,02										
Psma5	79	0,14							45	0,13										
Ptcd3									235	0,24										
Qser1	135	0,06																		
Rad21							57	0,05	51	0,05										
Rbbp5	144	0,27																		
Rbl1									62	0,06										
Rbpj	169	0,3	142	0,3							222	0,39	64	0,14	61	0,14				
Rfc3			71	0,09			74	0,19												
Rfc4	94	0,3	52	0,09	103	0,3			47	0,08	137	0,3								
Rnf2	212	0,6	52	0,1							176	0,6								
Rsf1	81	0,07	76	0,02	192	0,09	279	0,12	127	0,09	325	0,14								
Ruvbl1	676	1,71	448	1,04	603	1,71	444	0,9	320	0,83	281	0,53	89	0,61	55	0,61			70	0,14
Ruvbl2	729	2,3	240	0,42	497	1,02	716	2,8	291	0,49	170	0,32	130	0,32						
Samd1	206	0,3	139	0,22																
Sap130	215	0,22	183	0,14																
Sap30	136	0,57																		
Sap30l	53	0,18																		

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Sbno1									59	0,02										
Sephs1	83	0,28																		
Sf1	62	0,27																		
Sf3b1									150	0,07										
Shprh	84	0,04	75	0,04			46	0,02			118	0,06								
Sin3a	1717	1,85	1192	0,82	178	0,13	278	0,16	321	0,21	695	0,38								
Sin3b	115	0,09	118	0,09																
Skp1	110	0,45	71	0,45			61	0,45			61	0,2								
Smarca4	430	0,22	106	0,06	530	0,25	742	0,3	510	0,16	580	0,27	56	0,02	57	0,04				
					72	0,18	40	0,09			63	0,09								
	131	0,09	137	0,11	495	0,47	535	0,45	317	0,25	471	0,36	51	0,03	98	0,11				
					280	0,54	240	0,45	152	0,34	197	0,36								
Smc1a	126	0,11	56	0,05	240	0,19	66	0,05	452	0,3	180	0,13	141	0,11						
Smc6									82	0,05										
Smchd1					78	0,03			137	0,05			119	0,05	51	0,02				
Spin1	248	1,31	62	0,13	46	0,13	77	0,13	48	0,13	58	0,13								
Srrt	61	0,07			91	0,07			130	0,11	83	0,04								
Ssrp1	582	0,93	474	0,42	585	0,93	687	0,93	685	0,88	511	0,57	219	0,22	296	0,35	75	0,11	150	0,15
Suds3	234	0,6	105	0,21																
Supt4h2	77	0,67																		
Supt5h	386	0,31	265	0,13	191	0,16	281	0,16	72	0,06	155	0,06								
Supt6h	207	0,12	90	0,02	177	0,1	249	0,08												
Suv39h2													118	0,22	167	0,39				
Suz12	229	0,37	251	0,25	73	0,09	193	0,14	294	0,29	455	0,5			144	0,14				
Taf1	100	0,05																		
Taf2	149	0,12			47	0,03														
Taf3	79	0,04			68	0,04														
Taf4a	144	0,22																		

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Taf5	65	0,09																		
Taf6	230	0,29	136	0,11	97	0,11	168	0,17												
Taf7	84	0,2	70	0,11	48	0,11	59	0,11												
Taldo1							50	0,1			55	0,21								
Tbrg4									67	0,17										
Tcea1	78	0,23			133	0,52			47	0,1										
Tcerg1			57	0,03			125	0,06			63	0,03								
Tcof1													75	0,03						
Tead1							56	0,08			50	0,08								
Tlk2									67	0,08										
Trim24					265	0,24	343	0,25			90	0,07								
Trim33					123	0,09	131	0,06												
Trp53							52	0,11												
Trrap	204	0,06	148	0,04	98	0,04		0,05			52	0,01								
Uba1							64	0,06			83	0,06								
Ube2h									56	0,47										
Ubtf	732	1,07	457	0,5	374	0,5	359	0,38	342	0,35	457	0,5	81	0,13						
Usp48	195	0,13	162	0,16			143	0,16	457	0,29	223	0,13	52	0,05	53	0,05				
Utp14b									110	0,13										
Vm-n2r100			59	0,04							63	0,12								
Wdr18	52	0,08							119	0,24			68	0,08	49	0,08				
Wdr5	202	0,57	170	0,62	86	0,21	133	0,33			115	0,33								
Wdr55			57	0,09																
Xrcc6									70	0,1	56	0,15								
Zfp280c			116	0,09									161	0,24	190	0,19				
Zfp281	74	0,08																		
Zic5	64	0,06																		

Protein	Mascot H3K4 me3 2.0	emPAI H3K4 me3 2.0	Mascot H3K4 me3 2.1	emPAI H3K4 me3 2.1	Mascot H3K27 Ac 2.0	emPAI H3K27 Ac 2.0	Mascot H3K27 Ac 2.1	emPAI H3K27 Ac 2.1	Mascot H3K4 me1 2.0	emPAI H3K4 me1 2.0	Mascot H3K4 me1 2.1	emPAI H3K4 me1 2.1	Mascot H3K9 me3 2.0	emPAI H3K9 me3 2.0	Mascot H3K9 me3 2.1	emPAI H3K9 me3 2.1	Mascot GFP 2.0	emPAI GFP 2.0	Mascot GFP 2.1	emPAI GFP 2.1
Zmynd8					57	0,03	101	0,06			55	0,07								
Znf512			153	0,25					160	0,3	91	0,18	191	0,32	161	0,18				
Zscan10	60	0,08									93	0,17								

Supplementary table 3

Complexes identified by ChIP-MS

Complex / subunit	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
BAF complex	0,04	0,25	0,17	0,01		Enhancer
Arid1a (Baf250a)		0,07	0,06			Enhancer
Dpf2 (Baf45d)		0,22	0,12			Enhancer
Pbrm1 (Baf180)	0,02	0,11	0,11			Enhancer
Smarca4 (Brg1)	0,14	0,28	0,22	0,03		Enhancer
Smarcb1 (Baf47)		0,14	0,05			Enhancer
Smarcc1 (Baf155)	0,10	0,46	0,31	0,07		Enhancer
Smarcd1 (Baf60a)		0,50	0,35			Enhancer
Sin3 complex	0,31	0,01	0,03			Promoter
Sin3a	1,34	0,15	0,30			Promoter
Sin3b	0,09					Promoter
Arid4a (Rbbp1)	0,12					Promoter
Brms1	0,14					Promoter
Brms1l	0,47					Promoter
Fam60a	0,24					Promoter
Ing2	0,11					Promoter
Sap130	0,18					Promoter
Sap30	0,29					Promoter
Sap30l	0,09					Promoter
Suds3	0,41					Promoter
PRC1 complex	0,15		0,14			Promoter
Cbx7	0,10		0,10			Promoter
Phc1			0,04			Enhancer
Rnf2	0,35		0,30			Promoter
PRC2 complex	0,17	0,07	0,23	0,04		Enhancer
Ezh2	0,07	0,04	0,09			Enhancer
Jarid2	0,16	0,05	0,21	0,04		Enhancer
Mtf2	0,16	0,06	0,22	0,06		Enhancer
Suz12	0,31	0,12	0,40	0,07		Enhancer
TFIID complex	0,09	0,04				Promoter
Taf1	0,03					Promoter
Taf2	0,06	0,02				Promoter
Taf3	0,02	0,02				Promoter
Taf4a	0,11					Promoter
Taf5	0,05					Promoter
Taf6	0,20	0,14				Promoter
Taf7	0,16	0,11				Promoter
Aurora Kinase complex	0,19	0,08	0,11	0,69		Heterochromatin
Aurkb	0,22	0,10	0,10	0,65		Heterochromatin
Cdca8	0,28	0,13	0,24	1,17		Heterochromatin
Incenp	0,06			0,26		Heterochromatin

Complex / subunit	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Pol II complex	0,44	0,17	0,06			Promoter
Polr2a	0,15	0,13	0,06			Promoter
Polr2b	0,43	0,33	0,07			Promoter
Polr2c	0,57	0,16	0,11			Promoter
Polr2e	0,85	0,24	0,08			Promoter
Polr2g	0,22					Promoter
NuRD complex	0,21	0,21	0,31	0,07	0,02	Enhancer
Chd4	0,27	0,29	0,44	0,17	0,02	Enhancer
Gatad2a	0,16	0,21	0,27	0,11		Enhancer
Mta2	0,19	0,21	0,26		0,07	Enhancer
Mta3	0,24	0,14	0,27			Enhancer
MLL complex	0,34	0,04	0,05			Promoter
Mll2	0,21		0,02		0,01	Promoter
Ash2l	0,25					Promoter
Bap18	0,63					Promoter
Dpy30	0,18					Promoter
Hcfc1	0,05					Promoter
Men1	0,32	0,03	0,15			Promoter
Rbbp5	0,14					Promoter
Wdr5	0,60	0,27	0,17			Promoter
Trrap/Ep400 complex	0,57	0,66	0,23	0,15	0,01	Promoter
Trrap	0,05	0,05	0,01			Promoter
Ep400		0,02				Unclear
Dmap1	0,07		0,04			Promoter
Ruvbl1	1,38	1,31	0,68	0,61	0,07	Promoter
Ruvbl2	1,36	1,91	0,41	0,16		Promoter
INO80 complex	0,02	0,03				Promoter
Ino80	0,05	0,05				Promoter
Actr5		0,03				Unclear
Nfrkb		0,03				Unclear
HBO1/MOZ/MORF complex	0,22	0,09	0,10	0,02		Promoter
Myst2	0,56	0,36	0,39	0,06		Promoter
Myst3	0,08					Promoter
Myst4	0,04					Promoter
Brd1	0,03	0,05	0,10			Enhancer
Brpf1	0,16					Promoter
Brpf3	0,03					Promoter
Meaf6	0,51	0,19	0,19			Promoter
Ing4	0,14		0,07			Promoter
Ing5	0,58	0,13	0,20			Promoter
Phf16	0,04					Promoter
Phf17	0,31	0,26	0,16	0,12		Promoter
Cohesin complex	0,04	0,07	0,12	0,03		Enhancer

Complex / subunit	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Rad21		0,03	0,03			Enhancer
Smc1a	0,08	0,12	0,22	0,06		Enhancer
DSIF complex	0,28	0,08	0,03			Promoter
Supt4h2	0,34					Promoter
Supt5h	0,22	0,16	0,06			Promoter

[a] **Prediction of genome localization based on our ChIP-MS criteria**

Supplementary table 4

Overlap identifications for H3K4me3 with Vermeulen *et al.*[1], Bartke *et al.*[2], Nikolov *et al.*[3] and Soldi *et al.*[4]

Overlap with Vermeulen *et al.* [1]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Bap18	0,63					Promoter
Bptf	0,08	0,02	0,02			Promoter
Chd1	0,26	0,24	0,14	0,02		Promoter
Phf8	0,23	0,02				Promoter
Sin3a	1,34	0,15	0,30			Promoter
Taf1	0,03					Promoter
Taf2	0,06	0,02				Promoter
Taf3	0,02	0,02				Promoter
Taf4a	0,11					Promoter
Taf5	0,05					Promoter
Taf6	0,20	0,14				Promoter
Taf7	0,16	0,11				Promoter

Overlap with Bartke *et al.* [2]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Chd1	0,26	0,24	0,14	0,02		Promoter
Phf8	0,23	0,02				Promoter
Sin3a	1,34	0,15	0,30			Promoter
Spin1	0,72	0,13	0,13			Promoter

Overlap with Nikolov *et al.* [3]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Bptf	0,08	0,02	0,02			Promoter
Brms1	0,14					Promoter
Brms1l	0,47					Promoter
Chd1	0,26	0,24	0,14	0,02		Promoter
Dpy30	0,18					Promoter
Emsy	0,02					Promoter
Fam60a	0,24					Promoter
Ing2	0,11					Promoter
Ing4	0,14		0,07			Promoter
Ing5	0,58	0,13	0,20			Promoter
Kdm2a	0,16	0,03	0,03			Promoter
Kdm5a	0,02					Promoter
Kdm5b	0,25	0,02	0,12			Promoter
Myst2	0,56	0,36	0,39	0,06		Promoter
Phf16	0,04					Promoter
Phf23	0,53		0,09			Promoter
Sin3a	1,34	0,15	0,30			Promoter
Sin3b	0,09					Promoter

Overlap with Nikolov *et al.*[3] continued

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Sap30	0,29					Promoter
Spin1	0,72	0,13	0,13			Promoter
Taf1	0,03					Promoter
Taf3	0,02	0,02				Promoter
Taf6	0,20	0,14				Promoter
Taf7	0,16	0,11				Promoter
Wdr5	0,60	0,27	0,17			Promoter

Overlap with Soldi *et al.*[4]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Ash2l	0,25					Promoter
Bptf	0,08	0,02	0,02			Promoter
Brd4	0,03	0,21	0,10			Enhancer
Chd4	0,27	0,29	0,44	0,17	0,02	Enhancer
Ctcf	0,07		0,05			Promoter
Glyr1	0,19	0,24	0,26	0,07		Enhancer
Hcfc1	0,05					Promoter
Nrf1	0,08					Promoter
Polr2a	0,15	0,13	0,06			Promoter
Polr2b	0,43	0,33	0,07			Promoter
Rbbp5	0,14					Promoter
Sap130	0,18					Promoter
Sin3a	1,34	0,15	0,30			Promoter
Smarca4	0,14	0,28	0,22	0,03		Enhancer
Spin1	0,72	0,13	0,13			Promoter
Ssrp1	0,68	0,93	0,73	0,29	0,08	Enhancer
Taf2	0,06	0,02				Promoter
Wdr5	0,60	0,27	0,17			Promoter

[a] Prediction of genome localization based on our ChIP-MS criteria

Supplementary table 5

Overlap identifications for H3K9me3 with Vermeulen *et al.*[1], Bartke *et al.*[2], Nikolov *et al.*[3] and Soldi *et al.*[4]

Overlap with Vermeulen *et al.* [1]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Lrwd1	0,14		0,11	0,35		Heterochromatin

Overlap with Bartke *et al.* [2]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Lrwd1	0,14		0,11	0,35		Heterochromatin

Overlap with Nikolov *et al.* [3]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Atrx				0,09		Heterochromatin
Chd4	0,27	0,29	0,44	0,17	0,02	Enhancer
Smchd1		0,02	0,03	0,04		Heterochromatin

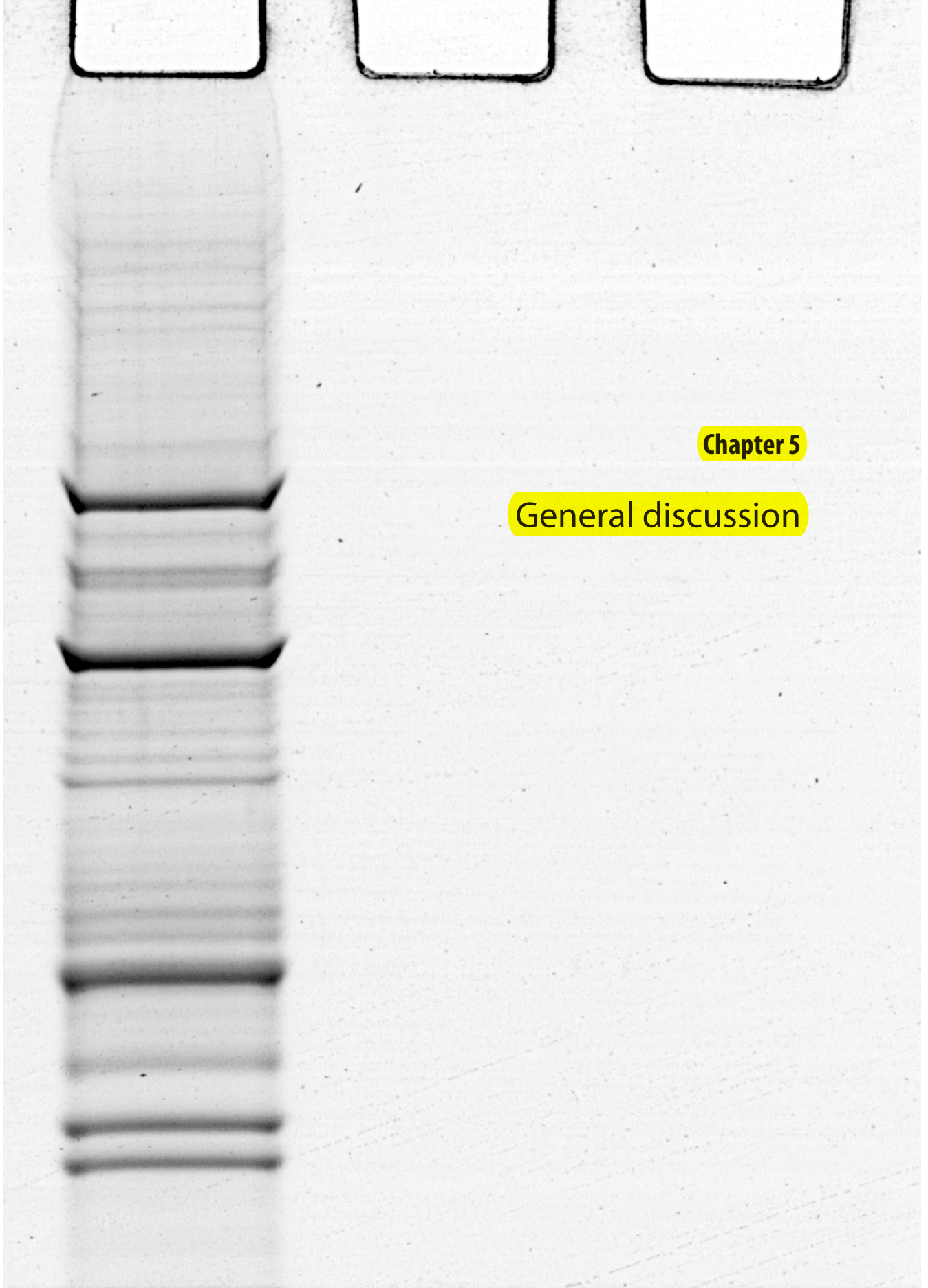
Overlap with Soldi *et al.* [4]

Protein	H3K4me3 Avrg emPAI	H3K27Ac Avrg emPAI	H3K4me1 Avrg emPAI	H3K9me3 Avrg emPAI	GFP Avrg emPAI	Prediction[a]
Atad2		0,03	0,04	0,04		Enhancer
Cdca8	0,28	0,13	0,24	1,17		Heterochromatin
Chd4	0,27	0,29	0,44	0,17	0,02	Enhancer
Hp1bp3	0,26		0,38	0,23		Enhancer
Incenp	0,06			0,26		Heterochromatin
Lmna	0,13	0,08	0,08	0,48		Heterochromatin
Lmnb2	0,12	0,24	0,30	0,98		Heterochromatin
Smc1a	0,08	0,12	0,22	0,06		Enhancer
Smchd1		0,02	0,03	0,04		Heterochromatin
Usp48	0,15	0,08	0,21	0,05		Enhancer

Supplementary table 6

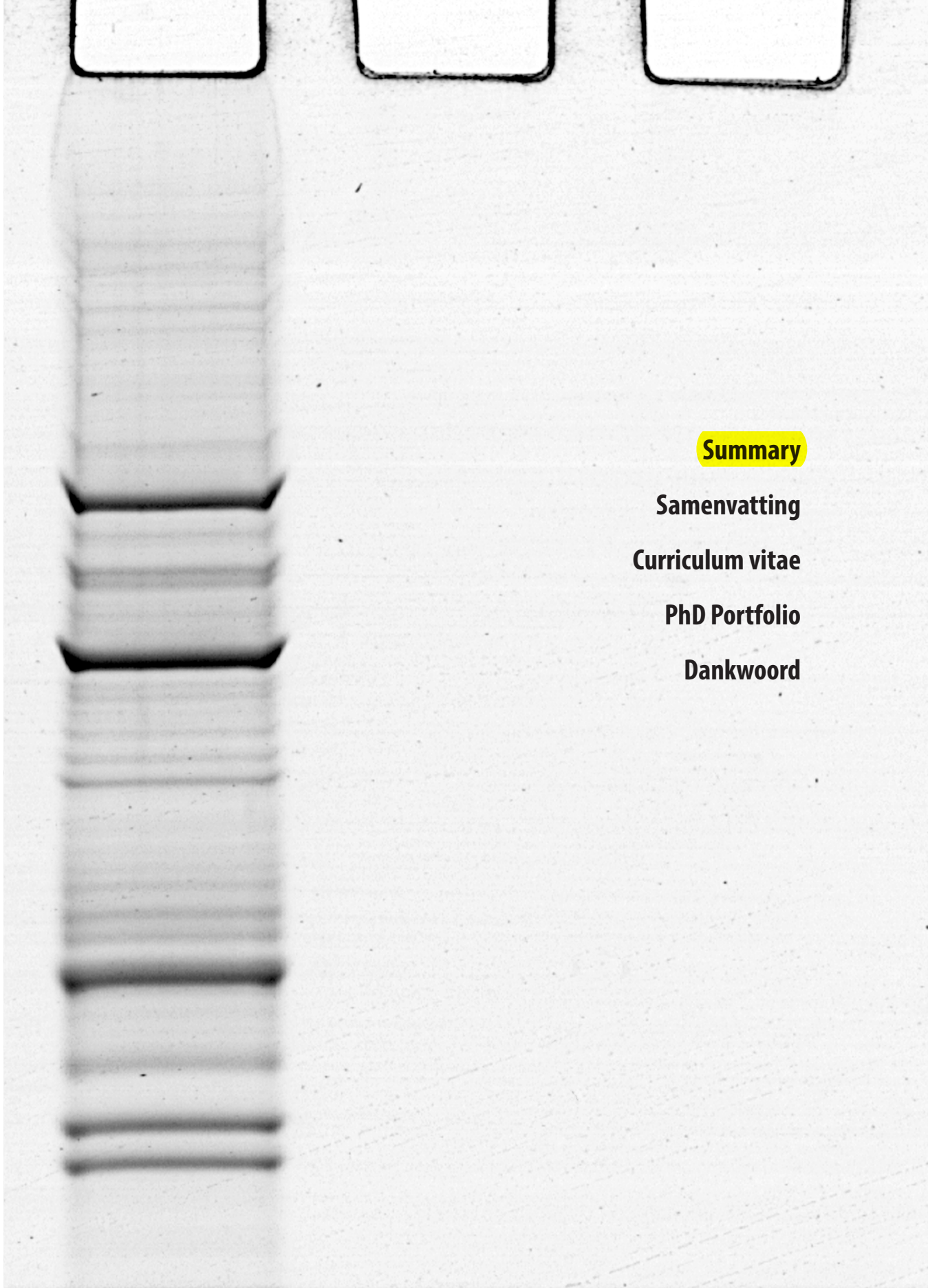
Accession numbers of studies used for genome-wide correlation

Histone modifications	GEO Dataset accesion number
H3K27Ac	GSE24165
H3K4me1	GSE11172
H3K4me3	GSE24165
H3K9me3	GSE12241
Protein	**GEO Dataset accesion number**
Atrx	GSE22162
Brd4	GSE36561
Cbx7	GSE42466
Cfct	GSE49847
Chd4	GSE27844
Ctr9	GSE20530
Esrrb	GSE11431
Ezh2	GSE49178
Hdac1	GSE27844
Hdac2	GSE27844
Jarid2	GSE19708
Kdm1a	GSE27844
Kdm2a	GSE21202
Kdm5b	GSE31968
Mll2	GSE48172
Mtf2	GSE16526
Polr2a	GSE49847
Oct4	GSE44286
Rad21	GSE33346
Rbbp5	GSE22934
Rnf2	GSE26680
Smarca4	GSE14344
Smc1a	GSE22557
Supt5h	GSE20485
Suz12	GSE48122
Taf1	GSE31270
Taf3	GSE30959
Wdr5	GSE22934
Protein	**Bioproject accesion number**
Tcea1	PRJEB2674
Control	GSE24165



Chapter 5

General discussion



Samenvatting

Het behoud van embryonale en neurale stamcellen wordt gestuurd door een specifiek netwerk van regulerende genen. Expressie van genen wordt op verschillende niveaus tijdens het transcriptie proces gereguleerd, waarbij een grote verscheidenheid aan eiwitten is betrokken. Het bepalen van eiwit-eiwit interacties onder transcriptie factoren en identificeren van nieuwe factoren die betrokken zijn bij transcriptionele regulatie zal ons begrip van de moleculaire mechanismen die genexpressie controleren helpen verbeteren.

Een veel gebruikte methode voor het bepalen van onbekende eiwit-eiwit interacties is immunoprecipitatie gevolgd door massaspectrometrie analyse. In hoofdstuk 2 beschrijven we een gedetailleerd protocol voor het uitvoeren van een op FLAG affiniteit gebaseerde purificatie procedure. Het protocol legt stap-voor-stap de procedure uit vanaf de bereiding van eiwit extract uit de celkern tot het uitvoeren van de FLAG affiniteit zuivering en wijst daarbij op de kritische aandachtspunten. De efficiëntie en reproduceerbaarheid van deze methode maken het mogelijk een transcriptiefactor en de interacterende factoren op te zuiveren.

De transcriptiefactor Sox2 is een essentiële regulator van de zelf-vernieuwing en differentiatie processen in neurale stamcellen (NS cellen). We hebben de in hoofdstuk 2 beschreven methode gebruikt om Sox2 op te zuiveren en de gebonden factoren in NS cellen van de muis te bepalen. We hebben meer dan 50 Sox2 interacterende factoren geïdentificeerd die worden gepresenteerd in hoofdstuk 3. Onder die factoren hebben we een ATP-afhankelijke chromatine vervormer, Chd7, als een belangrijke interactie partner van Sox2 gevonden. Depletie experimenten waarin we shRNA tegen Sox2 en Chd7 gebruikten bleek dat Sox2-Chd7 een overlappende set genen reguleert. We hebben ook chromatine immunoprecipitatie (ChIP) experimenten gevolgd door sequencing uitgevoerd. Deze ChIP experimenten toonde aan dat de meeste van deze genen ook door Sox2 en Chd7 gebonden zijn. In de mens veroorzaken haploïde mutaties in het *SOX2* gen een SOX2 anophthalmia syndroom terwijl mutaties in het *CHD7* gen het CHARGE syndroom. Na literatuur onderzoek vonden we dat symptomen tussen deze afzonderlijke syndromen overlappen dat een indicatie is voor een betrokkenheid van SOX2 en CHD7 in soortgelijke processen. Verder hebben we aangetoond dat de genen *Jag1* , *Gli2, Gli3* en *Mycn* worden gereguleerd door Sox2 en Chd7. Opvallend is dat deze genen betrokken zijn bij andere erfelijke syndromen die soortgelijke symptomen vertonen zoals waargenomen bij het SOX2 anophthalmia en CHARGE syndroom. Hieruit concluderen wij dat Sox2 met Chd7 samenwerkt om genen te reguleren, waarvan sommige in menselijke syndromen zijn gemuteerd.

Veel factoren die een functie hebben in de regulatie van transcriptie handelen op genomische elementen zoals promotors en enhancers. Deze gebieden kunnen worden onderscheiden door bepaalde chromatine kenmerken, zoals specifieke histon-eitwit

modificaties. Deze modificaties worden tegenwoordig gebruikt om genomische elementen te annoteren op het genoom. In hoofdstuk 4 beschrijven we een nieuwe aanpak waarbij we ChIP voor histon-eiwit modificaties hebben gecombineerd met massaspectrometrie (ChIP-MS). Met deze techniek hebben we factoren die aan genomische elementen gebonden zijn kunnen detecteren in muis embryonale stamcellen (ES cellen). Dit stelde ons in staat een catalogus van meer dan 250 factoren samen te stellen met factoren die verrijkt zijn in een bepaalde chromatine fractie. Bovendien konden we van deze factoren hun associatie met promotoren, enhancers en heterochromatin voorspellen. De op ChIP-MS gebaseerde voorspelling van lokalisatie op deze genomische elementen kon voor een selectie van deze factoren worden bevestigd door analyse van hun genomische bindingsprofielen. Opvallend is dat meer dan een kwart van de geïdentificeerde factoren zijn beschreven belangrijk te zijn voor pluripotentie in ES cellen. Ook hebben we een ChIP gevolgd door sequencing voor de pluripotentie factor, Dppa2 uitgevoerd. Het genomisch bindingsprofiel van Dppa2 overlapt met dat van H3K4me3. De voorspelling van promoter-associatie door ChIP-MS wordt bevestigt door de verrijking van Dppa2 op promotoren van genen die in de testis tot expressie komen.

In het kort dragen de technieken en resultaten als beschreven in dit proefschrift bij aan het begrip van transcriptie factor netwerken in embryonale en neurale stam cellen. Zij zullen ook nader onderzoek naar deze netwerken ondersteunen en kunnen leiden tot nieuwe inzichten in hun werking.

Curriculum vitae

Personal details

Name:	Erik Marinus Petrus Engelen
Date of birth:	27 May 1982
Place of birth:	Roosendaal en Nispen, the Netherlands

Education

2007-2009	MSc Molecular Medicine Erasmus University, Rotterdam, the Netherlands
2000-2004	BSc Biomedical Research Avans Hogeschool, Breda, the Netherlands
1995-2000	HAVO Mgr. Frencken College, Oosterhout, the Netherlands

Research

2009-2013	PhD research Department of Cell Biology, Erasmus MC, Rotterdam, the Netherlands (Prof.dr. F.G. Grosveld and Dr. R.A. Poot)
2006-2009	Research technician Department of Cell Biology, Erasmus MC, Rotterdam, the Netherlands (Prof.dr. F.G. Grosveld and Dr. R.A. Poot)
2004-2006	Research technician Department of Genetics, Erasmus MC, Rotterdam, the Netherlands (Prof.dr. G.T. van der Horst and Dr. F. Tamanini)
2003-2004	BSc Graduation Project Department of Genetics, Erasmus MC, Rotterdam, the Netherlands (Prof.dr. G.T. van der Horst and Dr. F. Tamanini)
2003	BSc Internship Department of Pathology, University of Vermont, Burlington, Vermont, United States of America (Prof.dr. Y.M.W. Janssen-Heininger)

List of publications

Engelen, E., Janssens, R.C., Yagita, K., Smits, V.A., van der Horst, G.T., and Tamanini, F. (2013). Mammalian TIMELESS is involved in period determination and DNA damage-dependent phase advancing of the circadian clock. PLoS One *8*, e56623.

Kheradmand Kia, S., Verbeek, E., Engelen, E., Schot, R., Poot, R.A., de Coo, I.F., Lequin, M.H., Poulton, C.J., Pourfarzad, F., Grosveld, F.G., et al. (2012). RTTN Mutations Link Primary Cilia Function to Organization of the Human Cerebral Cortex. Am J Hum Genet *91*, 533-540.

Engelen, E.*, Akinci, U.*, Bryne, J.C., Hou, J., Gontan, C., Moen, M., Szumska, D., Kockx, C., van Ijcken, W., Dekkers, D.H., et al. (2011). Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat Genet *43*, 607-611.

Gontan, C., Guttler, T., Engelen, E., Demmers, J., Fornerod, M., Grosveld, F.G., Tibboel, D., Gorlich, D., Poot, R.A., and Rottier, R.J. (2009). Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors. J Cell Biol *185*, 27-34.

van den Berg, D.L., 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.

(* equal author contribution)

PhD Portfolio

Department of Cell Biology **September 2009 - September 2013**

Research school: Medical Genetics Centre South-West Netherlands (MGC)

Promotor: Prof. dr. F.G. Grosveld **Co-promotor: Dr. R.A. Poot**

1. PhD training	
General courses	Year
▪ From Development to Disease	2009
▪ ESF Next Generation Sequencing Meeting	2010
▪ EuTRACC Proteomics course	2010
▪ Epigenetic regulation in health and disease	2010
▪ Loopbaan oriëntatie training	2012
▪ NIHES Biostatistical Methods I: Basic principles part A	2012
▪ Biomedical English Writing and Communication	2013
Seminars and Workshops	
▪ Winter School Chromatin Changes in Differentiation and Malignancies, Kleinwalsertal, Austria (oral presentation)	2009-2012
▪ MGC PhD student workshop	2010-2012
organizing committee, poster presentation	2011
oral presentation	2012
National Conferences	
▪ DSSCR Stem Cell Meeting, Rotterdam	2009
▪ NIRM Meeting, Utrecht	2010
▪ 10th Dutch Chromatin Meeting, Amsterdam	2012
▪ Joint Dutch Chromatin Meeting & NVBMB Fall Symposium, Rotterdam	2013
International Conferences	
▪ EMBO meeting, Amsterdam, the Netherlands	2009
▪ Stem Cells, Development and Regulation, Amsterdam, the Netherlands	2009, 2010, 2012

▪ Chromatin Changes in Differentiation and Malignancies, Giessen, Germany (poster presentation)	2011
▪ EuTRACC Transcription Networks Symposium, Berlin, Germany	2011
▪ Keystone Symposium Chromatin Dynamics, Keystone, Colorado, USA (scholarship award, poster presentation)	2012
▪ Chromatin Changes in Differentiation and Malignancies, Egmond aan Zee, the Netherlands	2013
2. Teaching	
▪ Master thesis	2011

Dankwoord

Het moment is daar. Het schrijven van het “belangrijkste” deel (of in ieder geval het deel dat de meeste van jullie daadwerkelijk zullen lezen) van mijn proefschrift; het dankwoord. Na tien en een half jaar hier rond te hebben gelopen, heb ik veel slimme, leuke en vriendelijke mensen leren kennen. In dit dankwoord wil ik mij vooral richten op de mensen die belangrijk zijn geweest voor mijn laatste 5 jaar als PhD student. Dit is een hoofdstuk uit mijn leven waar ik nu al met veel genoegen op terugkijk. Dankzij jullie heb ik deze jaren veel geleerd over wetenschap en ben ik altijd met veel plezier naar het werk gekomen. Naast het onderzoek heb ik mijn vrije tijd ook vaak gedeeld met vele van jullie doormiddels van vele avonturen en leuke activiteiten, waar ik enorm van heb genoten.

Ik wil beginnen met het bedanken van mijn begeleider Raymond. De keuze om voor jou als analist te gaan werken is waarschijnlijk één van de belangrijkere keuzes in mijn wetenschappelijke carrière geweest. Vanaf het begin stimuleerde jij mijn zelfstandigheid en ik kreeg al vrij snel een eigen project om aan te werken. Je heb me daarna altijd aangemoedigd te groeien en gaf me de kans om mijn masters diploma te halen wat logischerwijs doorliep in mijn promotie onderzoek. We hebben dat samen tot een heel mooi einde weten te brengen en hopelijk komt daar binnenkort nog een mooi verhaal bij. Ik heb veel geleerd van jouw kijk op de wetenschap en hoe deze wordt bedreven. Ik kon altijd bij jou terecht voor een eerlijke en nuchtere conversatie, of het nu over wetenschappelijke of over persoonlijke dingen ging. Raymond, bedankt voor jouw investering om mijn PhD tot een geslaagd einde te brengen.

Natuurlijk was mijn promotie onderzoek binnen de afdeling cel biologie niet mogelijk zonder mijn promotor, Frank Grosveld. Ik realiseer me maar al te goed dat de afdeling zoals jij die over de jaren hebt opgebouwd, met goede faciliteiten, de juiste mensen, een aangename werksfeer met het unieke pantomime, niet vanzelfsprekend is. Al deze dingen hebben veel bijgedragen aan een succesvolle afronding van mijn onderzoek. Ik wil je bedanken voor het leggen van dit solide fundament van waaruit ik heb mogen werken en alle input die je me hebt gegeven tijdens de monday morning meetings, op verschillende borrels en tijdens de KWT meeting.

Dan wil ik graag de leden van mijn kleine commissie Sjaak, Adriaan en Danny bedanken voor het lezen van mijn proefschrift. In het bijzonder wil ik Sjaak bedanken voor de vijf keer (geen record helaas) dat ik mee mocht naar de door jouw mede georganiseerde KWT meetings in Oostenrijk. De informele sfeer daar zorgt voor een perfecte mix van leerzame en gezellige momenten die ik voor geen goud had willen missen (helaas miste Kramer wel het goud tijdens de OS in 2010). Ook bedankt voor het delen van het vreugdevolle nieuws

van de acceptatie van het Nature Genetics artikel daar. Adriaan, bedankt voor de goede discussies die we hadden, of ze nu gingen over het leven of over Facebook. We zullen elkaar in ieder geval blijven volgen in de toekomst. Danny, bedankt dat je op het laatste moment wilde wisselen tussen grote en kleine commissie. De binnenkomst op de afdeling was memorabel met de Belgische bier proeverij.

Ook wil ik de overige leden van de commissie, Robert Hofstra, Fred van Leeuwen, Robbert Rottier en Derk ten Berge bedanken voor het lezen van mijn proefschrift en deel te nemen in de oppositie. Fred, wie had gedacht na elkaar in het vliegtuig richting Keystone, Colorado tegen te komen, jij twee jaar later in mijn promotie commissie zou zitten. Robbert, fijn dat jij als Sox2-expert en co-autheur op twee artikelen ook in mijn commissie zit. Derk, bedankt voor het begrip tijdens het wisselen van commissies.

Tijdens de laatste fase sta ik er gelukkig niet alleen voor en heb ik me voor versterking omringt met mijn paranimfen Sander, Maaike en Debbie. Sander, jij was toen ik bij de afdeling genetica begon gelijk een goede collega en voorbeeld voor mij als vooruitstrevende analist. Ondanks dat ik je baan inpikte nadat je terug kwam vanuit Australië, zijn we goede vrienden gebleven. Die vriendschap groeit nog steeds dankzij de verschillende festivals, snowboardvakanties en spelletjes avonden die ik met jou, Ilse en je schattige kereltje Luuk beleef. Ik vind het enorm fijn dat je naast me staat tijdens de verdediging en wens je voor daarna veel succes met de komst van de tweede!

Maaike, ik ben blij dat er zo'n slimme meid als jij naast mij staat tijdens de verdediging. Jij stimuleert het gevoel van samenhorigheid, door middel van het organiseren van Sinterklaas en spelletjes avonden, barbecues in het park of je dakterras, iets dat ieder lab nodig heeft. Ik bewonder jouw goede mensenkennis en ook jouw manier van het verwoorden van deze impressies. We kunnen het altijd goed met elkaar vinden en ik vind het cool zo'n stoere surfchick als paranimf te hebben. Ik wens je veel succes en doorzettingsvermogen voor het afronden van jou promotie.

Debbie, ik waardeer het dat je voor mij uit Londen komt om deze voor mij speciale gebeurtenis mee te maken. Met jouw succesvolle PhD hat je de lat wel hoog gelegd voor me, maar dat is gelukkig goed gekomen, iets wat zonder jouw input en altijd parate kennis niet mogelijk was geweest. Ik ben blij jou ook buiten het lab als goede vriendin te hebben en we hebben dan ook veel lol samen tijdens gezellige Boudewijn bier drink avonden en skivakanties. Je bent een zorgzame en gastvrije Brabander en ik geniet dan ook altijd enorm van de weekendjes in Londen. Ik heb al het vertrouwen dat jij je Postdoc goed zal afronden, waarna je de vrijheid hebt om te doen wat je wilt.

Dan zijn er natuurlijk ook de andere (oud) Poot groepsleden. Umut, teşekkür ederim for a successful cooperation on our cool Sox2-Chd7 adventure. I hope you find a new job soon and wish you luck in continuing a scientific career. Johan, tank voor de ijverige samenwerking aan het laatste hoofdstuk van dit proefschrift. Ik heb er vertrouwen in dat we het nog mooi gepubliceerd krijgen binnenkort. Marti, gràcies for the good vibe you brought to the lab. I wish you the best for your PhD. Mike, bedankt voor het helpen bij meerdere dingen tegelijk. Luca, grazie for helping me out on the ChIP-MS project. Good luck in your PhD on the tenth floor! En dan Ernie, onze lab moeder. Officieel hoor je dan wel niet bij het Poot lab, voor mij ben jij altijd de steunpilaar in het lab geweest. Bedankt voor jouw altijd luisterende oor en vele adviezen, of het nu over experimentele of alledaagse dingen ging. Er zijn dan ook nog veel mensen op lab 706 voorbij gekomen die allen bijgedragen hebben aan plezierige werksfeer, zoals studenten Tim en Chantal, maar ook Tiago (obrigado for coming to the defense), Katy, Marianne en Miyata.

Absoluut onmisbaar voor het onderzoek beschreven in dit proefschrift zijn de mensen van de “mass spec” faciliteit. Jeroen, Dick, Karel en Erik-Jan, enorm bedankt voor jullie goede en snelle afhandeling van onze vele gels en opdrachten. Ons lab draait om jullie expertise en gelukkig betaald dat zich uit in mooie verhalen die goed terecht komen. Karen, ik wens jou veel succes met de afronding van jou PhD en keuzes voor de toekomst.

De ondersteunende rol van de afdeling Biomics is ook essentieel geweest voor de resultaten die hier zijn beschreven. Wilfred, Christel, Edwin, Antoine, Zeliha, Rutger, Miriam en Xander bedankt voor al jullie harde werk.

In my time at the Erasmus I got to know some people within various departments that became close friends of mine. I would like to thank them for their listening ears and the fun moments they shared with me. My mate Karl. Thanks for all the great conversations and advise about almost everything from music to lifestyle during the many flat whites we had together. My man Eskeatnaf. Amesege’nallo’ for all the good times we had together, from chewing Khat to your first meters on the ski’s. Squash buddy’s Rick and Bjorn. Thanks for the introduction to the sport and for that “one” drink we always do. Filippo, thanks for your introduction into science and the great adventures we experienced together. Pocket rockets Eugin. What happens in Vegas stays in Vegas.

Ik ben ook erg dankbaar voor het hartelijke ontvangst dat ik kreeg van Frank en Maureen toen ik in Rotterdam kwam wonen. Jullie hebben letterlijk jullie deur voor mij opengezet en ik was altijd welkom om even een hapje mee te eten of gewoon voor een gezellige avond. Ik had me geen fijner thuis in Rotterdam kunnen wensen.

Of course I would like to thank everyone for their help with questions or experiments, just a joyful hello or daily conversation in the corridor or at drinks. Here's an attempt to thank most of you and I apologize if I've forgotten to mention you. Let's start close to home: The neighbors of 702; Ralph (het was gezellig als 'echte' buurman), Robert-Jan, Anita, Andrea, Guillaume, Petros, and old 702 folks; Eric, Charlotte, Daan, Erik (de eerste) and Petra (we go way back ;)). The other neighbours from 710; Dubi, Alex, Rien and Michael. The Philipsen lab; Thamar, Nynke, Harmen, Maria, Ileana and Sylvia. The other world upstairs, the department of Cell biology on the tenth floor; Dies, Linde (keep the haggis warm and the beer cold, see you soon in Edinburgh), Martine, Noori, Siska, Gert, Niels, Kris, Widia, Kerstin, Jessica, Lalini, Luna, Marta, Joshua, Dorota, Nesrin, Elaine, Thomas, Chris, Polynikis and Reinier. People from the in between (ninth) floor; Anton (bedankt dat ik als analist de master opleiding mocht volgen), Joost (bedankt voor de altijd goede raad), Willy, Cristina, Friedo, Agnese, Ruben, Bas, Fabrizia, Federica, Francesca, Selma and Annegien. Of course the people from my old "home", the department of Genetics. Jan, Bert (bedankt voor de mogelijkheid om in jou lab te mogen beginnen en kennis te maken met de wetenschap), Ines, Annelieke, Romana, Roel, César, Yanto, Renata, Wim, Arjan, Petra, Loes, Özge, Hervé, Natasja, Jeroen (enzymen doen met jou was altijd een leuk uitje), Nicolaas, Bert-Jaap (Daft Punk rules). Ook heb ik met veel plezier meegeholpen met het organiseren van het wetenschapscafé met dank aan Marja, Thijs, Ruud, Maikel, Eline.

Dit cluster draait niet zonder de steun van de vaste kern. Onmisbaar zijn jullie allen voor het succesvolle verloop van het onderzoek binnen deze afdeling en daarom wil ik bedanken: de secretaresses; Marike, Jasperina, Bep, Sonja en Mariëlle voor het altijd binnen kunnen lopen voor allerhande vragen en verzorgen van administratieve zaken, de inkoopafdeling; Melle, Leo, Koos, Annette en Enno voor het aanhoren van mijn gezeur over bestellingen, de dames van de spoelkeuken voor het altijd maar schoonmaken en autoclaveren van alle benodigdheden en niet te vergeten de IT afdeling; Ton, Jozef, Mario, Leo, Pim en Niels voor het gelijk oplossen van alle computer problemen. Van het clusterbureau van de afdeling immunologie wil ik Gelof van Steensil bedanken voor zijn gastvrijheid om op de verlaten achtste verdieping rustig aan dit proefschrift te mogen werken. Tom de Vries Lentsch, bedankt voor je hulp bij het gebruik van Indesign.

Mark, bedankt voor het bieden van die relax momenten, waar we gewoon een biertje dronken, filmpje keken en ouwehoerden over die goede oude tijd. Het was altijd ontspannend om na een dag op het lab even alles te kunnen vergeten en het eens even niet over wetenschap te hebben.

Ik wil natuurlijk ook mijn familie bedanken voor hun steun en interesse tijdens mijn studies. Het is fijn te weten dat ik zo'n goede terugval basis heb. Jullie liefde en warmhartigheid is

altijd daar wanneer ik het nodig heb. Ome Wim, ik kan me nog goed herinneren dat ik jou een rondleiding gaf op het lab. Die glinstering in uw ogen inspireert mij om de wetenschap onder de mensen te willen brengen.

Pa, ik had het fijn gevonden als jij bij mijn verdediging zou zijn geweest en dit met mij had kunnen delen.

Mama, bedankt dat je er altijd voor mij bent en in mij gelooft. De onvoorwaardelijke schouder waarop ik kon leunen tijdens het maken van moeilijke beslissingen en mij de vrijheid te geven in alles wat ik doe.

Liebe Judith. Du bist wirklich ein erstaunliches Mädchen und das Beste, was mir je passiert ist. Deine Unterstützung für diese Dissertation ist unmöglich in Worte zu fassen. Ich kann dir einfach nicht genug dafür danken und du sollst wissen, dass ich für dich da bin, wann immer du mich brauchst. Ich kann es kaum erwarten zu sehen, was unsere gemeinsame Zukunft bringen wird.