Dual Role of Med12

in PRC1-dependent Gene Repression &

ncRNA-mediated Transcriptional Activation

Dissertation zur Erlangung des Grades

"Doktor der Naturwissenschaften"

am Fachbereich Biologie

der Johannes Gutenberg-Universität in Mainz

Thaleia Papadopoulou

geb. am 25.07.1986 in Athen, Griechenland

Mainz, 2016

Tag der mündlichen Prüfung: 28.10.2016

Σα βγεις στον πηγαιμό για την Ιθάκη,
να εύχεσαι νάναι μακρύς ο δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον θυμωμένο Ποσειδώνα μη φοβάσαι,
τέτοια στον δρόμο σου ποτέ σου δεν θα βρεις,
αν μέν' η σκέψις σου υψηλή, αν εκλεκτή
συγκίνησις το πνεύμα και το σώμα σου αγγίζει.
Τους Λαιστρυγόνας και τους Κύκλωπας,
τον άγριο Ποσειδώνα δεν θα συναντήσεις,
αν δεν τους κουβανείς μες στην ψυχή σου,
αν η ψυχή σου δεν τους στήνει εμπρός σου.

As you set out for Ithaka
hope the voyage is a long one,
full of adventure, full of discovery.
Laistrygonians and Cyclops,
angry Poseidon—don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.
Laistrygonians and Cyclops,
wild Poseidon—you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

(Translation by Edmund Keeley/Philip Sherrard)

Brichst du auf gen Ithaka,
wünsch dir eine lange Fahrt,
voller Abenteuer und Erkenntnisse.
Die Lästrygonen und Zyklopen,
den zornigen Poseidon fürchte nicht,
solcherlei wirst du auf deiner Fahrt nie finden,
wenn dein Denken hochgespannt, wenn edle
Regung deinen Geist und Körper anrührt.
Den Lästrygonen und Zyklopen,
dem wütenden Poseidon wirst du nicht begegnen,
falls du sie nicht in deiner Seele mit dir trägst,
falls deine Seele sie nicht vor dir aufbaut.

(Tranlation by lyrikwelt.de/)

The poem Ithaka by Constantinos P. Cavafy very much describes the journey towards a PhD.

I dedicate this thesis to all those who contributed into "keeping my thoughts raised high" during this "voyage"

Table of Contents

INDEX OF ABBREVIATIONS

Abs: Antibodies
ASH2L: Absent Small Or Homeotic-Like
ATRA: All-Trans-Retinoic Acid
BCIP: 5-Bromo-4-chloro-3-indolyl phosphate
CBP: Creb Binding Protein
Cdk8: Cyclin-dependent kinase 8
ChIP: Chromatin ImmunoPrecipitation
CREs: Cis-Regulatory Elements
CSK: CytoSkeleton Buffer
CTCF: CCCTC-binding factor
CTD: C-Terminal Domain
DAPI: 4',6-diamidino-2-phenylindole
DE: Differentially Expressed
DMEM: Dubeco's Minimum Essential Medium
DMSO: Dimethyl sulfoxide
DNMT1: DNA (Cytosine-5-)-Methyltransferase 1
DTT: Dithiothreitol
EDTA: Ethylenediaminetetraacetic acid
EGTA: Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
eRNA: enhancer RNA
FACT: Facilitates Chromatin Trascription
FBS: Fetal Bovive Serum
FC: Fold Change
GMEM: Glasgow Minimum Essential Medium
GO: Gene Ontology
GTF: General Transcription Factor
HBS: HEPES Buffer Saline
HEK283T: Human Embryonic Kidney 293 Transformed
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ICM: Inner Cell Mass
IgG: Immunoglobulin G
IP: Immuno Precipitation
LIF: leukaemia Inhibiting Factor
lincRNA: long intergenic non-coding RNA
lncRNA: long non-coding RNA
Med1: Mediator subunit 1
Med12: Mediator subunit 12
MEL: Murine Erythroleukemia
mESCs: mouse embryonic stem cells
MLL: Mixed Lymphoblastic Leukemia
MPP11: M Phase Phosphoprotein 11
mRNA: messenger RNA
MS: Mass Spectrometry
NBT: Nitro-Blue Tetrazolium
NFW: Nuclease Free Water
NMC: Non Mammalian Control
NTMT: Alkaline Phospharase Buffer
PBS: Phosphate-buffered saline
PcG: Polycomb Group
PEI: Polyethylenimine
PIC: Pre-Initiation Complex
PIPES: 1,4-Piperazinediethanesulfonic acid
PRC: Polycomb Repressive Compex
PRE: Polycomb Responsive Elements
P-TEF2b: Positive- Transcription Elongation Factor 2b
qPCR: quantitatice Polymerase Chain Reaction
RA: Retinoic Acid
RAC: Ribosome Associated Complex
RNA pol: RNA polymerase
RT: Room Temperature
S2P: Serine 2 Phosphorylation
S5P: Sering 5 Phosphorylation
SDS: Sodium dodecyl sulfate
siRNA: short interferering RNA
TE: Trophoectoderm
TF: Transcription Factor
TFIIA-H: Transcription Factor IIA-H
TPA: 12-O-tetradecanoylphorbol-13-acetate
tRNA: transfer RNA
UCSC: University of California Santa Cruz (Database)
WT: Wild Type
ZRF1: Zuotin Related Factor 1

SUMMARY

Embryonic stem cells (ESCs) are cells derived from the inner cell mass of the blastocyst of the early-stage preimplantation embryo and constitute a unique cell population due to their capacity of self-renewal and differentiation. While self-renewal ensures the propagation of the stem cell population, with differentiation ESCs give rise to all three primary germ layers from which the tissues and organs of the developing embryo will be shaped. During this process, ESCs undergo stable changes in their epigenome that enables them to acquire new identities. Both in pluripotency and differentiation, ESCs are under a tight transcriptional program on the level of chromatin which controls the proper spatiotemporal gene expression. This epigenetic control requires the combinatorial action of gene-specific transcription factors (TFs), the general transcriptional machinery, ncRNAs and regulatory elements.

In an attempt to understand how genes which code for lineage-specification programs get activated in early differentiation, we studied the epigenetic changes which they acquire from pluripotency to early differentiation. We focused on genes which are silenced in pluripotency by Polycomb group complexes (PcGs) and more specifically by the PRC1 complex. This complex via Ring1b which is an E3 ligase, decorates the histone H2A with an ubiquitin mark, H2AK119ub, a histone modification considered to be a hallmark of gene repression. PRC1-repressed genes get activated in early differentiation via the displacement of Ring1b by the H2AK119ub-interacting protein, Zrf1. In this study we explored how PRC1 establishes the silencing of key developmental genes in pluripotency. We provide evidence of the interplay between Ring1b and the subunit Med12 of the Mediator co-activator complex, which restricts the expression of those genes in mouse ESCs. A set of these genes, get activated in early differentiation by Med12 in complex with ncRNAs. This step requires the assembly of Med12 with an additional Mediator subunit, Cdk8, a remodeling event which relies on the recruitment of the latter by Zrf1.

These findings contribute to a better understanding of how silencing by PRC1 is established in pluripotency and how this step sets the ground for the transition to early stem cell differentiation.

ZUSAMMENFASSUNG

Embryonale Stammzellen werden aus der inneren Zellmasse der Blastozyste gewonnen. Aus ihnen entstehen durch Vermehrung und weitreichende Differenzierung die vielen Zelltypen und Gewebe, die zur Bildung eines vollständigen Lebewesens erforderlich sind. Während der Pluripotenz sind Stammzellen durch ein Transkriptionsprogramm gekennzeichent, welches die raumzeitliche Expression bestimmter Gene ermöglicht. Während der Differenzierung wird das Epigenom der Zellen umprogrammiert, sodass neue Transkriptionsprogramme enstehen. Die zugrundeliegende epigenetischen Regulationsmechanismen beinhalten die Funktionen von Transkriptionsfaktoren, der allgemeinen Transkriptionsmaschninerie und epigenetischen Komponenten wie etwa ncRNAs und Gen-Regulationselemente.

Um zu verstehen wie Differenzierungs-relevante Gene aktiviert werden, wurden die epigenetischen Veränderungen untersucht, die vom Übergang der Pluripotenz zur frühen Differenzierungsphase auftreten. Dazu wurden primär Gene untersucht, die während der Pluripotenz durch die *Polycomb group complexes* (PRCs) und insbesondere den Polycomb-Komplex 1 (PRC1) abgeschaltet werden. PRC1 beinhaltet eine enzymatische Aktivität, die E3 Ubiquitin Ligase Ring1b, welche die Mono-ubiquitylierung des Histons H2A an dessen Lysine 119 katalysiert. Diese Histonmodifizierung wird mit einer Genabschaltung in Verbindung gebracht, allerdings wird die selbe Modifizierung auch durch das Protein Zrf1 gebunden, welches die Aktivierung des betreffenden Gens verursacht. In der vorliegenden Arbeit wurde untersucht wie PRC1 während der Pluripotenz Gene abschaltet, die für die Entwicklung essentiell sind. Es konnte diesbezüglich gezeigt werden, dass PRC1 und die Mediatorkomplex-Untereinheit Med12 gemeinsam an der Abschaltung dieser Gene beteiligt sind. Die Aktivierung

derselben Gene während der Differenzierung erfolgt einerseits durch ncRNAs, welche mit Med12 interagieren, andererseits durch einen Umbau des Mediatorkomplexes. Die Rekrutierung von Zrf1 an genregulatorische Elemente bewirkt dabei eine Loslösung des PRC1 vom Mediatorkomplex und den gleichzeitigen Einbau der Cdk8 Untereinheit in den Mediatorkomplex. Damit wird der Mediatorkomplex in einen Transkriptions-unterstützenden Faktor umgewandelt.

Zusammengenommen trägt die Arbeit zu einem besseren mechanistischen Verständis der PRC1-vermittelten Geneabschaltung und der Aktivierung von Polycomb-Genen bei.

INTRODUCTION

Stem cells are unique due to their capacity for self-renewal and differentiation (Keller, 1995; Morrison et al., 1997; Smith, 2001; Till and McCulloch, 1980; Weissman, 2000). The property of self-renewal ensures the propagation of the stem cell population through development and later in life is essential for its maintenance in adult tissues and expansion in the event of injury (He et al., 2009). By differentiating in response to external stimuli, stem cells give rise to every cell type of the developing organism.

Stem cell origin

A revolutionary point towards the understanding of stem cell biology was their establishment as an *in vitro* cell culture model. In 1981, undifferentiated cells, derived from blastocysts of early mouse embryos were successfully cultured without genetic transformation (Evans and Kaufman, 1981; Martin, 1981). Since then, embryonic stem cells (ESCs) can be expanded indefinitely in culture while being able to contribute in the development of chimeras when injected in mouse blastocysts (Martello and Smith, 2014). It is noteworthy that mESCs are pluripotent cells; they self-renew and differentiate in all cell types of the body except from the extraembryonic trophoblast lineage. Totipotency, the ability to generate all the lineages of the organism, is a property solely found in the zygote and early blastomeres (Tarkowski, 1959).

After the first cell divisions following egg fertilization, the 16-cell morula undergoes the first differentiation event in mammalian development; the external cells acquire trophoectodermal (TE) fate whereas the internal cells inner cell mass (ICM) fate. The blastula is formed during the following cell division resulting in the blastocyst sphere. The outer layer of the blastocyst is composed by TE cells which will form the chorion, the embryonic part of the placenta. The ICM harbors the pluripotent stem cells (ESCs) which give rise to the embryo proper (Boroviak et al., 2014; Morey et al., 2015) (Figure 1) through differentiation to the three germ layers during gastrulation (ectoderm, endoderm and mesoderm) (Chen and Dent, 2014). Once it is specified which cells will become TE or ICM, the resulting populations start expressing distinct

transcription factors (TFs) (Figure 2). TE cells express *Eomes* and *Cdx2*. Eomes activates the characteristic proteins of the trophoblast layer (Russ et al., 2000) and Cdx2 downregulates *Oct4* and *Nanog* (Strumpf et al., 2005). Oct4, Nanog and Stat3 are characteristic TFs of the ICM (Strumpf et al., 2005). Whereas at 8-cell stage all cells express *Cdx2*, *Eomes* and *Oct4*, in the blastocyst their expression is restricted to the newly specified lineages (Figure 1& 2) (Niwa et al., 2005).

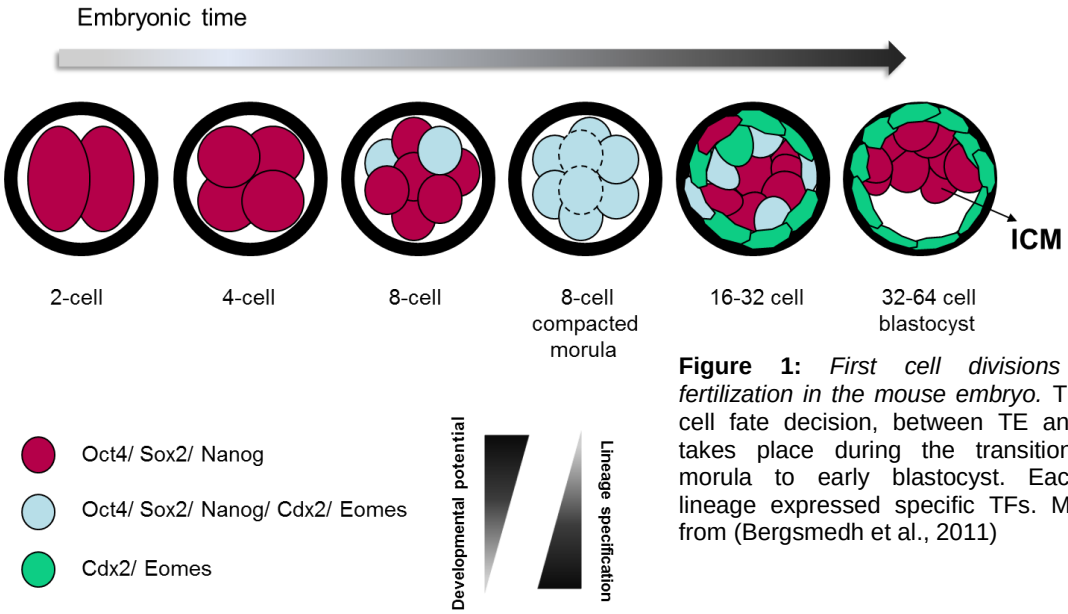


Figure 1: *First cell divisions post-fertilization in the mouse embryo.* The first cell fate decision, between TE and ICM takes place during the transition from morula to early blastocyst. Each cell lineage expressed specific TFs. Modified from (Bergsmedh et al., 2011)

Although mESCs were first derived in 1981, further efforts were required to ensure that their culture conditions resembled the natural conditions in the ICM. Early stem cell cultures required the co-culture of fibroblasts, as a supporting feeder layer, in the presence of calf serum. This combination created an artificial environment, highly unrelated to the physiological conditions of cells in preimplantation embryos (Martello and Smith, 2014). In the following years the basic goal was to identify feeder layer and serum factors, which are essential for the maintenance of ESC pluripotency. Research towards this end led to the identification of a cytokine, leukemia inhibitory factor (LIF), as the fibroblast secreted protein required for the effective suppression of mESC differentiation in culture (Smith and Hooper, 1987; Smith et al., 1988; Williams et al., 1988). The

molecular function of LIF requires its binding to the LIFR/gp130 receptor which activates the JAK-STAT pathway-dependent, STAT3 expression (Niwa et al., 1998). LIF is not sufficient to maintain self-renewal of ESCs in the absence of serum. It was much later revealed that the serum component BMP4, is essential for supporting pluripotency in the presence of LIF, by inducing the expression of Id proteins, required to repress differentiation TFs (Ying et al., 2003).

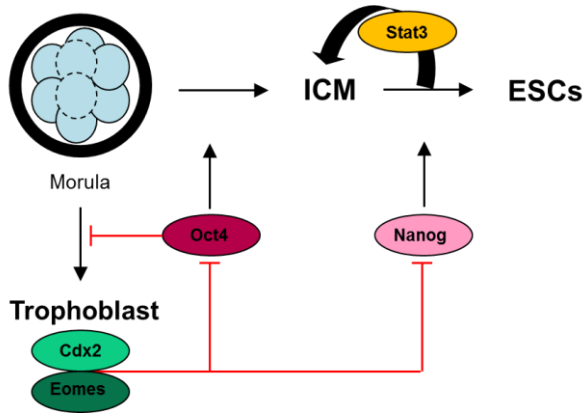


Figure 2: *Schematic representation for the functions of Oct4, Nanog, Cdx2, Eomes and Stat3 following the first differentiation event in mouse development*. Cdx2 downregulates Oct4 and Nanog in TE cells whereas Oct4, Nanog and Stat3 are essential for the maintenance of pluripotency. Oct4 inhibits ICM cells from acquiring trophoblastic fate. Modified from (Gilbert et al., 2006)

Transcriptional regulation in pluripotency and differentiation

Pluripotency as well as the transition to differentiation entail a very specific regulation. This is accomplished by the combinatorial function of TFs, chromatin regulators, non-coding RNAs (ncRNAs), DNA regulatory elements and the general transcriptional machinery. All these factors form an interconnected network that maintains pluripotency and establishes the proper environment for the subsequent induction of differentiation.

Embryonic Stem Cell Transcription Factors

The three TFs, Oct4, Nanog and Sox2 form the central core of the molecular circuitry that controls pluripotency (Figure 3). Oct4, expressed from the locus *Pou5f1*, is a homeodomain transcription factor which is considered to be the most upstream gene in the molecular circuitry of pluripotency (Jaenisch and Young, 2008). Its zygotic expression begins prior to the 8-cell stage (Palmieri et al., 1994; Yeom et al., 1991) and it is expressed uniformly throughout the morula (Wu and Scholer, 2014). It activates its own expression through a positive auto-regulatory loop with Sox2 (Okumura-Nakanishi et al., 2005). Nanog, which also harbors a

homeodomain, is essential for the stabilization of the pluripotent state rather than its maintenance (Chambers et al., 2007). The contribution of Sox2 in pluripotency is partially attributed to its function to regulate *Oct4* levels (Masui et al., 2007). These three factors bind altogether at each other's promoters (Boyer et al., 2005) in hESCs mainly to maintain their own expression (Jaenisch and Young, 2008). They also co-occupy their targets which can be divided in actively expressed genes as well as repressed genes, poised for differentiation (Boyer et al., 2005).

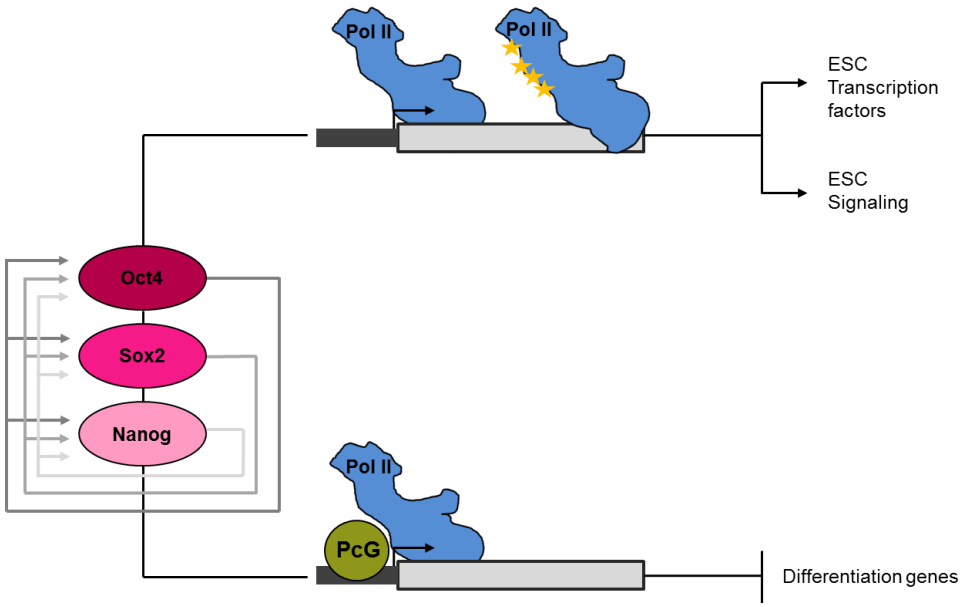


Figure 3: *Model of Core ESC Regulatory Circuitry*. The master transcription factors Oct4, Sox2 and Nanog occupy each other's promoters as well as those of actively transcribed and silenced genes in pluripotency. The actively transcribed genes include TFs and components of signaling pathways that are required for the ES state maintenance. The silenced genes include regulators of lineage specification and fate commitment which are also targeted by PcG complexes and are kept poised for activation upon a differentiation signal. PcGs prevent the phosphorylation (stars) of the RNA pol II and therefore the elongation of transcription. Modified from (Jaenisch and Young, 2008).

Chromatin Regulators

An additional layer of stem cell regulation lies at the level of chromatin. Promoters of genes which are silenced in pluripotency yet poised for activation upon differentiation, are co-occupied by the master TFs (Oct4, Nanog and Sox2) and epigenetic regulators (Bernstein et al., 2006; Lee et al., 2006) (Figure 3). Epigenetic regulators control gene expression through the deposition of post-transcriptional histone modifications that influence the chromatin state either

by affecting the compaction or by providing binding sites for additional proteins (Chen and Dent, 2014).

The most prominent epigenetic regulators that govern pluripotency are the Polycomb group (PcG) proteins. These multiprotein complexes regulate gene expression by repressing the transcription of differentiation genes, thereby controlling key fate decisions (Bracken and Helin, 2009; Sauvageau and Sauvageau, 2010; Schwartz and Pirrotta, 2007; Simon and Kingston, 2009).

Polycomb identification

PcG proteins were first identified in *Drosophila* as regulators of HOX gene expression (Geisler and Paro, 2015; Grossniklaus and Paro, 2014; Steffen and Ringrose, 2014), in genetic screens for homeotic transformations. They are conserved from plants to humans and their existence is largely connected with multicellularity (Whitcomb et al., 2007). PcG evolution is still under debate though, as various subunits are present in unicellular organisms as well but lost in certain cases as for example in *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* (Margueron and Reinberg, 2011; Shaver et al., 2010). PcG mutants exhibit defects in axis specification and body patterning (Jungers, 1985; Schwartz and Pirrotta, 2007) which highlights the importance of PcG-mediated repression in the spatiotemporal control of gene expression. The silencing of HOX expression outside their specific expression domain and its maintenance throughout development rendered PcGs as an attractive system for cellular memory (Ringrose and Paro, 2004; Schuettengruber and Cavalli, 2009). In the recent years, advances in genomic research have shown that PcG-mediated repression spreads beyond HOX genes, regulating numerous developmental factors and signaling pathways throughout the genome (Simon and Kingston, 2013).

Over the last decade two PcG Repression Complexes (PRCs) have been extensively studied: PRC2 (Morey and Helin, 2010; Schuettengruber et al., 2007), a catalyzer of histone H3 trimethylation on lysine 27 (H3K27me3) (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et

al., 2002; Muller et al., 2002) and PRC1, which contains an E3 ligase subunit (Ring1A or Ring1B) (Wang et al., 2004) and is responsible for the ubiquitylation of histone H2A on lysine 119 (H2AK119ub) (Cao et al., 2005).

Polycomb recruitment

These two complexes act successively or independently of each other. The hierarchical or canonical model of PRC action (Simon and Kingston, 2013) involves the recruitment of PRC1 as a function of PRC2. More specifically, the PRC2 complex gets first recruited to chromatin via a yet not fully understood mechanism (Figure 4A) (Morey and Helin, 2010; Morey et al., 2015). The deposition of H3K27me3 via the PRC2 serves as a docking site for the recruitment of the PRC1 complex through the chromodomain of the Cbx subunits (Bernstein et al., 2006; Luis et al., 2012; Morey and Helin, 2010). This model is associated with the canonical PRC1 harboring Cbx (Cbx2, 4, 6, 7 or 8), PHC (PHC1, 2 or 3) and PCGF (PCGF2 or 4) subunits (Comet and Helin, 2014). This hierarchy seems to be reverted in certain cases, where PRC1 variants are sufficient to recruit PRC2 in a mechanism which depends in DNA methylation (Blackledge et al., 2014; Cooper et al., 2014) (Figure 4B). In this case PRC1 exists in a variant form containing PCGF1, RYBP or YAF2 and KDM2B (Comet and Helin, 2014). Furthermore, H2Aub nucleosomes seem to provide a preferential binding site for PRC2 which enhances its catalytic activity (Kalb et al., 2014). On the other hand there is growing evidence for the existence of a non-canonical mechanism of PRC1 recruitment (Leeb and Wutz, 2007; Schoeftner et al., 2006) which is reinforced by the existence of a PRC1 variant, harboring RYPB instead of Cbx proteins, that deposits H2Aub independently or PRC2 (Tavares et al., 2012).

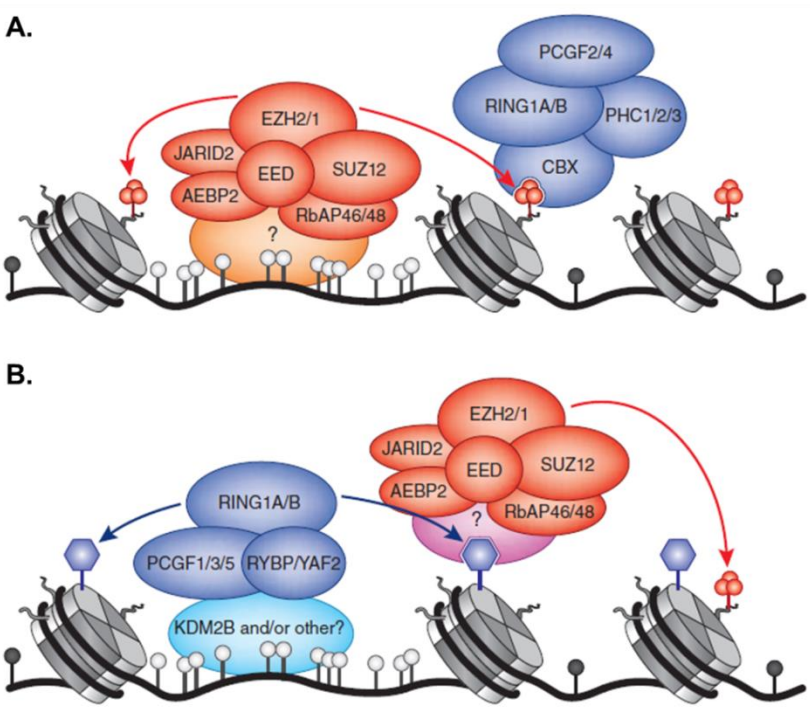


Figure 4: *Hierarchical models for recruitment of PcGs to chromatin*. A) Canonical model for PcG recruitment. PRC1 gets recruited to chromatin through recognition of H3K27me3 via its CBX subunit. B) Reverted hierarchical model (new hierarchical model) for PcG recruitment. PRC1 variants are recruited to unmethylated CpG islands, leading to H2AK199ub1 deposition and recruitment of PRC2 via a yet unresolved mechanism (Comet and Helin, 2014)

Despite the fact that many models exist to describe how PcGs are recruited in respect to each other, it is not fully understood how their *de novo* targeting on genomic targets is accomplished. This question has been quite well explored in *Drosophila*, where Polycomb Responsive Elements (PREs) and DNA-binding recruiters have been identified (Oktaba et al., 2008; Papp and Muller, 2006). In mammals though such DNA elements do not exist and the PcG-recruitment issue remains highly unresolved. With the increased complexity in PcG subunit composition from flies to mammals and the emergence of DNA methylation, it is proposed that recruitment has also evolved in a mechanism of higher complexity (Simon and Kingston, 2013). The most attractive speculations on that matter involve the employment of ncRNAs (Simon and Kingston, 2013), CpG islands (Ku et al., 2008; Mendenhall et al., 2010) as well as short RNAs which are transcribed from PcG-repressed loci, are enriched in H3K4me3 but their transcription is independent of PcG activity (Kanhere et al., 2010).

Polycomb Molecular function

The ultimate function of PRC2 is to create the H3K27me3 which is considered as the effective mark of PcG silencing. The catalytic subunit for the deposition of this mark lies within the SET domain of Ezh2 although the methyltransferase gets activated only upon assembly with Suz12 and Eed (Cao and Zhang, 2004; Ketel et al., 2005; Pasini et al., 2004). Methylation of H3K27 is accomplished in a processive manner (Margueron and Reinberg, 2011) and H3K27me3 is a stable mark (Zee et al., 2010), highly abundant in ESCs. Genome-wide studies show that at least 10% of ES genes are enriched with H3K27me3 (Mohn et al., 2008) which primarily localizes around the transcription start site (TSS) of mammalian promoters (Boyer et al., 2006; Lee et al., 2006; Pan et al., 2007; Zhao et al., 2007) which are rich in CpG islands (Ku et al., 2008; Meissner et al., 2008; Mikkelsen et al., 2007; Mohn et al., 2008) lacking methylated cytosine (Bantignies and Cavalli, 2011; Williams et al., 2011). To date, the main function of this histone modification appears to be the recruitment of additional factors that maintain repression, although it is likely that it also prevents the binding of proteins on chromatin which would suggest an indirect regulatory role in the process of transcriptional silencing (Margueron and Reinberg, 2011). In contrast to H3K27me3, the exact role and localization of H2AK119ub1 in ESCs is not fully resolved (Endoh et al., 2012; Vissers et al., 2008). It was initially thought to be linked with transcriptional activation (Levinger and Varshavsky, 1982) however its functional proximity with PcG has rendered it to a hallmark of repression (Richly and Di Croce, 2011). Like H3K27me3, H2AK119ub1 is also a very abundant mark, being found in around 10% of the total H2A of mammalian cells (Richly and Di Croce, 2011). A distinct set of genes which represent central PcG targets in mESCs, were found to be enriched in both marks and the deposition of H2AK119ub1 on these genes was Ring1-dependent (Endoh et al., 2012). Furthermore, the enzymatic function of PRC1 is required for gene silencing but is dispensable for its biding to target genes as well as for the propagation of chromatin compaction (Endoh et al., 2012; Eskeland et al., 2010). *In vitro* PRC1-compacted chromatin is resistant to remodeling via the SWI/SNF complex (Francis et al., 2001; King et al., 2002; Simon and Kingston, 2013) and this

function is evolutionary conserved and independent of Ring1b's catalytic activity (Eskeland et al., 2010; Simon and Kingston, 2013). Recently, the E3 ligase activity of Ring1b, was found to be dispensable for early mouse embryonic development (Illingworth et al., 2015), providing additional evidence that PRC1-mediated repression probably lies beyond H2AK119ub.

It is evident that PRC1-dependent repression requires multiple effector mechanisms which insure not only the robustness of silencing but also the plasticity of the system in response to differentiation stimuli. Amongst others, a rather attractive model proposes transcriptional repression by PRC1 via direct interaction with the transcriptional machinery. The first indications for the existence of such a regulatory mechanism were indirect, focusing in lower levels of promoter-associated RNA Polymerase II (RNA Pol II) at genes targeted by PRC1 and blocking of transcription initiation (Min et al., 2011). PRC1 is also connected with inhibition of transcriptional elongation which is H2AK199ub-dependent and this phenomenon is mostly associated with silenced, yet poised PcG targets (Stock et al., 2007; Zhou et al., 2008). Nonetheless, the finding that PRC1 inhibited the binding of Mediator to nucleosome templates proposed a more direct role for this complex in postponing transcriptional activation (Lehmann et al., 2012). Although PRC1 blocks Mediator, TFIID, a general transcription factor of the transcriptional machinery, could still bind to these templates, implying that Mediator subunits might provide an interphase for PRC1-mediated silencing (Simon and Kingston, 2013).

ncRNAs

One of the major and more intriguing findings of the Human Genome Project was the discovery of many genes that get transcribed but do not have protein-coding potential. Later it became clear that the majority of the genome is being actively transcribed (pervasive transcription) (Consortium et al., 2007). Nonetheless, the notion that RNA molecules can be indeed functional and not just a byproduct of faulty RNA pol II initiation events (Ebisuya et al., 2008; Struhl, 2007) was still controversial (Rinn and Chang, 2012). More recent advances in chromatin research, such as Chromatin Immunoprecipitation followed by massively parallel sequencing (ChIP-seq),

enabled the genome-wide identification of active transcription sites. These sites are characterized by distinct signatures of RNA pol II binding and active histone modifications, such as H3K4me3 at promoters and H3K36me3 at gene bodies (Guttman et al., 2009; Marson et al., 2008; Mikkelsen et al., 2007). The presence of such signatures across the mouse and human genome, in regions devoid of known protein-coding genes led to the identification of thousands of long intergenic non-coding transcripts lincRNAs (Guttman et al., 2009). Since then, long non-coding RNAs (lncRNAs) have evolved as new players in transcriptional control with emerging functions in association with proteins and DNA elements (Rinn, 2014).

LncRNAs, are transcripts that originate from RNA genes and are longer than 200bp with little or no coding potential (Rinn, 2014). Besides this very general classification, they can also be grouped according to their genomic origin or their function. Antisense lncRNAs are being transcribed from the opposite direction in respect to their protein-coding gene. Divergent lncRNAs get transcribed from bivalent promoters that also control protein-coding genes. Intronic lncRNAs originate from introns within protein-coding genes whereas transcripts that overlap protein-coding genes are often termed as overlapping lncRNAs and lincRNAs are transcribed from intergenic regions between protein-coding genes (Figure 5).

Despite of the fact that all these transcripts have different genomic origins, in many cases they share common mechanisms to exert their specific functions. The regulatory functions of lncRNAs are often linked to their binding with protein partners. Many well described RNA-protein interactions serve as examples for classifying lncRNAs depending on their functions (Figure 6). *Gas5* is an example of a non-coding transcript which gets induced upon growth factor starvation and acts as a decoy to inhibit the binding of the glucocorticoid receptor to DNA and subsequent expression of metabolic genes (Kino et al., 2010).

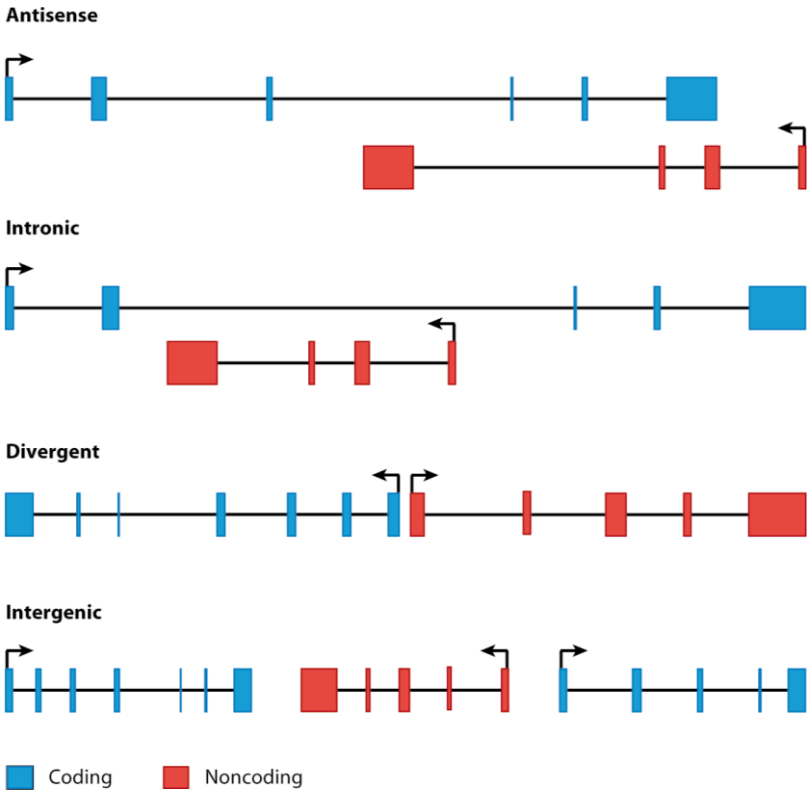


Figure 5: *Classification of lncRNAs as defined by the position of their locus in respect to protein-coding genes.* **Antisense** lncRNAs are transcribed from the 3' end of protein-coding genes. **Intronic** lncRNAs initiate their transcription within introns of protein-coding genes. **Divergent** lncRNAs get transcribed from bilavent promoters at a reverse direction in respect to the protein-coding gene. **Intergenic** lncRNAs are transcribed from genomic locations between protein-coding genes. These intergenic regions can often be defined as gene deserts (Rinn and Chang, 2012)

Other lncRNAs within this group include the PANDA lncRNA which prevents p53-mediated apoptosis by associating with the transcription factor NF-YA (Hung et al., 2011) and RNAs which bind to DNMT1 to inhibit DNA methylation of certain genomic loci like CEBPA (Di Ruscio et al., 2013). Many lncRNAs serve as scaffolds that bridge proteins into regulatory complexes, therefore facilitating their functions. Amongst the most well characterized scaffold lncRNAs are the TERC which enables the assembly of the telomerase complex (Zappulla and Cech, 2006) and the ncRNAs HOTAIR (Rinn et al., 2007) and ANRIL (Kotake et al., 2011; Yap et al., 2010).

HOTAIR is one of the most well studied lncRNAs mainly because of its overexpression in many types of cancer and its clinical use as a biomarker (Gupta et al., 2010; Kim et al., 2013; Wu et al., 2014; Zhang et al., 2014). It is expressed from the HOXC locus and it interacts with the PRC2 complex, enabling its localization and induction of silencing at the HOXD locus (Rinn, 2014; Rinn et al., 2007). Although this function would only attribute a guide role to HOTAIR, the subsequent finding that it binds simultaneously the PRC2 and the LSD1-CoREST complex to

induce both H3K27me3 and H3K4me2 demethylation (Tsai et al., 2010) revealed that it acts as a scaffold to coordinate silencing by these two protein complexes.

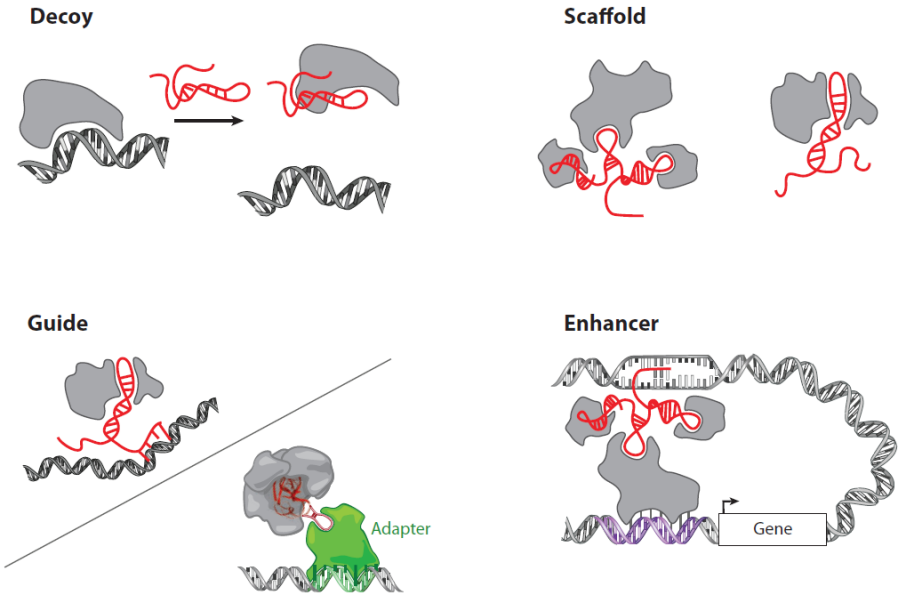


Figure 6: *Classification of lncRNAs in respect to their function;* **Decoy** lncRNAs inhibit the association of DNA-binding factors to their regulatory sequences to prevent transcription of downstream genes. **Scaffold** lncRNAs bind to multiple protein partners enabling the formation of protein complexes that control transcription in a coordinated manner. **Guide** lncRNAs bind to protein partners to facilitate their localization to selective regions of the genome. **Enhancer** lncRNAs get transcribed in extragenic enhancer regions and promote chromatin architecture by interacting with cell-type specific TFs and the Mediator complex (Rinn and Chang, 2012).

ANRIL also acts as a lncRNA-scaffold that facilitates the PRC1 and PRC2-mediated silencing of the tumor suppressor INK4b/ARF/INK4a locus (Kotake et al., 2011; Yap et al., 2010) in normally proliferating cells. Guide lncRNAs combine the binding to a protein partner with the targeting of the protein to selective regions in the genome (Rinn and Chang, 2012). Apart from HOTAIR, other lncRNAs that serve as guides are involved in dosage compensation and imprinting including Xist (Jeon and Lee, 2011; Khalil et al., 2009), Kcnq1ot1 (Pandey et al., 2008) and Air (Nagano et al., 2008). At last but not least, a special case of lncRNAs are the enhancer-RNAs (eRNAs) which get transcribed from enhancer regions and their transcription results in the activation of nearby genes (Kim et al., 2010; De Santa et al., 2010). The mechanism by which eRNAs enhance the activation of neighboring genes involves the facilitation of enhancer-promoter interactions, at least in cases of stimulus-induced enhancers (Shiekhattar, 2013). This

type of lncRNAs appear to be involved in differentiation processes whereby their transcription within regulatory regions of differentiation-specific TFs promotes chromatin remodeling that contributes to cell type-specific transcriptional control as an additional regulatory layer (Mousavi et al., 2013). It has been also shown that eRNAs contribute to chromatin architecture by interacting with the Mediator complex (Lai et al., 2013), known to enhance transcription by acting as a bridge between TFs and the general transcriptional machinery (Conaway and Conaway, 2011).

The General Transcriptional Machinery, Co-Activators & cis-Regulatory Elements

Transcription factors, chromatin regulators and lncRNAs are indispensable components of the molecular circuitry of pluripotency. Nonetheless, they only act as regulatory signals that insure the proper transcriptional control of this process and their function is always coupled with the main drivers of gene expression (Malik and Roeder, 2010; Roeder, 2005); the transcriptional machinery and the DNA regulatory elements.

Control Modes of the Eukaryotic Transcriptional Machinery

Gene expression, in the highly complex eukaryotic genome, requires multiple levels of control which coordinate this process in a synergistic manner (Roeder, 2003). The first level of control involves different classes of RNA polymerases (Roeder and Rutter, 1969) which specifically catalyze the transcription of distinct types of genes; RNA Pol I transcribes large ribosomal genes, RNA Pol II protein-coding and some structural RNA genes (such as lncRNAs) and RNA Pol III tRNA, small ribosomal genes (5S) and small structural RNA genes (Roeder, 2005; Weinmann and Roeder, 1974; Weinmann et al., 1974). The specificity of each of the eukaryotic RNA polymerases has emerged from the combination of common subunits with the prokaryotic RNA pol, with unique ones (Roeder, 2005; Sklar et al., 1975). The second control level ensures the accurate transcription initiation on gene promoters and includes RNA Pol – specific general initiation factors (Matsui et al., 1980; Ng et al., 1979; Parker and Roeder, 1977; Segall et al., 1980; Weil et al., 1979). The RNA Pol II-specific factors include TFIIA, TFIIB, TFIID, TFIIE and

TFIIH (Roeder, 1996). The assembly of RNA Pol II and initiation factors on core promoters constitutes the pre-initiation complex (PIC) which is the first step for the subsequent transcription initiation. The PIC is formed as a response to gene-specific regulatory factors -usually TFs- and co-factors that bind at distal elements in respect to the core promoter (Figure 7). These regulatory factors provide the third level of control in eukaryotic transcription (Roeder, 2005) and they exert their function by direct binding on DNA elements via recognition of specific nucleotide motifs which are 6-12bp long degenerate DNA sequences (Spitz and Furlong, 2012). In many cases these motifs are found in clusters within *cis*-regulatory elements that enables the binding of several TFs at the same time (Spitz and Furlong, 2012). This combinatorial binding of TFs provides an additional level of transcriptional control which in the processes of differentiation and development results in precise patterns of gene expression (Halfon et al., 2000; Lettice et al., 2012; Sandmann et al., 2007; Small et al., 1992; Yuh et al., 1994).

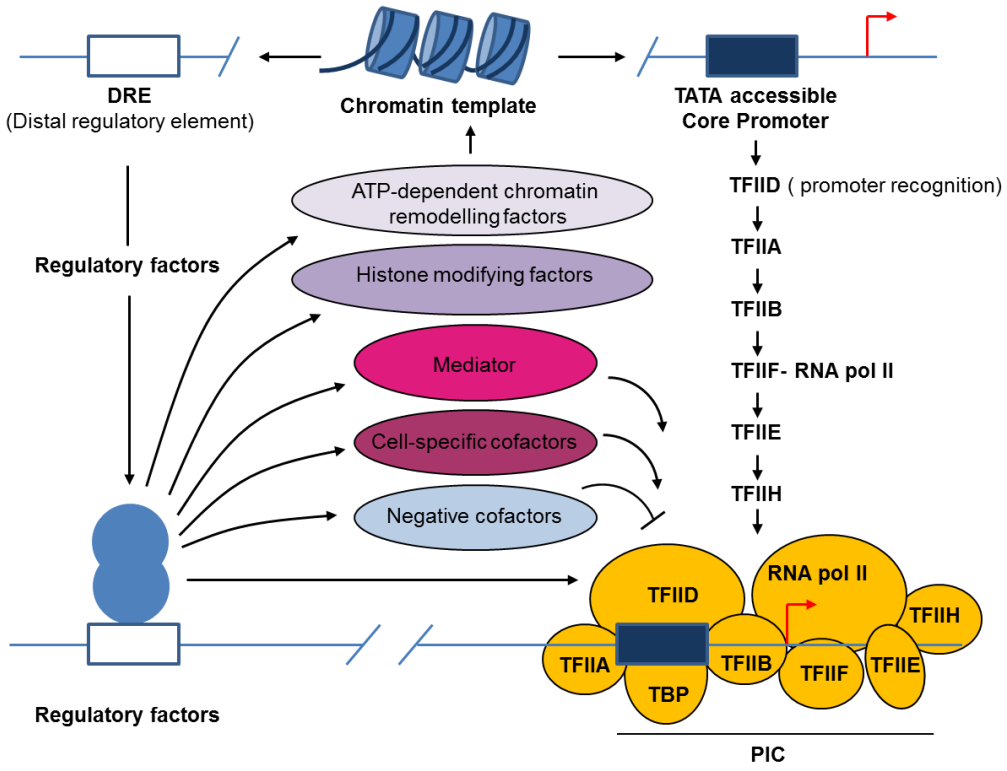


Figure 7: *The PIC assembly, in response to regulatory factors bound at distal regulatory elements, at promoters of RNA Pol II –transcribed genes.* The assembly of the PIC harboring RNA Pol II and general initiation factors (mustard) gets nucleated by the binding of TFIID to the TATA box of the core promoter (blue box). The regulation of PIC assembly includes cues from regulatory factors that bind to distal regulatory elements (white box), regulatory factor

interactions with chromatin modifying enzymes that remodel chromatin for the recruitment of additional factor interactions (Mediator) and cell type-specific co factors. Modified from (Roeder, 2005).

The regulatory factors and co-factors, such as the Mediator complex, interfere with negative cofactors and packaged chromatin to enable the PIC formation. Such a communication triggers the recruitment of chromatin remodeling complexes and histone modifiers that render the chromatin accessible for the binding of the multitude of peptides required for the PIC assembly (Figure 7 and 8A).

Promoter-proximal pausing

Following the PIC assembly, the next step is transcription initiation. In many cases, especially for genes that need to be rapidly induced in response to a stimulus, the PIC formation is followed by synthesis of very short mRNA fragments and subsequent polymerase pausing (Figure 8C). During pausing the RNA Pol II is stalled at the promoter (promoter-proximal pausing) and is not able to proceed to transcriptional elongation. This step is another mode of transcriptional regulation commonly found in metazoans. The first observations of this phenomenon arose from studies on single genes; the *Hsp70* gene in *Drosophila* (Gilmour and Lis, 1986; Rougvie and Lis, 1988) as well as the mammalian *c-myc* and *c-fos* genes (Fort et al., 1987; Krumm et al., 1992). With the employment of genome-wide analyses it then became clear that many genes which are subject to developmental regulation (Guenther et al., 2007; Muse et al., 2007; Zeitlinger et al., 2007) share this feature. The pausing index, which is the ratio of promoter to gene body-bound RNA Pol II, greatly varies among different genes (Adelman and Lis, 2012). In mESCs the pausing indices range from 30% to 90% (Min et al., 2011; Rahl et al., 2010) but the proportion of genes with RNA Pol II pausing is thought to be constant across species and developmental stages (Adelman and Lis, 2012).

Pausing involves the interaction between the polymerase and the pausing factors DSIF and NELF (Adelman and Lis, 2012; Sanso and Fisher, 2013) (Figure 8C). RNA pol II is regulated at a great extent by the phosphorylation status of the C-terminal domain (CTD) of its largest subunit.

The CTD contains several copies of a heptapeptide repeat with the consensus YSPTSPS (Corden et al., 1985) and it gets targeted by kinases for phosphorylation. Unphosphorylated

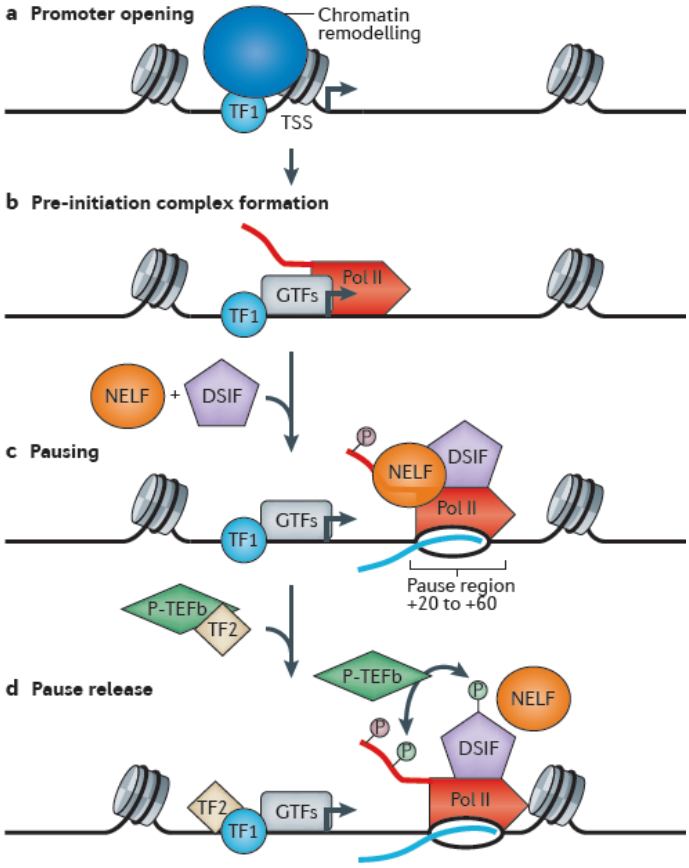


Figure 8: *Recruitment, pausing and release of RNA Pol II*. A) Promoter opening usually starts with the binding of a TF (TF1, blue circle) that recruits chromatin remodelers (dark blue oval) to make the DNA template accessible to the transcriptional machinery, by removal of nucleosomes around the TSS. B) Pre-initiation complex (PIC) formation involves the recruitment of general transcription factors (GTFs, grey oval) and the RNA Pol II (red) to the TSS. C) Pausing of the RNA Pol II occurs shortly after transcription initiation and the synthesis of a short mRNA molecule. It involves the recruitment of the pausing DSIF/NELF (orange and violet respectively) complex and the phosphorylation (pink circle) of the CTD of the polymerase. D) Pause release requires the recruitment of the P-TEFb (green) which phosphorylates both the CTD of RNA Pol II and DSIF (green circle), leading to the dissociation of the DSIF/NELF complex. Once phosphorylated DSIF gets transformed into a positive elongation factor that remains bound to RNA Pol II throughout the gene body (Adelman and Lis, 2012).

CTD corresponds to PIC-associated RNA pol II (Chesnut et al., 1992; Kang and Dahmus, 1993; Laybourn and Dahmus, 1990; Lu et al., 1991) (Figure 8B). The phosphorylation of the CTD is usually the readout for the transition of the polymerase to a stable elongation complex (Dahmus, 1996). Phosphorylation on the serine at the position 5 (S5P) by the Cdk7 kinase of the TFIIH is indicative of transcription initiation whereas phosphorylation at the position 2 (S2P) by the Cdk9 of the positive transcription elongation factor (P-TEFb) (Czudnochowski et al., 2012; Peterlin and Price, 2006) is required for the removal, by phosphorylation, of the DSIF/NELF complex which leads to pause-release and the transition to elongation (Marshall and Price, 1992, 1995; Wada et al., 1998) (Figure 8D).

Functionality of paused RNA Pol II

Paused RNA Pol II at TSSs has important regulatory functions including: a) promoter clearance by nucleosome displacement that allows the subsequent recruitment of additional factors required for gene expression (Workman, 2006), b) the accessibility of promoters to activators for rapid activation especially in the cases of genes which are responsive to stimuli and c) the integration of multiple regulatory signals that are needed for the combinatorial control of expression levels (Adelman and Lis, 2012).

Although RNA polymerases and GTFs have the ability to successfully transcribe DNA from promoter elements *in vitro*, this ability is hindered *in vivo* due to the tight packaging of the DNA to chromatin and the presence of negative factors (Roeder, 2005). This observation highlights the importance of the local remodeling of chromatin by nucleosome displacement around the gene promoter. For genes that are subject to promoter-proximal pausing, the nucleosome displacement happens before and independently of active transcription (Costlow and Lis, 1984; Wu, 1980) which is suggested to facilitate the binding of regulatory TFs in a stimulus-responsive manner (Lee et al., 1992; Shopland et al., 1995). In these genes, there is competition for nucleosome or RNA Pol II binding and the DNA sequence appears to be a critical factor that favors or not the nucleosomal presence (Adelman and Lis, 2012). Interestingly, gene promoters highly abundant in CpG islands are more enriched in paused RNA Pol II (Core et al., 2008) and the surrounding chromatin is usually in an open conformation (Jones, 2012). CpG-rich gene promoters are often observed in developmentally regulated genes which are also targeted by PcG proteins and are poised for expression upon differentiation. Co-occupancy by poised polymerase and PcGs at gene promoters is a mammalian-specific event which has probably emerged as an additional regulatory step.

The presence of paused RNA Pol II with an already established PIC at gene promoters enables the rapid transition to the elongation phase (Yudkovsky et al., 2000) upon gene activation and facilitates TF and co-factor binding. The need for quick switch to productive elongation, provided

by RNA Pol II pausing, is not a sole property for inducible genes; it is also observed at promoters of constitutively expressed genes with roles in signal transduction pathways in *Drosophila* and mESCs (Gilchrist et al., 2012) and is thought to fine-tune responsiveness to external cues (Adelman and Lis, 2012). At last but not least pausing is an event that allows the integration of synchronous regulatory signals, required for the combinatorial control of expression levels (Blau et al., 1996; Nechaev and Adelman, 2008). Such control is accomplished by the parallel binding of several TFs some of which recruit co-activators or GTFs while others establish pausing (Adelman and Lis, 2012; Blau et al., 1996).

Transcriptional co-activators

Apart from the basal transcriptional machinery (RNA Pol II and the GTFs) and gene-specific TFs, gene activation often requires the presence of additional co-factors. One type of co-factors is the transcriptional coactivators which are often found as multisubunit complexes and convey the function of DNA-bound transcriptional activators to drive gene expression (Bonnet et al., 2014). The features that facilitate the function of coactivators in transcriptional activation include the multitude of contact surfaces with TFs, provided by their modular nature, the activator-specific distinct conformations that they acquire which are then sensed accordingly by the basal transcriptional machinery as well as their enzymatic activities that alter the chromatin conformation (Fong et al., 2012). Coactivators can be classified in three families according to the mode of their function: 1) histone modifiers, such as the acetyltransferase complexes p300 or CBP (CREB Binding Protein) that loosen the chromatin around promoters while functioning as adaptor molecules for TF binding (Hermanson et al., 2002), 2) members of the TRAP/DRIP/Mediator/ARC complex which bind TFs, recruit the RNA Pol II but also interact with the transcriptional apparatus and 3) members of the SWI/SNF family of complexes which remodel chromatin in an ATP-dependent manner (Spiegelman and Heinrich, 2004).

The Mediator coactivator complex

Mediator is a multi-subunit assemblage, exclusively found in eukaryotes (Boube et al., 2002; Bourbon, 2008; Sato et al., 2004) which functions as a critical co-regulator of RNA Pol II transcription. The Mediator complex promotes activation of RNA pol II transcription via direct interactions with both DNA binding TFs and the PIC (Balamotis et al., 2009; Malik and Roeder, 2005, 2010; Myers and Kornberg, 2000). Mediator is also implicated in post-initiation stages of RNA Pol II transcription mostly by: a) bypassing or overcoming the activities of factors that negatively regulate elongation (Cheng et al., 2012; Jishage et al., 2012; Malik et al., 2007), b) recruiting Pol II transcription elongation factors and pre-mRNA processing factors (Donner et al., 2010; Huang et al., 2012; Mukundan and Ansari, 2011; Takahashi et al., 2011) and c) controlling phosphorylation of the Pol II CTD (Boeing et al., 2010; Donner et al., 2010; Jiang et al., 1998; Takahashi et al., 2011).

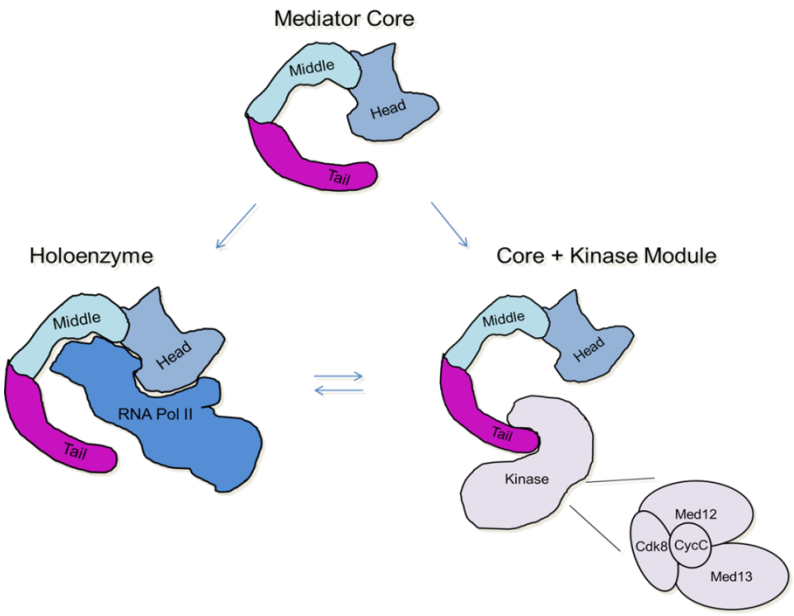


Figure 9: *Multiple forms of the Mediator complex that is organized in a modular fashion*. The Mediator Core represents the minimal complex including the Head, Middle and Tail modules. The Mediator Core can be found in complex with the RNA Pol II forming the Holoenzyme complex or with the Kinase Module harboring Cdk8, CycC, Med12 and Med13. The subunit composition of Mediator is interchangeable providing different modes of transcriptional regulation. Modified from (Conaway and Conaway, 2013)

Mediator can be found in many different forms that self-organize in modules. The minimal complex, named Mediator Core is composed of over 20 different proteins that group in three modules: the head, middle and tail (Conaway and Conaway, 2013) (Figure 9). Additionally, facultative modules may interact with the Mediator Core, such as the “Kinase Module”. The latter

is a comparatively small module consisting of Cdk8 kinase, Cyclin C, Med12 and Med13 proteins. The presence of facultative modules suggests that the Mediator complex may exert different functions. Early studies proposed a negative effect in transcription after the addition of the Kinase Module to the Mediator Core (Green and Johnson, 2004; Holstege et al., 1998; Samuelsen et al., 2003; Song et al., 1996; Wahi and Johnson, 1995). Meanwhile, further studies revealed the components of the Kinase Module to be involved in both activation and repression of transcription, illustrating the complexity of the issue.

The kinase activity of Cdk8 is required for the transcriptional activation of various genes (Belakavadi and Fondell, 2010; Donner et al., 2007, 2010) and indirectly activates RNA Pol II transcription by β-catenin (Morris et al., 2008). Med12 and Med13 have also be shown to interact and/or be required for activation by various TFs including β-catenin (Kim et al., 2006), Nanog (Tutter et al., 2009) and members of the *Gata* and *Runx* families (Gobert et al., 2010). Med12 is also part of an ES-specific interaction network harboring ES cell regulators and cohesin that maintains pluripotency (Apostolou et al., 2013; Kagey et al., 2010) and can enhance chromatin architecture via associating with eRNAs to activate gene expression (Lai et al., 2013). Interestingly this protein was also found to have a Polycomb-like phenotype in a genetic screen in *Drosophila* (Gaytan de Ayala Alonso et al., 2007). More specifically, the fly homologues for Med12 and Med13 were required to repress the classic PcG target *Ubx* in imaginal discs and individuals mutant for these genes failed to silence PcG target genes.

Cis-regulatory elements in transcriptional activation

Cis-regulatory elements (CREs) are DNA sequences (usually non-coding) that provide binding sites for TFs or other factors (originating from *trans*-regulatory elements) to activate gene expression (Ong and Corces, 2011). Their central role in transcriptional regulation is highlighted by the observation that mutations within these elements generate stronger phenotypes in comparison to mutations affecting the expression of TFs (Carroll, 2008; Stern and Orgogozo, 2008). Amongst the best characterized CREs are promoters and enhancers (Bulger and

Groudine, 2011; Levine, 2010). Although promoter mutations often cause human disease (Savinkova et al., 2009), they do not represent the primary drivers of *cis*-regulatory divergence (Brown and Feder, 2005) as, alone, they can produce only basal levels of mRNA transcripts (Wittkopp and Kalay, 2012). Enhancers on the other hand are the primarily responsible CREs for *cis*-regulatory divergence (Wray, 2007).

Enhancers

Enhancers are DNA regulatory elements, essential for transcriptional activation that do not exhibit specific sequence composition. They can drive gene expression in an orientation-independent manner and their distance in respect to their target promoter can vary from few kb to 1Mb (Erokhin et al., 2015). Enhancers can be found in intragenic or intergenic regions but also within gene introns or exons (Birnbaum et al., 2012; Ritter et al., 2012). They provide binding sites for TFs and they are characterized by low nucleosome density which is facilitated by the presence of histone variants such as H2A.Z or H3.3 (Barski et al., 2007; Jin et al., 2009; Wang et al., 2008). Once TFs are bound to enhancers, transcriptional activation is initiated by the recruitment of coactivators, release of RNA Pol II pausing and progression to elongation. An intriguing feature of enhancers is their restricted activity within distinct domains, controlled by the binding of specific TFs and the presence of insulators (Chetverina et al., 2014; Core and Lis, 2009; Gorkin et al., 2014; Kyrchanova and Georgiev, 2014) and Topologically Associated Domain (TAD) boundaries (Ciabrelli and Cavalli, 2015; Schwarzer and Spitz, 2014). Although the current knowledge about the relationship between histone modifications and enhancer activity is limited, the presence of H3K4me1/ H3K27me3 is usually associated with poised enhancers whereas H3K4me1/ H3K27ac with active ones (Figure 10B).

Enhancer-dependent transcriptional activation is often driven by the acquisition of distinct chromatin architectures such as DNA looping that enables functional contacts between enhancers and target promoters (Figure 10A). These contacts are usually established prior to

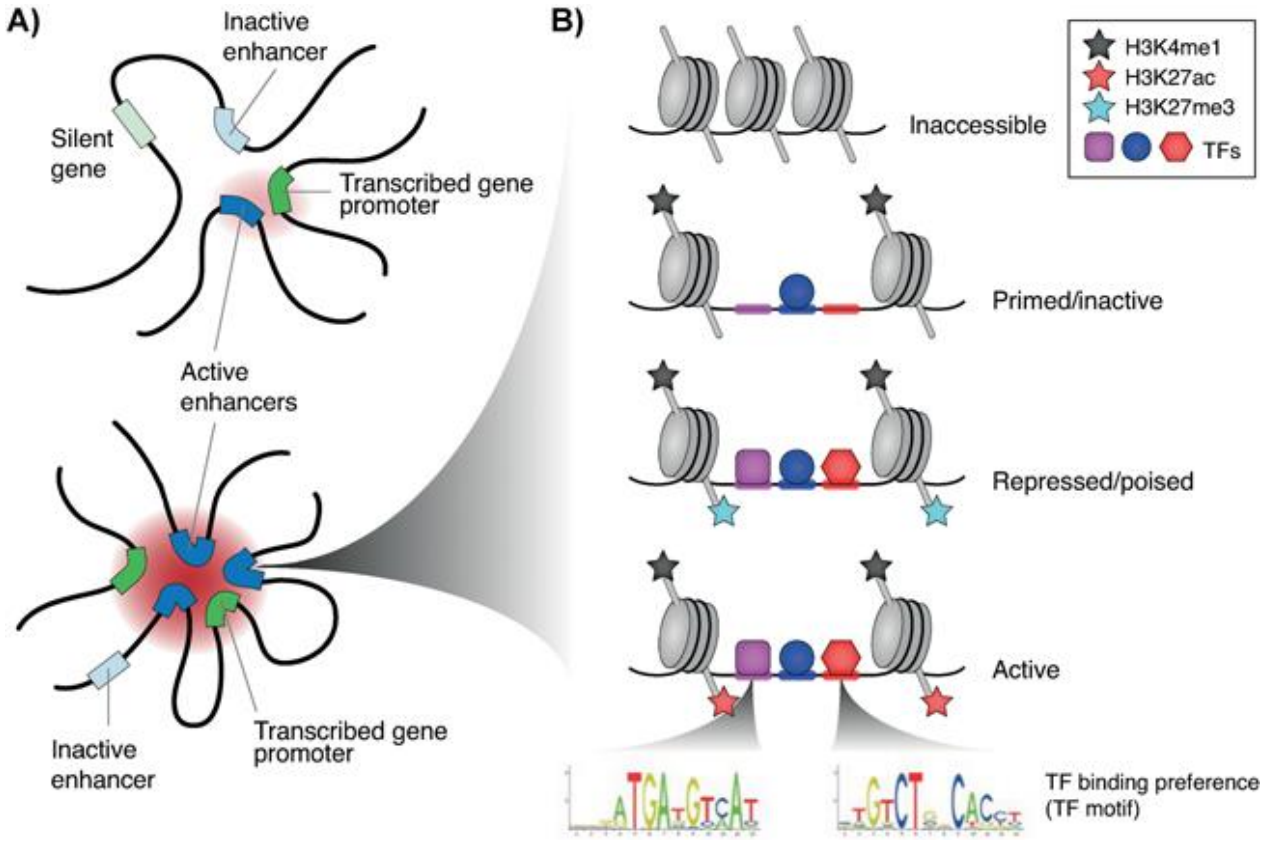


Figure 10: *Regulation of enhancer activity and enhancer-mediated regulation of gene expression*. A) Schematic representations of single enhancer-promoter architecture (upper panel) and a multi-player regulatory architecture involving many active enhancers and promoters (lower panel). B) Epigenetic landscape of different enhancer states. Inaccessible enhancers have increased chromatin compaction which hinders TF binding resulting in decreased activity. Primed enhancers are still inactive although the local chromatin is open and marked with H3K4me1.At this state enhancers might also be bound by TFs. Repressed/ poised enhancers have open chromatin, decorated by both PcG marks (H3K27me3) and active marks (H3K4me1) and are targeted by TFs. Active enhancers have open chromatin (H3K4me1 and H3K27ac) and bound TFs (Andersson, 2015).

interactions (Erokhin et al., 2015). Often enhancer-promoter loops are controlled or stabilized by protein complexes such as the Mediator and cohesin complexes which act in a synergistic manner to drive gene expression (Kagey et al., 2010) (Figure 11). Mediator conveys the activating signals from enhancers to promoters via interacting with both TFs and the basal transcriptional machinery. The regulatory role of Mediator on this respect is particularly important in mESCs. In pluripotency Mediator is co-bound with the master regulators Oct4, Sox2, Nanog, Klf4 and Esrrb in large enhancer domains, called super-enhancers, thereby controlling the maintenance of this state (Kagey et al., 2010). PRC1 is an additional regulator of the mESC

genome architecture by organizing an interaction network encompassing repressed developmental regulator genes and poised enhancers (Schoenfelder et al., 2015). Although there is no direct evidence about a possible communication between the governing interaction networks of pluripotency, those including active and repressed genes, their centrality in regulating ESCs makes it an a rather attractive scenario (Cavalli, 2015).

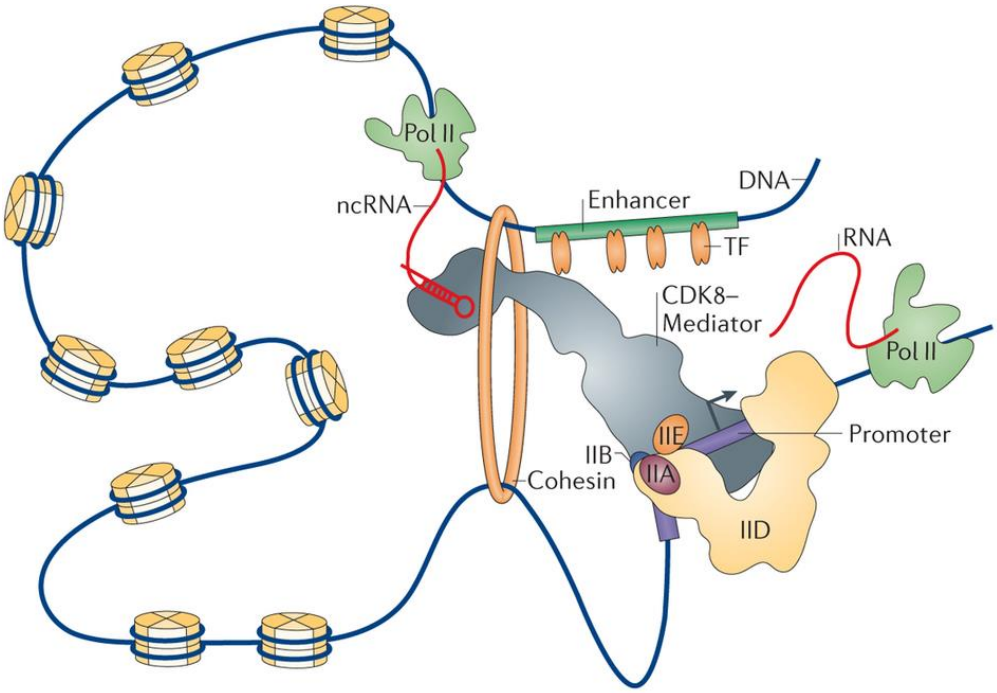


Figure 11: *Mediator, harboring the Kinase module, regulates chromatin architecture by interacting with TFs, cohesin and ncRNAs*(Allen and Taatjes, 2015). The association between ncRNAs and the Kinase module is thought to promote the binding of the latter to Core Mediator, which would release RNA Pol II to the gene body, required for productive elongation

With the majority of the genome being transcribed, it is of no surprise that enhancers can also produce ncRNA transcripts. The first described case of enhancer transcription was for the human beta-globin locus, where an ncRNA could only get synthesized when the enhancer was active (Tuan et al., 1992). A novel class of lncRNAs originating from enhancers, eRNAs, are lately being described and appear to have functional roles in the control of transcriptional activation (Erokhin et al., 2015). Interestingly, one type of eRNAs, ncRNA-a (ncRNA-activating), associate with Mediator harboring the Kinase module and via chromatin looping activate the transcription of target genes *in cis* (Lai et al., 2013) .

Genome organization

As many other cellular functions, gene expression is as well influenced by the topological organization of the genome within the nucleus (Vietri Rudan and Hadjur, 2015). During interphase, loci are distributed in a non-random fashion within the chromosomes which organize in diverse territories within the nucleus (Bickmore and van Steensel, 2013; Cremer et al., 2006). While active loci tend to be localized in gene-dense regions, those being inactive are mostly found within gene-poor regions (Caron et al., 2001; Croft et al., 1999). On a different level, individual genes are regulated by local chromatin conformations that physically bridge them with their respective enhancers (Lettice et al., 2003; Ptashne, 1986; Vietri Rudan et al., 2015). Architectural proteins such as CTCF and cohesin are key drivers of genome organization that define chromatin domains via coordinating long-range interactions (Merkenschlager and Odom, 2013) (Figure 12).

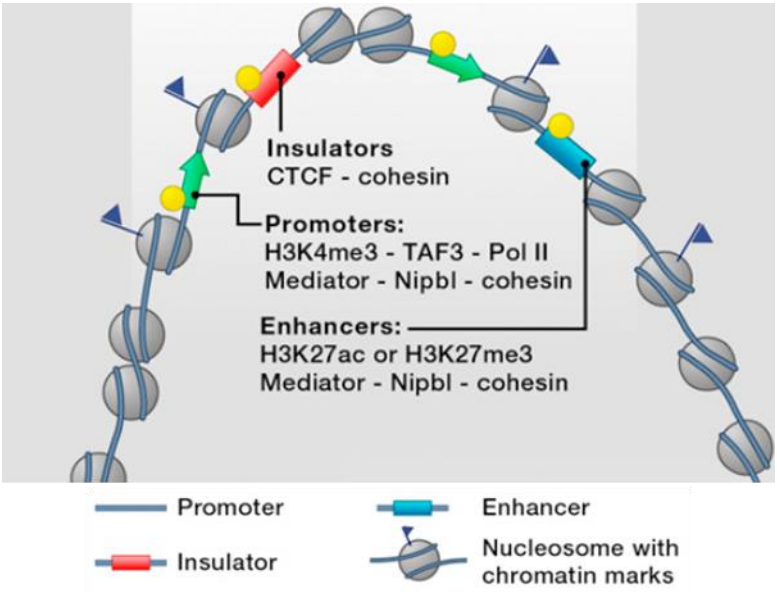


Figure 12: *CTCF and cohesin regulate complex loci*. Insulators regulate gene expression via blocking enhancer-promoter contacts. Modified from (Merkenschlager and Odom, 2013)

CTCF, which was first identified as a regulator of the c-*myc* oncogene (Baniahmad et al., 1990; Filippova et al., 1996; Lobanenkov et al., 1990) binds DNA via its 11 zinc fingers and is considered as the unique component to establish vertebrate insulators (Bell et al., 1999). The

binding of CTCF at thousands of sites within the chromosome arms of the vertebrate genome (Kim et al., 2007; Xie et al., 2007) has raised the possibility that it could act as a global insulator that segregates the genome in expression domains (Vietri Rudan and Hadjur, 2015) by organizing intra-chromosomal as well and inter-chromosomal developmentally regulated contacts (Phillips and Corces, 2009). CTCF is required for the loading of cohesin at enhancer elements (Parelho et al., 2008; Wendt et al., 2008) and the presence of both appears important for defining interactions between enhancers but have no (Seitan et al., 2013) or very limited (Sofueva et al., 2013; Zuin et al., 2014) effect in the maintenance of chromosome compartmentalization. Indeed the proposed combinatorial function of CTCF and cohesin in non-replicative cells is to demarcate and punctuate enhancer landscapes with the latter regulating expression of genes residing near conventional enhancers without affecting the maintenance of enhancer elements (Ing-Simmons et al., 2015). In contrast, it has also been suggested that CTCF and cohesin might also affect genome organization and hence gene expression separately as defined by the position of genes affected after depletion of the one or the other factor (Zuin et al., 2014). Remarkably, cohesin depletion can cause gene deregulation independently of its function in the maintenance of enhancer-promoter interactions in ES cells, at least for the pluripotency genes *Klf4* and *Lefty1* (Lavagnolli et al., 2015).

AIM OF THE STUDY

Given the complexity of mechanisms regulating gene expression during pluripotency and differentiation as well as the remaining open questions in respect to PRC1-mediated silencing, we sought to investigate the following:

- The possibility of a connection between PRC1 and the Mediator complex due to their common phenotypes in developing *Drosophila* embryos
- The molecular mechanism by which the H2AK119ub- interacting protein, Zrf1, leads to the transcriptional activation of PRC1-repressed genes during early ESC differentiation

MATERIALS & METHODS

Cell culture

mESCs

E14Tg2A ES cells were cultured in feeder-free conditions in cell culture dishes coated with 0.1% gelatin. Gelatin solution was produced by overnight stirring of powder gelatin (Sigma) in MQ water at 75°C, followed by sterile filtering. Before plating of the cells, dishes were coated with the required volume of gelatin solution, enough to cover the dish surface (5 ml in 10 cm dishes), and incubated for at least 30 min at 37 °C in Forma™ Steri-Cult™ CO_2 incubator (Thermo Scientific). Before plating of the cells, excess gelatin was removed from dishes and fresh medium was immediately added.

Cells were cultured in Glasgow Minimum Essential Medium (Sigma) supplemented with 15% PanSera ES Bovine Serum (PAN-Biotech), Non-Essential Amino Acids (Gibco), Sodium Pyruvate (Gibco), L-Glutamine (Gibco), Penicillin/Streptomycin (Gibco), β-mercaptoethanol (Sigma) and Leukemia Inhibitory Factor (LIF). For *in vitro* differentiation, cells were cultured in medium without LIF, in the presence of 10-6 M all-trans-Retinoic acid (ATRA) (Sigma). Cells were passaged every second day and their medium was exchanged on a daily basis. Depending on the cell line, the passage ratio varied from 1/8 to 1/4.

LIF was either purchased from Millipore or made in the lab by transfecting HEK293T cells with LIF plasmid using PEI as a transfection reagent. 24 hrs post-transfection, DMEM was replaced with stem cell medium without LIF. 3 days post-transfection, medium containing LIF was harvested, sterile filtered and stored at – 80 °C in 1 mL aliquots. The working concentration of LIF was estimated by titration. Cells were cultured for at least 15 days in different concentrations of LIF, using as a reference medium containing purchased LIF. Cells were harvested and RNA was extracted for cDNA synthesis and qPCR analysis. The relative levels of pluripotency

markers were calculated as fold change to the commercial LIF – treated sample. The concentration of LIF in the sample that showed no changes in all marker genes tested, as compared to the control, was used for future stem cell cultures.

Freezing/thawing cells

Confluent cells were washed twice with 1x sterile PBS and incubated for 1min with 2ml (for a 10 cm dish) of 1x Trypsin (1/10 dilution in 1x PBS from 0.5% Trypsin EDTA stock – Gibco) in the incubator. Trypsinized cells were collected by adding 4ml of culture medium and transferred in 15 ml falcon tubes. Cells were pelleted by centrifugation for 4 min at 900 rpm at RT. Pellets were resuspended in 0,5 ml medium and transferred to cryotubes. Equal volume of freezing medium (60% GMEM / 20% serum/ 20% DMSO) was added dropwise to cells and the tube was inversed gently for mixing. Cells were allowed to freeze gradually in stereofoam boxes at -80 °C overnight and the following day were transferred at a -150 °C freezer for long term storage.

Cells frozen in cryovials were thawed by the addition of pre-warmed medium and gentle mixing by pipetting. Cell suspension was transferred to 15 ml tubes and cells were pelleted by centrifugation for 4 min at 900 rpm at RT. Medium was removed completely and cells were resuspended in 0.5 ml fresh medium and then plated in 10 cm gelatinized dishes containing 9.5 ml medium. The following day cells were washed with 1x PBS to remove dead cell debris and the medium was exchanged. Cells were recovered for at least 2 passages before being used for experiments.

HEK293T

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (Life Technologies), supplemented with 10% FBS Gold (PAA Laboratories), L-glutamine and Penicillin/Streptomycin in 15 cm dishes. They were passaged every 2nd to 3rd day by removing the medium and washing the cells once with 1 x PBS. After the addition of 3 ml 1x Trypsin and incubation at 37 °C, cells were collected by the addition of 7 ml fresh medium and centrifuged in 15 ml tubes at 1200 rpm for 4 min at RT. Cells were resuspended in fresh medium and plated in new dishes containing

18 ml medium at a ratio 1/5. For experiments, 5, 10 cm dishes could be acquired from 1 fully confluent 15 cm dish.

Transfections

siRNA Transfections in mESCs

ESCs were grown overnight in 6-well plates and transfection in suspension was carried out the next day using 5 µl Lipofectamine (Life Technologies) in 250 µl OPTIMEM and 100 pmol siRNA in 250 µl OPTIMEM. Cells were plated in 10 cm dishes in medium without antibiotics, and the lipofectamine- siRNA mixture was added dropwise to the cell suspension. 7 hours post-transfection, the medium was replaced with medium containing antibiotics and cells were collected 48 hours later for RNA extraction. RA was added to cells 24 hours prior to harvesting.

Calcium phosphate Transfections in HEK293T cells

Cells were seeded in 10 cm dishes and transfections were carried out the day post-seeding, when cells acquired a confluency of around 50-60%. For each transfection a master mix was prepared as described in Table 1. First VE H_2O was mixed with the plasmid and $CaCl_2$ was added dropwise to mixture. This mixture was added dropwise to a new tube containing the appropriate volume of 2x HBS by gently flicking the tube. After 10 min incubation of the mixture at RT, 1mL was added to the cells. 24 hours (hrs) post-transfection the medium was replaced with fresh DMEM and 48 hrs post-transfection the cells were collected by resuspension and pelleted by centrifugation at 1200 rpm for 4 min at RT. Cell pellets were washed once with 1x PBS and then stored at -20 °C or directly processed for experiments.

Table 1: Concentrations and volumes for HEK293T Calcium phosphate transfections

$CaCl_2$ 2M	Plasmid	VE H_2O	2x HBS
62 µL	20 µg	Up to 0.5 mL	0.5 mL

PEI Transfections in HEK293T cells

Cells were seeded in 15 cm dishes and transfected when they reached 80-90% confluency. 18 mL of fresh medium was added to each plate at least 3 hrs prior to transfection. 12 µg of plasmids were mixed with 2 mL OPTIMEM or DMEM without serum and polyethylenimine (PEI) reagent (stock concentration 1 mg/mL) was added at 1:3 ratios (1:3 DNA: PEI). Mixture was vortexed and incubated for 30 min at RT. After incubation the mixture was added dropwise to cells. Cells were harvested 48 hrs post-transfection by trypsinization and pelleted by centrifugation at 1200 rpm 4 min at RT. Pellets were washed once with 1x PBS and either processed further for experiments or stored at -20°C.

Generation of stable ESC Knockdown Lines

For the production of lentiviruses containing shRNA, HEK293T cells were grown overnight in 10 cm dishes and the next day the medium was replaced with OPTIMEM low serum medium (Gibco), two hours prior to transfection. For the transfection, 15 µg shRNA, 5 µg pMDLg/pRRe, 5 µg pRSV-Rev and 5 µg pMD2.G plasmids were mixed with 500 µl OPTIMEM and 43 µg lipofectamine 2000 (Life Technologies) with OPTIMEM. 5 min later plasmids were mixed with lipofectamine and 10 min later the mixture was added to the HEK293T cells. 9 hours post-transfection, the OPTIMEM was replaced with 6ml complete stem cell medium and the day after, the medium containing the viral particles was collected and passed through a 0,45 µm filter (Millipore). 2mL of filtered medium containing the viral particles was added for transduction to $2x10^5$ stem cells in a 6 well plate (pre-incubated with 8 µg/ml polybrene (Sigma) for 2 hours) and the plates were centrifuged for 90 min at 1200 rpm at room temperature. 2ml fresh fresh medium was added per well and cells were incubated overnight. The same procedure was repeated 3 more times after which the cells were left to recover for 48 hrs. Knock down cells were selected with 2 µg/mL puromycin (Sigma) for 3 days and then frozen or analyzed for RNA or protein expression. After thawing, cells were selected with up to 8 µg/mL puromycin.

Alkaline Phospatase Staining

10^4 cells were seeded in 24-well plates and grown for 5 days. The colonies were washed once with PBS and incubated in 4% PFA for 2 min at RT. The fixed colonies were washed once with TBSX buffer (2 mM KCl/ 150 mM NaCl/ 100 mM Tris-HCl pH 7.5/ 0.1 % Triton-X-100) and incubated for 30 min on a rocking platform in TBSX, followed by 10 min incubation in NTMT buffer (100 mM NaCl/ 100 mM Tris-HCl pH 9.5/ 1% Tween/ 5 mM $MgCl_2$). The colonies were stained in 500 µL NTMT buffer containing 0.5 µL NBT (Roche) and 1.75 µL BCIP (Roche). As soon as staining was detectable (~1 min later), colonies were washed 3 times with PBS and pictures were acquired using a DM2500 upright Leica microscope.

Site directed mutagenesis in Zrf1 SANT domain

Oligos for site-directed mutagenesis of the SANT domain of ZRF1 were designed and purchased from Sigma (Table 2).

Table 2: Oligo sequences for site-directed mutagenesis

Name	Sequence (5' to 3') FWR/REV
W554G	CCCTATACAGACTTCACCCCTGGGACAACAGAAGAACAGAAGCTT
	AAGCTTCTGTTCTTCTGTTGTCCCAGGGGTGAAGTCTGTATATGG
W554F	CCATATACAGACTTCACCCCTTTTACAACAGAAGAACAGAAGCTT
	AAGCTTCTGTTCTTCTGTTGTAAAAGGGGTGAAGTCTGTATATGG
L561G	GGACAACAGAAGAACAGAAGGGTTTGGAACAAGCTTTGAAAAC
	GTTTTCAAAGCTTGTTCCAAACCCTTCTGTTCTTCTGTTGTCC

The mutagenesis was carried out using the QuikChange II XL Site-directed Mutagenesis Kit (Stratagene). For each PCR reaction the following master mix was prepared:

5x buffer	5 µL
Plasmid template	10 ng
dNTP mix	1 µL
QuikSolution	3 µL
ddH_2O	36,5 µL

125 ng of each oligo was added to each master mix and 1 µL *Pfu* Ultra HF DNA polymerase (2.5 U/µL). Samples were used for PCR reactions under the following cycling parameters:

	95 °C	60 sec
18 x	95 °C	50 sec
	60 °C	50 sec
	68°C	7 min
	68°C	7 min
	10 °C	1 min

PCR products were incubated for 1hr at 37 °C with 1 µL DpnI (10U/ µL) for digestion of the parental supecoiled dsDNA. The digestion product was used to transform XL10-Gold Ultracompetent Cells. For each transformation 45 µL of competent cells was thawed on ice. 2 µL of β-mercaptoethanol was added to each cell aliquot and the mixture was incubated for 10 min on ice. 2 µL of digestion reaction was added to the mixture and after 30 min incubation on ice samples were heat-shocked at 42°C for 30 sec and then placed on ice for 2 min. 0.5 ml pre-warmed SOC outgrowth medium (New England Biolabs) were added to the mixture and bacteria were recovered for 2 hrs at 37°C while shaking at 235 rpm. During cell recovery Kanamycin containing agar plates were left to dry at RT and 250 µL was used for plating. The plates were placed at 37 °C overnight and the following day single colonies were picked and used to inoculate 5mL LB cutlures supplemented with Kanamycin. Cultures were grown overnight at 37°C on a shaker and the following day they were used for plasmid purification by minipreps (Qiagen). Plasmid concentrations were measured and a portion was send for whole plasmid sequencing to verify the mutagenesis.

Chromatin Association Assays

Cells were grown in 10 cm dishes until 85% confluent and RA was added for differentiation 24 or 48 hrs prior to harvesting. After harvesting, cell pellets were resuspended in 10 mL 1x PBS and incubated while shaking for 10 min with 270 µL 37% formaldehyde (1% final concentration) at RT. Crosslinking was quenched by adding 0,125 M Glycine, diluted in 1x PBS, and the reaction

was incubated for 5 min at RT while shaking. Crosslinked cells were pelleted by centrifugation at 4°C for 4 min at 1200 rpm. Pellets were washed twice with ice-cold 1 x PBS to remove crosslinking debris and pelleted again under the same centrifugation conditions. Pellets were resuspended in 400 µL Buffer A (100 mM HEPES pH 7.9/ 5 mM $MgCl_2$/ 10 mM KCl/ 300 mM Sucrose/ 10% Glycerol/ 1% NP-40) supplemented with complete protease inhibitors (Roche) and 1 mM DTT and incubated for 10 min on ice to disrupt the cell membranes. Nuclei were pelleted by centrifugation at 3000 rpm for 5 min at 4°C. Nuclear pellet was resuspended in 500 µL Hypotonic Solution (3 Mm EDTA/ 0.2 mM EGTA) supplemented with protease inhibitors and 1 mM DTT and incubated for 30 min on ice. Nuclear lysate was pelleted by centrifugation for 5 min at 3000 rpm at 4 °C and the incubation with Hypotonic Solution was repeated, this time for 15 min on ice. Chromatin was pelleted by centrifugation at maximum (max) speed (14.8K rpm) at 4°C for 10 min. The supernatant was removed completely and the chromatin pellet was resuspended in 70 µL 2x reducing laemmli buffer. Samples were sonicated for 12 cycles (30 sec ON/ 30 sec OFF) at high setting using a Diagenode Bioruptor. After sonication samples were briefly vortexed and boiled at 95 °C for 30 min to revert crosslinking. Samples were centrifuged for 10 min at max speed at RT. For chromatin binding analysis 7 to 10 µL of the samples were loaded on 12 % SDS- PAGE and subjected to Western Blot analysis.

Chromatin Associations after RNaseA treatment

Two 10 cm dishes were prepared for each condition (-/+ RNaseA treatment) and cells here harvested as usual. Pellets were resuspended in 1 mL, 1x PBS and pre-fixation was carried out by adding Formaldehyde at 1% final concentration and by incubation for 3 min at RT. Pre-fixation was quenched by the addition of 0.125 M Glycine and incubation for 5 min at RT. Pre-fixed cells were pelleted by centrifugation at 2000 rpm for 2 min at RT. Cells were permeabilized by adding 1mL CSK buffer (100 mM NaCl/ 300 mM sucrose/ 3 Mm $MgCl_2$/ 10 mM PIPES pH 6.8)/ 0,25% TritonX100 and by incubation for 2 min at RT. Permeabilized cells were centrifuged as before and pellets were resuspended in 1 mL, 1x PBS. Half of the samples were incubated

for 10 min at RT with 10 µL RNaseA (100 µg/ µL stock concentration, Qiagen). Tubes were gently flicked to avoid pellet formation. RNaseA treated and untreated samples were pelleted and then washed once with 1x PBS. After this step, the protocol was carried out as described above for chromatin association (including the crosslinking step).

Cell Fractionation Assays

These assays were carried out based on the protocol "Chromatin Isolation by small-scale biochemical Fractionation" (Wysocka et al., 2001). Cells were grown in 10 cm dishes and harvested as usual when 80-90% confluent. Pellets were washed once with 1x PBS and centrifuged at 1000 rpm for 2 min at RT. Washed pellets were either frozen at -20 °C or used directly to isolate chromatin. Pellets were resuspended in 0.4 mL of Buffer A (10 mM HEPES pH 7.9/ 10 mM KCl/ 1.5 mM $MgCl_2$/ 0.34 M Sucrose/ 10% Glycerol/ 1 mM DTT) supplemented with complete protease inhibitors and Triton X-100 was added to a final concentration of 0.1%. Cells were incubated for 8 min on ice and then centrifuged for 5 min at 4 °C, at 1300 g. Pellets were washed once with 0.5 mL Buffer A and after centrifugation, lysed by adding 0.25 ml Buffer B (3 mM EDTA/ 0.2 mM EGTA/ 1 mM DTT) supplemented with complete protease inhibitors. Lysis was carried out by incubation for 30 min on ice. Chromatin pellet was washed two times with Buffer B and pelleted by centrifugation for 10 min at max speed at 4 °C. The supernatant was completely removed and pellets were resuspended in 0.1 mL 2x reducing laemmli buffer and boiled for 10 min at 95 °C. After 10 min centrifugation at max speed RT, 5 to 10 µL were loaded on 12 % SDS-PAGE for chromatin analysis by Western blot.

RNA extraction, cDNA synthesis and RTqPCR

RNA was extracted by lysis with 1 mL Trizol Reagent (Ambion) for pellets acquired from 10 cm dishes. Lysis was carried out by proper mixing and homogenization by passing the lysate through a syringe with a 0.9 x 40 mm 20G needle. Homogenized lysate was transferred in phase lock Heavy 2 mL tubes and incubated for 5 min at RT. 0.2 mL Chloroform (Sigma) was added and samples were shacked vigorously for 15 sec and incubated at RT for 3 min. Samples were

centrifuged for 15 min at 4 °C, at 12.000 g and aqueous phase was transferred to a new tube. 0.5 mL analytical grade Isopropanol was added and RNA was precipitated by incubation at RT for 10 min followed by centrifugation at 4 °C, 12.000 g for 10 min. The RNA pellet was washed once with 75% Ethanol and centrifuged for 5 min at 4 °C at 7500 g. After Ethanol removal the pellet was resuspended in 60 µL Nuclease Free Water (NFW) and incubated for 5 min at 57 °C. RNA was re-precipitated to remove any Trizol traces by adding 6 µL NaOAc 3M and 1 mL Ethanol. Precipitation was carried out at -80 °C from 1 hr to overnight. Samples were centrifuged for 20 min at max speed at 4 °C and washed twice with 1 mL ice- cold 75% Ethanol. Pellets were air-dried and resuspended in 60 µL NFW. For RNA- sequencing, samples were purified with the Qiagen RNeasy Mini Columns after Trizol extraction.

2.5 µg of RNA was used for cDNA synthesis by using the First Strand cDNA Synthesis Kit (Fermentas) with random hexamer primers. cDNA was diluted with NFW and 2 µL were used for qPCR. For each qPCR reaction, 2 µL cDNA and 2 µL NFW were combined with the following master mix at a final volume of 10 µL per reaction:

Cyber Green Mix 2x (Life Technologies)	5 µL
Forward primer (10 µM)	0.5 µL
Reverse primer (10 µM)	0.5 µL

qPCRs, were carried out using a Viia7 cycler (Life Technologies) at normal settings and expression was analyzed by using the ddCt method. The oligos (Table 3) were designed to span exon-exon junctions of transcripts and the results were normalized by using Actin, GAPDH or 18S oligos as reference.

Table 3: RT PCR oligo sequences

Oligo Name	**Sequence (5' to 3') FWR/REV**
18S	GATCCATTGGAGGGCAAGTCT
	CCAAGATCCAACTACGAGCTTTTT
Gapdh	CCCTTAAGAGGGATGCTGCC
	TACGGCCAAATCCGTTCACA
Actin	ACGGCCAGGTCATCAC TATTG
	CAAGAAGGAAGGCTGGAAAA
Hand1	AAAGGGAGTTGCCTCAGC
	CTGGTCTCACTGGTTTAGCTC
Eomes	TGTGACGGCCTACCAAAACA
	AGCCGTGTACATGGAATCGT
Snai1	TGTGTGGAGTTCACCTTCCAG
	AGAGAGTCCCAGATGAGGGT
Meis1	GTTGGAATTAGAGAAGGTACACGA
	GCTGATCTTGTTACATCTTCACTG
Id2	CATCCCACTATCGTCAGCCTGC
	GGGAATTCAGATGCCTGCAAGGAC
Msx1	CCGATCTAGTTTCTCGGGGC
	TCTCGGCCATTTCTCAGTCG
Jun	TGGGCACATCACCACTACAC
	TCTGGCTATGCAGTTCAGCC
Krt8	GTCGAGGAGCAACATGGACA
	GCTTCGTCCACATCCTTCTTG
Lmx1a	AACCAGCGAGCCAAGATGAA
	CCCGCATTCCCACTACCATT
Igfbp3	TAAGAAGAAGCAGTGCCGCC
	TTTCCCCTTGGTGTCGTAGC
Fgfr2	CCCTGCGGAGACAGGTAAC
	ACGCGTTGTTATCCTCACCA
Oct4	GAGGAGTCCCAGGACATGAA
	AGATGGTGGTCTGGCTGAAC
Sox2	CTGCAGTACAACTCCATGACCAG
	GGACTTGACCACAGAGCCCAT
Nanog	AGGCTGATTTGGTTGGTGTC
	CCAGGAAGACCCACACTCAT
Klf4	GATGCAGTCACAAGTCCCCT
	CCTTCTTCCCCTCTTTGGCT
Fgf5	GAAAACTCCATGCAAGTGCCA
	ATCGCGGACGCATAGGTATT
Hoxd11	GAAAAAGCGCTGTCCCTACA

	CAGACGGTCCCTGTTCAGTT
Dlx1	AGTTTGCAGTTGCAGGCTTT
	ACTTGGAGCGTTTGTTCTGG
Dlx2	CCTCACCCAAACTCAGGTCAA
	TATCTCGCCGCTTTTCCACA
Itga6	GGAGCAGATTGTTTGTGAGCC
	TCAGGAAGTTCCCGTTTCTTTCT
Sp3	AGCGACAGGTGATTTGGCTT
	TGTAGCTGACAAAACCTCCCA
Sp9	GCAGCCCAAAACCTGCTATG
	CGATCTTGTTGCAGGTCGC
Scrn3	CTGGAGACGTCGGGGAAGTA
	GTCTGGATGTTCCCGGTCAA
Chn1	GGGACCTAACACCTTCCAGA
	GGACCATCTTGGAACACTGCT
Atf2	GCTTCACTGATAAAACACGGACC
	GATCCTCGTTGGTAAAACGCTG
Lnp	GAGAACGAGGTGCTCTGGAT
	CTGCTGACATATGAGGGCGT
Hoxd13	GGTGTACTGTGCCAAGGATCAG
	TTAAAGCCACATCCTGGAAAGG
Hoxd12	CTCTTGCCTGCGATCTTCACT
	GAATTCATTGACCAGGAATTCGTT
Hoxd10	TGCCTGGCTGAGGTTTCC
	TCAGACTTGATTTCCTCTTTGCTTT
Hoxd9	GCAGCAACTTGACCCAAACA
	GGTGTAGGGACAGCGCTTTTT
Hoxd8	TGTTTCCCTGGATGAGACCAC
	CCTCGATTCTCCTCTTCCTGG
Hoxd4	GAGGCTTTACCCCAGTGGGAAC
	ACTAAGCCACACAGCTCCCTGAA
Hoxd3	AATCCCGACAGAACTCCAAGC
	TGGGCTCTTGTCCTCACAGT
Hoxd1	CCCCAAGAAAAGCAAACTGTCC
	GCTCTGTCAGTTGCTTGGTG
Hotairm (linc1548)	GGGAGGGAAAGGAAGAGTTG
	CGAGTGGAGTGGAATGGAGT
Dlx1as **(Dinger et al., 2008)**	GCAGACAGAATTGGGTCGTT
	CTCAACTACCGCCTGCAAA
Rian	CGTTGCGTACACCTTGGTTG
	AGTGCCAATCCCATTGAGGG
Cdk8	CCTCCGACTATCAGCGTTCC

	GCTGAGTATCCCATGCTGCT
Ring1b	GCAGGAGCCGAAATGTCTCA
	ATTGCCTCCTGAGGTGTTCG
Zrf1 (Dnajc2)	CTGTTCGGTTGGCTAAGGAG
	CCGAGAAGTGATTCCAGCTC
Med12	AGGTTCTGAATCGCAAGGGG
	GCCCATCTTCCCCACCTAAG
Wnt8a	GAGAAAGGGAAGGATGCCAGA
	CACTTGCAGGTCCTTTTCGTG
Sox7	AGCAAGATGCTGGGAAAGTC
	TCTGCCTCATCCACATAGGGT
Hoxb4	CACTCCGCGTGCAAAGAGC
	TAATTGGGGTTTACCGTGCTCA
Hoxa4	GTCCTCGTCCTCGTTACTGC
	CAGGTGTCCAATCCTGGCAA
Lefty1	GATGAAGTGGGCCGAGAACT
	TGAAAGGCACATCCTTGGGG

Immunoprecipitations (IPs)

Preparing mESCs for IPs

Cells were grown for 48 hrs (for endogenous IPs, 4x15 cm dishes and for FLAG IPs, 2x15 cm dishes) and harvested by adding 3ml 1x Trypsin (Gibco). After trypsin was inactivated, cells were pelleted by centrifugation at 900 rpm for 4min at RT. Pellets were washed twice in 1x PBS and either stored at -20 °C or used for immunoprecipitations.

Preparing HEK293Ts for IPs

Cells were transfected with plasmids either using $CaCl_2$ or PEI as a transfection reagent. Transfected cells were harvested 48 hrs post-transfection by disruption of the cell monolayer using a serological pipette and the culture medium. Cells were pelleted by centrifugation at 1200 rpm for 4 min at RT, washed twice in 1x PBS and either stored at -20 °C or used directly for immunoprecipitations.

Preparation of cell extracts for IP

Pellets were resuspended in 5 mL Buffer A (10mM HEPES pH 7.9/ 10 mM KCl/ 1.5 mM $MgCl_2$) and homogenized in a pre-cooled glass dounce homogenizer. Lysate was incubated for 5 min on ice and nuclei were pelleted by centrifugation at 3000 rpm for 5min at 4 °C. Nuclear pellets were washed once with Buffer A and resuspended in 1,5 mL Lysis Buffer (20mM HEPES pH 7.5/ 150mM NaCl/ 2.5mM EGTA/ 2mM EDTA/ 0.1% Triton-X-100) supplemented with complete protease pnhibitors. The lysate was sonicated for 30 cycles (30'' ON/ 30''OFF) using a Bioruptor Plus (Diagenode). Lysates were centrifuged for 15 min at 4 °C, at max speed and clear supernatant was transferred in new tubes. Protein concentration was measured (Pierce Coomassie, Thermo Scientific) and samples were diluted with Lysis buffer to obtain the same protein concentation. 50 µL of lysate was kept as input and stored at -20 °C. IPs were carried out overnight by incubation on a rotating wheel at 4 °C.

Endogenous IP

For endogenous IPs, samples with the same volumes and concentrations were incubated overnight with 5 µg primary antibodies (Abs) and 1µg of isotype controls (IgG), on a rotating wheel at 4 °C. Following the overnight incubation, 30 µL of pre-washed (with Lysis Buffer) sepharose beads (A for rabbit Abs, G for mouse or goat Abs) were added to the IP reaction to capture immunocomplexes, by incubation for 2 hrs at 4 °C on a rotating wheel. Captured immunocomplexes on beads were washed 4 times with 1 mL Lysis buffer for 10 min at 4 °C on a rotating wheel and centrifuged for 1 min at 1000 rpm at 4 °C. After the washes, the lysis buffer was removed completely and 30 to 40 µL of 2x reducing laemmli buffer were added to beads. 50 µL 4x reducing laemmli were added to inputs and all samples were boiled for 10 min at 95 °C and then centrifuged for 10 min at RT at max speed. Samples were then loaded on 8, 9 or 10% SDS-PAGE to be analyzed by western blot.

FLAG IP

For IPs from either mESCs stably expressing Zrf1-FLAG or from transiently transfected HEK293Ts, samples with equal volume and concentrations, were incubated overnight at 4 °C on

a rotating wheel, with 50 µL of pre-washed a-FLAG M2 affinity gel (Sigma). Instead of IgG isotype control, these IPs were carried out with either wild type cells as control for mESCs or with cells transfected with a no-FLAG containing plasmid. After the overnight incubation all steps for washing and elution were carried out as described for endogenous IPs.

IPs followed by RNase Treatment

For IPs with RNaseA treatment, all steps until bead washing were carried out as described above. During the last wash, beads were split in two tubes and resuspended in 200 µL 100mM Tris pH 7.5/ 10mM $MgCl_2$. In one of the samples (+RNase), 140 µg RNaseA was added and all samples were incubated for 45 min at 37 °C at 400 rpm. Beads were washed again 4 times with Lysis Buffer and samples were prepared for Western Blot Analysis as usual.

Chromatin Immunoprecipitations (ChIP)

ChIP for all Abs

For all Abs except from FLAG, ChIPs were carried out as described in (Morey et al., 2012). For each experiment cells were seeded in 3, 15 cm dishes at a confluency of 80%. Dishes were washed twice with 1x PBS crosslinking was carried out by the addition of 2 mL formaldehyde solution (11% freshly added formaldehyde/ 0.1 M NaCl/ 1 mM EDTA/ 50 mM HEPES-KOH pH 7.6) in 20 mL of 1x PBS and by incubation for 10 min at RT. Crosslinking was quenched by the addition of Glycine at a 0.125 M final concentration and incubation of the dishes for 4 min on ice. Dishes were washed twice with 1x PBS ice-cold and then cells were scraped by the addition of 6 mL of SDS buffer (100mM NaCl/ 50Mm Tris-HCl pH 8/ 5mM EDTA/ 0.5% SDS) containing complete protease inhibitors, for all three dishes. Lysate was snap frozen in liquid nitrogen and either stored at -80 °C or thawed directly at RT. Lysate was centrigufed at RT, at 600 g for 6 min and resuspended in 1.3 mL of ChIP buffer (1 volume SDS Buffer, 0.5 volume Triton Dilution Buffer (100 mM Tris-HCl pH 8.6/ 100 mM NaCl/ 5 mM EDTA/ 5% Triton-X-100) supplemented with complete protease inhibitors. Lysate was sonicated for 15 cycles (30 sec ON/ 30 sec OFF) at high setting in a Diagenode Bioruptor. A maximum of 3 samples was sonicated at a time.

Lysates were transferred in 1.5 mL Eppendorf tubes and centrifuged for 20 min at 4 °C, at max speed to remove debris. Clear chromatin was transferred in new tubes and protein concentration was estimated by Bradford (Pierce Coomassie, Thermo Scientific). Because of the high concentration of SDS in the buffer, an aliquot of each sample was diluted 33 times prior to measurement. After dilution of the samples to acquire the same concentrations, 1 mg of protein was used for IPs with non-histone Abs and 0.1 mg for IPs with histone Abs. Triton Dilution Buffer, supplemented with complete protease inhibitors, was added to each IP to decrease the final concentration of SDS to < 0.1 %. The suggested amount of primary Abs were added to IPs and 1 μg of isotype controls (IgG) and the IPs were carried out overnight at 4 °C on a rotating wheel. 1 or 10% inputs were kept and stored at -20 °C. After overnight incubation, 40 μL of pre-washed protein A or G agarose beads/ salmon sperm DNA, were added to each IP and immunocomplexes were captured by incubation at 4 °C on a rotating wheel for 2 hrs. Beads were centrifuged for 4 min at 4 °C, at 1000 rpm and washes were performed with 1 mL of each of the buffer described in Table 4, by incubation at 4 °C for 10 min.

Table 4: ChIP Wash Buffers

Times	Buffer Name	Buffer Recipe
2x	Low Salt Wash Buffer	0.1 % SDS/ 1% Triton-X-100/ 2 mM EDTA/ 20 mM Tris-HCl pH 8.1/ 150 mM NaCl
1x	High Salt Wash Buffer	0.1 % SDS/ 1% Triton-X-100/ 2 mM EDTA/ 20 mM Tris-HCl pH 8.1/ 500 mM NaCl
1x	LiCl Wash Buffer	250 mM LiCl/ 1% IGEPAL/ 1% Na-deoxycholate/ 1 mM EDTA/ 10 mM Tris-HCl pH 8.1
1x	TE Buffer	10 mM Tris-HCl pH 8.1/ 1 mM EDTA

After the washes, immunoprecipitated material was eluted by the addition of 110 μL of freshly made Elution Buffer (1% SDS/ 100 mM $NaHCO_3$), to the beads. The same volume of Elution Buffer was also added to the input samples. All samples were incubated overnight (or at least for 7 hrs) at 65 °C, at 1000 rpm to revert the crosslinking. After crosslinking reversal, 1 μL RNaseA (100 mg/ mL stock, Qiagen) was added to the samples and they were incubated at 37 °C for 1hr. Proteins were degraded by the incubation of the samples with 20 μg of Proteinase K at 56

°C for 1hr. The DNA was purifired with phenol: chlorophorm: isoamyl alcohol in phase lock tubes and chloroform. The purified DNA was precipitated overnight at – 80 °C by adding 2 µL Glycoblue (Ambion), 10% of sample volume NaOAc 3M and 1 mL Ethanol. Precipitated DNA pellets were acquired by centrifugation at max speed at 4 °C for 20 min. Pellets were washed twice with cold 75% Ethanol and after they were air dried, resuspended in 200 µL NFW. 2 µL of the DNA was used for qPCR reactions. Reactions were carried out as described for the RT-qPCR, using oligos specifically designed to amplify promoter or gene body regions of genes.

ChIP for FLAG-tagged lines

ChIPs for the mESC line, stably expressing Zrf1-FLAG, were carried out as described in (Chen et al., 2013). Cells were grown in 3, 10 cm dishes at a 80% confluency. Crosslinking, quenching and cell scraping were carried out as for the normal ChIP. Crosslinked and washed samples were resuspended in 1.5 mL Lysis Buffer 1 (50 mM HEPES-KOH pH 7.6/ 140 mM NaCl/ 1 Mm EDTA/ 10% v/v Glycerol/ 0.5% NP-40/ 0.25% Triton X-100) supplemented with complete protease inhibitors and samples were incubated for 10 min at 4 °C on a rotating wheel. Nuclei were pelleted with centrifugation at 4 °C, at 1350 g for 5 min. Nuclei were resuspended in 1.5 mL Lysis Buffer 2 (10 mM Tris-HCl pH 8/ 200 Mm NaCl/ 1 mM EDTA/ 0.5 mM EGTA/), supplemented with complete protease inhibitors, and incubated for 10 min at 4 °C on a rotating wheel. Nuclei were pelleted as before and resuspended in 650 µL Lysis Buffer 3 (10 mM Tris-HCl pH 8/ 100 mM NaCl/ 1 mM EDTA/ 0.5 mM EGTA/ 0.1% Na-deoxycholate/ 0.5% N-lauroylsarcosine), supplemented with complete protease inhibitors. After incubation for 10 min at 4 °C on a rotating wheel, samples were transferred in 15 mL tubes and sonicated for 3 x 10 cycles (30 sec ON/ 30 sec OFF) at high setting. After each cycle the samples were incubated for 5 min on ice. Debris was removed by centrifugation at max speed at 4 °C for 15 min and clear supernatant was transferred in new tubes. Samples were diluted 3 fold in ChIP Buffer (20 mM Tris-HCl pH 8/ 150 mM NaCl/ 2 mM EDTA and Triton X-100 at a final concentration of 1% in the sample), supplemented with complete protease inhibitors. Protein G agarose Magna ChIP beads

(Millipore) were pre-blocked by incubation with PBS/ 5 mg/ml BSA for at least 1 hr at 4 °C on a rotating wheel. After blocking, beads were washed with PBS/ BSA and resuspended in ChIP buffer. 30 µL of washed beads were added per IP and 10 µL a-FLAG M2 Ab (Sigma) or 1 µL a-mouse IgG. IPs were carried out overnight at 4°C on a rotating wheel. 10% inputs were kept at -20 °C. After overnight incubation, bead-captured immunocomplexes were washed with the buffers described in Table 5, with 1 mL buffer and separatation on a magnetic rack.

Table 5: FLAG ChIP Wash Buffers

Times	Buffer Name	Buffer Recipe	Wash Duration
2x	ChIP Buffer	20 mM Tris-HCl pH 8/ 150 mM NaCl/ 2 mM EDTA	Instant
1x	ChIP Buffer 500 mM NaCl	20 mM Tris-HCl pH 8/ 2 mM EDTA/ 500 mM NaCl	Instant
4x	RIPA Buffer	10 mM Tris-HCl pH 8/ 0.25 M LiCl/ 1 mM EDTA/ 0.5 % NP-40/ 0.5% Na-deoxycholate	5 min
1x	TE Buffer	10 mM Tris-HCl pH 8.1/ 1 mM EDTA	5 min

Elution of the immunocomplexes from the magnetic beads was carried out by the addition of 100 µL 2x STOP Buffer (20 mM Tris-HCl pH 8/ 100 mM NaCl/ 20 mM EDTA/ 1% SDS) and incubation of samples in a thermomixer at 65 °C for 15 min. Elution was carried out two times and eluates were combined in new tubes. 2x STOP Buffer was also added to input samples and all samples were incubated for at least 6 hrs at 65 °C, at 1000 rpm. After reversal of the crosslinking, 200 µL TE were added to each sample and DNA was purified and resuspended as described in the normal ChIP protocol. qPCRs were carried out under the same conditions, as for the normal ChIP protocol. Oligos used for all ChIP qPCRs are described in Table 6. The ChIP enrichments were calculated as recovery % over the inputs.

Table 6: Names and sequences of oligos used for ChIP qPCRs

Oligo Name	Sequence (5' to 3') FWR/REV
Eomes	GCTCACAAGTTTCCAAGCGG
	ACCTCTCAGGACCCCTACAC
Pax3	GCAGCTTCGCTCGCAAATTA
	AGATCCGGAGAGTTCCCGAG
Msx2	TGGCAACTCTCAAGCGAAGA
	CAGTTGGTTGAGCCGAGTCT
Fgf9	ATATCGCCCAGGCACTTACG
	GGTCTTCATCCCATCCGACC
Nanog	GGGTAGGGTAGGAGGCTTGA
	AGCCTTCCCACAGAAAGAGC
Hoxd11	AAATCCCTACCAGCCGGAAC
	ATTGCGACACTCTCTAGGGC
Meis1	CCGTGCGTGTGTAAAGTGTG
	GCATTGTGTAAGACGCGACC
Gli2	TGCAATCCATCAGCGTCTCT
	AGCTAAGGACGCGACGATCA
Snai1	AGTTGAAGACTCGAAGGCGG
	GAAGGTGAACTCTGCGGGAA

RNA- Chromatin Immunoprecipitation (RNA- ChIP)

RNA- ChIPs were carried out according to the protocol by Bryan K. Sun and Jeannie T. Lee (Protocol 28), as described in the Epigenesys network. Cells were seeded in 10 cm dishes at at 80-90% confluency. Cells were washed twice with 1x PBS, trypsinized and pelleted as described before. Cell pellets were resuspended in 10 mL 1x PBS and 37% formaldehyde was added at a final concentration of 1%. Crosslinking was carried out by incubation for 10 min at RT with occasional manual mixing. Reactions were quenched by the addition of Glycine at a 0.125 M final concentration and by incubation for 5 min at RT with occasional mixing. Cells were pelleted by centrifugation at 900 rpm for 4 min at RT and washed twice with 1x cold PBS supplemented with complete protease inhibitors. Washed pellets were transferred on ice, resuspended in 200 µL Buffer A (5 mM PIPES pH 8/ 85 mM KCl/ 0.5% NP40/ 50 U/mL RNase Inhibitors) supplemented with complete protease inhibitors and incubated on ice for 10 min. The crude

nuclear fraction was pelleted by centrifugation for 5 min at 5000 rpm, at 4°C. The pellet was washed once with Buffer A without NP40, resuspended in 0.5 mL Buffer B (1% SDS/ 10 mM EDTA/ 50 mM Tris-HCl pH 8/ 50 U/mL RNase Inhibitors) supplemented with complete protease inhibitors and incubated on ice for 10 min. Samples were sonicated for 15 cycles (30 sec ON/ 30 sec OFF) and then centrifuged for 15 min at 4°C, at max speed. Clear supernatant was transferred in new tubes and diluted 10 fold by adding IP Buffer (0.01% SDS/ 1.1% Triton-X-100/ 1.2 mM EDTA/ 16.7 mM Tris-HCl pH 8.1/ 167 mM NaCl/ 50 U/mL RNase Inhibitors), supplemented with complete protease inhibitors. 30 µL of prewashed protein A agarose beads/ salmon sperm DNA were added and samples were pre-cleared by incubation for 2 hrs at 4 °C on a rotating wheel. Samples were centrifuged for 5 min at 1000 rpm at 4 °C and supernatant was transferred in new tubes. IPs were carried out overnight at 4 °C, on a rotating wheel with the required amount of primary antibodies and isotype controls (IgG) and 1% inputs were stored at -80 °C. After the incubation, 30 µL pre-washed protein A agarose beads/ salmon sperm DNA were added to each IP and immunocomplexes were captured by incubation at 4 °C for 2 hrs on a rotating wheel. Beads were washed for 10 min with 1mL ice-cold buffers (Table 7) by rotation at 4 °C and centrifugation for 1 min at 1000 rpm at 4 °C.

Table 7: RNA-ChIP Wash Buffers

Times	Buffer Name	Buffer Recipe
1x	Low Salt Wash Buffer	0.1 % SDS/ 1% Triton-X-100/ 2 mM EDTA/ 20 mM Tris-HCl pH 8.1/ 150 mM NaCl
1x	High Salt Wash Buffer	0.1 % SDS/ 1% Triton-X-100/ 2 mM EDTA/ 20 mM Tris-HCl pH 8.1/ 500 mM NaCl
1x	LiCl Wash Buffer	250 mM LiCl/ 1% IGEPAL/ 1% Na-deoxycholate/ 1 mM EDTA/ 10 Mm Tris-HCl pH 8.1
2x	TE Buffer	10 mM Tris-HCl pH 8.1/ 1 mM EDTA

After the beads were washed, immune complexes were eluted by the addition of 250 µL Elution Buffer (1% SDS/ 100 mM $NaHCO_3$/ 50 U/mL RNase Inhibitors) to the beads and 15 min rotation at RT. Elution was carried out twice and supernatants were combined in a new Eppendorf tube. To revert crosslinking NaCl was added to all samples (including inputs) at a final concentration

of 200 mM and samples were incubated for 3 hrs at 65 °C, at 400 rpm. Proteins were digested by incubating the samples for 45 min at 42 °C, with 20 µg Proteinase K, 20 µL of 1 M Tris-HCl pH 6.5 and 10 µL of 0.5 M EDTA. Samples were purified by phenol:chlorophorm:isoamyl alcohol and precipitated as described in the ChIP experiments. Pellets were resuspended in NFW and incubated with Turbo DNaseI (Ambion) according to manufacturer's protocol. Samples were purified again and following precipitation, RNA pellet was resuspended in 10 µL NFW. All RNA was used for reverse transcription (as described before, using the Fermentas first strand cDNA synthesis Kit) and then diluted 4 times with NFW. 2µL were used for RT-qPCR reactions with oligos described in Table 3 and the enrichments were calculated as recovery over input.

RNA-Immunoprecipitations

For RNA Immunoprecipitations 6, 10 cm dishes of HEK293T cells were transfected with $CaCl_2$ as previously described. Harvested cells were pooled and centrifuged at 1200 rpm for 6 min at RT. Pellets were resuspended in 30 mL 1x PBS and crosslinked for 10 min at RT while shaking with 1% formaldehyde. Crosslinking was quenched for 5 min at RT while shaking, by 0.125 M Glycine. Samples were centrifuged for 4 min at 2000 rpm, at RT and pellets were washed twice with 1x PBS. Crosslinked pellets were either stored at -80 °C or used directly for IPs. IPs were carried out as described previously (FLAG IPs) with the difference that all buffers were supplemented with 50 U/mL of RNase Inhibitors. After washing of the FLAG bead-captured immune complexes, 75% of the beads were transferred to new tubes. The rest 25 % was boiled with 2x reducing laemmli buffer and analyzed by Western Blot to confirm immunoprecipitation. The 75% of beads were resuspended in 200 µL NFW and incubated with DNaseI at 37 °C for 30 min. The crosslinking was reverted by incubation at 65 °C, at 400 rpm for 3 hrs in the presence of 20 µg Proteinase K. The decrosslinked material was precipitated with Ethanol as previously described and RNA pellets were acquired by centrifugation at max speed, at 4 °C for 20 min. After pellets were washed twice with 75% Ethanol and air-dried, they were resuspended in 40 µL

of NFW. Samples were split in two tubes and 1 µL RNaseA (100 µg/µL stock) was added in one of the fractions. All samples were incubated at 37 °C for 1 hr and then loaded on a 1% agararose gel in TE and RNA was visualized with Sybr Safe.

Table 8: Sequences and code numbers of RNAi

Name	Company	Sequence sense	Sequence antisense
siRNA Dlx1as 1: Silencer Pre-designed siRNA	Ambion	GCACAAUGCAAUAACAAUATT	UAUUGUUAUUGCAUUGUGCTG
siRNA Dlx1as 2: Silencer Pre-designed siRNA	Ambion	GGACUAUUGGAAGAGGCUATT	UAGCCUCUUCCAAUAGUCCTA
MISSION® siRNA Universal Negative Control #1 SIC001	Sigma		
shRNAs pLKO.1			
shRing1b : SHCLNG-NM_011277 TRCN0000226018	Sigma		
shMed12: SHCLNG-NM_021521 TRCN0000376081	Sigma		
shCdk8: SHCLNG-NM_153599 TRCN0000023107	Sigma		
Control: Mission pLKO.1-puro Non-Mammalian shRNA control	Sigma		

Table 9: Antibody list

Name	Company	Code	Application
Med12	Bethyl	A300-774A	WB / IP
	Novus Biologicals	NB100-2357	WB/IP/ChIP/ RIP
	SANTA Cruz	sc-8998	WB
Cdk8	Santa Cruz	dc-1521	WB/IP/ChIP
	Cell Signaling	g398	WB
Dnajc2 (Zrf1)	Novus Biologicals	NBP2-12802	WB/IP
	Abcam	ab134572	WB/ IP
	Serum Homemade		WB
Med1	Santa Cruz	sc-8998	WB
Flag	Sigma	F1804	WB/ChIP/RIP
	Sigma	A2220	IP
Ring1b	Cell Signaling	5694	WB/IP/ChIP/RIP
H2AK119ub	Cell Signaling	D27C4	WB
ASH2L	Cell Signaling	D93F6	WB
MLL (H-10)	Santa Cruz	sc-374392	WB

Immunofluorescent staining

For immunofluorescent staining mESCs were seeded in 6-well tissue culture dishes, each containing a UV- treated and gelatinized coverslip. For differentiation induction, medium was changed 6 hours post-seeding, replaced with differentiation medium and RA was added at 10^{-6}M. A day post-cell seeding, the medium was removed and cells were washed twice with 1X PBS and then fixed for 10 min with 4%PFA. The fixed cells were washed twice with 1X PBS and incubated with the following concentrations of Triton X for permeabilization:

- 5 min in 0.1% Triton in 1X PBS
- 7 min in a 1:1 mixture of 0.1% and 0.5% Triton in 1X PBS
- 10 min in 0.5% Triton in 1X PBS

After permeabilization, the cells were washed twice with 1X PBS and blocked for 1 hour at RT in 1mL of 0.1% Triton/ 20% FBS in 1X PBS. The blocked cells were washed twice with 1X PBS and incubated over night at 4°C with primary antibodies diluted in 0.1% Triton/ 5% FBS (Table 10).

Table 10: Primary antibody dilutions for immunostaining

Antibody	Company	Dilution
Med12	Novus Biologicals	1:250
Cdk8	Santa Cruz	1:200
Ring1b	Serum	
ZRF1	Novus Biologicals	
Oct4	Abcam	
H3K4me1	Abcam	1:100
H3K27ac	Cell Signalling	

After overnight incubation, stained cells were washed thrice with PBS/ 5% BSA/ 0.1% Triton and incubated with secondary antibodies (1:1000 dilution) in PBS/5% BSA/ 0.1% Triton/ 5% FBS, one hour at RT in dark. Cells were washed 3 times for 10 min in dark with PBS/ 5% BSA/ 0.1% Triton, mounted with VECTASHIELD Antifade Mounting Medium with DAPI (Vector) and sealed.

Images were acquired with an SP5 Leica Confocal Microscope using the manufacturer's software and then processed with ImageJ.

Genome-wide approaches

RNA-sequencing

For RNA-sequencing, RNA was extracted as described above and samples were submitted to the IMB Genomics Facility for library preparation. The obtained reads varied between 65 and 94 million reads per sample and were mapped in iGenomes. The analysis of the sequencing data was carried out by Sergi Sayols from the IMB Bioinformatics Core Facility. Detailed information about the methods used for the analysis are described in Papadopoulou et al., 2016. The data of the RNA-seq analysis were deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) under the GEO (Gene Expression Omnibus) Series accession number GSE73352.

ChIP-sequencing

For the ChIP-sequencing analysis data were downloaded from publicly available datasets (GSE22562, GSE42466 and GSE44288) and were analyzed by Sergi Sayols from the IMB Bioinformatics Core Facility. Detailed information about the analysis can be found in Papadopoulou et al., 2016.

Mass Spectrometry

For the Mass Spectrometry analysis, the samples were prepared as described above (Section FLAG-IP). A fraction of the immunoprecipitated material was submitted to the Proteomics Core Facility for downstream analysis. The Core Facility provided us with tables that were filtered using the MaxQuant software for removal of common contaminants and reverse database entries as well as identification of proteins with at least two peptides.

Results

PART 1: Common functions of PRC1 and Med12-Mediator in pluripotent mESCs

PRC1 and Mediator occupy similar chromatin regions in mESCs

We sought to investigate the potential relationship between PRC1 and Mediator, based on their essential roles in stem cell regulation and the hypothesis that Mediator subunits might provide an interphase for PRC1-mediated silencing (Simon and Kingston, 2013). We started by assessing the global chromatin binding profiles of PRC1, PRC2 and Mediator subunits, via exploring publically available ChIP-seq datasets from mESCs (Boyer et al., 2006; Kagey et al., 2010; Morey et al., 2012, 2013; Whyte et al., 2013). By comparing the ChIP profiles of the PRC1 component Ring1b and the two Kinase Module subunints Med12 and Cdk8, we found 486 common target genes in mESCs. Interestingly, Med12 co-occupies many more genes with either Cdk8 (2131) or Ring1b (1670) corresponding to ~26% and ~20% of the Med12 targets respectively. This observation shows that Med12 is bound to distinct classes of genes in mESCs, those co-bound by Cdk8 or Ring1b (Figure 13A), suggesting that Med12 might have an additional role in mESCs, outside the Kinase Module, along with Ring1b. Ring1b and Cdk8 have very few common targets (177), suggesting that they are probably functionally unrelated. We then compared the common Med12-Ring1b targets with those of the Core Mediator subunit, Med1. 1339 of Med1 targets were co-bound by Med12-Ring1b which corresponds to 80% of Med12-Ring1b targets (Figure 13B). Hence, the majority of genes co-bound by Med12 and Ring1b in mESCs are also occupied by Core Mediator, as Med1 chromatin binding profiles imply.

In order to gain insight into the nature of the two identified classes of genes in mESCs, we performed Gene Ontology (GO) analysis of the common Med12-Ring1b and Med12-Cdk8 targets. The most significantly enriched GO terms for Med12-Ring1b targets are related to

differentiation and development (Figure 13C) whereas those for Med12-Cdk8 are related to metabolic processes (Figure 13D). This analysis further supports the idea that Med12 regulates distinct sets of genes along with either Ring1b or Cdk8.

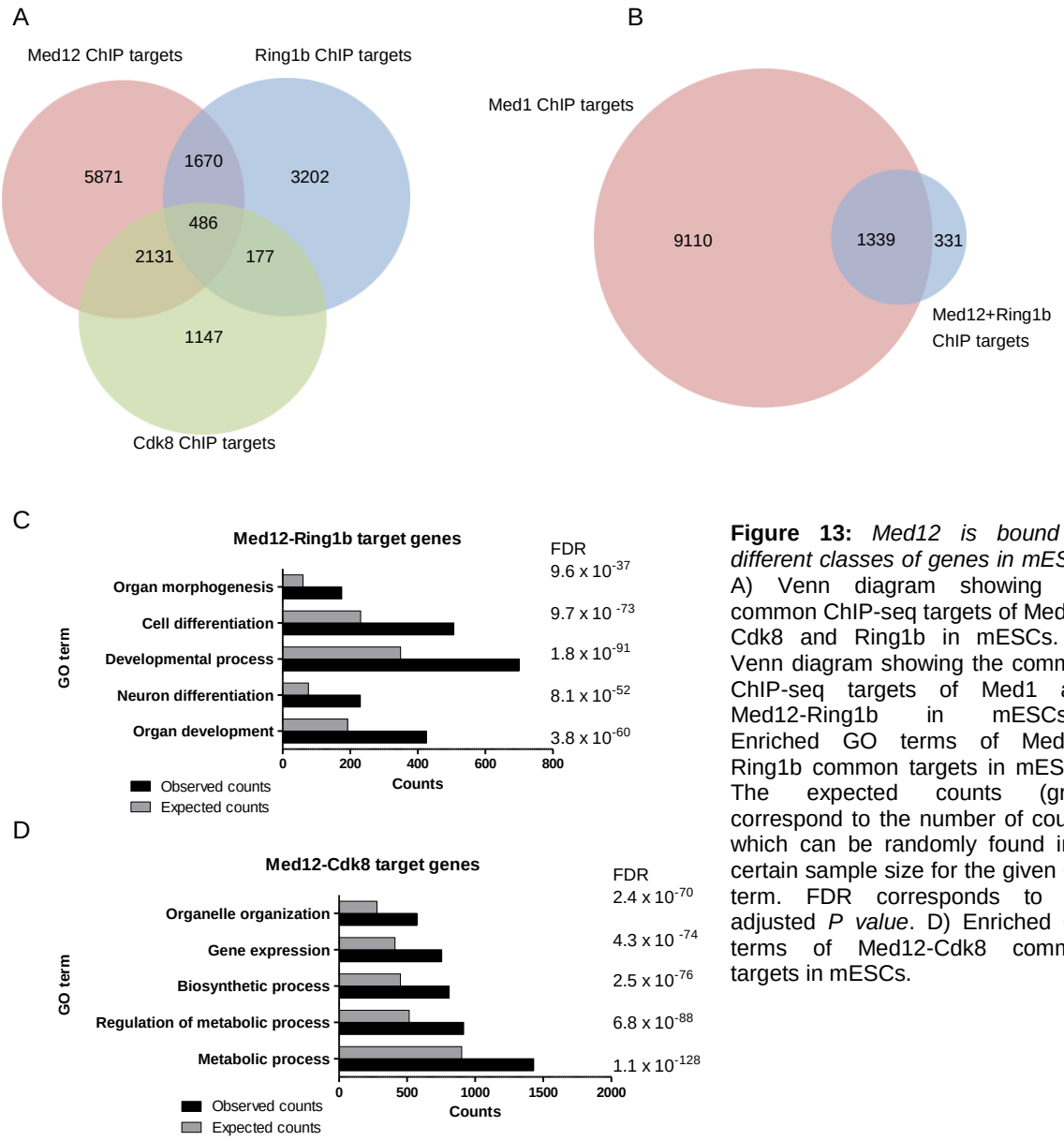


Figure 13: *Med12 is bound to different classes of genes in mESCs* A) Venn diagram showing the common ChIP-seq targets of Med12, Cdk8 and Ring1b in mESCs. B) Venn diagram showing the common ChIP-seq targets of Med1 and Med12-Ring1b in mESCs.C) Enriched GO terms of Med12-Ring1b common targets in mESCs. The expected counts (grey) correspond to the number of counts which can be randomly found in a certain sample size for the given GO term. FDR corresponds to the adjusted *P value*. D) Enriched GO terms of Med12-Cdk8 common targets in mESCs.

To substantiate this classification we then looked for the presence of additional subunits of PRCs on these two sets of genes. The promoters of Med12-Ring1b targets, such as *Msx2*, *Jun* and *Eomes*, are also enriched for the PRC1 subunit Cbx7 (Figure 14A, C) (677 genes) and the

PRC2 subunit Suz12 (Figure 14A, C) (1409 genes), suggesting the presence of a canonical PRC1 complex on those genes. Furthermore, genes co-bound by Med12-Ring1b are also enriched in additional PRC1 and PRC2 subunits, such as RYBP, Pch1, Eed as well as in their respective histone modifications H2AK119ub and H3K27me3 (Figure 14C). On the contrary, the promoters of Med12-Cdk8 targets, such as *Nanog*, *Gli2* and *Fzd5*, are not enriched in neither Cbx7, nor Suz12 (Figure 14B).

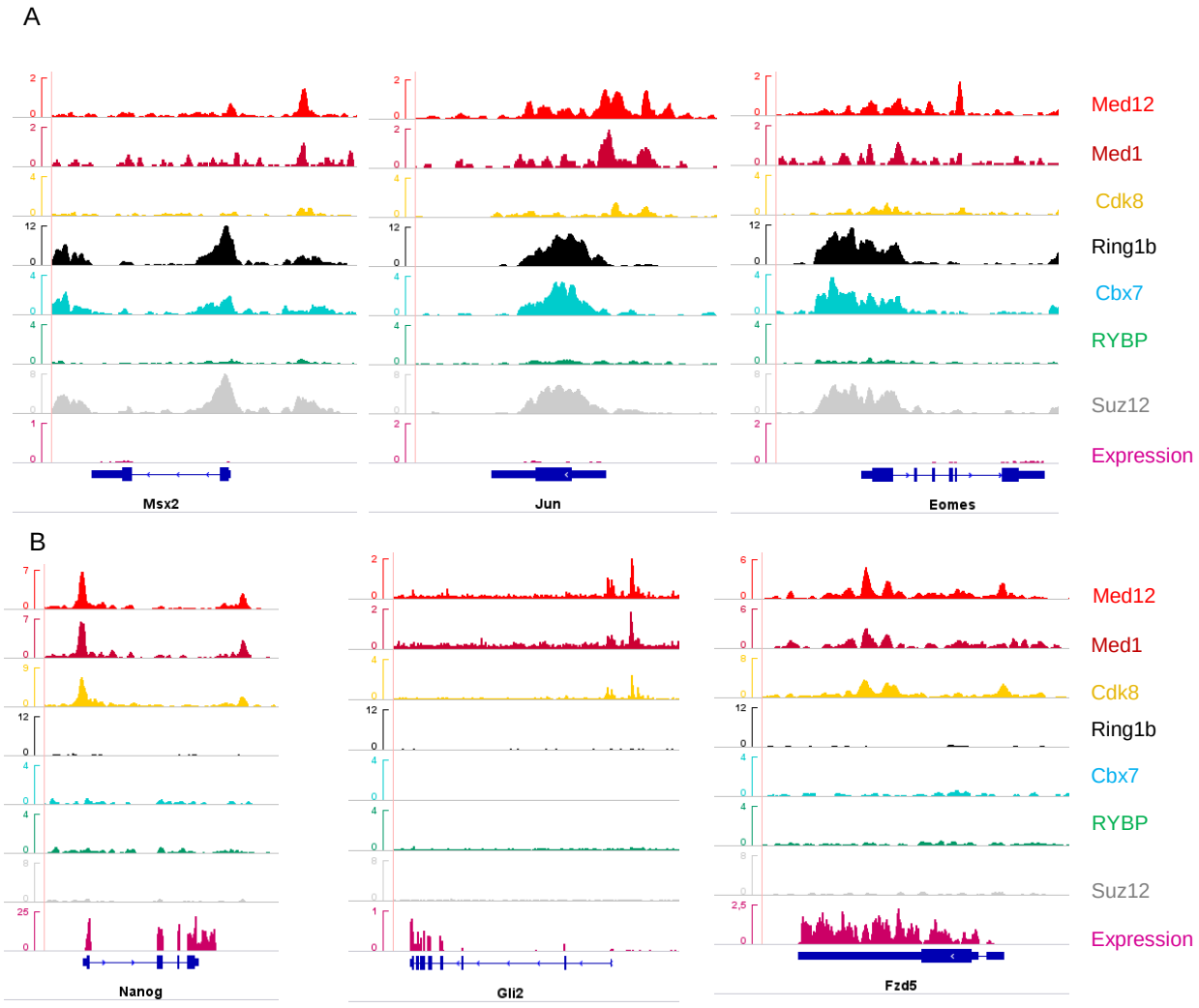


Figure 14: *Med12-Ring1b targets are regulated by the canonical PRC1*. Screenshots generated by the IGV browser showing the ChIP-seq profiles of Med12, Med1, Cdk8, Ring1b, Cbx7, RYBP, Suz12 as well as the expression (RNA-seq) in selected Med12-Ring1b (A) or Med12-Cdk8 (B) targets. (C) Venn diagrams showing the overlapping ChIP–seq targets of Med12-Ring1b with subunits of the PRC1 (Cbx7, RYBP, Phc1) and PRC2 (Suz12, Eed) complexes and their respective histone modifications (H2AK199ub and H3K27me3).

Genes which are bound by Med12 and other PRC1/PRC2 components or by both Med12-Ring1b and other PRC1/PRC2 components have functions in development, differentiation, morphogenesis and cell fate commitment as depicted from GO term analysis (Figure 15A, C). These genes are repressed in mESCs, a function which is regulated by the PRCs. In agreement, we observed that all the Med12-Ring1b common targets have either no or very little expression in pluripotency (Figures 14A and 15B (blue)). On the other hand, genes enriched for Med12-Cdk8 are expressed in mESCs (Figure 14B and 15B (green)), in line with their function in metabolic and biosynthetic processes (Figure 13D). The few genes which are co-bound by Ring1b-Cdk8 (177) appear to be involved in the regulation of biosynthetic and differentiation processes and are expressed in mESCs (Figure 15B (purple)). Similarly, genes targeted by Med12-Ring1b-Cdk8 are also moderately expressed in pluripotency (Figure 15B (red)).

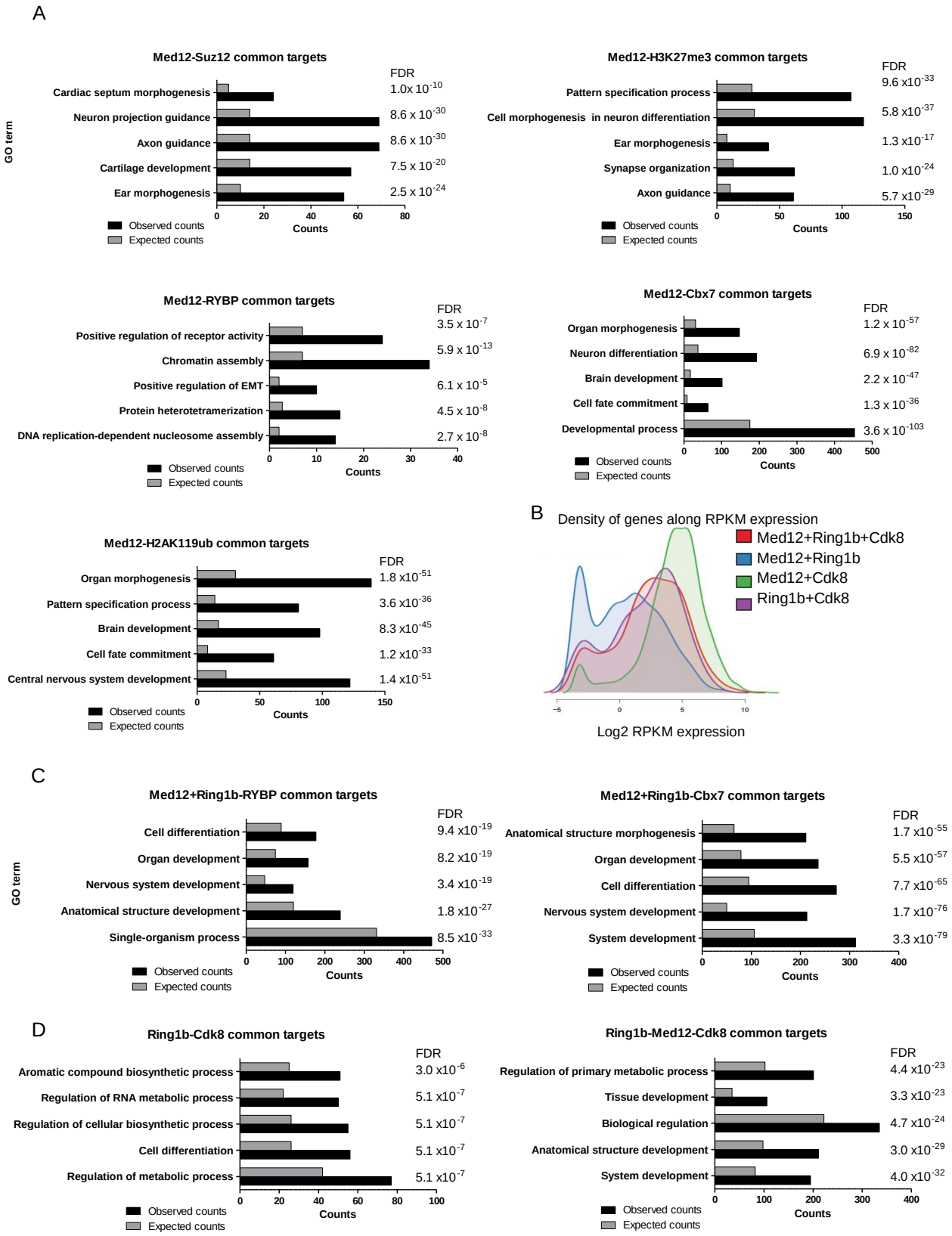


Figure 15: *Med12 and Med12-Ring1b targets co-bound by PRC1 and PRC2 components are genes related to differentiation and development that are not expressed in mESCs*. (A) Enriched GO terms for Med12-Suz12, Med12-H3K27me3, Med12-RYBP, Med12-Cbx7 and Med12-H2AK119ub common targets. (B) Expression levels of genes co-bound by Med12-Ring1b-Cdk8 (red), Med12-Ring1b (blue), Med12-Cdk8 (green) and Ring1b-Cdk8 (purple). (C) Enriched GO terms for genes co-bound by Med12-Ring1b and subunits of the non-canonical (RYBP) and the

canonical (Cbx7) PRC1 complex. (D) Enriched GO terms for Ring1b-Cdk8 and Ring1b-Med12-Cdk8 common target genes

Collectively, these data reveal a potential novel role for Med12 in mESCs, outside the Kinase Module of Mediator. In particular, Med12 is co-bound with Ring1b at the promoters of silenced genes in pluripotency which have functions related to differentiation and development. These genes are also enriched in other components of the PRC1 and PRC2 complexes such as Cbx7 and Suz12 respectively. A second class of genes, primarily metabolic, which are co-bound by Med12 and Cdk8 are expressed in pluripotency and devoid of PRC1 and PRC2 components.

Ring1b-dependent recruitment of Med12 at chromatin

Having identified this novel class of silenced genes bound by Med12-Ring1b in mESCs we sought to explore a potential relationship between those proteins. To assess the latter, we generated mESCs stably expressing shRNA against a non-mammalian target (control shNMC), Med12 (shMed12), Ring1b (shRing1b). These knockdown lines were used for cell fractionations from which the chromatin fraction was analyzed by western blot to evaluate the binding profile of the proteins from pluripotency to early differentiation (Figure 16A). Med12 is bound at chromatin in pluripotency and early differentiation and the levels drop 48hrs post-differentiation induction. Both Ring1b and its respective histone mark H2AK119ub, are found at high levels at chromatin in pluripotency and are gradually dislodged as differentiation proceeds. On the other hand, the Cdk8 subunit of the Kinase Module gets successively recruited at chromatin upon differentiation with a peak at 48hrs RA. Upon Med12 depletion, the levels of Ring1b and H2AK119ub at chromatin remain unchanged, however Cdk8 recruitment is abolished. Upon Ring1b depletion, we observed, as expected, reduced levels of H2AK119ub at chromatin. Surprisingly, both Cdk8 and Med12 are not recruited at chromatin in shRing1b cells (Figure 16A). To rule out the possibility that Cdk8 and Med12 are regulated at the transcriptional or post-transcriptional level by Ring1b, their total mRNA and protein levels were measured in the respective knockdown lines. The levels of Cdk8 and Med12, both the RNA (Figure 16B) and protein (Figure 16C) were

not affected upon Ring1b depletion suggesting a functional role for Ring1b in their recruitment at chromatin. In the same experimental set up the knockdown levels in all the respective cell lines were confirmed, allowing us to use them for further assays.

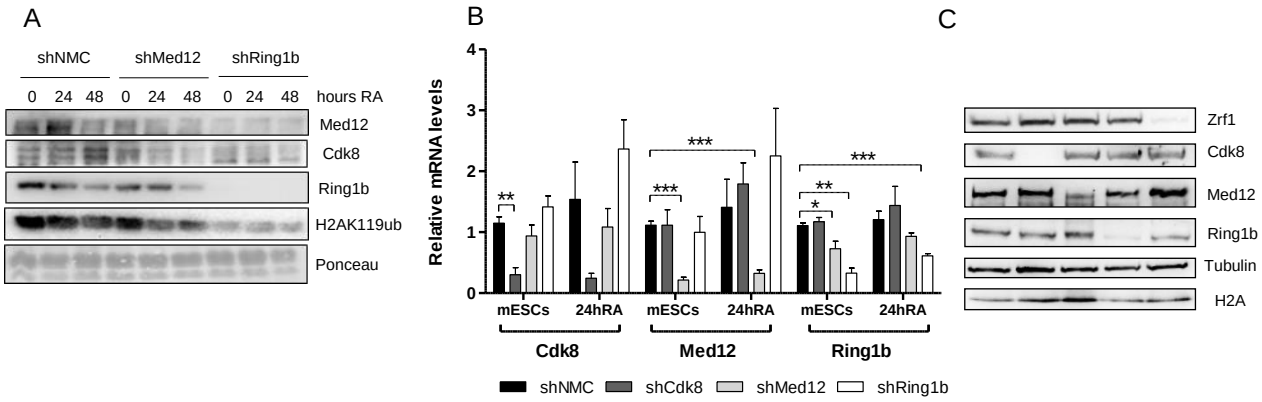


Figure 16: *Ring1b is functionally required to recruit Med12 and Cdk8 at chromatin in pluripotent and differentiating mESCs*. (A) The chromatin fraction from cell fractionations of the indicated cell lines was used for western blot analysis. After incubation with the indicated antibodies the chromatin binding profile of Med12, Cdk8, Ring1b and H2AK119ub was analyzed in shNMC (control), shMed12 and shRing1b pluripotent and early differentiating (24 and 48hRA) mESCs, using ponceau staining as a loading control. (B) RNA was purifired from shNMC, shCdk8, shMed12 and shRing1b pluripotent and early differentiating mESCs (24hRA) and was used for cDNA synthesis. The cDNA was used for qPCR using the indicated primer pairs and GAPDH as a reference gene. The expression levels were calculated with the ddCt method and represented as fold change to the shNMC mESCs sample +/- S.E.M. *** P-value ≤ 0.0009, ** P-value ≤ 0.005, * P-value = 0.04 as calculated by two-tailed unpaired t test, n=3. (C) Whole cell extracts were prepared from mESCs of all the indicated cell lines and used for western blot analysis. The blots were incubated with the indicated antibodies using H2A and Tubulin as loading controls.

After showing that Ring1b has a functional role in recruiting Med12 and Cdk8 at chromatin in early differentiation, as indicated by the cell fractionation assays, we sought to confirm this regulatory role in a gene-specific context. To that end we carried out Chromatin Immunoprecipitation assays (ChIPs) from shNMC, shMed12 and shRing1b mESCs for Med12, using IgG as an isotype control. We then measured the enrichment of Med12 in both Med12-Ring1b (*Msx2*, *Fgf9* and *Pax3*) and Med12-Cdk8 (*Nanog* and *Gli2*) target genes (Figure 17A). We observed that in all Med12-Ring1b target genes, the enrichment of Med12 is significantly reduced in both shMed12 and shRing1b lines as compared to shNMC. This finding further confirmed that Ring1b is an essential factor for the recruitment of Med12 at its target genomic loci in mESCs. We also observed reduced Med12 levels at the promoters of Med12-Cdk8 target genes in shMed12 and shRing1b cells, in agreement to the findings from the cell fractionations

that showed overall reduced levels of Med12 at chromatin upon Ring1b depletion (Figure 16A). We carried out a similar experiment, this time using Ring1b and did not observe any significant changes at the enrichment levels on Med12-Ring1b targets upon Med12 depletion (Figure 17B). In line with the findings from the cell fractionations, Med12 had no effect in the recruitment of Ring1b at chromatin. Furthermore we observed no enrichment for Ring1b at the promoters of Med12-Cdk8 target genes, confirming the presence of separate classes of genes; those targeted by either Med12-Cdk8 or Med12-Ring1b (Figure 17B).

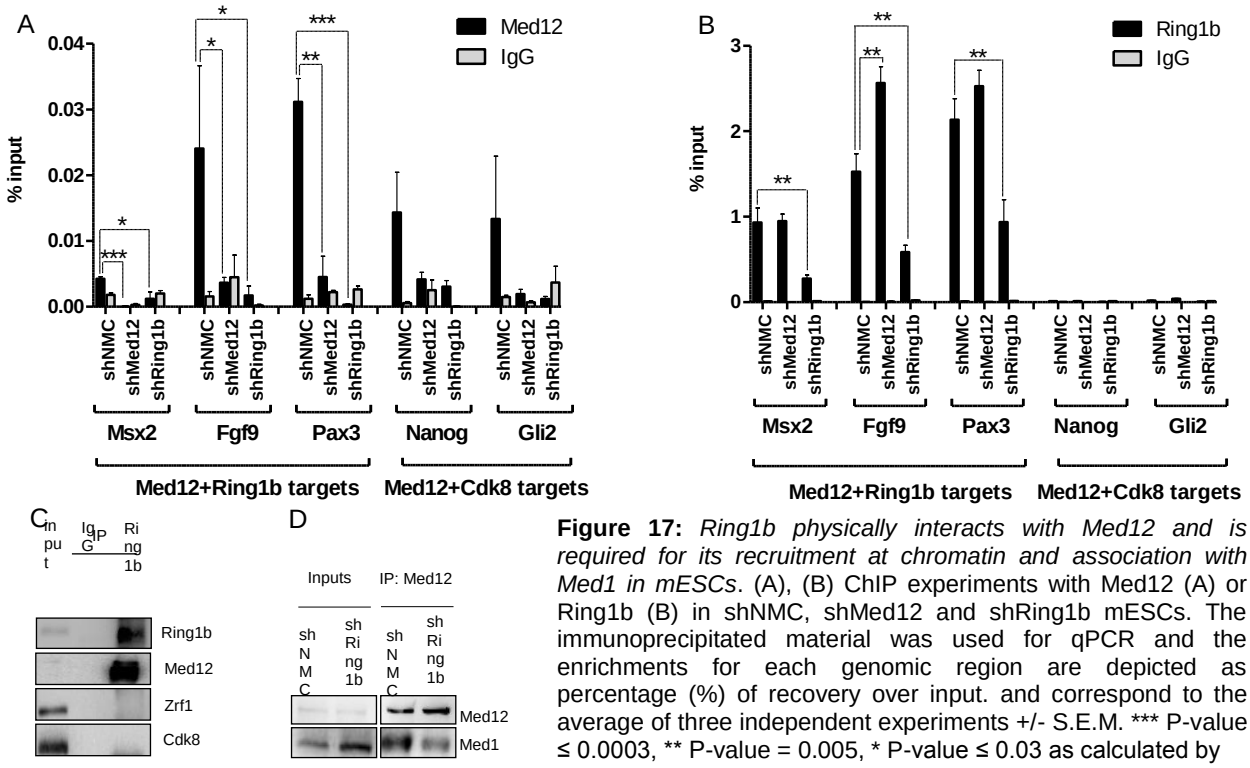


Figure 17: *Ring1b physically interacts with Med12 and is required for its recruitment at chromatin and association with Med1 in mESCs*. (A), (B) ChIP experiments with Med12 (A) or Ring1b (B) in shNMC, shMed12 and shRing1b mESCs. The immunoprecipitated material was used for qPCR and the enrichments for each genomic region are depicted as percentage (%) of recovery over input. and correspond to the average of three independent experiments +/- S.E.M. *** P-value ≤ 0.0003, ** P-value = 0.005, * P-value ≤ 0.03 as calculated by two-tailed unpaired t-test. (C) Endogenous IPs with Ring1b antibodies and nuclear extracts from mESCs. The precipitated material was analyzed by western blotting with the indicated antibodies. Inputs represent 7% of the material used for each IP. (B) Ring1b controls the association of Med12 with the core Mediator protein Med1. Endogenous immunoprecipitations (IPs) with Med12 antibodies and protein extracts from control (shNMC) and Ring1b (shRing1b) knockdown mESCs. The precipitated material was analyzed by western blotting with the indicated antibodies. Inputs represent 7% of the material used for each IP.

The newly found regulatory role of Ring1b in recruiting subunits of the Kinase Module of Mediator at chromatin, prompted us to examine the possibility of a physical interaction between those factors in mESCs. We tested the latter by performing endogenous Ring1b

Immunoprecipitations (IPs) from mESC nuclear extracts (Figure 17C). Ring1b co-precipitated with Med12 but neither with Cdk8, nor the H2AK119ub-binding protein, Zrf1. This finding combined with the genome wide ChIP data (Figures 13A and B, 14A) and the gene-specific ChIP data (Figure 16A), further suggest a functional link between Ring1b and Med12 in mESCs. Having shown that a fraction of Ring1b physically interacts with Med12 but not Cdk8 (Figure 17C), we then asked whether Ring1b is involved with the interaction of Med12 and the Core Mediator. This question was raised after identifying a large portion of Med12-Ring1b targets (80%) to be also targeted by the Core Mediator subunit Med1 (Figure 13B). By carrying out endogenous Med12 IPs in shNMC and shRing1b mESCs, we observed a reduced Med1 co-precipitation upon Ring1b depletion (Figure 17D). This experiment confirmed our hypothesis that Ring1b regulates the association between Med12 and the Core Mediator subunit Med1.

In summary, we have identified a class of genes which are targeted by Ring1b, Med12 and Med1 in mESCs. These genes are devoid of Cdk8 suggesting that a fraction of Med12 might exert a function in mESCs outside the Kinase Module, along with Ring1b and Med1 from the Core Mediator. Ring1b not only physically interacts with Med12, but also is required for the recruitment of the protein at the promoters of its target genes (both Med12-Ring1b and Med12-Cdk8), as well as for its association with the Core Mediator.

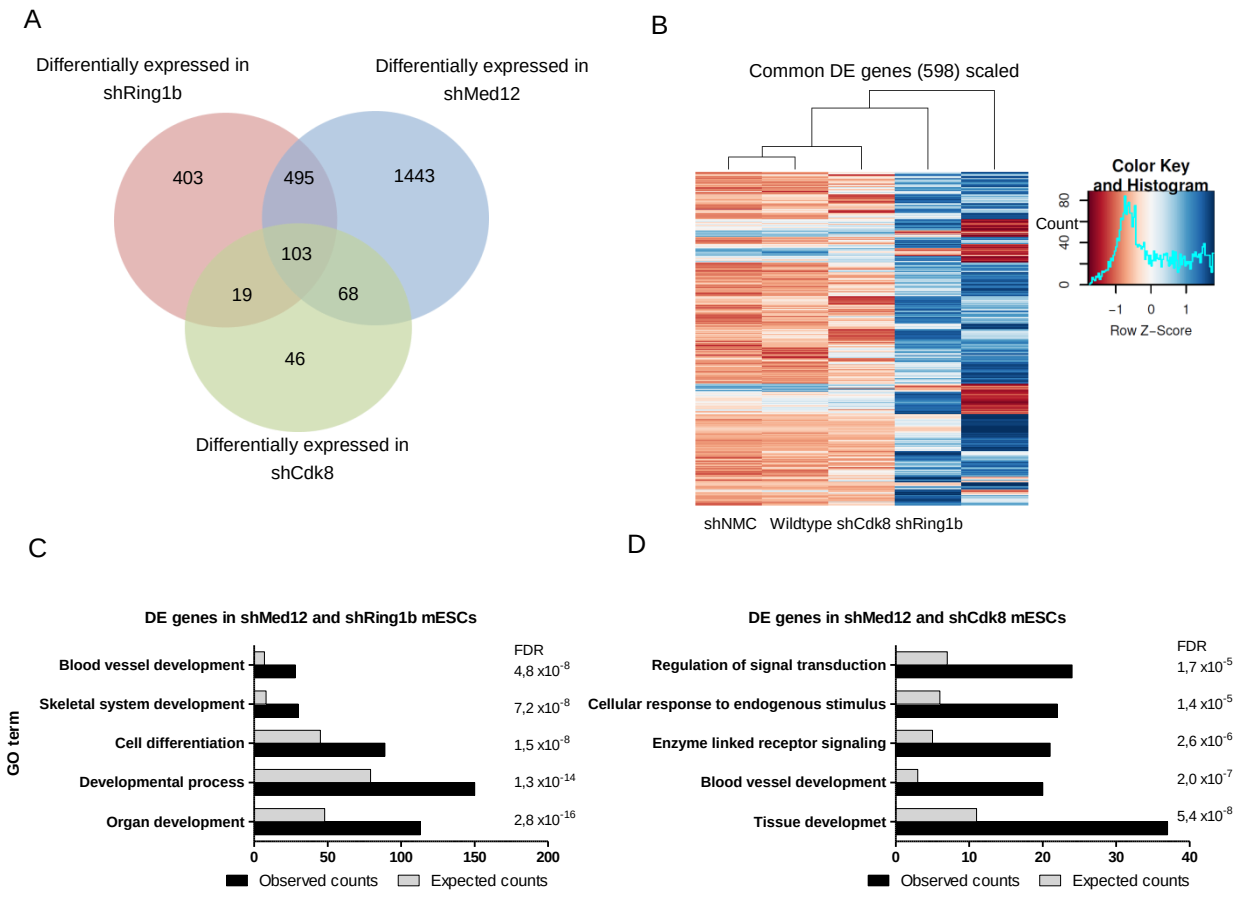
Ring1b and Med12 repress key developmental genes in pluripotency

Having shown that Ring1b and Med12 physically interact and because of the important roles of these factors in stem cell regulation, the next question to be asked was whether their functional interplay impacts gene expression. To assess the latter we used wildtype, shNMC, shCdk8, shRing1b and shMed12 mESCs, in biological triplicates, to perform RNA-sequencing (RNA-seq). We obtained between 64 and 95 million reads per sample, which were then mapped to the mouse reference genome (mm9 version from UCSC browser). From the obtained gene lists, we analyzed further those which were differentially expressed (DE) in each line compared to the shNMC or wildtype samples. When comparing all the DE genes, we found 103 to be common

between shCdk8, shMed12 and shRing1b samples (Figure 18A). Such a number is considered to be quite low for the respective experiment, which could be attributed to the low impact of Cdk8 depletion in gene expression (only 236 genes in total). On the contrary, depletion of either Med12 or Ring1b leads to a much higher number of DE genes (2109 and 1020 respectively), in agreement with their previously described roles in stem cell regulation. In addition, we observed 495 DE genes which were common in the shMed12 and shRing1b samples (Figure 18A). We then added to these common genes the 103 which were common for all three shCdk8, shMed12 and shRing1b (598), and created a heatmap showing their scaled expression levels (Figure 18B). When looking at the expression profile of these genes we did not observe any major differences between the wildtype, shNMC and shCdk8 samples. As expected, the profiles of the shRing1b and shMed12 samples are similar to each other and rather different when compared to the rest of the samples. The majority of genes which are common DE in shRing1b and shMed12, show higher expression levels (Figure 18B, blue) as compared to all three other samples. We then performed GO analysis with these common DE genes and the most significantly enriched terms were related to developmental and differentiation processes (Figure 18C). Such genes are usually not expressed in mESCs, a function largely associated with PRCs. After doing the same analysis, this time for the common DE genes in shMed12 and shCdk8 samples, we observed that the majority of the enriched terms were connected with signal transduction pathways (Figure 18D). The GO term “blood vessel development” was enriched in both of the groups tested (Figure 18C, D). Similarly, plots depicting the fold change in expression, show that the majority of common DE genes between shMed12 and shRing1b are upregulated upon depletion of the factors (Figure 18E, left quadrants) whereas the common DE genes between shMed12 and shCdk8 are either upregulated or downregulated (Figure 18F).

To confirm that the depletion of either Med12 or Ring1b causes upregulation to the DE genes identified in the RNA-seq, we carried out RT-qPCRs to check the expression of a selected fraction of them. In line with previous reports and the profile from the RNA-seq, most of the

tested genes showed elevated mRNA levels in shRing1b mESCs when compared to shNMC (Figure 18G, light grey). We observed the same pattern in most of the genes tested also in shMed12 mESCs (Figure 18G, dark grey), further suggesting that Med12 and Ring1b contribute to the repression of developmental genes in pluripotency. In agreement with the genome-wide data, Cdk8-depleted mESCs showed very similar mRNA levels of the tested genes to the shNMC sample (Figure 18G, white).



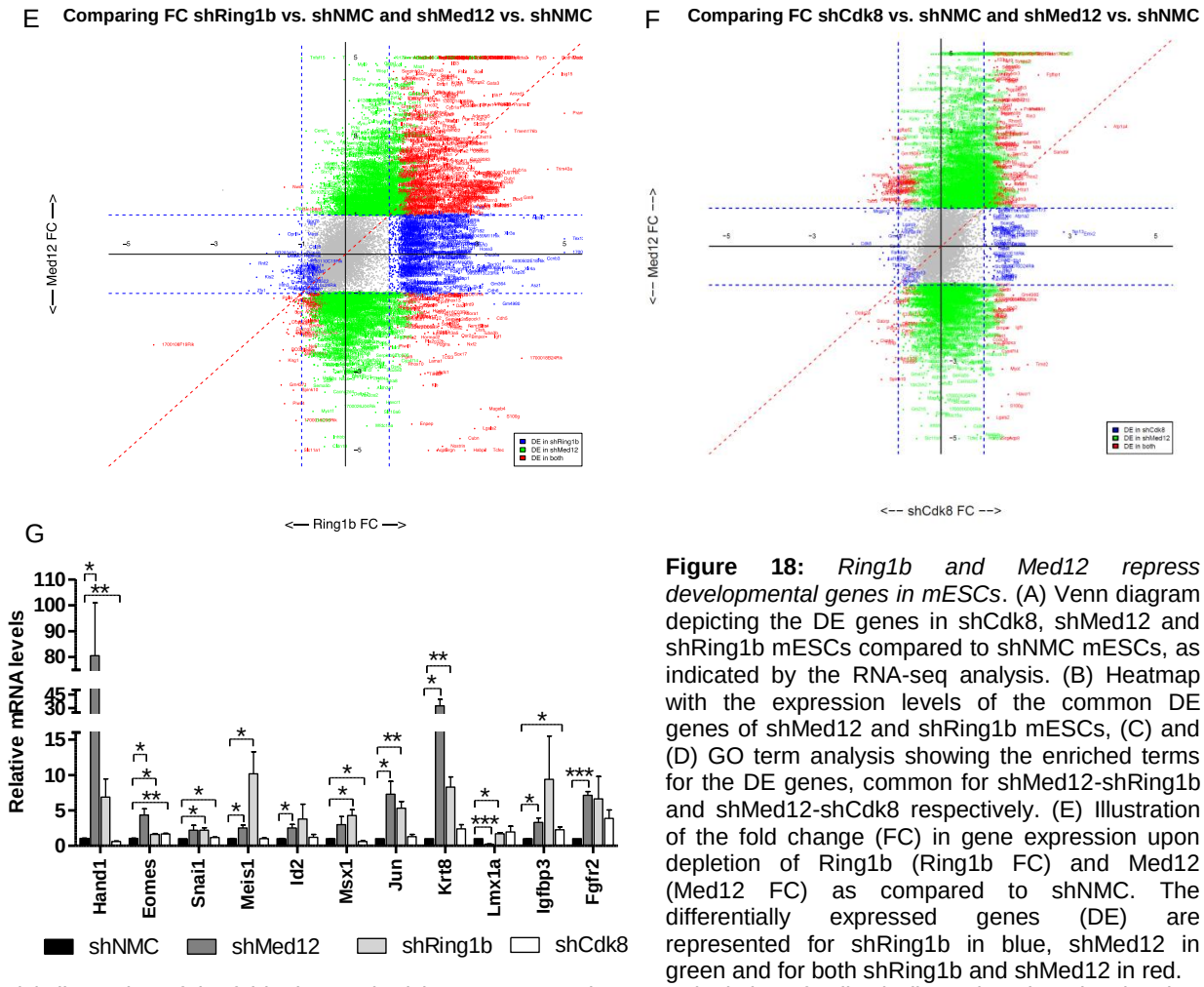


Figure 18: *Ring1b and Med12 repress developmental genes in mESCs*. (A) Venn diagram depicting the DE genes in shCdk8, shMed12 and shRing1b mESCs compared to shNMC mESCs, as indicated by the RNA-seq analysis. (B) Heatmap with the expression levels of the common DE genes of shMed12 and shRing1b mESCs, (C) and (D) GO term analysis showing the enriched terms for the DE genes, common for shMed12-shRing1b and shMed12-shCdk8 respectively. (E) Illustration of the fold change (FC) in gene expression upon depletion of Ring1b (Ring1b FC) and Med12 (Med12 FC) as compared to shNMC. The differentially expressed genes (DE) are represented for shRing1b in blue, shMed12 in green and for both shRing1b and shMed12 in red. (F) Illustration of the fold change (FC) in gene expression upon depletion of Cdk8 (Cdk8 FC) and Med12 (Med12 FC) as compared to shNMC. The differentially expressed genes (DE) are represented for shCdk8 in blue, shMed12 in green and for both shCdk8 and shMed12 in red. (G) Knockdown of either Ring1b or Med12 causes de-repression of their common target genes in mESCs. Shown are the relative expression levels of selected Ring1b/Med12 target genes in the aforementioned cell lines. The data are represented as the average values of three biological replicates +/- S.E.M. *** P-value < 0.0001, ** P-value < 0.001, * P-value < 0.02, as calculated by two-tailed, unpaired t test.

After showing that Med12 and Ring1b are targeted to the promoters of 1670 genes in mESCs (Figure 13A) which are related with differentiation and development and their depletion results in the upregulation of genes with similar functions, we then asked which genes are directly regulated by these factors. In order to find the common, direct targets of Ring1b and Med12 in mESCs, we carried out a combined analysis of the ChIP-seq and the RNA-seq datasets (Figure 19).

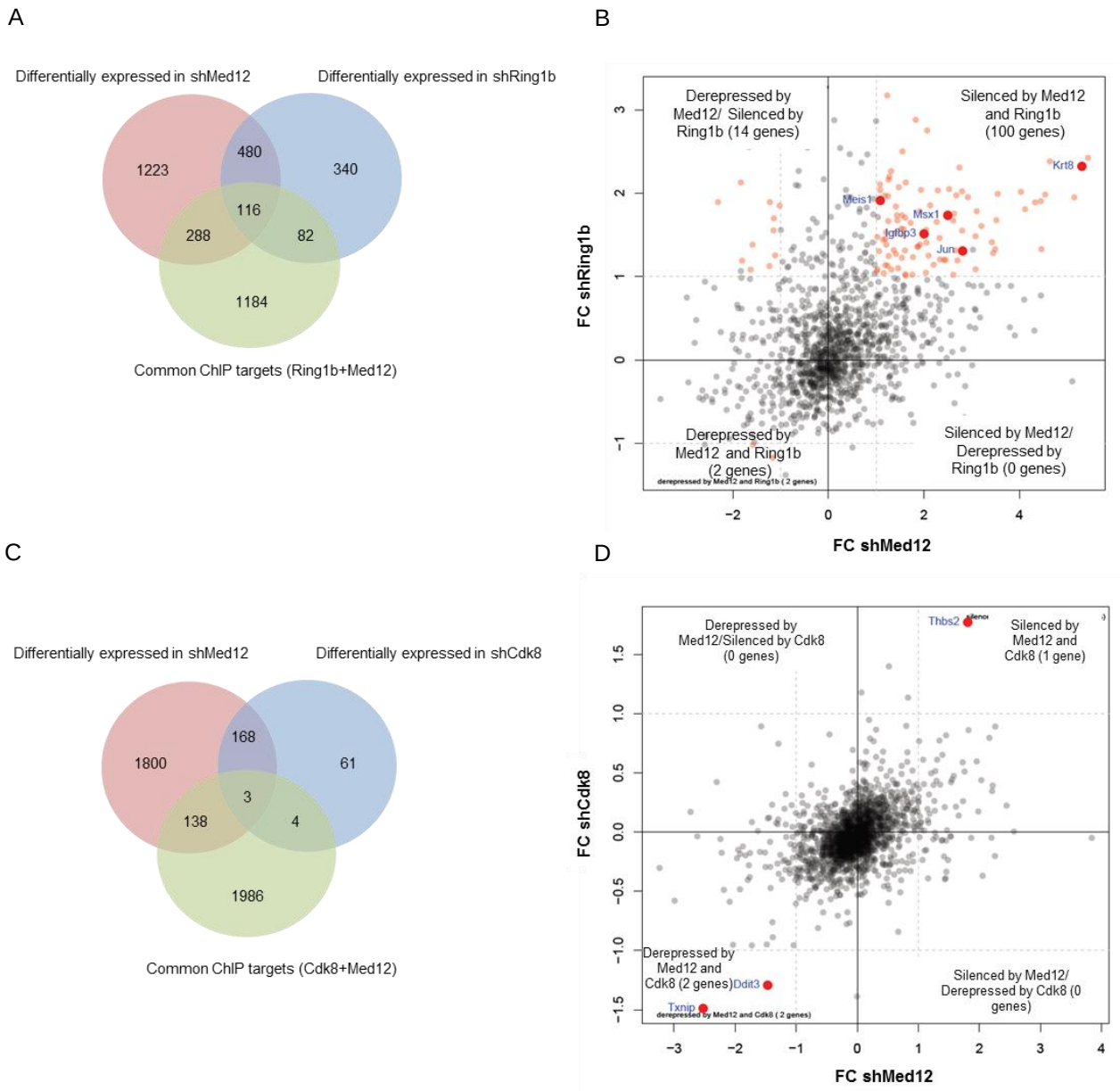


Figure 19: *Med12 and Ring1b directly regulate developmental genes in mESCs*. (A) Venn diagram depicting the direct Med12-Ring1b targets in mESCs; 116 genes are bound by both factors and the expression of which is altered in shMed12 and shRing1b mESCs. (B) Med12 and Ring1b repress their common target genes in mESCs. The plot depicts the expression Fold Change (FC) of the 116 direct target genes (Figure 7A). 100 out of 116 target genes (~ 86%) are de-repressed in both shRing1b and shMed12 mESCs. (C) Med12 and Cdk8 directly control the expression of 3 genes. Venn diagram illustrating the direct Med12-Cdk8 targets in mESCs; 3 genes are bound by both factors and their expression is altered in shMed12 and shCdk8 mESCs. (D) Med12 and Cdk8 either repress or activate their common target genes in mESCs. The plot shows the expression Fold Change (FC) of the 3 direct target genes (Figure 7C). 2 genes are derepressed (lower, left quadrant) and 1 gene is silenced (upper, right quadrant) in shMed12 and shCdk8 mESCs.

From this analysis, 116 genes were identified as direct targets; genes which are targeted by both proteins in pluripotency and differentially expressed in both shMed12 and shRing1b mESCs

(Figure 19A). From these 116 direct targets, 100 were found to be genes which are silenced in pluripotency by Med12 and Ring1b, as their expression is higher in shMed12 and shRing1b compared to shNMC mESCs (Figure 19B, upper-right quadrant). The same combined analysis for Med12 and Cdk8 revealed only 3 direct targets (Figure 19C). 2 of these genes are activated in (Figure 19D, lower-left quadrant), whereas 1 gene gets silenced by both factors (Figure 19D, upper-right quadrant) in mESCs.

In summary, it was shown that the majority of Med12-Ring1b targets in pluripotency are developmental genes. Both proteins regulate the expression of these genes, ensuring their silenced state in mESCs. A second class of genes which are related to metabolic and signal transduction pathways is targeted by Med12-Cdk8. Although the direct targets of the Kinase Module of Mediator in mESCs (Med12-Cdk8 targets) are only 3, as indicated by the combined ChIP-seq and RNA-seq analysis, this is majorly attributed to the low impact of Cdk8 alone in gene expression.

Med12 and Ring1b depleted mESCs fail to properly differentiate

After showing that Ring1b and Med12, which functionally associate, are required for the silencing of developmental genes in mESCs, whereas Med12 and Cdk8 regulate a distinct class of genes at the same state, we then studied the role of all three factors in stem cell self-renewal. The latter was addressed by measuring the relative mRNA levels of all four pluripotency markers, *Oct4*, *Sox2*, *Nanog* and *Klf4* with qRT-PCR from shNMC, shCdk8, shMed12 and shRing1b mESCs (Figure 20A). None of the tested pluripotency markers were significantly affected upon depletion of Cdk8. On the other hand, all four markers were significantly downregulated in shMed12 mESCs, showing the importance of this factor in maintaining pluripotency as earlier described (Kagey et al., 2010). Notably the impact of Med12 in the expression of *Klf4* in mESCs has not been discussed in the past. In shRing1b mESCs, while the levels of *Oct4*, *Nanog* and *Klf4* were similar to shNMC, we observed reduced levels of the transcription factor *Sox2*, in line with previous reports (Leeb and Wutz, 2007; van der Stoop et

al., 2008). We also performed immunofluorescent stainings with Oct4, but the heterogeneity in the knockdown levels of each protein within each cell population, made it hard to distinguish intensity differences between the samples (Figure 20B).

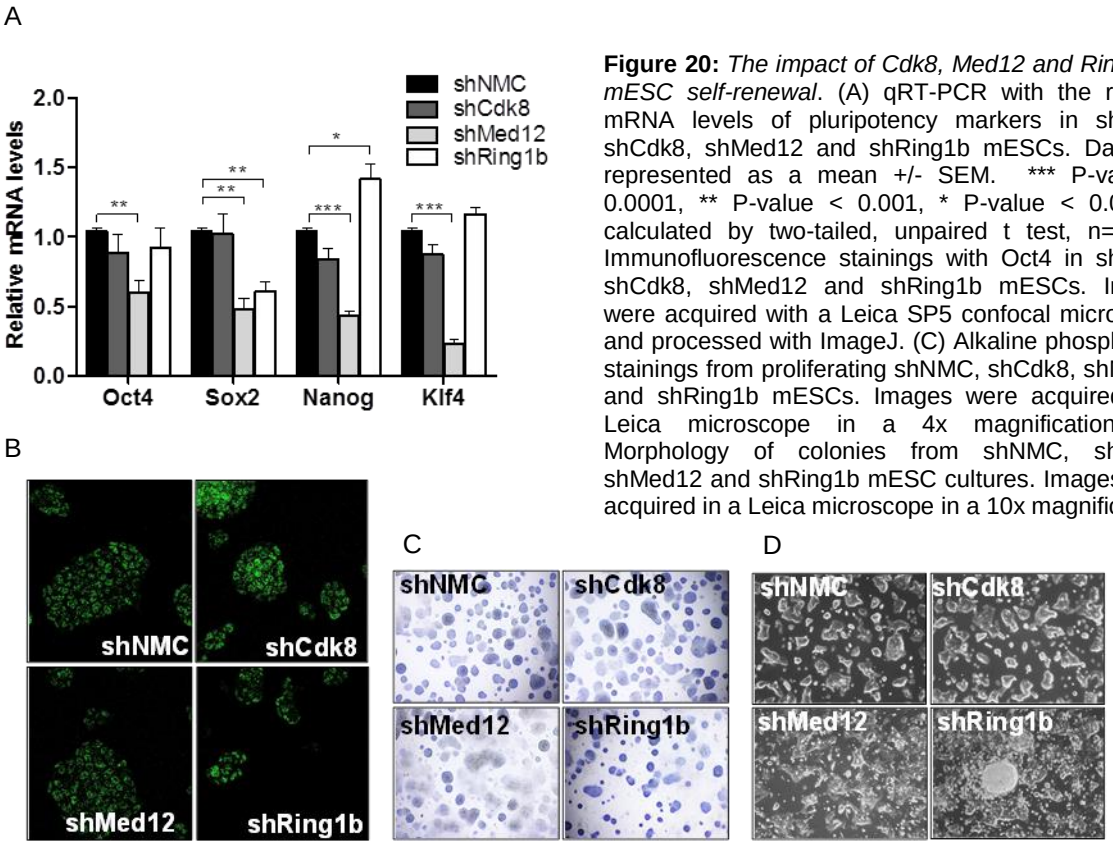


Figure 20: *The impact of Cdk8, Med12 and Ring1b in mESC self-renewal*. (A) qRT-PCR with the relative mRNA levels of pluripotency markers in shNMC, shCdk8, shMed12 and shRing1b mESCs. Data are represented as a mean +/- SEM. *** P-value < 0.0001, ** P-value < 0.001, * P-value < 0.02, as calculated by two-tailed, unpaired t test, n=3. (B) Immunofluorescence stainings with Oct4 in shNMC, shCdk8, shMed12 and shRing1b mESCs. Images were acquired with a Leica SP5 confocal microscope and processed with ImageJ. (C) Alkaline phosphatase stainings from proliferating shNMC, shCdk8, shMed12 and shRing1b mESCs. Images were acquired in a Leica microscope in a 4x magnification. (D) Morphology of colonies from shNMC, shCdk8, shMed12 and shRing1b mESC cultures. Images were acquired in a Leica microscope in a 10x magnification.

After seeing the effect of Med12 depletion in the expression of pluripotency markers, we then asked whether these mESCs spontaneously differentiate. By performing alkaline phosphatase stainings in proliferating mESCs, we observed that the majority of the shMed12 colonies had lower levels of the enzyme as compared to shNMC, an indication of spontaneous differentiation (Figure 20C). This phenotype is also depicted by the morphology of the shMed12 colonies, which appear flattened (Figure 20D). On the other hand, the proliferation potential of these cells was not significantly lower compared to shNMC, at least in respect to the passaging ratio (data not shown). Spontaneous differentiation was not observed after Cdk8 or Ring1b depletion (Figure 20C) and in some cases shRing1b colonies had stronger alkaline phosphatase staining

when compared to shNMC. Although shNMC and shCdk8 mESCs had the same morphology, shRing1b colonies were in some cases flattened with several large cell aggregates in the center (Figure 20D), reminiscent of embryoid bodies. Besides the fact that shRing1b mESCs had a lower proliferation rate compared to all other cell lines (data not shown), the alkaline phosphatase staining (Figure 20C) and the expression of pluripotency markers (Figure 20A) would not suggest that these cells spontaneously differentiate.

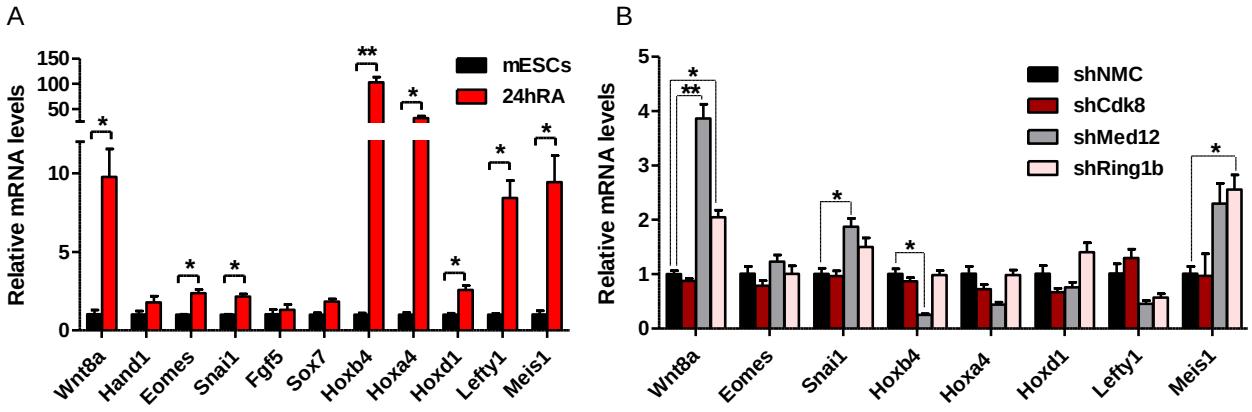


Figure 21: *Med12 and Ring1b have diverse effects in gene expression of differentiating mESCs*. (A) qRT-PCR with the relative mRNA levels of genes which get upregulated in differentiation in shNMC pluripotent and early differentiating (24hRA) mESCs. Data are represented as fold change to shNMC mESCs, +/- S.E.M., n=3. ** P-value = 0.009, * P-value ≤ 0.05 as calculated by two-tailed, unpaired t test. (B) qRT-PCR with shNMC, shCdk8, shMed12 and shRing1b, early differentiating (24hRA) mESCs. Data are represented as fold change to the shNMC sample, +/- S.E.M., n=3. ** P-value = 0.008, * P-value ≤ 0.08 as calculated by two-tailed, unpaired t test.

The great impact of Med12 and Ring1b in the expression of developmental genes in pluripotency, prompted us to investigate the effect of these factors in early differentiation. By removing the LIF from the mESC culture medium and incubating the cells with Retinoic Acid (RA) for 24hrs, we acquired early differentiating cells. We then compared the expression levels between pluripotency and early differentiation of several early-induced differentiation genes (Figure 21A) and observed that the majority of the genes tested were rapidly induced for expression upon treatment with RA. We then selected the genes with significantly higher mRNA levels in differentiating compared to pluripotent mESCs and measured by qRT-PCR the impact

of Cdk8, Med12 and Ring1b depletion in their expression after 24 hours of RA treatment (Figure 21B). In 3 of the selected genes we observed higher mRNA levels upon depletion of Med12 or Ring1b compared to shNMC cells. These genes are probably prematurely upregulated in early differentiation in the respective cells, as two of them, *Snai1* and *Meis1*, are also targeted for silencing by Med12 and Ring1b in pluripotency. Thus, for a fraction of differentiation genes, Med12 and Ring1b are required to ensure their silenced state in pluripotency and their proper expression in early differentiation. Interestingly, the mRNA levels in early differentiation of several genes including *Hoxb4*, *Hoxa4* and *Hoxd1*, were affected in a different way by either Med12 or Ring1b depletion. Although Med12 seems to be required for the expression of these genes at this stage, as indicated by the lower mRNA levels in shMed12 cells when compared to shNMC, Ring1b depletion has either no impact (*Hoxb4* and *Hoxa4*) or leads to overexpression of the genes (*Hoxd1*).

Collectively we observe that although Med12 and Ring1b act in a concerted fashion to silence differentiation genes in mESCs, Med12 but not Ring1b is essential for the maintenance of pluripotency. Besides that the interplay between Med12 and Ring1b is required for the silencing of differentiation genes in mESCs, later in early differentiation, these two factors acquire distinct roles in the transcriptional regulation of the same class of genes.

PART 2: Remodeling of Med12-Mediator enhances gene activation in early stem cell differentiation in a ncRNA-dependent manner

Med12 acts in concert with ncRNAs as a transcriptional activator of developmental genes in RA-mediated differentiation

The distinct effects of Med12 and Ring1b in *Hox* gene expression after RA-induced differentiation of mESCs, prompted us to study the underlying mechanism for this function. In order to do that we analyzed the regulation of a well-described *Hox* gene Polycomb target, *Hoxd11*. In pluripotency, the *Hoxd11* locus is silenced by being targeted with Cbx7-PRC1 and the PRC2 complex (Figure 22A). At the same timepoint, the promoter of *Hoxd11* is enriched in Med12-Mediator devoid of Cdk8. *Hoxd11* gets upregulated upon RA treatment of mESCs (Figure 22D) and Med12 depletion impairs this upregulation (Figure 22B). On the contrary, Ring1b depletion leads to higher levels of *Hoxd11* 24hrs post RA treatment of mESC (Figure 22B), in line with its role in regulating the proper spatiotemporal expression of developmental genes. As shown before for other *Hox* genes (Figure 21B), the expression of *Hoxd11* is influenced by both Med12 and Ring1b, but in the opposite way. More specifically while Ring1b restricts the expression in early differentiation, just as in pluripotency, Med12 appears to be required for the induction of the gene.

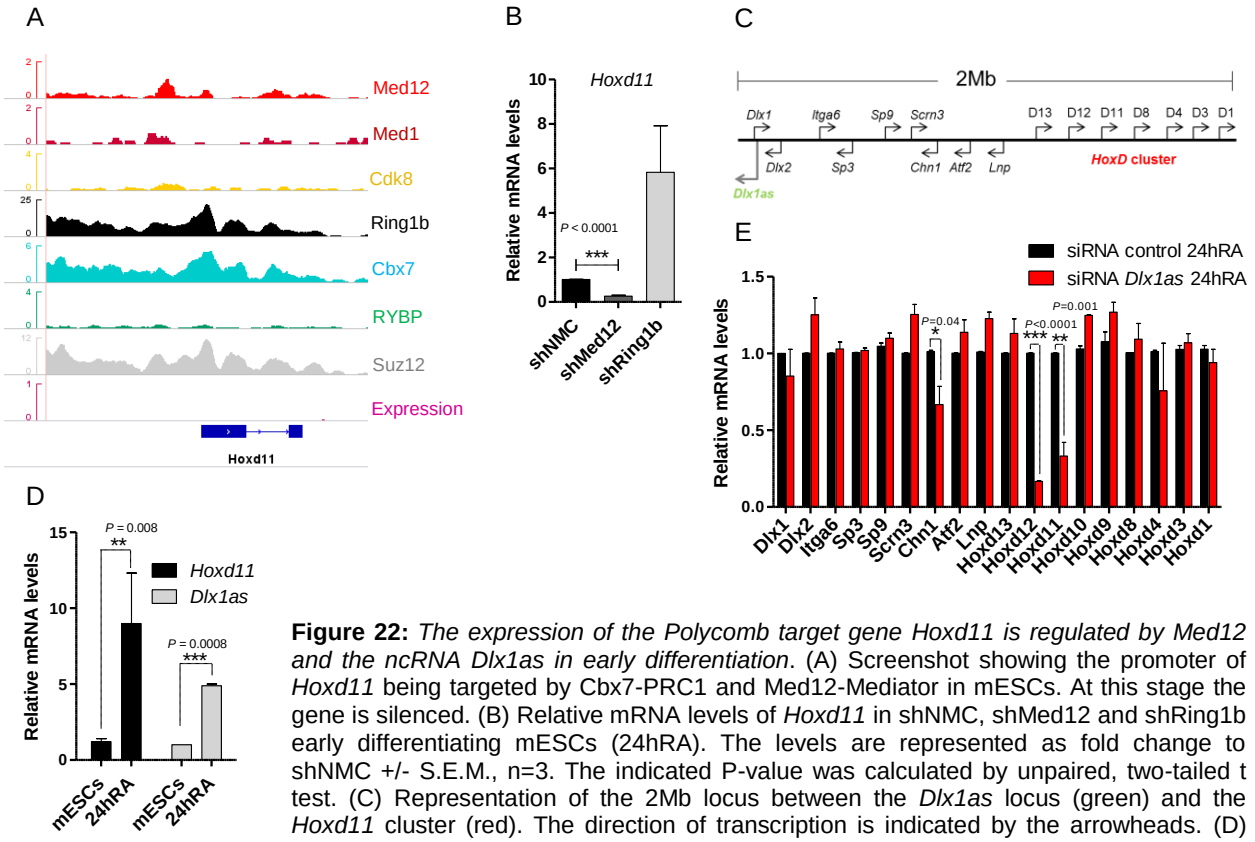


Figure 22: *The expression of the Polycomb target gene Hoxd11 is regulated by Med12 and the ncRNA Dlx1as in early differentiation*. (A) Screenshot showing the promoter of *Hoxd11* being targeted by Cbx7-PRC1 and Med12-Mediator in mESCs. At this stage the gene is silenced. (B) Relative mRNA levels of *Hoxd11* in shNMC, shMed12 and shRing1b early differentiating mESCs (24hRA). The levels are represented as fold change to shNMC +/- S.E.M., n=3. The indicated P-value was calculated by unpaired, two-tailed t test. (C) Representation of the 2Mb locus between the *Dlx1as* locus (green) and the *Hoxd11* cluster (red). The direction of transcription is indicated by the arrowheads. (D) *Hoxd11* and *Dlx1as* are upregulated upon RA-induced differentiation. qRT-PCR with the relative mRNA levels of *Hoxd11* and *Dlx1as* in pluripotent and early differentiating mESCs. The levels are represented as fold change to mESCs +/- S.E.M., n=3. The indicated P-values were calculated by unpaired, two-tailed t test. (E) The expression of genes of the *HoxD* cluster in early RA-induced differentiation is regulated in cis by the ncRNA Dlx1as. mESCs were transfected with siRNA against control or Dlx1as and then induced for differentiation with RA. 24 hours post RA adrministration, RNA was used for qRT-PCR. The relative mRNA levels were represented as fold change to siRNA control-treated samples +/- S.E.M., n=3. The indicated P-values were calculated by two-tailed, unpaired t test.

Med12 was previously reported to act as an enhancer of transcription in concert with activating lncRNAs (Lai et al., 2013). To investigate whether Med12 activates *Hoxd11* in an ncRNA-dependent manner, we searched for lncRNA candidates around the locus. *Dlx1as* is a lncRNA which is expressed from a locus around 2Mb upstream of the *HoxD* cluster (Figure 22C) and gets also induced after RA treatment of mESCs (Figure 22D). To test whether *Dlx1as* is also contributing in the expression of *Hoxd11* in early differentiation, like Med12, we transiently

depleted it from mESCs (Figure 23G). After 24hrs of RA treatment, the *Dlx1as* depleted cells (siRNA *Dlx1as*) were used to measure the relative mRNA levels of the downstream genes (Figure 22E). Both *Hoxd11* and *Hoxd12* as well as non-Polycomb target, *Chn1*, had reduced mRNA levels after depletion of *Dlx1as* when compared to the control samples.

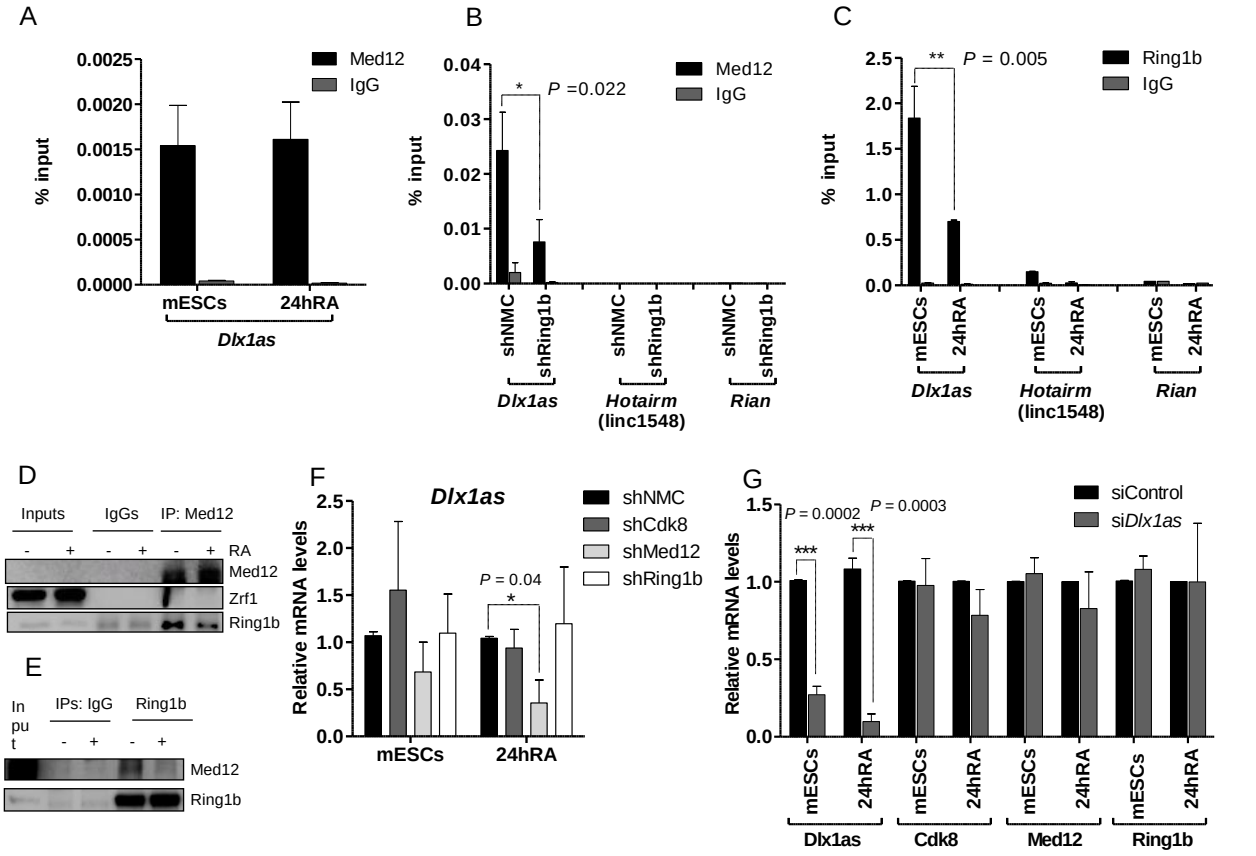


Figure 23: *The gene-activating Med12-Dlx1as association pre-exists in pluripotency and is dependent on Ring1b*. (A) RNA IP (RIP) for Med12 and IgG as isotype control, in pluripotent and early differentiating mESCs. The level of enrichment of *Dlx1as* is represented as percentrage of recovery over input. (B) RIP for Med12 and IgG as an isotype control in shNMC and shRing1b pluripotent mESCs. The levels of enrichment of the ncRNAs tested are represented as percentage of recovery over input, +/- S.E.M., n=2. The indicated P value was calculated by F test to compare the differences. (C) RIP for Ring1b and IgG as isotype control in pluripotent and early differentiating mESCs. The levels of enrichment of the tested ncRNAs are represented as percentage of recovery over input, +/- S.E.M., n=2. The indicated P value was calculated by F test to compare the differences. (D) Endogenous Med12 IP with nuclear extracts from pluripotent and early differentiating mESCs. The IPs were used for Western blotting and incubation with the indicated antibodies. Inputs represent 7% of the material used for the IPs. (E) Endogenous Ring1b IP with nuclear extracts from shNMC mESCs. The precipitated material was split in two fractions and one fraction was subjected to RNaseA treatment prior to Western blotting and incubation with the indicated antibodies. Inputs represent 5% of the material used for the IPs. (F) qRT-PCR with the relative mRNA levels of *Dlx1as* in shNMC, shCdk8, shMed12 and shRing1b pluripotent and early differentiating mESCs. The levels for each timepoint are represented as fold change compared to shNMC, +/- S.E.M., n=3. The indicated P value was calculated with two-tailed, unpaired t test. (G) qRT-PCR with the relative mRNA levels of *Dlx1as*, *Cdk8*, *Med12* and *Ring1b* after transfection with control siRNA or siRNA against *Dlx1as* in pluripotent and early differentiating mESCs. The levels at each timepoint are represented as fold change compared to the sample treated with siRNA control, +/- S.E.M., n=3. The indicated P values were calculated with two-tailed, unpaired t test.

After showing that both the ncRNA *Dlx1as* and Med12 are essential for the proper upregulation of *Hoxd11* in early RA-mediated differentiation of mESCs, we tested whether they exert this function by interacting with each other. RIP experiments with Med12 in pluripotent and early differentiating mESCs showed an interaction with the ncRNA at both timepoints (Figure 23A). Thus, the Med12-*Dlx1as* "complex" not only exists in early differentiation to activate *Hoxd11*, but it already forms in pluripotency, while Med12 is bound to Ring1b and has a repressive-like function (Figures 15B, 17C, 18B, 18G and 19B). As both Med12-*Dlx1as* and Med12-Ring1b physical interactions exist in pluripotency, we then asked if Ring1b influences the formation of the activating, ncRNA-containing complex. To that end we performed Med12 RIPs in shNMC and shRing1b mESCs and observed a reduced interaction with *Dlx1as* upon depletion of Ring1b (Figure 23B). In the same experiment we confirmed the specificity of the association between Med12 and *Dlx1as,* as there was no observed enrichment for two other lncRNAs with described roles in development; *Hotairm* (Diaz-Beya et al., 2015; Lin et al., 2011; Zhang et al., 2009) or *Rian* (Hagan et al., 2009; Han et al., 2014; Stadtfeld et al., 2010).

The Ring1b dependency of the Med12-*Dlx1as* interaction, prompted us to investigate the possibility of the former interacting with the ncRNA. We carried out Ring1b RIPs and observed a specific interaction with *Dlx1as* which decreases upon RA-induced differentiation (Figure 23C). Notably, the interaction of Ring1b with Med12 decreases as well in early differentiation (Figure 23D) further substantiating that at this time point the two proteins acquire opposing functions. Interestingly, the interaction between Med12 and Ring1b in pluripotency is RNA-dependent, as upon RNA depletion, less Med12 is co-precipitated in a Ring1b IP (Figure 23E).

In summary, it was shown that in pluripotency Ring1b is required to assemble and interacts with Med12-*Dlx1as* "complexes". These complexes activate the expression of the developmental gene *Hoxd11* in early RA-induced differentiation by reducing their affinity with Ring1b. While neither Ring1b nor Med12 are required for the expression of the ncRNA in pluripotency, in early differentiation, *Dlx1as* gets activated in a Med12-dependent manner (Figure 23F). These data

collectively show that an activating Med12-ncRNA complex, which is required to upregulate differentiation genes in early differentiation, pre-exists in pluripotency under the regulation of Ring1b and while these genes are silenced.

Zrf1 associates with activating ncRNAs during differentiation

The finding that Ring1b gradually loses its affinity to *Dlx1as* during the transition from pluripotency to differentiation was reminiscent of its dislocation from chromatin by the H2AK119ub-binding protein, Zrf1 (Richly et al., 2010). In order to investigate whether these two events are connected we examined the RNA-binding capacity of Zrf1.

The two SANT domains at the C' of ZRF1 (Figure 24A), known to bind nucleic acids, appeared as good candidates for RNA binding. We therefore performed FLAG pull downs from human cells (HEK293T) transfected with either control plasmids, full length ZRF1 ($ZRF1^{FLAG}$) or a truncated ZRF1 with a deletion in both SANT domains ($ZRF1\Delta SANT^{FLAG}$). 20% of the FLAG eluate was analyzed by immunoblot to confirm the IP efficiency (Figure 24B, upper panel) and 80% was used for RNA extraction. The RNA extract was then split in two fractions, one of which was incubated with RNaseA. Both fractions were loaded on an agarose gel to visualize RNA in the FLAG co-precipitate (Figure 24B, lower panel). We detected a smear only in the $ZRF1^{FLAG}$ fraction which was not incubated with RNase, indicating that ZRF1 can pull down RNA. The absence of a smear in the $ZRF1\Delta SANT^{FLAG}$ fraction implies that the RNA-binding ability of ZRF1 indeed lies within the SANT domains.

Having identified ZRF1 as a novel RNA-interacting protein and in order to further prove the importance of the SANT domains for this association, we then sought to map the RNA binding domain. To this end we mutated two amino acids (Saikumar et al., 1990) within the second SANT domain (Figure 24A) and used the mutants for FLAG pull downs (Figure 24C, upper panel).

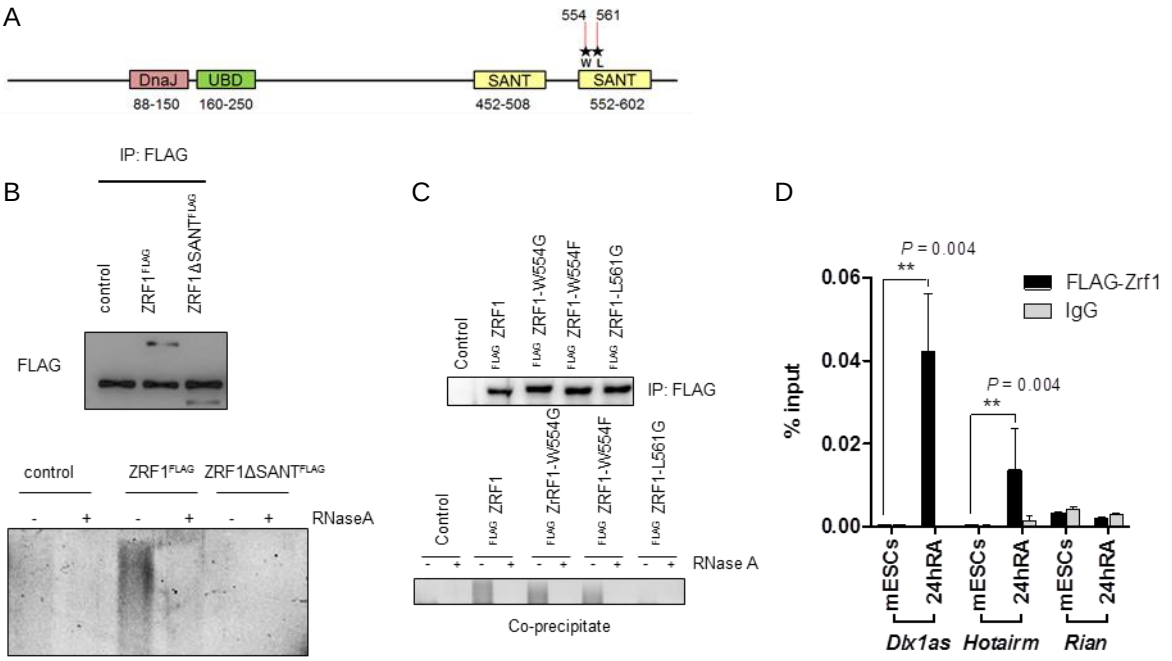


Figure 24: *Zrf1 associates with ncRNAs via its SANT domain*. (A) Cartoon depicting the functional domains of ZRF1. The asterisks correspond to the amino acids which were converted by site-directed mutagenesis. (B) HEK293T cells were transfected with control, ZRF1^{FLAG} or ZRF1ΔSANT^{FLAG} plasmids and the nuclear extracts were used for a FLAG pull down. A fraction of the eluate was analyzed by Western blotting using FLAG antibodies and another fraction was used for an RNaseA assay and loaded on an agarose gel to detect RNA in the co-precipitate. (C) HEK293T cells were transfected with control, ZRF1^{FLAG} and each of the three single amino acid mutants (ZRF1 ^{FLAG}-W554G, ZRF1 ^{FLAG}-W554F and ZRF1 ^{FLAG} -L561G) and the nuclear extracts were used for a FLAG pull down. A fraction of the eluate was analyzed by Western blotting using FLAG antibodies and another fraction was used for an RNaseA assay and loaded on an agarose gel to detect RNA in the co-precipitate. (D) RIP with FLAG and IgG as an isotype control from pluripotent and early differentiating mESCs, stably expressing FLAG-Zrf1. The levels or recovery of the ncRNAs *Dlx1as*, *Hotairm* and *Rian* are represented as percentage of input +/-S.E.M., n=3. The indicated P values were calculated with F test to compare the variances between pluripotent and 24hRA mESCs.

In a similar experimental set up as before (Figure 24B), we observed that leucine 561 (Figure 24C, lower panel) is essential for the RNA binding capacity of ZRF1. In order to connect the RNA binding of Zrf1 to the displacement of Ring1b from chromatin and the parallel dissociation from *Dlx1as*, we then performed RIP experiments for Zrf1 in pluripotent and early differentiating mESCs (Figure 24D). Notably, Zrf1 associates with *Dlx1as* as well as *Hotairm*, only in early differentiation. This association appears to be specific as there was no significant enrichment for a third ncRNA, *Rian*, after the pull down.

These data reveal Zrf1 as a novel ncRNA-interacting protein which not only displaces Ring1b from chromatin in early differentiation but also associates with *Dlx1as*. This activating ncRNA

forms a complex with Ring1b and Med12 and is required in early differentiation of the activation of *Hoxd11*. Here we observed that during the transition from pluripotency to differentiation, *Dlx1as* loses its affinity to Ring1b but associates with Zrf1, while remaining bound to Med12.

Zrf1 is implicated in the Med12-*Dlx1as*-dependent activation of *Hoxd11*

The association of Zrf1 with *Dlx1as* in early differentiation suggested that this H2AK119ub-binding protein might be implicated in the interplay between Ring1b and Med12 and more specifically in the dissociation of Med12-*Dlx1as* from Ring1b. According to this hypothesis Zrf1 would lead to the transcriptional activation of *Hoxd11* in a *Dlx1as*-dependent manner through Ring1b displacement. To investigate the latter we carried out ChIP experiments using a mESC line stably expressing Zrf1-flag, which showed recruitment of Zrf1 at the promoter of *Hoxd11* in early differentiation (Figure 25A). To confirm that this recruitment leads to transcriptional activation by Zrf1, we measured the relative mRNA levels of the genes between the *Dlx1as* locus and the *HoxD* cluster in early differentiation, after Zrf1 depletion (Figure 25B). Six out of the eighteen genes checked, including *Hoxd11*, had reduced mRNA levels in early differentiation after Zrf1 depletion. Thus, the recruitment of Zrf1 at the promoter of *Hoxd11* in early differentiation seems to be required for the activation of this gene, amongst others, just like Med12 and *Dlx1as*.

We then analyzed the effect of Zrf1 in the expression of genes which are subject to the Ring1b-Med12 repression in pluripotency. The majority of the tested genes get upregulated in response to RA and this upregulation is impaired upon Zrf1 depletion (Figure 25C). In addition, Zrf1 gets recruited to the promoters of a fraction of these genes in early differentiation as shown by ChIP experiments (Figure 25D). These findings collectively reveal that Zrf1, apart from being involved in the transcriptional activation of *Hoxd11* in a Med12-*Dlx1as* manner through the displacement of Ring1b from chromatin, also activates genes which are targeted by Ring1b and Med12 for silencing in pluripotency, suggesting a more global role for Zrf1 in activating PcG-repressed genes.

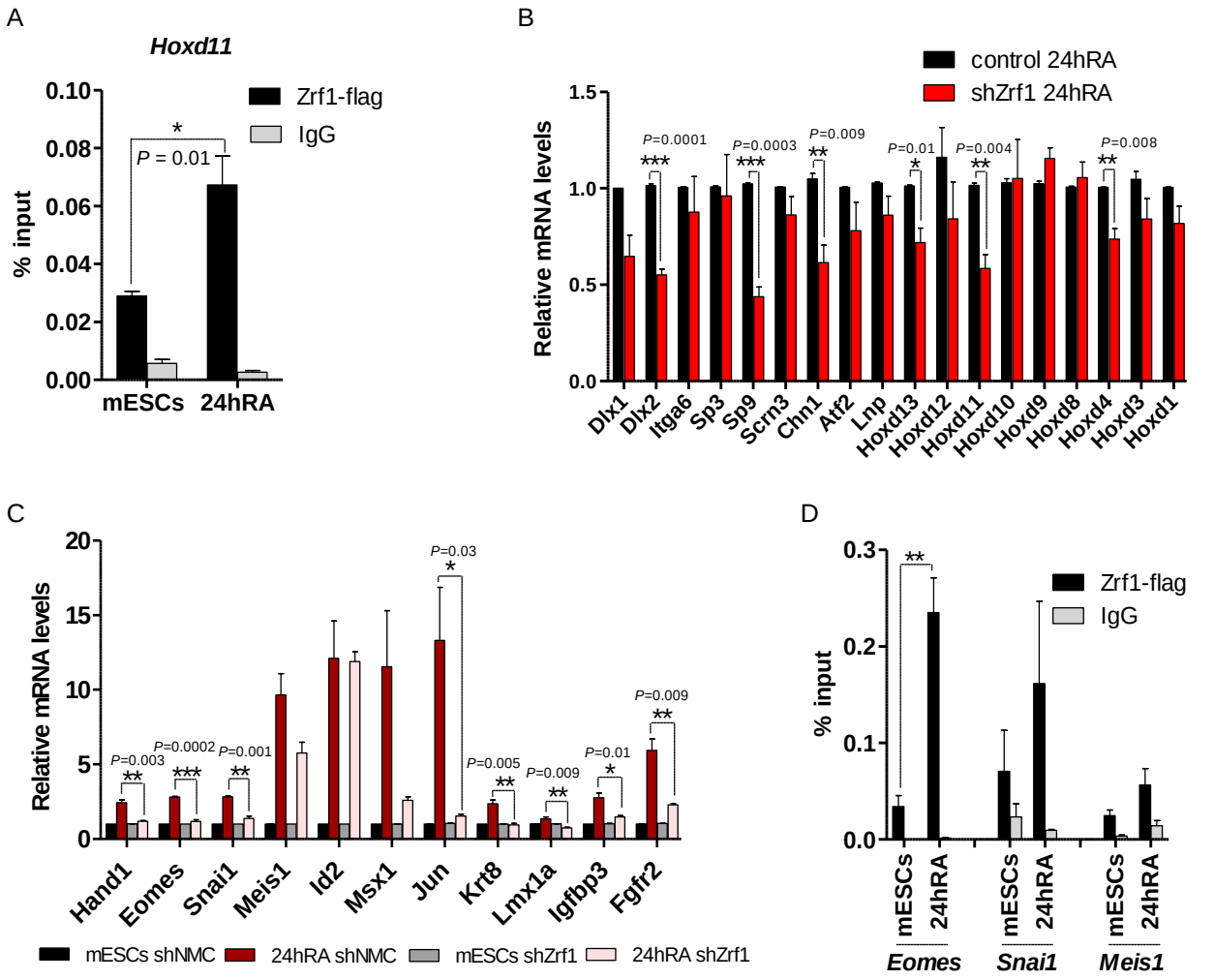


Figure 25: *Zrf1 gets recruited at promoters of developmental genes in early differentiation to activate their transcription*. (A) FLAG ChIP with pluripotent and early differentiating mESCs stably expressing Zrf1-flag. The enrichment at the promoter of *Hoxd11* is represented as percentage of recovery over input +/-S.E.M. The indicated P value was calculated with two-tailed, unpaired t test. (B) qRT-PCR with the relative mRNA levels of genes between the *Dlx1as* locus and the *HoxD* cluster in early differentiation in control and shZrf1. The mRNA levels of shZrf1 samples are represented as fold change to control samples +/- S.E.M., n=3. The indicated P values were calculated by unpaired, two-tailed t test. (C) qRT-PCR with the relative mRNA levels of genes which are repressed in pluripotency by Med12 and Ring1b using pluripotent and early differentiating shNMC and shZrf1 mESCs. The mRNA levels of the early differentiating samples are represented as fold change to the pluripotent samples +/- S.E.M., n=3. The indicated P values were calculated by unpired, two-tailed t test. (D) FLAG ChIP with pluripotent and early differentiating mESCs stably expressing Zrf1-flag. The enrichments at the promoters of *Eomes*, *Snai1* and *Meis1* are represented as percentage of recovery over input +/- S.E.M., n=3. The indicated P values were calculated by two-tailed, unpaired t test.

The functional role of Zrf1 in ncRNA-mediated gene activation during early differentiation is further supported by the fact that neither the mRNA (Figure 26A, 26B) nor the protein levels (Figure 16C) of the other involved factors are affected upon its depletion. Notably, Zrf1 is

required for the expression of *Dlx1as* in early differentiation (Figure 26C) as was shown before for Med12 (Figure 23F). The association of Zrf1 and Med12 with *Dlx1as* and their effects in gene expression of developmental genes in early differentiation indicate that the former might be implicated in the remodeling of Med12-*Dlx1as* complexes by displacing PRC1 from chromatin.

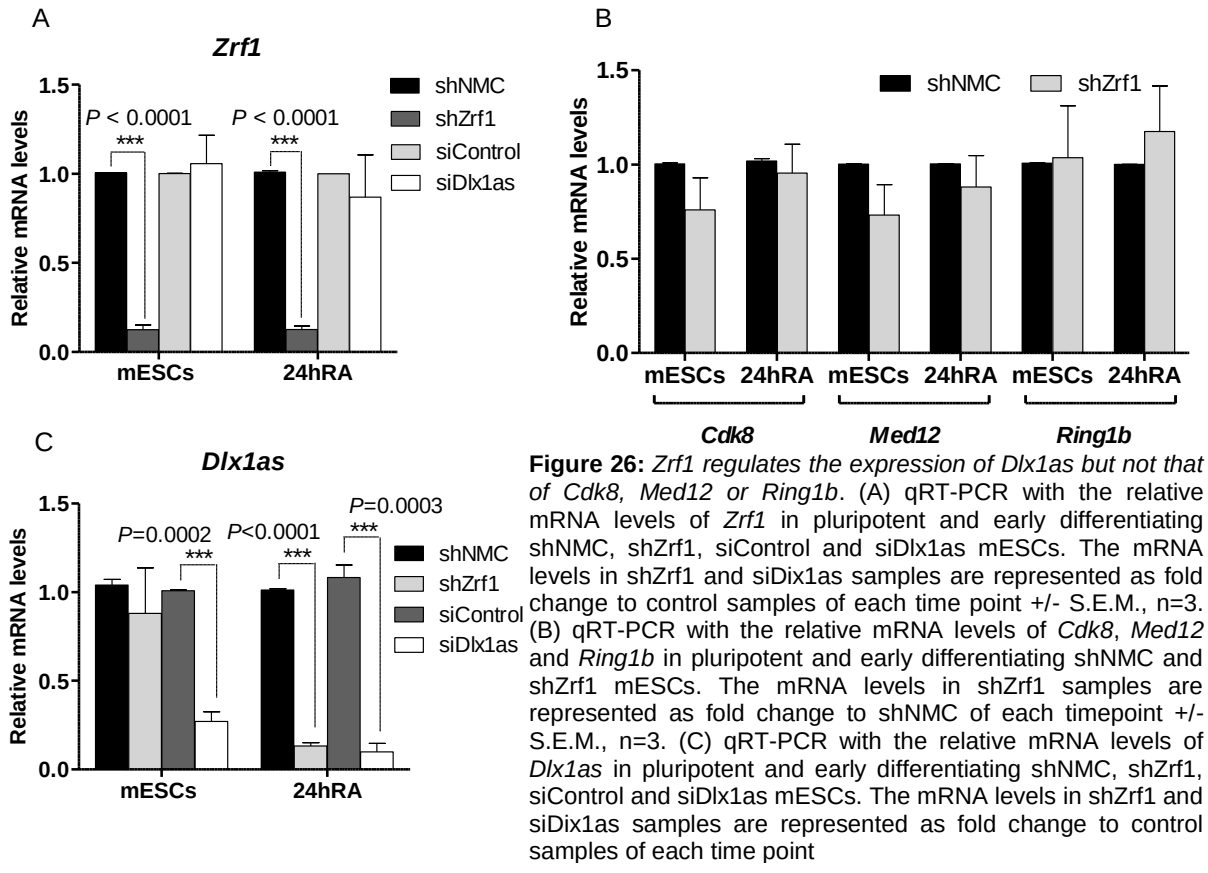


Figure 26: *Zrf1 regulates the expression of Dlx1as but not that of Cdk8, Med12 or Ring1b*. (A) qRT-PCR with the relative mRNA levels of *Zrf1* in pluripotent and early differentiating shNMC, shZrf1, siControl and siDlx1as mESCs. The mRNA levels in shZrf1 and siDix1as samples are represented as fold change to control samples of each time point +/- S.E.M., n=3. (B) qRT-PCR with the relative mRNA levels of *Cdk8*, *Med12* and *Ring1b* in pluripotent and early differentiating shNMC and shZrf1 mESCs. The mRNA levels in shZrf1 samples are represented as fold change to shNMC of each timepoint +/- S.E.M., n=3. (C) qRT-PCR with the relative mRNA levels of *Dlx1as* in pluripotent and early differentiating shNMC, shZrf1, siControl and siDlx1as mESCs. The mRNA levels in shZrf1 and siDix1as samples are represented as fold change to control samples of each time point

Zrf1 assembles the Kinase Module of Mediator during differentiation

The finding that Zrf1 associates with *Dlx1as* in early differentiation and activates *Hoxd11* in a ncRNA-dependent manner, just like Med12 does, raised the possibility that Zrf1 might promote the remodeling of the Med12-Ring1b-*Dlx1as* association that is formed in pluripotent mESCs. In order to test this hypothesis we first analyzed the chromatin dynamics of Zrf1 from pluripotency to early differentiation. By doing cell fractionations and analyzing the chromatin fraction we found Zrf1 to peak at chromatin in early RA-induced differentiation (Figure 27A). As expected, upon Zrf1 depletion, both Ring1b and H2AK119ub were retained at chromatin after RA treatment of

mESCs. Although Zrf1 had no effect on the levels of Med12 at chromatin, we noticed that the recruitment of Cdk8 was significantly reduced upon Zrf1 depletion. Not only Zrf1 reduces the recruitment of Cdk8 globally (Figure 27A), but also at the promoter of *Hoxd11* (Figure 27C). On the other hand, Cdk8 depletion had no effect on the recruitment of Zrf1 at chromatin, highlighting the regulatory role of the latter in early differentiation of mESCs (Figure 27B).

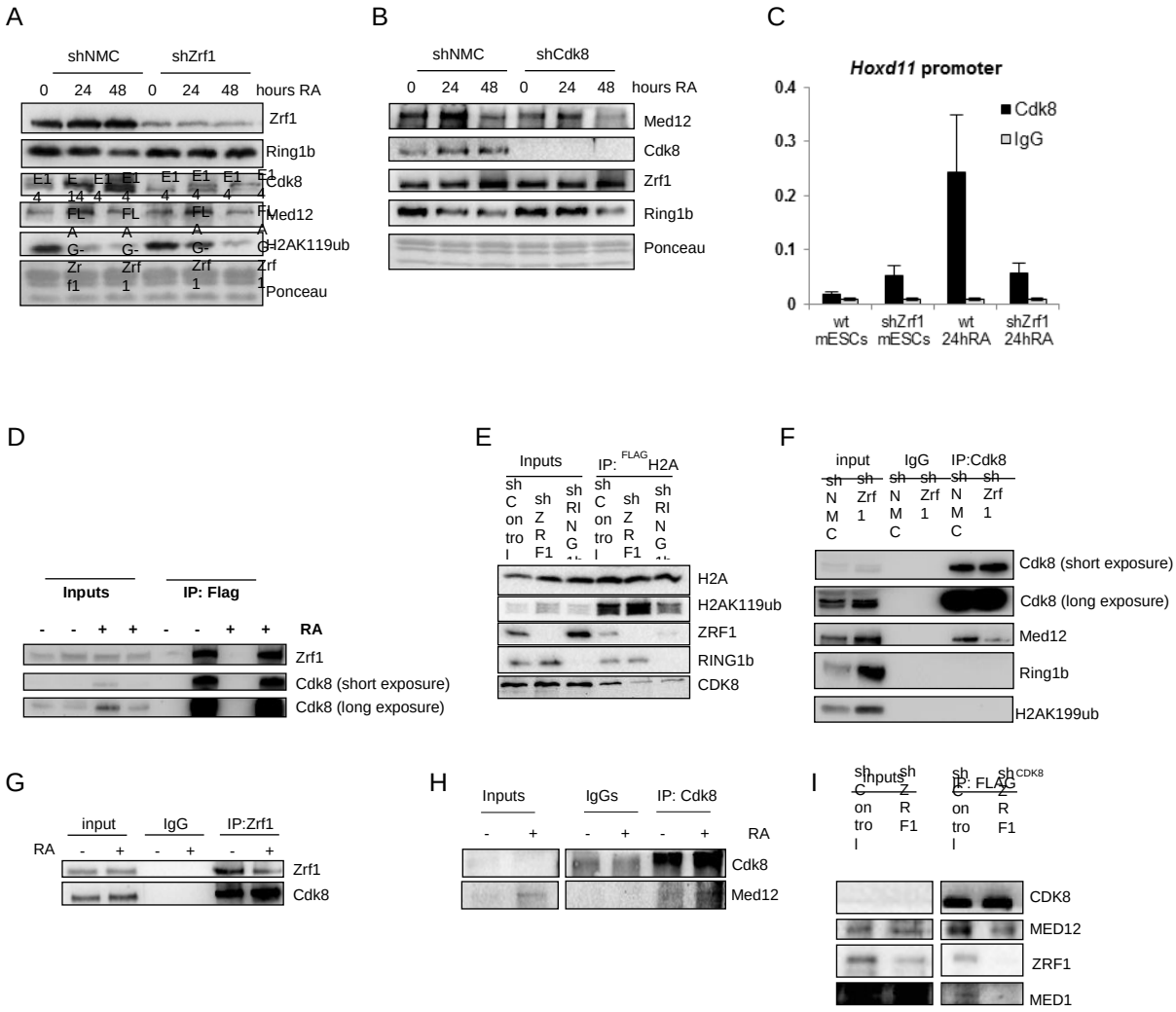


Figure 27: *Zrf1 interacts with Cdk8 and recruits it at chromatin in early mESC differentiation*. (A), (B) Cell fractionations were carried out from pluripotent and early differentiating shNMC, shZrf1 and shCdk8 mESCs. The chromatin fraction was used for western blot analysis with the indicated antibodies and ponceau as a histone balancing control. (C) ChIP with Cdk8 and IgG as isotype control in wt and shZrf1 pluripotenct and early differentiating mESCs. The enrichment at the *Hoxd11* promoter is depicted as percentage of recovery over the input sample +/- S.E.M. (D) Flag IP from pluripotent and early differentiating wt (E14) and FLAG-Zrf1 (E14 FLAG-Zrf1) expressing mESCs. Inputs represent 7% of the material used for the IP. The samples were used for western blot analysis and incubation with the indicated antibodies. (E) Flag IP from shControl, shZRF1 and shRING1B HEK293T cells,

transiently transfected with FLAG-H2A. The inputs correspond to 7% of the material used for each Flag IP. The samples were used for western blot analysis and incubation with the indicated antibodies. (F) Endogenous Cdk8 IP from early differentiating (24hRA) shNMC and shZrf1 mESCs, using IgG as an isotype control. Inputs correspond to 7% of the material used for each IP. Samples were used for western blot analysis and incubation with the indicated antibodies. (G), (H) Endogenous Zrf1 and Cdk8 IPs, using IgG as isotype control, from pluripotent and early differentiating mESCs. Inputs represent 7% of the material used for each IP. Samples were analyzed by western blot with the indicated antibodies. (I) Flag IP from shControl and shZRF1 HEK293T cells transfected with FLAG-CDK8. The inputs correspond to 7% of the material used for each IP. The samples were used for western blot with the indicated antibodies.

Because of the role of Zrf1 in the disassembly of the Ring1b-Med12-*Dlx1as* association that results in the activation of *Hoxd11* in early differentiation, we then asked whether the recruitment of Cdk8 by Zrf1 is also involved in this process. We observed a robust interaction of Zrf1 with Cdk8 both in pluripotent and early differentiating mESCs, using Flag-tagged cells (Figure 27D) or by endogenous IPs with Zrf1 antibodies (Figure 27G). Moreover, CDK8 was also found to interact with ZRF1 in human cells (HEK293T) (Figure 27I), showing that this interaction is conserved in mice and humans. By doing FLAG IPs is cells transfected with H2A-FLAG plasmids, we observed that the recruitment of CDK8 at chromatin (H2AK119ub-containing nucleosomes) is dependent on ZRF1 (Figure 27E).

As we earlier observed, the association between Ring1b and Med12 decreases in the course of differentiation (Figure 27D). By doing endogenous IPs it was shown that although Med12 interacts with Cdk8 already in pluripotency (Figure 27H), the interaction increases in early differentiation. Because of the recruitment of Cdk8 at the same timepoint on chromatin, it was speculated that Zrf1 might be involved in the assembly of Med12-Cdk8 in differentiating mESCs. This was confirmed by endogenous Cdk8 IPs from shNMC and shZrf1 differentiating cells (Figure 27F) as well as HEK293T cells (Figure 27I), both showing a decreased interaction between Med12/MED12 and Cdk8/CDK8 upon Zrf1/ZRF1 depletion. Notably, in HEK293T cells CDK8 does not co-precipitate with MED1, from the Core Mediator, upon ZRF1 depletion, indicating that the assembly of the Holoenzyme Mediator possibly lies in the presence of ZRF1 on chromatin.

Epigenetic regulation by Zrf1 on the *Hoxd11* locus

After showing that Zrf1 is crucial for the activation of PcG targets and specifically of *Hoxd11* in early stem cell differentiation, we sought to address how it influences the chromatin dynamics during this process. We therefore carried out ChIP-qPCR experiments for the presence of activating histone modifications on the promoter of *Hoxd11* (Figure 28A-C). We observed that *Hoxd11* is enriched for both H3K4me1 (Figure 28A) and H3K27ac (Figure 28B) marks which are indicative of active and/or poised enhancers (Shlyueva et al., 2014). Notably the enrichment for both marks was highly compromised upon depletion of Zrf1. We observed a similar reduction

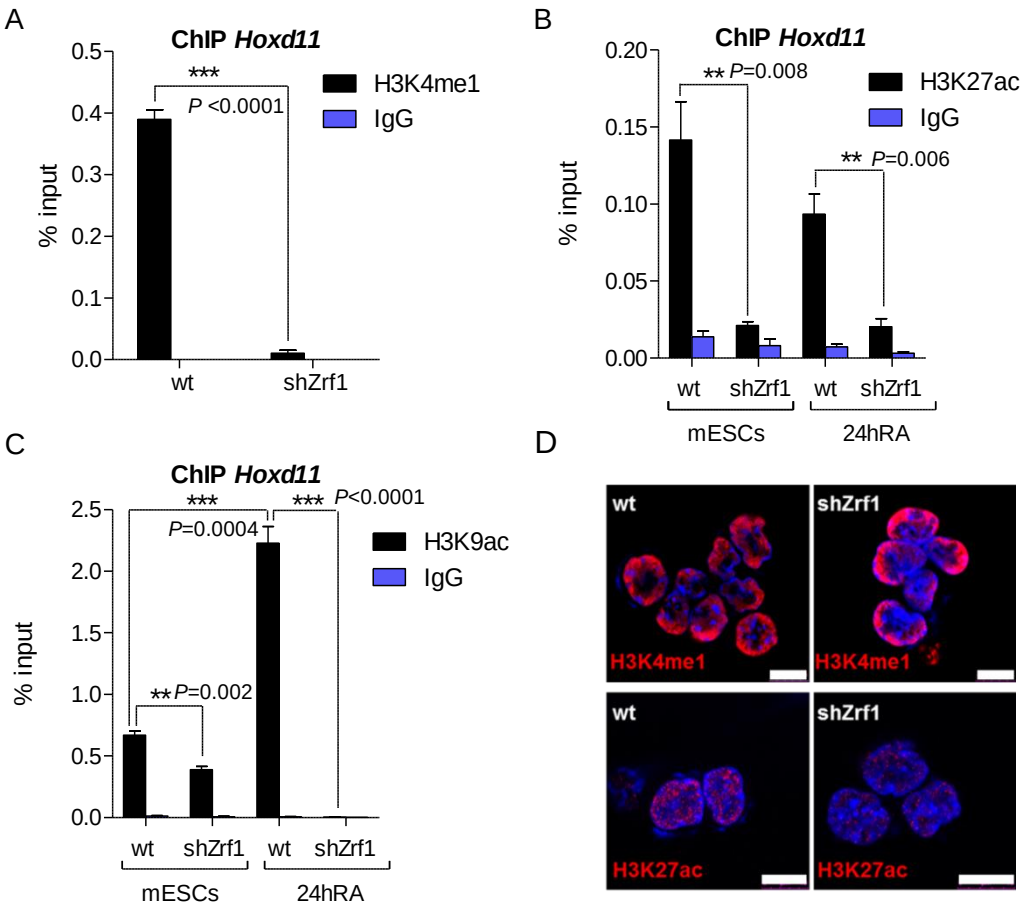


Figure 28: *The deposition of activating histone marks on the promoter of Hoxd11 is Zrf1-dependent*. ChIP-qPCR for H3K4me1 (A), H3K27ac (B) and H3K9ac (C), using wt and Zrf1-depleted pluripotent and early differentiating mESCs.The levels of the histone modifications and IgG, as isotype control, are represented as percentage of recovery over input. The indicated *P* values were calculated by two-tailed, unpaired t test. (D) Immunofluorescent staining with H3K4me1 (upper panel) and H3K27ac (lower panel) of wt and shZrf1 mESCs, followed my confocal microscopy. The pictures were acquired with the same leiser intensity and settings and processed with the same parameters using ImageJ. The white bars correspond to 10µm.

in the levels of both histone marks after immunostaining with the respective antibodies (Figure 28D), suggesting that Zrf1 has a more global role in gene activation. The fact that both marks tested are primarily indicative of poised/active enhancers could imply that Zrf1 leads to gene activation in early differentiation by regulating the activity of respective enhancers already in pluripotency.

We then checked for the presence of H3K9ac at the same promoter, which is a mark highly connected with transcriptional activation (Zhou et al., 2010). We noticed that the levels of H3K9ac increased more than twofold at the promoter of *Hoxd11* from pluripotency to differentiation (Figure 28C), correlating with the upregulation of *Hoxd11* expression at the same timepoint (Figure 22D). More importantly, in Zrf1-depleted early differentiating cells, the enrichment of this activating mark was completely abolished, indicating the importance of Zrf1 in the epigenetic control of transcriptional activation.

After showing that Zrf1 is a novel RNA-binding protein and participates in the activation of *Hoxd11* in a ncRNA-dependent manner, we asked whether RNA is also involved in the recruitment of Zrf1 at chromatin. To this end, pluripotent and early differentiating cells were pre-fixed, permeabilized and then treated or not with RNaseA. After a chromatin association assay which is a variation of cell fractionation including a crosslinking step, the chromatin fraction was analyzed by immunoblot (Figure 29A). We observed a substantial reduction in the chromatin-bound levels of Zrf1 after RNaseA treatment, indicating that its recruitment at chromatin is at least partially dependent in the presence of RNA. The pattern of Cdk8 recruitment was similar however the levels of Ring1b were not altered in the RNA-depleted samples.

We further investigated the impact of RNA in the recruitment of Zrf1 to chromatin by overexpressing in HEK293T cells full-length FLAG-ZRF1, FLAG-ZRF1ΔSANT and each of the three point mutants in the SANT domain and checked for chromatin modifications in the co-precipitate after performing FLAG immunoprecipitations (Figure 29B).

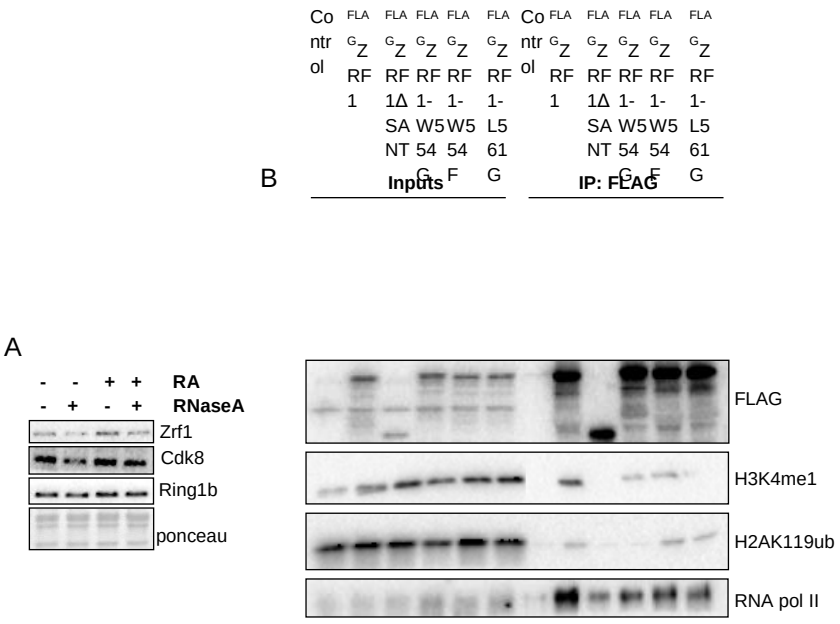


Figure 29: *The recruitment of Zrf1 to chromatin is RNA-dependent*. (A) Chromatin association assay after RNaseA treatment of of pluripotent and early differentiating mESCs. The chromatin pellet was used for immunoblot with the indicated antibodies using ponceau staining as a balancing control. (B) Immunoblots with FLAG-IPs after transfection of HEK293T cells with control, FLAG-ZRF1, FLAG-ZRF1ΔSANT, FLAG-ZRF1W554G, FLAG-ZRF1W554F or FLAG-ZRF1L561G plasmids. Inputs represent 7% of the material used for each IP.

Deletion of the whole SANT domain seems to abolish the capacity of ZRF1 to interact with chromatin as indicated by the absence of both H3K4me1 and H2AK119ub from the co-precipitate. Furthermore, in the absence of the SANT domain, ZRF1 co-precipitates with much lower levels of RNA poll II in comparison to the full-length protein. Each of the three point mutants of the SANT domain affect the binding of ZRF1 to histone modifications or RNA pol II in diverse ways. All three decrease the interaction with RNA pol II, however the substitution of tryptophan (W) to phenylalanine (F) does not abolish completely the binding to chromatin. The conversion of (W) to glycine (G), reduces the interaction of ZRF1 with H2AK119ub and more importantly, the L561G mutant, which abolishes the RNA-binding capacity of ZRF1, leads to loss of H3K4me1 from the eluate. These observations collectively indicate that the binding of ZRF1 to chromatin, as shown via the interactions with histone modifications, is highly correlated with its ability to bind to RNA. This property also influences the recruitment of downstream factors to chromatin by Zrf1, as shown for Cdk8, as well as the interaction with the transcriptional machinery, as observed for RNA pol II.

ZRF1 interacts with transcription and mRNA processing and factors

The FLAG immunoprecipitations with the different constructs for ZRF1 (Figure 29) showed that its RNA-binding capacity affects the interaction with chromatin as well as with other proteins. To investigate this further, the eluates of the FLAG IP were used for mass spectrometry. The obtained data were analyzed by first excluding the peptides which were also identified in the control eluate, being considered as background. Following this step the proteins were ranked according to the identified peptide number. This ranking was carried out for each of the samples separately and the top hits of ZRF1 were then compared with those of the mutant samples. This comparison would enable us to estimate the protein interactions that got lost upon abolishment of the RNA binding capacity of ZRF1. However such analysis proved to be misleading as the experiment was carried out without labelling (label-free MS) making it impossible to draw any unbiased conclusions.

Thus, the analysis was focused only on the top hits of the full length ZRF1 co-IP (Table 11). The 15 top hits, in respect to the number of unique peptides, are proteins mostly related with the cytosolic function of ZRF1 as chaperone, as they are mostly ribosomal or ribonucleoprotein complex components. In order to assess how ZRF1 is related with these co-precipitated proteins, we used the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) tool (Szklarczyk et al., 2015) to create an interacting network (Figure 30A). ZRF1 (DNAJC2) is predicted to interact with confidence with the heatshock protein, HSPA14 and the nucleolar protein, NOP56. As this analysis did not reveal any new insights into the function of ZRF1, apart from the fact that it is highly correlated with the nucleolus, we broadened our approach by focusing in all unique identified peptides.

Table 11: Top 15 hits of ZRF1 protein interactors as identified by mass spectrometry analysis

Protein	Symbol	Control FLAG IP (unique peptides)	ZRF1 FLAG IP (unique peptides)
Dnaj homologue subfamily C member 2	DNAJC2 (ZRF1)	0	47
Developmentally-regulated GTP-binding protein 1	DRG1	0	16
60S ribosomal protein L4	RPL4	0	16
Putative ATP-dependent RNA helicase DHX30	DHX30	0	14
Myb-binding protein 1A	MYBBP1A	0	13
60S ribosomal protein L7a	RPL7A	0	13
Heat shock 70 kDa protein 14	HSPA14	0	12
Protein RRP5 homolog	PDCD11	0	11
40S ribosomal protein S2	RPS2	0	11
40S ribosomal protein S9	RPS9	0	11
N-acetyltransferase 10	NAT10	0	10
Nucleolar protein 56	NOP56	0	9
Ribosomal RNA processing protein homolog B	RRP1B	0	9
Splicing factor 3B subunit 2	SF3B4	0	9
60S ribosomal protein L5	RPL5	0	9

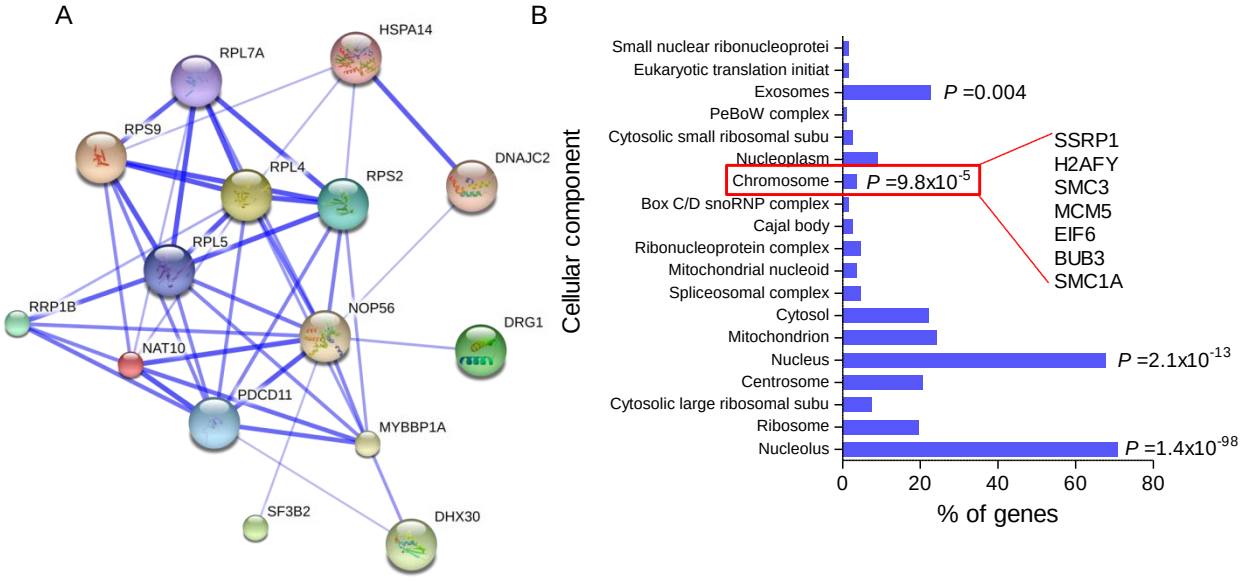


Figure 30: *Protein-protein interaction analysis of ZRF1*. (A) The top 15 hits of unique peptides identified by mass spectrometry after ZRF1-flag immunoprecipitation were used for interaction analysis via the STRING tool (Szklarczyk et al., 2015). The interaction network shows a confidence view of the predicted interactions. The stronger associations are represented by thicker lines. (B) Cellular component enrichment analysis from the ZRF1-interacting unique peptides (192). The bas represent the percentage (%) of genes from the dataset which are enriched for each term. The plot contains the most significantly enriched terms as calculated by Bonferroni Hochberg P-value correction.

We used 192 unique peptides for functional enrichment analysis by the FunRich tool (Pathan et al., 2015) and plotted the most significantly enriched cellular component terms (Figure 30B). The significance was automatically calculated by the tool parameters using the Bonferroni Hochberg P-value correction method. The most significantly enriched term in respect to cellular component was nucleolus, in agreement with the previous analysis with the 15 top hits of the list. Another significant term was nucleus and most importantly, chromosome. Within the chromosome term, the identified peptides included proteins of the cohesin complex (SMC1A and SMC3), proteins related with cell cycle regulation (MCM5 and BUB3) as well as the FACT complex protein, SSRP1.

The identification of two cohesin complex proteins among the interactors of ZRF1, is of great importance, as this complex is not only highly connected with genome architecture and gene expression but is also a described interactor of Mediator (Kagey et al., 2010). More importantly, Smc1a is co-bound with Med12 at enhancer and promoter regions of highly expressed genes in mESCs (Kagey et al., 2010), implying that ZRF1 could be participating in gene expression through interacting with proteins that connect the latter with chromatin architecture. Based on this observation and in an attempt to find more common interactors between ZRF1 and MED12, we searched for described interactions of Med12 using the online tool HIPPIE (Schaefer et al., 2012). By comparing the two lists we found two common proteins; POLR2E and CDC5L. POLR2E codes for RPABC1 subunit of the DNA-directed RNA polymerases I, II and III, and CDC5L which acts as a positive regulator of cell cycle progression and is a component of a non-snRNA spliceosome.

The observation that MED12 and ZRF1 have common interactors in HEK293T cells raised the possibility that these interactors are also related to CDK8. This hypothesis was based on the regulatory role of Zrf1/ZRF1 towards Cdk8/CDK8 in respect to its recruitment at chromatin. To explore this we compared the 192 unique peptides identified in the ZRF1-flag IP with known

interactors of CDK8, as listed in the HIPPIE database. The common proteins were then analyzed in FunRich and the most significantly enriched terms in respect to cellular component were plotted (Figure 31A). Interestingly, the two enriched terms cohesin core heterodimer and kinetochore include the proteins SMC1A, SMC3 and SMARCA4.

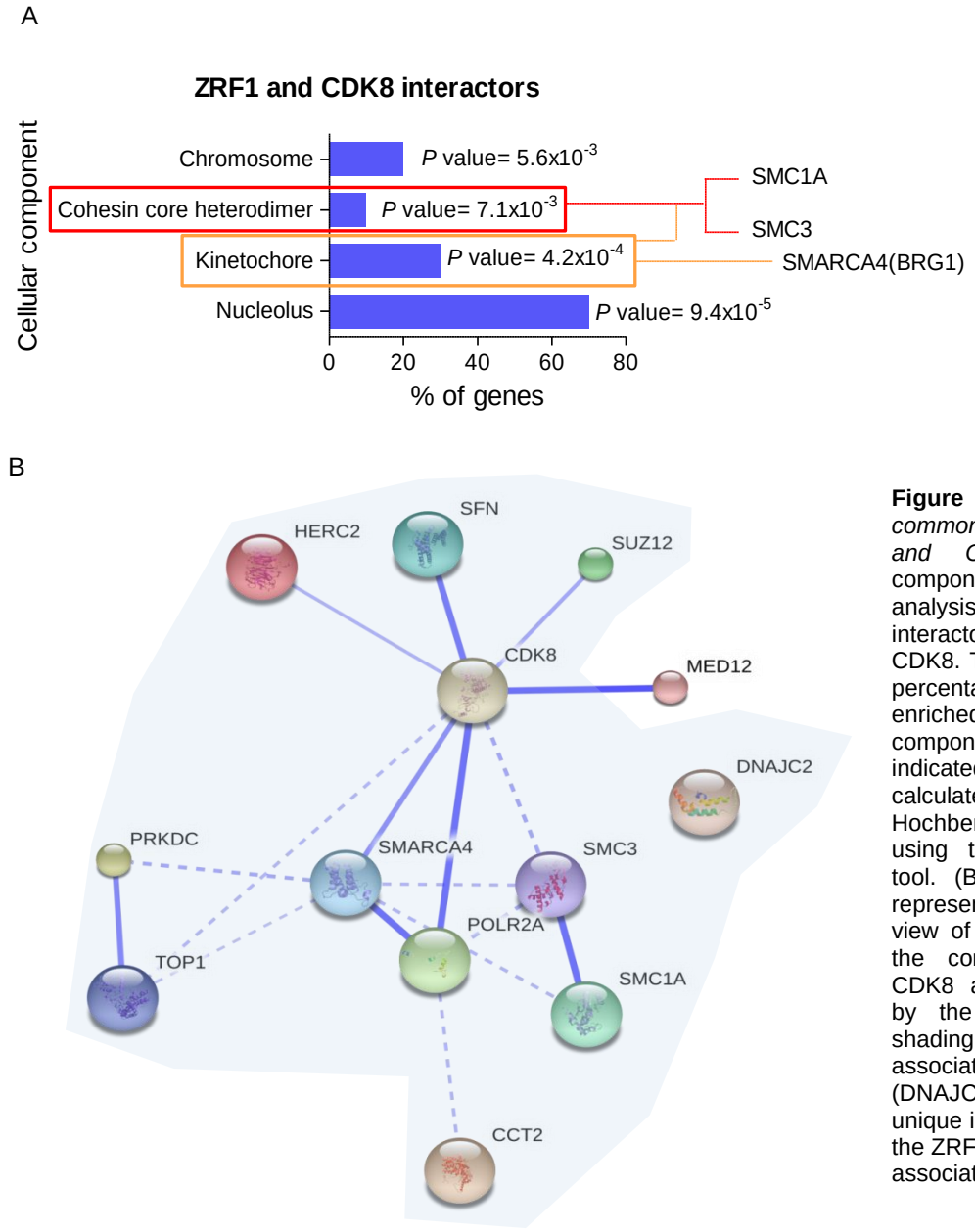


Figure 31: *Analysis of the common interactors of ZRF1 and CDK8*. (A) Cellular component enrichment analysis of the common interactors of ZRF1 and CDK8. The bars represent the percentage (%) of genes enriched in each cellular component term. The indicated *P* values were calculated by Bonferroni Hochberg P-value correction using the FunRich analysis tool. (B) Interaction network representing the confidence view of associations between the common interactors of CDK8 and ZRF1, generated by the STRING tool. The shading encompasses the associations of ZRF1 (DNAJC2), based on the unique identified peptides from the ZRF1-flag IP as well as the association with CDK8.

SMARCA4, also known as BRG1, has an ATPase activity and regulates transcription by forming the catalytic core of various chromatin modifying complexes including the SWI/SNF, NCoR, mSin3A/HDAC (Khavari et al., 1993; Trotter and Archer, 2008). BRG1, as well as SUZ12 which

is another common interactor of ZRF1 and CDK8, were identified as CDK8-interacting proteins in a yeast two-hybrid screen (Fukasawa et al., 2012).

In an attempt to understand whether cohesin components and other CDK8-ZRF1 interactors might provide a link between ZRF1, CDK8 and MED12, the common interactors of ZRF1 and CDK8 as well as ZRF1, CDK8 and MED12 were analyzed via STRING to predict a possible network between these proteins (Figure 31B). In the confidence view of the interaction network, the stronger associations are represented as thicker blue lines. Notably MED12 interacts with confidence only with CDK8. ZRF1 (DNAJC2), can not be associated with confidence to any of the proteins of the network, however its association with CDK8 and the unique peptides, identified by mass spectrometry, are highlighted with blue shading.

Based on the protein interaction analysis of ZRF1, CDK8 and MED12 and in combination with the previous findings, it can be concluded that ZRF1 might be participating in gene activation via regulating the subunit composition of the Mediator Complex in a mechanism which appears to be coordinated with chromatin architecture and the action of chromatin remodeling complexes.

Cdk8 has a very small impact in gene expression during differentiation

The RNA-seq of pluripotent mESCs showed a quite small impact in gene expression after Cdk8 depletion (Figure 18A, B) which was also confirmed by the qPCR validations for the Med12-Ring1b targets (Figure 18G). Likewise, the expression levels of developmental genes in shCdk8 early differentiating mESCs (24hRA) were similar to those of the shNMC samples (Figure 21B). These observations seem to contradict a possible role of Cdk8 in the expression of differentiation genes after its recruitment by Zrf1 to chromatin, especially because its depletion did not affect significantly the upregulation of *Hoxd11* (Figure 32A) at the same timepoint. In order to get more insight into the function of Cdk8 in early differentiation, we carried out RNA-

seq in shCdk8 early differentiating mESCs. We then obtained the DE genes in comparison to shNMC cells both in pluripotency and early differentiation and analyzed the molecular function of the 20 top hits of each timepoint (Table 12, Figure 32C). Amongst the enriched terms we found

Table 12: Top 20 genes with differential expression in shCdk8 cells compared to shNMC cells from RNA-seq.

Gene ID	Fold Change	log2 Fold Change	p value
mESCs			
Atp1a4	14.13	3.82	0.00
Cdk8	0.24	-2.06	0.00
Emx2	8.85	3.15	0.00
Sh3bgrl	2.78	1.47	0.00
Napsa	2.35	1.23	0.00
Fgfbp1	4.8	2.26	0.00
Prtg	2.14	1.1	0.00
Anxa3	2.47	1.3	0.00
Sohlh2	2.05	1.04	0.00
Parva	2.21	1.14	0.00
Pea15a	2.25	1.17	0.00
Ephx2	0.51	-0.96	0.00
Hhip	2.09	1.07	0.00
Htra1	3	1.59	0.00
Smyd1	2.12	1.09	0.00
Rcsd1	2.83	1.5	0.00
Nlrp4f	2.33	1.22	0.00
Lef1	0.37	-1.43	0.00
Ano9	0.54	-0.9	0.00
Clcnkb	0.33	-1.62	0.00

Gene ID	Fold Change	log2 Fold Change	p value
24hRA			
Cdk8	0.22	-2.21	0.00
Nsg2	6.52	2.7	0.00
Trh	5.01	2.33	0.00
Il33	3.11	1.64	0.00
Klhdc7a	3.22	1.69	0.00
Prtg	2.72	1.44	0.00
Pcdh1	2.92	1.54	0.00
Ddx58	2.83	1.5	0.00
Tex13	4.64	2.21	0.00
Atp1a4	10.2	3.35	0.00
Fbn2	2.51	1.33	0.00
Sulf1	2.41	1.27	0.00
Pdha2	0.3	-1.72	0.00
Thbs2	2.98	1.57	0.00
Oasl2	2.98	1.58	0.00
Nlrp4f	2.44	1.29	0.00
Aqp3	0.47	-1.10	0.00
Pcsk9	3.02	1.59	0.00
Zfy1	0.39	-1.35	0.00
Dusp4	2.31	1.21	0.00

molecular functions associated with binding of proteins and nucleic acids (Figure 32C), implying that Cdk8 is involved in gene expression.

Because of the association of Cdk8 with Zrf1 and the Zrf1-dependent recruitment to chromatin, we then sought to investigate whether these two factors affect at all gene expression in a similar manner, conversely to what we observed for *Hoxd11*. To that end we carried out RNA-seq in shZrf1 pluripotent and early differentiating mESCs and then compared the DE genes as a fold change from pluripotency to differentiation (Figure 32B). This analysis enabled us to assess the

effect of Cdk8 and Zrf1 depletion in genes which are expressed at different levels in comparison to control upon transition from pluripotency to differentiation. Although the genes which were DE in both shZrf1 and shCdk8 cells (Figure 32B, red) were much fewer when compared to the DE after either Zrf1 or Cdk8 depletion (Figure 21B blue and green, respectively), the majority of them were either upregulated or downregulated in both samples (Figure 32B upper-right and lower-left quadrants).

In an attempt to connect these common DE genes between Zrf1 and Cdk8 to the target genes of Ring1b-Med12, we compared the two lists and observed no significant similarities (data not shown). This led us to the conclusion that although Cdk8 depletion does not affect gene activation upon stem cell differentiation, in a similar manner as Zrf1 or Med12 depletion does, its recruitment by Zrf1 is required for the remodeling of the Mediator Complex during this transition.

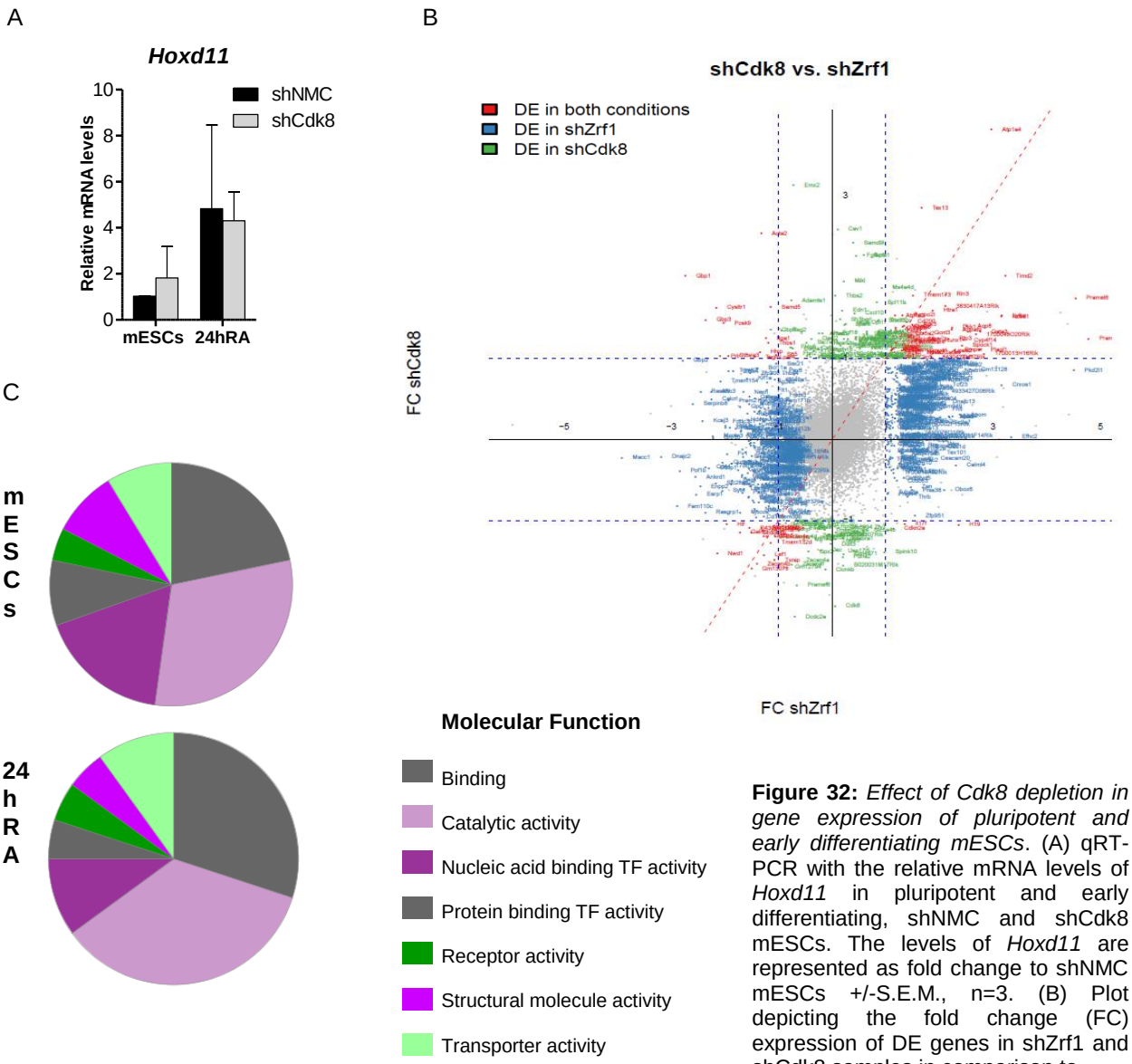


Figure 32: *Effect of Cdk8 depletion in gene expression of pluripotent and early differentiating mESCs*. (A) qRT-PCR with the relative mRNA levels of *Hoxd11* in pluripotent and early differentiating, shNMC and shCdk8 mESCs. The levels of *Hoxd11* are represented as fold change to shNMC mESCs +/-S.E.M., n=3. (B) Plot depicting the fold change (FC) expression of DE genes in shZrf1 and shCdk8 samples in comparison to shNMC. The FC corresponds to the change in the expression levels from pluripotency to differentiation. (C) Pie charts depicting the molecular functions of the 20 best hits of DE genes in shCdk8 pluripotent and early differentiating mESCs (Table 1). The charts were generated by the online database Panther using the functional classification analysis tool (Thomas et al., 2003).

Discussion

PART 1: Common functions of PRC1 and Med12-Mediator in pluripotent mESCs

In this study we identified a class of genes, which code for differentiation and development factors, to be regulated in mESCs by the canonical PRC1 complex in concert with Med12. This finding was the result of our attempt to gain a better understanding of how PRC1-mediated silencing is connected with the transcriptional machinery. Although the molecular functionality of PRC2 and H3K27me3 in epigenetic silencing is highly resolved (Margueron and Reinberg, 2011), that of PRC1 is still under debate (Endoh et al., 2012; Vissers et al., 2008). The E3 ligase function of Ring1 and thus H2AK119ub is required for gene silencing, however PRC1 gets recruited to chromatin and propagates chromatin compaction in a H2AK119ub-independent manner (Endoh et al., 2012; Eskeland et al., 2010). Amongst the various effector mechanisms that have been proposed to describe PRC1-mediated repression, direct inhibition of transcription, either initiation or elongation has been attributed to PRC1. In this line, it was recently suggested that subunits of the Mediator co-activator complex might coordinate repression by PRC1 (Simon and Kingston, 2013).

Apart from Ring1b and Med12, the identified class of silenced genes coding for essential developmental regulators, were also enriched for Cbx7, another PRC1 subunit, as well as for the PRC2 component Suz12. Cbx7-containing PCR1 is the prevalent form of PRC1 that maintains gene silencing in pluripotency (Morey et al., 2012; O'Loghlen et al., 2012) and its recruitment at chromatin is PRC2 and H3K27me3-dependent (Morey et al., 2012). This would imply that the genes regulated by Med12 and Ring1b are under the regulation of a canonical PRC1 complex (Simon and Kingston, 2013). However, the large variability in PRC1 composition (Gil and O'Loghlen, 2014; Vandamme et al., 2011) highlights the importance of identifying additional subunits, that participate in this mode of regulation. Notably, a recent study reported Med12

amongst the Bmi1 interactors in an endogenous immunoprecipitation experiment followed by mass spectrometry, in prostate cancer cells (Cao et al., 2014). Although this finding was not obtained in mESCs, it implies that PRC1 could be interacting with Med12 via another subunit, distinct of Ring1b.

Mediator is a multisubunit complex which enhances transcription via bridging enhancer-bound transcription factors with the general transcriptional machinery. Two subunits of Mediator, Med1 and Med12, regulate chromatin architecture in mESCs by participating in a pluripotency interactome involving cohesin and pluripotency factors such as Nanog and Oct4 (Apostolou et al., 2013; Kagey et al., 2010). Here we report that Ring1b (PRC1) and Med12 are both enriched around promoter regions of classic PcG targets which are silenced in pluripotency. This gene group is much smaller when compared to the group of expressed genes which are targeted by cohesin, Mediator as well as the pluripotency master transcription factors. Nevertheless, the presence of Med12 on these genes might explain the PcG-like phenotype of Med12 mutant flies as described in an earlier study (Gaytan de Ayala Alonso et al., 2007). Notably Med12 is one of the four subunits of the Kinase module, which is not always in complex with the Core Mediator. The Kinase module harbors apart from Med12, Med13, Cdk8 and CycC. The PRC1-Med12 targeted genes in pluripotency were devoid of Cdk8 from the Kinase module, whereas Med12-Cdk8 were enriched in a separate class of genes, at the same stage. These genes are highly expressed and mostly related with metabolic processes. This analysis raised the hypothesis that Med12 might have functions in pluripotency, besides the Kinase Module, related to PRC1-mediated repression.

We explored this possibility and showed that Ring1b associates with Med12 in pluripotency and regulates its presence at chromatin. The latter was shown at a global level by cell fractionations as well as with ChIP experiments on the common Ring1b-Med12 targets. Furthermore, Ring1b depletion compromised the levels of Med12 around the promoter regions of Cdk8-Med12

targets. This implies that Ring1b regulates the binding of Med12 at chromatin on a more global level however this issue could not be addressed by genome-wide mapping of Med12 occupancy after Ring1b depletion, mostly due to technical issues. It is noteworthy that only one study (Kagey et al., 2010) was successful to date, to perform Med12 ChIP-seq. Although the unavailability of ChIP-grade Med12 antibodies appears as a possible scenario, the multiple interactions of the protein with the transcriptional machinery, other Mediator subunits as well as TFs could be a reason why, upon crosslinking, Med12 cannot be effectively immunoprecipitated.

Ring1b and Med12 seem to be only functionally linked as they do not regulate each other on a transcriptional or post-transcriptional level. Our immunoprecipitation experiments reveal that Med12 can interact with Ring1b and Cdk8, however these interactions are independent to each other, suggesting the presence of two “complexes”; one between Ring1b and Med12 that regulates silenced genes and a second one between Cdk8 and Med12 that is enriched at promoters of highly expressed genes. To clarify the latter additional biochemical studies would be needed. Although the endogenous immunoprecipitations imply that these proteins associate directly or indirectly with each other, *in vitro* approaches would contribute into a further understanding of the nature of these interactions and would indicate whether they correspond to distinct complexes that exist in pluripotency. Nonetheless both approaches need to be combined as *in vitro* biochemical approaches alone, might only reveal static complexes that do not necessarily reflect the heterogeneity of transient complexes that might exist *in vivo* (Clark et al., 2015).

Our hypothesis that Ring1b and Med12 are functionally linked was reinforced by the analysis of differentially expressed genes upon depletion of these factors from mESCs. Although Med12 depletion resulted in upregulated as well as downregulated genes, the former group is related to differentiation and development and corresponds to already described PRC1-silenced genes (Boyer et al., 2006). To understand the common regulatory effects of Med12 and Ring1b, we

integrated the data of their common targets (ChIP-seq) with those from the differentially expressed genes after Ring1b or Med12 depletion (RNA-seq).This analysis resulted in 116 genes which are directly regulated by these factors. The vast majority of these genes are de-repressed upon depletion of either Ring1b or Med12, highlighting the important functions of these factors in regulating the silencing of differentiation genes in pluripotency. Notably, de-repression of developmental TFs after Med12 depletion has been reported before (Kagey et al., 2010), however it was mostly interpreted as an indirect effect due to loss of stem cell identity rather as a direct regulation by Med12 on these genes. In order to understand which of the two scenarios lead to de-repression upon Med12 depletion the question should be addressed in an experimental set up which would separate the two functions.

Noteworthily, the number of direct common targets between Ring1b and Med12, as revealed from our integrative ChIP-seq and RNA-seq analysis, appears as rather small. The ChIP-seq targets for Med12 and Ring1b are ~10.000 and ~5000 respectively. On the other hand, approximately 2000 genes are differentially expressed upon Med12 depletion and around 1000 in shRing1b mESCs. The number of differentially expressed genes corresponds to 1/5 of the targeted genes for each factor which can be attributed to the great complexity of regulation that drives gene expression. As a result when both analyses are integrated in order to correlate the regulatory function of both factors, the number of regulated genes decreases even more. This can be attributed to the fact that after each individual analysis, many genes are discarded post-quantification and filtering as well as *P*-value correction, in order to avoid false positive candidates.

Apart from the regulatory role that has been attributed to Med12 in respect to maintaining ES cell architecture (Apostolou et al., 2013; Kagey et al., 2010), PRC1 and specifically Ring1b has emerged as a master regulator of the genome architecture as well (Cavalli, 2015; Schoenfelder et al., 2015). More specifically PRC1 is essential for properly organizing contacts between

promoters of repressed genes in the three-dimensional nuclear space, which is thought to serve in confining their premature expression while maintaining their poised state (Schoenfelder et al., 2015). Based on these studies, a plausible scenario would involve a communication between those two networks that drive pluripotency; the network of contacts encompassing the highly expressed genes via ES cell regulators, cohesin and Mediator and that of repressed yet poised for activation genes via PRC1. The latter has been partially addressed (Schoenfelder et al., 2015) by comparing the binding profiles of all these factors in mESCs but this study did not identify a substantial overlap between these networks. Nevertheless this question remains open, as the comparison carried out in this study, was based in the analysis of ChIP-seq datasets which were carried out in rather different experimental set ups and by employing sequencing platforms that correspond to different technologies. Moreover, PRC1 subunits were found to physically interact with cohesin in *Drosophila* embryonic nuclear extracts (Strubbe et al., 2011) and to be functionally linked for the regulation of both active and repressed genes in fly imaginal disks (Schaaf et al., 2013). Although in this study cohesin was shown to facilitate the loading of PRC1 at active genes, the presence of those complexes on silenced genes was reported as being mutually exclusive (Schaaf et al., 2013). While the functional interplay between PRC1 and cohesin remains elusive and not studied to date in vertebrates, it appears likely that together with Med12 these complexes establish chromatin architectures around repressed genes that keep repressed genes in a poised state. In line with this hypothesis, PRC1 promotes RNA pol II pausing in repressed loci in mESCs which is thought to contribute in the poised state of these genes (Brookes et al., 2012; Stock et al., 2007).

In our study, we observed that Ring1b is essential for the proper recruitment of Med12 at target genes. Furthermore, we tested whether Med12 could also regulate Ring1b in the same respect but did not detect any significant changes on the levels of Ring1b at chromatin upon Med12 depletion. Interestingly, individual Core Mediator subunits of all three submodules, where shown to be essential for maintaining the intensity of Polycomb foci in *Drosophila* cells (Gonzalez et al.,

2014). Although Med12 was not among the Mediator subunits that regulate the organization of these foci, this study raises the possibility that PRC1 might be regulated by the Core Mediator, not in respect to chromatin occupancy but rather in respect to its localization within the nuclear space. In the same organism, cohesin is thought to promote looping between Polycomb Responsive Elements (PREs) that establish silencing of the developmentally important *inverted-engrailed* gene complex (Schaaf et al., 2013). Although cohesin is not enriched at the promoter regions of silenced genes, it remains to be studied whether it can influence the architecture of PcG-repressed loci by targeting distal regulatory elements such as enhancers.

PART 2: Remodeling of Med12-Mediator enhances gene activation in early stem cell differentiation in a ncRNA-dependent manner

Our results link the functions of Ring1b and Med12-Mediator in the regulation of gene silencing in mESCs. As a consequence of this function, the expression of Med12-Ring1b target genes is altered in early differentiation after depletion of these factors. Med12 and Ring1b knock down mESCs show premature expression of a subset of genes in early differentiation (*Wnt8a*, *Snai1* and *Meis1*) which is an indication of a concerted action of these proteins also at this state. On the other hand a separate subset of genes, including mostly *Hox* genes, has reduced expression in early differentiation upon Med12 depletion while Ring1b depletion results in increased mRNA levels. This finding suggested that in early differentiation a subset of differentiation genes is regulated by Med12 and Ring1b but each protein affects gene expression in opposing ways; Med12 is required for the activation of these genes whereas Ring1b restricts their expression.

In order to gain insight into the disconcerted regulation by Med12 and Ring1b in early differentiation we focused on one known PcG target gene, *Hoxd11*. Med12 depleted early differentiating cells, have reduced levels of *Hoxd11*. Notably Med12 was earlier reported to be required for the expression of *Hoxd11* in the pectoral fin of *zebrafish* (Muto et al., 2014). This

finding seems rather interesting as *Hoxd11* is essential for the formation of the tetrapod forelimb, an organ homologous to the pectoral fin of fishes (Zuniga, 2015). Although our analysis was carried out in an *in vitro* system, our expression data agree with findings supported by experiments on an organismal level, showing that such type of regulation can be recapitulated in cell culture and is evolutionarily important for vertebrate development.

Based on findings that describe Med12 to enhance gene expression by association with eRNAs (Lai et al., 2013), we identified a lncRNA, *Dlx1as*, as a promising candidate to regulate the activation of *Hoxd11*. Depletion of this lncRNA is linked with abnormal skeletal and neurological phenotypes in mouse and proposed to regulate the protein-coding transcript *Dlx1* in an inverse manner (Kraus et al., 2013). Conversely in a separate study, *Dlx1* and *Dlx1as* were found to be expressed in a similar manner in differentiating EBs as well as in adult mouse neural tissues (Dinger et al., 2008). Interestingly, *Hoxd11* is also involved in the development of the vertebrate skeletal system, as it is essential for the proper formation of sacral vertebrae during early mouse development (Favier et al., 1995). Notably, upon depletion of *Dlx1as* we observed a reduction in the mRNA levels of *Hoxd11* in early RA-induced differentiation but not at the levels of *Dlx1*. The common effect of *Dlx1as* and Med12 depletion in the expression of *Hoxd11*, as well as the fact that *Dlx1as* and Med12 interact, suggested that this locus could be regulated in a Med12-*Dlx1as*- dependent manner. As *Hoxd11* is targeted by PRC1 for silencing in pluripotency, we asked whether Ring1b is also implicated in the ncRNA-dependent regulation of this locus. Ring1b associates with *Dlx1as* in pluripotency, however this association decreases upon differentiation. We concluded that Ring1b is required to assemble Med12-ncRNA containing complexes in pluripotency, while *Hoxd11* is repressed, which are later required for the activation of the locus in early differentiation. This conclusion is based on the following observations: 1) Ring1b and Med12 are enriched on the promoter of *Hoxd11* in pluripotency, while this gene is silenced, 2) Ring1b associates with Med12 in pluripotency in an RNA-dependent manner, 3) while Med12 remains bound to *Dlx1as* in early differentiation Ring1b loses its affinity for the

lncRNA and 4) in Ring1b depleted mESCs, the association between Med12 and *Dlx1as* decreases.

Thus, Ring1b not only silences this gene in pluripotency, but is also required for setting the "ground" for the proper activation of *Hoxd11* in early differentiation by recruiting Med12 to chromatin as well as by bridging it to *Dlx1as*. In order to fully prove this mode of regulation on *Hoxd11*, it should be asked whether the recruitment of Med12 and Ring1b at the promoter of *Hoxd11* is affected upon depletion of *Dlx1as*. Furthermore, if Med12 and *Dlx1as* regulate *Hoxd11* as proposed by Lai et al., 2013, it would be required to confirm the formation of chromatin looping between the expression units of *Dlx1as* and *Hoxd11* which ideally would be abolished upon Med12 depletion.

Lately the definition of lncRNAs and eRNAs has been challenged (Espinosa, 2016; Paralkar et al., 2016). More specifically it was reported that the lncRNA *Lockd* is not required as a molecule for the regulation of its nearby gene *Cdkn1b* and it is rather the locus that acts as an enhancer, providing a platform for the binding of TFs (Paralkar et al., 2016). Therefore it still has to be proven whether *Dlx1as* as an RNA molecule or the *Dlx1as* locus is required for enhancing the expression of *Hoxd11* in early differentiation. Interestingly the intron of *Dlx1as* and the region upstream of its first exon, contain well-conserved enhancer elements (Gonzales-Roybal and Lim, 2013), making it plausible that *Dlx1as* is an eRNA rather than a functional ncRNA. Besides the need of further experimental proof, the idea that PRC1 establishes Med12-ncRNA complexes which are thought to activate genes by looping, would be in line with the proposed roles of PRC1 in regulating the 3D architecture of repressed loci in pluripotency as well as with its physical interaction with cohesin in *Drosophila*.

It was shown earlier that Zrf1 gets recruited on chromatin in early differentiation in an H2AK119ub-dependent manner, which leads to the displacement of Ring1b and eventually to de-repression of PRC1-repressed genes (Richly et al., 2010). Here we showed that Zrf1

participates in the activation of *Hoxd11* by associating in early differentiation with *Dlx1as*. Notably, at the same time that Ring1b loses its affinity for the ncRNA and chromatin, Zrf1 gets recruited to chromatin and binds *Dlx1as*. We showed that Zrf1 participates in the remodelling of the Kinase Module of Mediator by recruiting Cdk8 to chromatin in early differentiation and more specifically on the promoter of *Hoxd11*. This recruitment is H2AK119ub-dependent and in the absence of Zrf1, we observed reduced association between Med12 and Cdk8 in early differentiation. The dependence of the recruitment of Cdk8 on Zrf1, can also explain the reduced levels of Cdk8 on chromatin in shRing1b cells; in the absence of Ring1b and thus H2AK119ub, there is no recruitment of Zrf1 to chromatin and as a consequence no recruitment of Cdk8.

On the other hand, depletion of Cdk8 did not alter the expression levels of *Hoxd11* as would be expected. Moreover we noticed a very small impact for Cdk8 in global gene expression. Although the latter could be attributed to compensation of the function of Cdk8 by its paralog Cdk19 (Conaway and Conaway, 2013), it is also a possible scenario that the effects of Cdk8 depletion in expression were not measurable in our experimental set up. For transcriptional regulators, including Cdk8 (Andrau et al., 2006) and the SAGA complex (Bonnet et al., 2014), there is often little overlap between genome-wide occupancy and effects in gene expression. In the case of the SAGA complex, which is only transiently bound to DNA, it was only showed by nascent RNA-seq that it indeed affects global gene expression. Although Cdk8 might be a different case, it should not be ruled out that it could behave in a similar fashion as the SAGA complex in respect to gene expression. Thus, it would be worth re-assessing the role of Cdk8 in gene expression by taking into account both the existence of Cdk19 as well as the speed of the association of Cdk8 with DNA.

Although we could not experimentally verify the importance of the recruitment of Cdk8 by Zrf1 to chromatin, at least in respect to gene expression, it can be speculated that the assembly of the Kinase Module by Zrf1, is required to bypass PRC1-mediated repression. As previously

discussed, PRC1 promotes RNA pol II pausing (Brookes et al., 2012; Stock et al., 2007) and this is considered to be a rate-limiting step in transcriptional activation. Since the transition from pausing to transcriptional elongation is highly dependent on the recruitment of P-TEFb, Zrf1 could participate in the latter by facilitating the recruitment of Cdk8 at chromatin. This hypothesis is based on the fact that Cdk8 gets recruited to chromatin by the HIFa (Hypoxia Induced Factor 1) to promote elongation by recruiting P-TEFb (Galbraith et al., 2013). According to this scenario, like HIFa upon hypoxia bypasses pausing via the recruitment of Cdk8, Zrf1 in response to RA could act in a similar fashion.

We described ZRF1 as a novel RNA-binding protein and showed the importance of the SANT domains in the interaction with RNA and more specifically of L561 which resides within the second SANT domain. Although the tryptophan mutations to either glycine (W554G) or phenylalanine (W554F) did not compromise the ability of ZRF1 to precipitate RNA, they reduced the interaction of the protein with H3K4me1 as well as H2AK119ub. These findings are in agreement with the reduced recruitment of Zrf1 to chromatin after RNase treatment as well as with an earlier study that suggests a role for the tryptophan repeats and flanking amino acids in interactions of Myb domains with DNA (Saikumar et al., 1990). Myb-related proteins harbor DNA-binding domains which are highly similar to the SANT domain (Grune et al., 2003). However, it has been proposed that these two domains might be functionally divergent, mainly because conserved amino acids essential for the binding of Myb to DNA are absent from the SANT domain (Boyer et al., 2004). Structural domain analysis (Lemak at al., 2013) suggests that the second SANT domain of ZRF1, participates in a secondary structure which involves three helices (Figure 33A).

As leucine is one of the most typical residues found in alpha-helices and glycine is a known “breaker” of such structures it can be speculated that the L561G mutation abolishes the RNA-binding capacity of ZRF1 by disrupting this first alpha helix of the second SANT domain. FoldX

prediction (De Baets et al., 2012) which calculates differences in free energy caused by mutations, suggests that L561G causes a 2.59 kcal/mol ddG (delta delta G) increase, implying a reduction in protein stability. Interestingly the W554G mutation leads to a much higher degree of destabilization of the peptide, 6,29 kcal/mol ddG, although this disruption appears not to involve the nearby alpha helix (Figure 33C). At last but not least the W554F mutation affects at a very low level the stability of the protein (Figure 33D). The *in silico* analysis of the mutation effects on ZRF1 imply that the first alpha helix of the second SANT domain of the molecule is required for the binding to RNA. Although the mutation of a tryptophan is predicted to heavily destabilize the peptide and affect its structure, this residue does not appear essential for the binding of ZRF1 to RNA, yet affects its affinity towards histone modifications. Further experimental structural studies are required in order to fully elucidate whether the stability of this alpha-helix within the second SANT domain of ZRF1 is essential for the ability of the protein to bind RNA. Future experiments should be carried out by taking into account the increased conservation of the second ZRF1 SANT domain when compared to the first (Chen et al., 2014) as well as the fact that the two SANT domains act synergistically (Boyer et al., 2004) but also carry out separate functions at least in HDACs (Histode De-Acetylase Complexes) (Guenther et al., 2001; Yu et al., 2003).

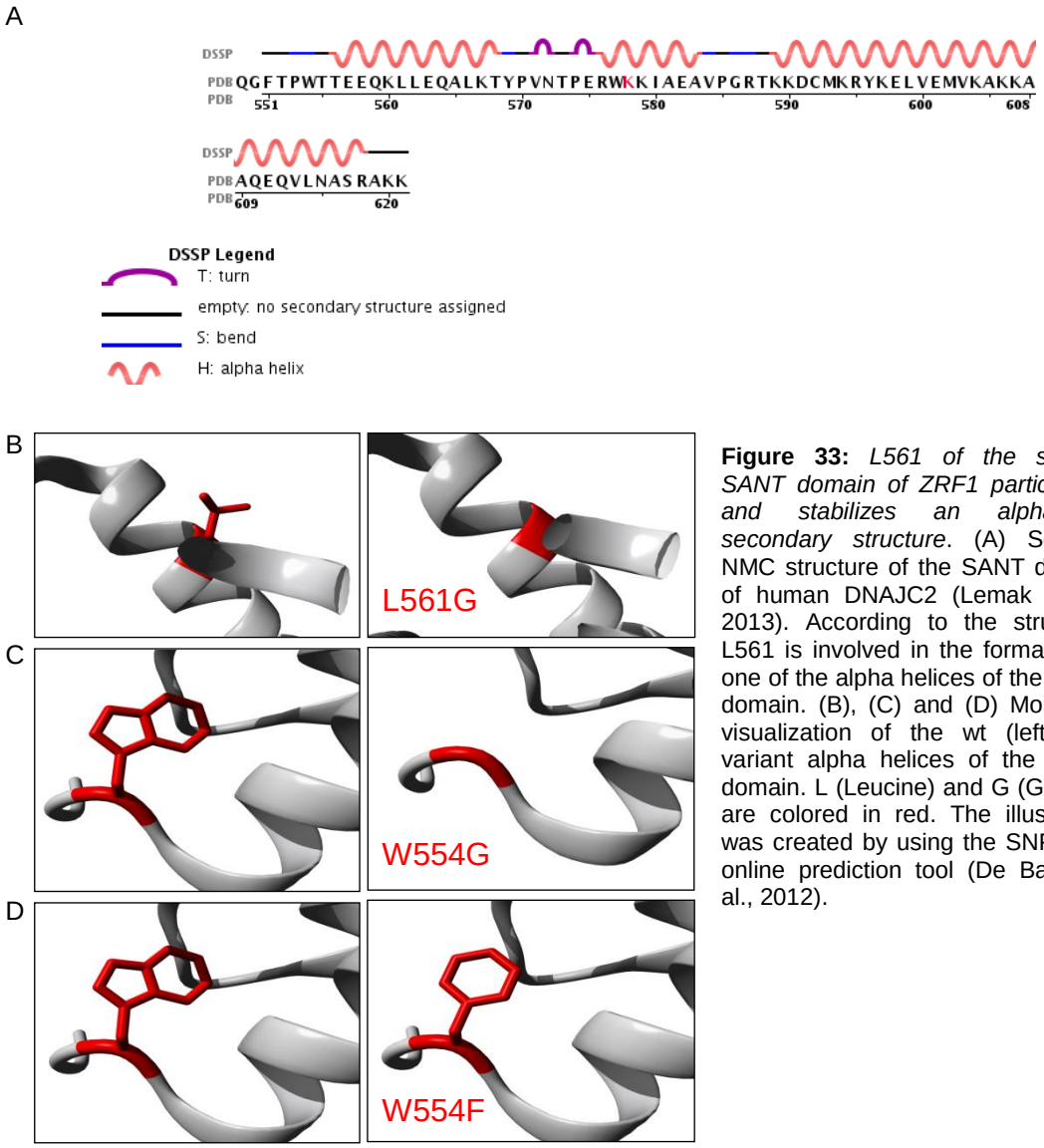


Figure 33: *L561 of the second SANT domain of ZRF1 participates and stabilizes an alpha-helix secondary structure*. (A) Solution NMC structure of the SANT domain of human DNAJC2 (Lemak at al., 2013). According to the structure, L561 is involved in the formation of one of the alpha helices of the SANT domain. (B), (C) and (D) Molecular visualization of the wt (left) and variant alpha helices of the SANT domain. L (Leucine) and G (Glycine) are colored in red. The illustration was created by using the SNPeffect online prediction tool (De Baets et al., 2012).

Apart from *Hoxd11*, Zrf1 is responsible for the activation of additional Med12-Ring1b targets in early differentiation, including *Hand1*, *Eomes*, *Snai1*, *Jun*, *Krt8*, *Lmx1a*, *Igfbp3* and *Fgfr2*. The majority of these genes code for developmental transcription factors that participate in differentiation programs of all three germ layers, highlighting the important regulatory function of Zrf1 in this process. We suggested that Zrf1 participates in the activation of *Hoxd11* by interacting with the lncRNA *Dlx1as*. It is possible that Zrf1 leads to the activation of additional genes in a ncRNA-dependent manner. Notably, one of the targets of Zrf1, Med12 and Ring1b, *Snai1*, is positively regulated by the eRNA ncRNA-a7 (Orom et al., 2010). Although this study

was carried out in human cells, it is likely that *Snai1* might be regulated during stem cell differentiation in a similar manner that also involves Zrf1. This hypothesis can be further supported by the fact that both *Dlx1as* and the ncRNA-a7 affect the expression of nearby genes while being induced themselves for expression upon differentiation induction; in our study by RA, in the study by Orom et al., 2010 by TPA (12-O-tetradecanoylphorbol-13-acetate). Therefore it remains to be further investigated whether the global role of Zrf1 in gene activation during differentiation involves its association with differentiation-induced ncRNAs. In support of this scenario, we showed that Zrf1 interacts during differentiation with the lncRNA *Hotairm* which is located between the genes HOXA1 and HOXA2 in humans (Sasaki et al., 2007; Sessa et al., 2007) but is also expressed in mESCs (Guttman et al., 2011). Zrf1 could potentially activate genes of the *HoxA* cluster by interacting with *Hotairm*, as the latter has been described to positively regulate the induction of HOXA1 and HOXA4 in RA-induced differentiation towards the myeloid lineage (Zhang et al., 2009). More importantly, this lncRNA was found to interact with the PRC1 complex in mESCs (Guttman et al., 2011) in agreement with our Ring1b RIP experiments. These findings suggest a more global role for Zrf1 in activating PcG-repressed genes by associating with lncRNAs which would have to be addressed by combining a Zrf1 RNA IP and genome-wide sequencing.

In order to gain further insight into the mechanism by which Zrf1/ZRF1 participates in gene activation we analyzed the interacting proteins by mass spectrometry after immunoprecipitation. The unique peptides identified as top hits have functions mostly related to ribosome biogenesis and code for components of ribonucleoprotein complexes. This finding was of no surprise as Zrf1 is a member of the Hsp40 family of proteins with conserved functions as molecular co-chaperones (Chen et al., 2014). The N' ZUO1 domain harbors a highly conserved DnaJ domain which enables members of this family to contribute into nascent polypeptide folding by enhancing the ATPase activity of the chaperone Hsp70 and stabilize its association to specific partners (Chen et al., 2014). Notably this cytosolic function of ZUO1/ZRF proteins is considered

as the ancestral function for this family of proteins whereas the nuclear functions, mostly connected with the emergence of the SANT domain at the C' in plants and animals are viewed as acquired "neofunctions". In line with the role of ZRF1 as a co-chaperone, amongst the identified peptides we observed the protein HSPA14, also known as HSP70L1. These two proteins have been previously described to form a stable complex which represents the mammalian RAC (Ribosome-associated complex) complex (Otto et al., 2005), involved in the regulation of translational fidelity as well as *de novo* protein folding, first identified in yeast (Gautschi et al., 2001). Furthermore we identified several ribosomal proteins (RPL4, RPL7A, RPS2, RPS9, RRP1B and RPL5), consistent with the association of the RAC with the ribosomes.

In order to understand further the nuclear function of ZRF1 we sought to analyze the co-precipitated peptides which have known functions in the nucleus. By performing cellular component analysis, we observed that ZRF1 associates with proteins of the cohesin complex (SMC1A, SMC3) as well as cell cycle regulators (MCM5, BUB3). It remains to be proven experimentally whether these associations of ZRF1 are related to its function in regulating transcriptional activation via Mediator or with its already described implication in the control of cell cycle. In *C. elegans* the ZRF1 orthologue, DNAJ11, is involved in the asymmetric cell division of the neuroblast by regulating the position of the mitotic spindle and therefore the orientation of the division axis (Hatzold and Conradt, 2008). However, this study did not describe any interaction of the ZRF1 orthologue with the mitotic spindle or associated proteins but rather its involvement in regulating asymmetric division by participating in the repression of the snail-related factor ces-1, which then enables apoptosis of one of the daughter cells, suggesting a developmental role for this protein. A similar function has been described for the mouse orthologue, MIDA1, which promotes cell growth and inhibits differentiation of MEL cells (murine erythroleukemia cells) and controls neural differentiation of mESCs (Aloia et al., 2015) by inhibiting the activity of the Id protein (Shoji et al., 1995). On the other hand ZRF1 was first

identified as MPP11 (M-phase phosphoprotein 11), which co-localizes with the mitotic spindle in dividing HeLa cells but also found in “punctuate cytoplasmic foci” during interface (Matsumoto-Taniura et al., 1996). Thus it remains to be further investigated whether Zrf1/ZRF1 participates directly in the control of cell cycle progression by associating with MCM5 and BUB3. This would require an experimental set up that would allow the separation of the function of Zrf1 in transcriptional regulation and the putative function in cell cycle as previously described in Lavagnolli et al., 2015.

Apart from its essential role during M phase, cohesin has emerged as an important regulator of gene expression during interphase by participating in the formation of long range chromosomal interactions (Apostolou et al., 2013; Dorsett and Merkenschlager, 2013; Hadjur et al., 2009; Kagey et al., 2010; Lavagnolli et al., 2015; Seitan et al., 2011, 2013; Wei et al., 2013; Zhang et al., 2013). Cohesin-Mediator complexes are required for DNA looping between enhancers and promoters that ensure proper gene expression and appear to be cell-type specific (Kagey et al., 2010). Remarkably, the cohesin loading factor Nibl and Med12 are essential and act synergistically to regulate limb development in zebrafish by bridging enhancers with the hoxda cluster. Interestingly the most affected genes upon depletion of either factor are located at the 5’ end of the cluster, including *Hoxd11* (Muto et al., 2014). This finding is in line with our conclusion that Zrf1 regulates *Hoxd11* in a ncRNA-Med12 dependent manner. In support of this model, the analysis of the co-precipitated peptides after ZRF1 immunoprecipitation and mass spectrometry revealed the subunits of cohesin, SMC1A and SMC3, as interactors. Further experimental studies would be required in order to substantiate a role for Zrf1 in cohesin-Mediator-mediated gene regulation.

At last but not least, ZRF1 co-precipitated with SUZ12 and SMARCA4 (BRG1) which were described previously as CDK8 interactors (Fukasawa et al., 2012). The study from Fukasawa and colleagues suggested that subunits of the Kinase Module might serve as a communication

platform between transcriptional and chromatin regulators that could link expression with epigenetic control. This suggestion is similar with the proposed function of Mediator subunits in respect to PRC1-mediated silencing (Simon and Kingston, 2013). In order to clarify the mechanism by which Zrf1 and Cdk8 regulate gene expression by interacting with known factors, further studies need to address this issue in a context-specific manner but with a global perspective. The context-specificity of such a study is essential as the known mechanisms that regulate chromatin events in respect to nuclear architecture are in most cases cell type-specific (Apostolou et al., 2013; Denholtz et al., 2013; Phillips-Cremins et al., 2013; Wei et al., 2013; de Wit et al., 2013), yet involve interactions of genomic regions with great distance between them, the position of which does not change in different cell types and is rather evolutionary conserved (Bonora et al., 2014; Dixon et al., 2012; Nora et al., 2012).

CONCLUSION

This study tried to address how PRC1-mediated repression in pluripotency sets the ground for the proper transition to gene expression during early differentiation. We identified a class of key developmental genes which are co-regulated by the PRC1 component Ring1b and the subunit Med12 of the Mediator complex in pluripotency. At this stage Ring1b is essential for the recruitment of Med12 at the promoters of developmental genes, a function which insures their repressed state. Particularly in the case of the classical PcG target, *Hoxd11*, Ring1b apart from regulating its silencing, also participates in the formation of associations between Med12 and the ncRNA *Dlx1as* on the *Hoxd11* promoter. After RA-induced differentiation and the displacement of PRC1 from chromatin by the H2AK119ub-binding protein Zrf1, Med12-*Dlx1as* remain bound at the promoter of *Hoxd11* and appear to be essential for the induction of its expression. This induction is highly enhanced by Zrf1 presumably via favoring the transition from pausing to elongation through *on site* remodeling of the Kinase Module of Mediator.

Part of this thesis was recently published under the title: '*Dual Role of Med12 in PRC1-dependent Gene Repression and ncRNA-mediated Transcriptional Activation*' (Papadopoulou et al., 2016). A summary of the key findings, hypotheses and proposed research for the future were included in a review article which is currently in revision in the journal *Bioessays*, under the title: '*On-site remodelling at chromatin: How multiprotein complexes are rebuilt during DNA repair and transcriptional activation*' (Papadopoulou T. and Richly H.).

References

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720–731.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155–166.

Aloia, L., Gutierrez, A., Caballero, J.M., and Di Croce, L. (2015). Direct interaction between Id1 and Zrf1 controls neural differentiation of embryonic stem cells. EMBO Rep. *16*, 63–70.

Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. Bioessays *37*, 314–323.

Andrau, J.C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M.G., van de Peppel, J., Werner, M., and Holstege, F.C. (2006). Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. Mol Cell *22*, 179–192.

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

De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., Schymkowitz, J., and Rousseau, F. (2012). SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. Nucleic Acids Res *40*, D935–D939.

Balamotis, M.A., Pennella, M.A., Stevens, J.L., Wasylyk, B., Belmont, A.S., and Berk, A.J. (2009). Complexity in transcription control at the activation domain-mediator interface. Sci Signal *2*, ra20.

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

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends Genet *27*, 454–464.

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

Belakavadi, M., and Fondell, J.D. (2010). Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation. Mol Cell Biol *30*, 2437–2448.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396.

Bergsmedh, A., Donohoe, M.E., Hughes, R.A., and Hadjantonakis, A.K. (2011). Understanding the molecular circuitry of cell lineage specification in the early mouse embryo. Genes (Basel) *2*, 420–448.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental

genes in embryonic stem cells. Cell *125*, 315–326.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270–1284.

Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., et al. (2012). Coding exons function as tissue-specific enhancers of nearby genes. Genome Res *22*, 1059–1068.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445–1459.

Blau, J., Xiao, H., McCracken, S., O'Hare, P., Greenblatt, J., and Bentley, D. (1996). Three functional classes of transcriptional activation domain. Mol Cell Biol *16*, 2044–2055.

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. J. Biol. Chem. *285*, 188–196.

Bonnet, J., Wang, C.Y., Baptista, T., Vincent, S.D., Hsiao, W.C., Stierle, M., Kao, C.F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes Dev *28*, 1999–2012.

Bonora, G., Plath, K., and Denholtz, M. (2014). A mechanistic link between gene regulation and genome architecture in mammalian development. Curr Opin Genet Dev *27*, 92–101.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat Cell Biol *16*, 516–528.

Boube, M., Joulia, L., Cribbs, D.L., and Bourbon, H.M. (2002). Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. Cell *110*, 143–151.

Bourbon, H.M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. Nucleic Acids Res *36*, 3993–4008.

Boyer, L.A., Latek, R.R., and Peterson, C.L. (2004). The SANT domain: a unique histone-tail-binding module? Nat Rev Mol Cell Biol *5*, 158–163.

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

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

Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nat Rev Cancer *9*, 773–784.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell *10*, 157–170.

Brown, R.P., and Feder, M.E. (2005). Reverse transcriptional profiling: non-correspondence of transcript level variation and proximal promoter polymorphism. BMC Genomics *6*, 110.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell *15*, 57–67.

Cao, Q., Wang, X., Zhao, M., Yang, R., Malik, R., Qiao, Y., Poliakov, A., Yocum, A.K., Li, Y., Chen, W., et al. (2014). The central role of EED in the orchestration of polycomb group complexes. Nat. Commun. *5*, 3127.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (80-.). *298*, 1039–1043.

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

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., et al. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science (80-.). *291*, 1289–1292.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Cavalli, G. (2015). PRC1 proteins orchestrate three-dimensional genome architecture. Nat Genet *47*, 1105–1106.

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

Chen, T., and Dent, S.Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nat Rev Genet *15*, 93–106.

Chen, D.-H., Huang, Y., Liu, C., Ruan, Y., and Shen, W.-H. (2014). Functional conservation and divergence of J-domain-containing ZUO1/ZRF orthologs throughout evolution. Planta *239*, 1159–1173.

Chen, P.B., Hung, J.H., Hickman, T.L., Coles, A.H., Carey, J.F., Weng, Z., Chu, F., and Fazzio, T.G. (2013). Hdac6 regulates Tip60-p400 function in stem cells. Elife *2*, e01557.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A., et al. (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol. Cell *45*, 38–50.

Chesnut, J.D., Stephens, J.H., and Dahmus, M.E. (1992). The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. J Biol Chem *267*, 10500–10506.

Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P., and Schedl, P. (2014). Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. Bioessays *36*, 163–172.

Ciabrelli, F., and Cavalli, G. (2015). Chromatin-driven behavior of topologically associating domains. J Mol Biol *427*, 608–625.

Clark, A.D., Oldenbroek, M., and Boyer, T.G. (2015). Mediator kinase module and human tumorigenesis. Crit. Rev. Biochem. Mol. Biol. *50*, 393–426.

Comet, I., and Helin, K. (2014). Revolution in the Polycomb hierarchy. Nat Struct Mol Biol *21*, 573–575.

Conaway, R.C., and Conaway, J.W. (2011). Function and regulation of the Mediator complex. Curr Opin Genet Dev *21*, 225–230.

Conaway, R.C., and Conaway, J.W. (2013). The Mediator complex and transcription elongation. Biochim Biophys Acta *1829*, 69–75.

Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep *7*, 1456–1470.

Corden, J.L., Cadena, D.L., Ahearn Jr., J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A *82*, 7934–7938.

Core, L.J., and Lis, J.T. (2009). Paused Pol II captures enhancer activity and acts as a potent insulator. Genes Dev *23*, 1606–1612.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (80-.). *322*, 1845–1848.

Costlow, N., and Lis, J.T. (1984). High-resolution mapping of DNase I-hypersensitive sites of Drosophila heat shock genes in Drosophila melanogaster and Saccharomyces cerevisiae. Mol Cell Biol *4*, 1853–1863.

Cremer, T., Cremer, M., Dietzel, S., Muller, S., Solovei, I., and Fakan, S. (2006). Chromosome territories--a functional nuclear landscape. Curr Opin Cell Biol *18*, 307–316.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol *145*, 1119–1131.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185–196.

Czudnochowski, N., Bosken, C.A., and Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat Commun *3*, 842.

Dahmus, M.E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. J Biol Chem *271*, 19009–19012.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602–616.

Diaz-Beya, M., Brunet, S., Nomdedeu, J., Pratcorona, M., Cordeiro, A., Gallardo, D., Escoda, L., Tormo, M., Heras, I., Ribera, J.M., et al. (2015). The lincRNA HOTAIRM1, located in the HOXA genomic region, is expressed in acute myeloid leukemia, impacts prognosis in patients in the intermediate-risk cytogenetic category, and is associated with a distinctive microRNA signature. Oncotarget *6*, 31613–31627.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Solda, G., Simons, C., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res *18*, 1433–1445.

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

Donner, A.J., Szostek, S., Hoover, J.M., and Espinosa, J.M. (2007). CDK8 is a stimulus-specific positive coregulator of p53 target genes. Mol Cell *27*, 121–133.

Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nat Struct Mol Biol *17*, 194–201.

Dorsett, D., and Merkenschlager, M. (2013). Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. Curr Opin Cell Biol *25*, 327–333.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. Nat Cell Biol *10*, 1106–1113.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., et al. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genet *8*, e1002774.

Erokhin, M., Vassetzky, Y., Georgiev, P., and Chetverina, D. (2015). Eukaryotic enhancers: common features, regulation, and participation in diseases. Cell Mol Life Sci *72*, 2361–2375.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol Cell *38*, 452–464.

Espinosa, J.M. (2016). Revisiting lncRNAs: How Do You Know Yours Is Not an eRNA? Mol. Cell *62*, 1–2.

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

Favier, B., Le Meur, M., Chambon, P., and Dolle, P. (1995). Axial skeleton homeosis and forelimb malformations in Hoxd-11 mutant mice. Proc Natl Acad Sci U S A *92*, 310–314.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V. V (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol *16*, 2802–2813.

Fong, Y.W., Cattoglio, C., Yamaguchi, T., and Tjian, R. (2012). Transcriptional regulation by coactivators in embryonic stem cells. Trends Cell Biol *22*, 292–298.

Fort, P., Rech, J., Vie, A., Piechaczyk, M., Bonnieu, A., Jeanteur, P., and Blanchard, J.M. (1987). Regulation of c-fos gene expression in hamster fibroblasts: initiation and elongation of transcription and mRNA degradation. Nucleic Acids Res *15*, 5657–5667.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a functional core polycomb repressive complex. Mol Cell *8*, 545–556.

Fukasawa, R., Tsutsui, T., Hirose, Y., Tanaka, A., and Ohkuma, Y. (2012). Mediator CDK subunits are platforms for interactions with various chromatin regulatory complexes. J Biochem *152*, 241–249.

Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D., et al. (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell *153*, 1327–1339.

Gautschi, M., Lilie, H., Funfschilling, U., Mun, A., Ross, S., Lithgow, T., Rucknagel, P., and Rospert, S. (2001). RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. Proc Natl Acad Sci U S A *98*, 3762–3767.

Gaytan de Ayala Alonso, A., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D., and Muller, J. (2007). A genetic screen identifies novel polycomb group genes in Drosophila. Genetics *176*, 2099–2108.

Geisler, S.J., and Paro, R. (2015). Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. Development *142*, 2876–2887.

Gil, J., and O'Loghlen, A. (2014). PRC1 complex diversity: where is it taking us? Trends Cell Biol. *24*, 632–641.

Gilbert, S.F., Singer, S.R., Tyler, M.S., and Kozlowski, R.N. (2006). Developmental biology (Sunderland, Mass.: Sinauer Associates, Inc. Publishers).

Gilchrist, D.A., Fromm, G., dos Santos, G., Pham, L.N., McDaniel, I.E., Burkholder, A., Fargo, D.C., and Adelman, K. (2012). Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. Genes Dev *26*, 933–944.

Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol *6*, 3984–3989.

Gobert, V., Osman, D., Bras, S., Auge, B., Boube, M., Bourbon, H.M., Horn, T., Boutros, M., Haenlin, M., and Waltzer, L. (2010). A genome-wide RNA interference screen identifies a differential role of the mediator CDK8 module subunits for GATA/ RUNX-activated transcription in Drosophila. Mol Cell Biol *30*, 2837–2848.

Gonzales-Roybal, G., and Lim, D.A. (2013). Chromatin-based epigenetics of adult subventricular zone neural stem cells. Front. Genet. *4*, 1–11.

Gonzalez, I., Mateos-Langerak, J., Thomas, A., Cheutin, T., and Cavalli, G. (2014). Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. Mol. Cell *54*, 485–499.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell *14*, 762–775.

Green, S.R., and Johnson, A.D. (2004). Promoter-dependent roles for the Srb10 cyclin-dependent kinase and the Hda1 deacetylase in Tup1-mediated repression in Saccharomyces cerevisiae. Mol Biol Cell *15*, 4191–4202.

Grossniklaus, U., and Paro, R. (2014). Transcriptional silencing by polycomb-group proteins. Cold Spring Harb Perspect Biol *6*, a019331.

Grune, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F., Becker, P.B., and Muller, C.W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. Mol Cell *12*, 449–460.

Guenther, M.G., Barak, O., and Lazar, M.A. (2001). The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Mol Cell Biol *21*, 6091–6101.

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

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

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

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410–413.

Hagan, J.P., O'Neill, B.L., Stewart, C.L., Kozlov, S. V, and Croce, C.M. (2009). At least ten genes define the imprinted Dlk1-Dio3 cluster on mouse chromosome 12qF1. PLoS One *4*, e4352.

Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jimenez, F., Baylies, M.K., and Michelson, A.M. (2000). Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell *103*, 63–74.

Han, J., Zhang, J., Chen, L., Shen, B., Zhou, J., Hu, B., Du, Y., Tate, P.H., Huang, X., and Zhang, W. (2014). Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. RNA Biol *11*, 829–835.

Hatzold, J., and Conradt, B. (2008). Control of apoptosis by asymmetric cell division. PLoS Biol *6*, e84.

He, S., Nakada, D., and Morrison, S.J. (2009). Mechanisms of stem cell self-renewal. Annu Rev Cell Dev Biol *25*, 377–406.

Hermanson, O., Glass, C.K., and Rosenfeld, M.G. (2002). Nuclear receptor coregulators: multiple modes of modification. Trends Endocrinol Metab *13*, 55–60.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Huang, S., Holzel, M., Knijnenburg, T., Schlicker, A., Roepman, P., McDermott, U., Garnett, M., Grernrum, W., Sun, C., Prahallad, A., et al. (2012). MED12 controls the response to multiple cancer drugs through regulation of TGF-beta receptor signaling. Cell *151*, 937–950.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet *43*, 621–629.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev. *29*, 1897–1902.

Ing-Simmons, E., Seitan, V.C., Faure, A.J., Flicek, P., Carroll, T., Dekker, J., Fisher, A.G., Lenhard, B., and Merkenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. Genome Res *25*, 504–513.

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

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Jiang, Y.W., Veschambre, P., Erdjument-Bromage, H., Tempst, P., Conaway, J.W., Conaway, R.C., and Kornberg, R.D. (1998). Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. Proc. Natl. Acad. Sci. U. S. A. *95*, 8538–8543.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. Nat Genet *41*, 941–945.

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Mol. Cell *45*, 51–63.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484–492.

Jungers, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature *316*, 153–155.

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

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

Kang, M.E., and Dahmus, M.E. (1993). RNA polymerases IIA and IIO have distinct roles during transcription from the TATA-less murine dihydrofolate reductase promoter. J Biol Chem *268*, 25033–25040.

Kanhere, A., Viiri, K., Araujo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell *38*, 675–688.

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

Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol Cell Biol *25*, 6857–6868.

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

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., and Crabtree, G.R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature *366*, 170–174.

Kim, K., Jutooru, I., Chadalapaka, G., Johnson, G., Frank, J., Burghardt, R., Kim, S., and Safe, S. (2013). HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. Oncogene *32*, 1616–1625.

Kim, S., Xu, X., Hecht, A., and Boyer, T.G. (2006). Mediator is a transducer of Wnt/beta-catenin signaling. J Biol Chem *281*, 14066–14075.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V. V, and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

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

King, I.F., Francis, N.J., and Kingston, R.E. (2002). Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. Mol Cell Biol *22*, 7919–7928.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal *3*, ra8.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene *30*, 1956–1962.

Kraus, P., Sivakamasundari, V., Lim, S.L., Xing, X., Lipovich, L., and Lufkin, T. (2013). Making sense of Dlx1 antisense RNA. Dev Biol *376*, 224–235.

Krumm, A., Meulia, T., Brunvand, M., and Groudine, M. (1992). The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. Genes Dev *6*, 2201–2213.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

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

Kyrchanova, O., and Georgiev, P. (2014). Chromatin insulators and long-distance interactions in Drosophila. FEBS Lett *588*, 8–14.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Lavagnolli, T., Gupta, P., Hormanseder, E., Mira-Bontenbal, H., Dharmalingam, G., Carroll, T., Gurdon, J.B., Fisher, A.G., and Merkenschlager, M. (2015). Initiation and maintenance of pluripotency gene expression in the absence of cohesin. Genes Dev *29*, 23–38.

Laybourn, P.J., and Dahmus, M.E. (1990). Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. J Biol Chem *265*, 13165–13173.

Lee, H., Kraus, K.W., Wolfner, M.F., and Lis, J.T. (1992). DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev *6*, 284–295.

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

Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J Cell Biol *178*, 219–229.

Lehmann, L., Ferrari, R., Vashisht, A.A., Wohlschlegel, J.A., Kurdistani, S.K., and Carey, M. (2012). Polycomb repressive complex 1 (PRC1) disassembles RNA polymerase II preinitiation complexes. J Biol Chem *287*, 35784–35794.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet *12*, 1725–1735.

Lettice, L.A., Williamson, I., Wiltshire, J.H., Peluso, S., Devenney, P.S., Hill, A.E., Essafi, A., Hagman, J., Mort, R., Grimes, G., et al. (2012). Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. Dev Cell *22*, 459–467.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr Biol *20*, R754–R763.

Levinger, L., and Varshavsky, A. (1982). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. Cell *28*, 375–385.

Lin, M., Pedrosa, E., Shah, A., Hrabovsky, A., Maqbool, S., Zheng, D., and Lachman, H.M. (2011). RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS One *6*, e23356.

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

Lu, H., Flores, O., Weinmann, R., and Reinberg, D. (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proc Natl Acad Sci U S A *88*, 10004–10008.

Luis, N.M., Morey, L., Di Croce, L., and Benitah, S.A. (2012). Polycomb in stem cells: PRC1 branches out. Cell Stem Cell *11*, 16–21.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256–263.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet *11*, 761–772.

Malik, S., Barrero, M.J., and Jones, T. (2007). Identification of a regulator of transcription elongation as an accessory factor for the human Mediator coactivator. Proc. Natl. Acad. Sci. U. S. A. *104*, 6182–6187.

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

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol Cell Biol *12*, 2078–2090.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem *270*, 12335–12338.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. Annu Rev Cell Dev Biol *30*, 647–675.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A *78*, 7634–7638.

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

Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem *255*, 11992–11996.

Matsumoto-Taniura, N., Pirollet, F., Monroe, R., Gerace, L., and Westendorf, J.M. (1996). Identification of novel M phase phosphoproteins by expression cloning. Mol Biol Cell *7*, 1455–1469.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet *6*, e1001244.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285–1297.

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

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev *25*, 742–754.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755–766.

Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. Trends Biochem Sci *35*, 323–332.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S.A., and Di Croce, L. (2012). Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell *10*, 47–62.

Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A., and Di Croce, L. (2013). RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell Rep *3*, 60–69.

Morey, L., Santanach, A., and Di Croce, L. (2015). Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. Mol Cell Biol *35*, 2716–2728.

Morris, E.J., Ji, J.Y., Yang, F., Di Stefano, L., Herr, A., Moon, N.S., Kwon, E.J., Haigis, K.M., Naar, A.M., and Dyson, N.J. (2008). E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. Nature *455*, 552–556.

Morrison, S.J., Shah, N.M., and Anderson, D.J. (1997). Regulatory mechanisms in stem cell biology. Cell *88*, 287–298.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell *51*, 606–617.

Mukundan, B., and Ansari, A. (2011). Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. J. Biol. Chem. *286*, 37053–37057.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet *39*, 1507–1511.

Muto, A., Ikeda, S., Lopez-Burks, M.E., Kikuchi, Y., Calof, A.L., Lander, A.D., and Schilling, T.F. (2014). Nipbl and mediator cooperatively regulate gene expression to control limb development. PLoS Genet *10*, e1004671.

Myers, L.C., and Kornberg, R.D. (2000). Mediator of transcriptional regulation. Annu Rev Biochem *69*, 729–749.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science (80-.). *322*, 1717–1720.

Nechaev, S., and Adelman, K. (2008). Promoter-proximal Pol II: when stalling speeds things up. Cell Cycle *7*, 1539–1544.

Ng, S.Y., Parker, C.S., and Roeder, R.G. (1979). Transcription of cloned Xenopus 5S RNA genes by X. laevis RNA polymerase III in reconstituted systems. Proc Natl Acad Sci U S A *76*, 136–140.

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

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

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

O'Loghlen, A., Munoz-Cabello, A.M., Gaspar-Maia, A., Wu, H.A., Banito, A., Kunowska, N., Racek, T., Pemberton, H.N., Beolchi, P., Lavial, F., et al. (2012). MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. Cell Stem Cell *10*, 33–46.

Oktaba, K., Gutierrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E., and Muller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Dev Cell *15*, 877–889.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem *280*, 5307–5317.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet *12*, 283–293.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46–58.

Otto, H., Conz, C., Maier, P., Wolfle, T., Suzuki, C.K., Jeno, P., Rucknagel, P., Stahl, J., and Rospert, S. (2005). The chaperones MPP11 and Hsp70L1 form the mammalian ribosome-associated complex. Proc Natl Acad Sci U S A *102*, 10064–10069.

Palmieri, S.L., Peter, W., Hess, H., and Scholer, H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. Dev Biol *166*, 259–267.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299–312.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232–246.

Papadopoulou, T., Kaymak, A., Sayols, S., and Richly, H. (2016). Dual Role of Med12 in PRC1-dependent Gene Repression and ncRNA-mediated Transcriptional Activation. Cell Cycle.

Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev *20*, 2041–2054.

Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A. V, Prasad, R., Luan, J., Davies, J.O.J., Hughes, J.R., Hardison, R.C., et al. (2016). Unlinking an lncRNA from Its Associated cis Element. Mol. Cell *62*, 104–110.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422–433.

Parker, C.S., and Roeder, R.G. (1977). Selective and accurate transcription of the Xenopus laevis 5S RNA genes in isolated chromatin by purified RNA polymerase III. Proc Natl Acad Sci U S A *74*, 44–48.

Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J *23*, 4061–4071.

Pathan, M., Keerthikumar, S., Ang, C.S., Gangoda, L., Quek, C.Y., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., et al. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics *15*, 2597–2601.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297–305.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature *322*, 697–701.

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

Richly, H., and Di Croce, L. (2011). The flip side of the coin: role of ZRF1 and histone H2A ubiquitination in transcriptional activation. Cell Cycle *10*, 745–750.

Richly, H., Rocha-Viegas, L., Ribeiro, J.D., Demajo, S., Gundem, G., Lopez-Bigas, N., Nakagawa, T., Rospert, S., Ito, T., and Di Croce, L. (2010). Transcriptional activation of polycomb-repressed genes by ZRF1. Nature *468*, 1124–1128.

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

Rinn, J.L. (2014). lncRNAs: linking RNA to chromatin. Cold Spring Harb Perspect Biol *6*.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev Biochem *81*, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Ritter, D.I., Dong, Z., Guo, S., and Chuang, J.H. (2012). Transcriptional enhancers in protein-coding exons of vertebrate developmental genes. PLoS One *7*, e35202.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase IIRoeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21, 327–335. Trends Biochem Sci *21*, 327–335.

Roeder, R.G. (2003). Lasker Basic Medical Research Award. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. Nat Med *9*, 1239–1244.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett *579*, 909–915.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234–237.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795–804.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. Nature *503*, 371–376.

Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C., et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature *404*, 95–99.

Saikumar, P., Murali, R., and Reddy, E.P. (1990). Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. Proc Natl Acad Sci U S A *87*, 8452–8456.

Samuelsen, C.O., Baraznenok, V., Khorosjutina, O., Spahr, H., Kieselbach, T., Holmberg, S., and Gustafsson, C.M. (2003). TRAP230/ARC240 and TRAP240/ARC250 Mediator subunits are functionally conserved through evolution. Proc Natl Acad Sci U S A *100*, 6422–6427.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev *21*, 436–449.

Sanso, M., and Fisher, R.P. (2013). Pause, play, repeat: CDKs push RNAP II's buttons. Transcription *4*, 146–152.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol *8*, e1000384.

Sasaki, Y.T., Sano, M., Kin, T., Asai, K., and Hirose, T. (2007). Coordinated expression of ncRNAs and HOX mRNAs in the human HOXA locus. Biochem Biophys Res Commun *357*, 724–730.

Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybailov, B., Swanson, S.K., Banks, C.A., Jin, J., Cai, Y., Washburn, M.P., et al. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. Mol Cell *14*, 685–691.

Sauvageau, M., and Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell *7*, 299–313.

Savinkova, L.K., Ponomarenko, M.P., Ponomarenko, P.M., Drachkova, I.A., Lysova, M. V, Arshinova, T. V, and Kolchanov, N.A. (2009). TATA box polymorphisms in human gene promoters and associated hereditary pathologies. Biochem. *74*, 117–129.

Schaaf, C.A., Misulovin, Z., Gause, M., Koenig, A., Gohara, D.W., Watson, A., and Dorsett, D. (2013). Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS Genet. *9*, e1003560.

Schaefer, M.H., Fontaine, J.F., Vinayagam, A., Porras, P., Wanker, E.E., and Andrade-Navarro, M.A. (2012). HIPPIE: Integrating protein interaction networks with experiment based quality scores. PLoS One *7*, e31826.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J *25*, 3110–3122.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat Genet *47*, 1179–1186.

Schuettengruber, B., and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development *136*, 3531–3542.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9–22.

Schwarzer, W., and Spitz, F. (2014). The architecture of gene expression: integrating dispersed cis-regulatory modules into coherent regulatory domains. Curr Opin Genet Dev *27*, 74–82.

Segall, J., Matsui, T., and Roeder, R.G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J Biol Chem *255*, 11986–11991.

Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., et al. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature *476*, 467–471.

Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res *23*, 2066–2077.

Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. (2007). Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. RNA *13*, 223–239.

Shaver, S., Casas-Mollano, J.A., Cerny, R.L., and Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. Epigenetics *5*, 301–312.

Shiekhattar, R. (2013). Opening the Chromatin by eRNAs. Mol Cell *51*, 557–558.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272–286.

Shoji, W., Inoue, T., Yamamoto, T., and Obinata, M. (1995). MIDA1, a protein associated with Id, regulates cell growth. J Biol Chem *270*, 24818–24825.

Shopland, L.S., Hirayoshi, K., Fernandes, M., and Lis, J.T. (1995). HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev *9*, 2756–2769.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol *10*, 697–708.

Simon, J.A., and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol Cell *49*, 808–824.

Sklar, V.E., Schwartz, L.B., and Roeder, R.G. (1975). Distinct molecular structures of nuclear class I, II, and III DNA-dependent RNA polymerases. Proc Natl Acad Sci U S A *72*, 348–352.

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. EMBO J *11*, 4047–4057.

Smith, A.G. (2001). Embryo-derived stem cells: of mice and men. Annu Rev Cell Dev Biol *17*, 435–462.

Smith, A.G., and Hooper, M.L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. Dev Biol *121*, 1–9.

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

Sofueva, S., Yaffe, E., Chan, W.C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J *32*, 3119–3129.

Song, W., Treich, I., Qian, N., Kuchin, S., and Carlson, M. (1996). SSN genes that affect transcriptional repression in Saccharomyces cerevisiae encode SIN4, ROX3, and SRB proteins associated with RNA polymerase II. Mol Cell Biol *16*, 115–120.

Spiegelman, B.M., and Heinrich, R. (2004). Biological control through regulated transcriptional coactivators. Cell *119*, 157–167.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613–626.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat Rev Mol Cell Biol *15*, 340–356.

Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution (N. Y). *62*, 2155–2177.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol *9*, 1428–1435.

van der Stoop, P., Boutsma, E.A., Hulsman, D., Noback, S., Heimerikx, M., Kerkhoven, R.M., Voncken, J.W., Wessels, L.F., and van Lohuizen, M. (2008). Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells. PLoS One *3*, e2235.

Strubbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C., and Paro, R. (2011). Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. Proc Natl Acad Sci U S A *108*, 5572–5577.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol *14*, 103–105.

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development *132*, 2093–2102.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A.S., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., et al. (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92–104.

Tarkowski, A.K. (1959). Experiments on the development of isolated blastomers of mouse eggs. Nature *184*, 1286–1287.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664–678.

Thomas, P.D., Kejariwal, A., Campbell, M.J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., et al. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res *31*, 334–341.

Till, J.E., and McCulloch, E.A. (1980). Hemopoietic stem cell differentiation. Biochim Biophys Acta *605*, 431–459.

Trotter, K.W., and Archer, T.K. (2008). The BRG1 transcriptional coregulator. Nucl Recept Signal *6*, e004.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science (80-.). *329*, 689–693.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. Proc Natl Acad Sci U S A *89*, 11219–11223.

Tutter, A. V, Kowalski, M.P., Baltus, G.A., Iourgenko, V., Labow, M., Li, E., and Kadam, S. (2009). Role for Med12 in regulation of Nanog and Nanog target genes. J Biol Chem *284*, 3709–3718.

Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P., and Angrand, P.-O. (2011). Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. Mol. Cell. Proteomics *10*, M110.002642.

Vietri Rudan, M., and Hadjur, S. (2015). Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. Trends Genet *31*, 651–660.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep *10*, 1297–1309.

Vissers, J.H., Nicassio, F., van Lohuizen, M., Di Fiore, P.P., and Citterio, E. (2008). The many faces of ubiquitinated histone H2A: insights from the DUBs. Cell Div *3*, 8.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. EMBO J *17*, 7395–7403.

Wahi, M., and Johnson, A.D. (1995). Identification of genes required for alpha 2 repression in Saccharomyces cerevisiae. Genetics *140*, 79–90.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873–878.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897–903.

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes

long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. Cell Stem Cell *13*, 36–47.

Weil, P.A., Luse, D.S., Segall, J., and Roeder, R.G. (1979). Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. Cell *18*, 469–484.

Weinmann, R., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. Proc Natl Acad Sci U S A *71*, 1790–1794.

Weinmann, R., Raskas, H.J., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. Proc Natl Acad Sci U S A *71*, 3426–3439.

Weissman, I.L. (2000). Stem cells: units of development, units of regeneration, and units in evolution. Cell *100*, 157–168.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. Trends Genet *23*, 494–502.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J. V, Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

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

de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet *13*, 59–69.

Workman, J.L. (2006). Nucleosome displacement in transcription. Genes Dev *20*, 2009–2017.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet *8*, 206–216.

Wu, C. (1980). The 5’ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.

Wu, G., and Scholer, H.R. (2014). Role of Oct4 in the early embryo development. Cell Regen *3*, 7.

Wu, Y., Zhang, L., Wang, Y., Li, H., Ren, X., Wei, F., Yu, W., Wang, X., Zhang, L., Yu, J., et al.

(2014). Long noncoding RNA HOTAIR involvement in cancer. Tumour Biol *35*, 9531–9538.

Wysocka, J., Reilly, P.T., and Herr, W. (2001). Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. Mol Cell Biol *21*, 3820–3829.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A *104*, 7145–7150.

Yap, K.L., Li, S., Munoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell *38*, 662–674.

Yeom, Y.I., Ha, H.S., Balling, R., Scholer, H.R., and Artzt, K. (1991). Structure, expression and chromosomal location of the Oct-4 gene. Mech Dev *35*, 171–179.

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

Yu, J., Li, Y., Ishizuka, T., Guenther, M.G., and Lazar, M.A. (2003). A SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. EMBO J *22*, 3403–3410.

Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225–229.

Yuh, C.H., Ransick, A., Martinez, P., Britten, R.J., and Davidson, E.H. (1994). Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. Mech Dev *47*, 165–186.

Zappulla, D.C., and Cech, T.R. (2006). RNA as a flexible scaffold for proteins: yeast telomerase and beyond. Cold Spring Harb Symp Quant Biol *71*, 217–224.

Zee, B.M., Levin, R.S., Xu, B., LeRoy, G., Wingreen, N.S., and Garcia, B.A. (2010). In vivo residue-specific histone methylation dynamics. J Biol Chem *285*, 3341–3350.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet *39*, 1512–1516.

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

Zhang, S., Chen, S., Yang, G., Gu, F., Li, M., Zhong, B., Hu, J., Hoffman, A., and Chen, M. (2014). Long noncoding RNA HOTAIR as an independent prognostic marker in cancer: a meta-analysis. PLoS One *9*, e105538.

Zhang, X., Lian, Z., Padden, C., Gerstein, M.B., Rozowsky, J., Snyder, M., Gingeras, T.R., Kapranov, P., Weissman, S.M., and Newburger, P.E. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. Blood *113*, 2526–2534.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.K., Shahab,

A., Kuznetsov, V.A., et al. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286–298.

Zhou, J., Wang, X., He, K., Charron, J.B., Elling, A.A., and Deng, X.W. (2010). Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression. Plant Mol Biol *72*, 585–595.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell *29*, 69–80.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, Ij.W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A *111*, 996–1001.

Zuniga, A. (2015). Next generation limb development and evolution: old questions, new perspectives. Development *142*, 3810–3820.