PWWP2A – a novel H3K36me3 reading component of the NuRD complex

Tianyi Zhang

University College

University of Oxford

Thesis submitted for the degree of

Doctor of Philosophy

October 2016

ABSTRACT

Epigenetic regulation of gene expression is critical for lineage determination during development and the stable transmission of cellular identity through cell growth and division. Chromatin modifiers are a varied group of enzymes that regulate patterns of gene expression through the modification of DNA and histones. Epigenetic modifications alter the physical structure of chromatin or recruit effectors to promote or inhibit transcription. In this study, I examined the role of pre-existing histone modifications on the activity of Polycomb and Trithorax chromatin modifying enzymes. Additionally, identification of a novel H3K36me3-reader PWWP2A may constitute a new histone crosstalk pathway between H3K36me3 and histone deacetylation.

Polycomb and Trithorax proteins are two important families of transcriptional regulators that promote gene repression and activation respectively. In this study, I find that histone crosstalk between Polycomb complexes PRC1 and PRC2 promote one another's activity while inhibiting Trithorax function. Recombinant nucleosomes carrying the PRC2 modification H3K27me3 stimulates the activity of CBX-PRC1, and nucleosomes with the PRC1 modification H2AK119u1 stimulates the activity of JARID2-PRC2. H2AK119u1 negatively regulates Trithorax activity and inhibits both the H3K4 methyltransferases MLL1/4 and the H3K36

methyltransferase NSD2. Histone crosstalk appears to be an important mechanism in the regulation of the recruitment and activity of chromatin modifiers.

I identified a novel H3K36me3 binding protein PWWP2A in an unbiased screen for readers of H3K27me3 and H3K36me3. Biochemical characterisation of PWWP2A in HeLa cells and mouse embryonic stem cells (mESCs) showed that PWWP2A interacts with NuRD subunits MTA1-3, HDAC1/2, and RBBP4/7. Genome-wide mapping of PWWP2A targets in mESCs shows that it is enriched over the gene bodies of highly transcribed genes sharing many gene targets with HDAC1. PWWP2A localisation correlates highly with H3K36me3, and its localisation at gene bodies is dependent on the putative H3K36me3-binding PWWP domain. In budding yeast, H3K36me3-mediated recruitment of the yeast HDAC complex Rpd3S is important for maintaining transcriptional integrity over coding regions. PWWP2A-NuRD may represents an analogous pathway of transcriptional regulation in mammalian organisms.

ACKNOWLEDGEMENTS

I would like to thank all the wonderful people from Brockdorff and Klose labs who have given me so much help, advice, and encouragement over the years. Thank you Neil for your constant support and always challenging me with new ideas and projects. I feel like I have grown a lot as a scientist under your supervision. Special thanks to my fellow DPhil students Mafalda and Greta for being such amazing friends and scientists.

I would like to thank all the friends I've made in the UK, especially Martin and Nana for being sources of endless positivity. To my dearest friends in Canada, thank you for listening to my scientific ramblings, agreeing to skype at odd hours, and driving long distances to see me when I'm home.

Thank you Taha for your unwavering support and belief in me, and for being a constant source of inspiration.

Lastly to my parents for always being there for me and letting me live my dreams.

TABLE OF CONTENTS

6 List of Abbreviations

9 Introduction

41 Project Aims

42 Results Chapter 1 – Histone crosstalk between Polycomb and Trithorax

65 Results Chapter 2 – Identification of readers of H3K27me3 and H3K36me3

86 Results Chapter 3 – Biochemical characterisation of the PWWP2A-NuRD complex

105 Results Chapter 4 – Genome-wide mapping of PWWP2A and HDAC1 targets

124 Results Chapter 5 – Generation of PWWP2A and PWWP2A/B knockouts in mESCs

138 Discussion

147 Materials and Methods

165 References

182 Appendix

LIST OF ABBREVIATIONS

ADARB1 adenosine deaminase, RNA specific B1

AEPB2 AE binding protein 2

ASH2L absent, small, or homeotic like

BEND3 ben domain containing 3

BRE1 brefeldin-A sensitivity protein 1

CBX chromobox

CFP1 CXXC finger protein 1

ChIP-sequencing chromatin immunoprecipitation followed by DNA sequencing

CHD3/4 chromodomain helicase DNA-binding protein 3/4

CpG cytosine-phosphate-guanine

DNMT3A/B DNA methyltransferase 3A/B

DOT1 disruptor of telomeric silencing 1

DOT1L DOT1-like

DPY30 dumpy-30 protein homolog

EED embryonic ectoderm development

ESC embryonic stem cell

EZH1/2 enhancer of zeste homolog 1/2

GATAD2A/B GATA zinc finger domain containing 2A/B

G9a/GLP G9a and G9a-like protein

H2AK119u1 Histone H2A lysine 119 monoubiqutination

H2BK120u1 Histone H2B lysine 120 monoubiqutination

H2BK34u1 Histone H2B lysine 34 monoubiqutination

H3/4Ac acetylation of histone H3/4

H3K27me1/2/3 Histone H3 Lysine 27 mono/di/tri methylation

H3K36me1/2/3 Histone H3 Lysine 36 mono/di/tri methylation

H3K4me1/2/3 Histone H3 Lysine 4 mono/di/tri methylation

H3K9me1/2/3 Histone H3 Lysine 9 mono/di/tri methylation

HAT histone acetyltransferase

HBO1 Histone acetyltransferase bound to ORC 1

HDAC histone deactylase complex

HLTF helicase-like transcription factor

HMT histone methyltransferase

HP1 heterochromatin protein 1

JARID2 jumonji, AT rich interactive domain 2

KDM2A/B Lysine demethylase protein 2A/B

MBD2/3 methyl-CpG binding domain protein 2/3

MES-4 mesoderm expressed 4

MLL1/2/3/4 Complex mixed-lineage leukemia 1/2/3/4 complex

MTA1/2/3 metastasis associated protein 1/2/3

NSD1/2/3 nuclear receptor binding SET domain protein 1/2/3

NuA3/4 nucleosomal acetyltransferase of histone H3/H4

NuRD nucleosome remodelling and deacetylase complex

ORC origin recognition complex

P300/CBP P300 and CREB binding protein

PAF Complex polymerase associated factor complex

PCGF1/2/3/4/5/6 polycomb group ring finger 1/2/3/4/5/6

PCL1/2/3 Polycomb-like 1/2/3

PH polyhomeotic

PHD finger plant homeodomain finger

Pol II RNA polymerase II

PRC1 Polycomb repressive complex 1

PRC2 Polycomb repressive complex 2

RAD6 ras-related associated with diabetes protein 6

RBBP5 Retinoblastoma binding protein 5

RING1A/B Really Interesting New Gene 1A/B

RNF20/40 ring finger protein 20/40

RPA replication protein A

RpAb46/48 Rb-associated protein 46/48

RPD3S reduced potassium dependency 3S complex

RYBP RING1 and YY1 binding protein

SAGA Complex Spt-Ada-Gcn5 Acetyltransferase complex

SALL4 sal-like protein 4

SET Su(var)3-9, Enhancer-of-zeste and Trithorax.

SETD1A/B SET domain containing 1A/B

SETD2 SET domain containing 2

SETMAR SET Domain And Mariner Transposase Fusion containing protein 2

SMYD2 SET and MYND domain containing protein 2

SUV3-9 H1/H2 suppressor of variegation 3-9 homolog 1/2

SUZ12 Suppressor of zeste 12 homolog

TSS transcription start site

TTS transcription termination site

WDR5 WD repeat containing protein 5

YAF2 YY1 associated factor 2

ZPF423/521 zinc finger protein 423/521

INTRODUCTION

Eukaryotic genomes are packaged in the form of chromatin. The fundamental unit of chromatin is the nucleosome, which is formed by approximately 147 bp of DNA wrapped 1.7 turns around a histone octamer containing two copies of each histone H2A, H2B, H3, and H4 (Luger et al., 1997). For all vital cellular DNA-based processes such as replication, transcription and repair to occur, chromatin must remain both highly organised and dynamic. With the discovery of epigenetic modifications, including DNA methylation and a plethora of histone modifications, it has been recognised that the epigenetic landscape and the enzymes that catalyse these modifications can impart and regulate this dynamic characteristic of chromatin (reviewed in (Bannister and Kouzarides, 2011; Kouzarides, 2007)).

From genome-wide mapping of histone modifications, different modifications have been found to associate with different types of chromatin, broadly separated into active euchromatin and inactive heterochromatin. Epigenetic modifications like histone acetylation, H3K4me3, H3K36me3, H3K79me3, and H2BK120u1 are hallmarks of euchromatin found at actively transcribed genes (Heintzman et al., 2007; Hon et al., 2009). Transcriptionally repressed facultative heterochromatin is associated with the Polycomb modifications H2AK119u1 and H3K27me3, while constitutively

silenced heterochromatin is localised to the pericentric, centric and telomeric regions and enriched for DNA methylation, H3K9me2, and H3K9me3 (Ho et al., 2014; Hon et al., 2009; Saksouk et al., 2015).

Genic regions carry unique epigenetic features correlated with their transcriptional state. At active genes, enhancers are marked with H3K4me1 and H3K27ac, promoters with H3K4me3, H3/H4 acetylation, and gene bodies with H3K79me3 and H3K36me3 (Ho et al., 2014; Hon et al., 2009). H3K79me3 has also been identified at a subset of developmental enhancers in Drosophila embryos, pointing to a possible new role of this modification in embryogenesis (Bonn et al. 2012). Repressed gene promoters are enriched for H3K27me3, H2AK119u1, and H3K9me3. In early embryogenesis, the promoters of many key developmental genes exist in a state of bivalency and contain low levels of the active mark H3K4me3 and the repressive Polycomb marks H3K27me3 and H2AK119u1 (Azuara et al., 2006; Bernstein et al., 2006). Quantitative modelling studies show that both transcription status and gene expression levels can be predicted from the pattern of epigenetic modifications with a high level of accuracy across a range of cell types and organisms, suggesting that the correlative relationship between specific chromatin signatures and transcription is broadly conserved (Araki et al., 2009; Cheng et al., 2011; Dong et al., 2012; Karlic et al., 2010).

In this chapter, I will first summarise our current understanding of the function of histone modifications in the regulation of transcription, then discuss how the epigenetic landscape is maintained through various crosstalk pathways between epigenetic marks, their writers, readers, and erasers. Lastly, I will introduce the nucleosome remodelling and deacetylase complex NuRD, its structural organisation and role in development.

1.1 The function of histone modifications in transcriptional regulation

The function of histone modifications in active gene transcription

There exist many diverse mechanisms by which histone modifications regulate transcription. Transcriptional start sites of active genes are highly organised to promote the binding and passage of Pol II, this is facilitated by the coordinated action between chromatin remodellers, histone chaperones and histone modifications. Modifications such as acetylation of lysine residues neutralise its positive charge and weaken the electrostatic interaction between nucleosomes or between the histone octamer and the DNA generating a more open chromatin environment. Acetylation of lysines on histones H3 and H4 decondense chromatin arrays *in vitro* (Garcia-Ramirez et al., 1995; Tse et al., 1998), and regions rich in H3/H4ac have greater accessibility as measured by DNAseI digestion (Hebbes et al., 1994; Krajewski and Becker, 1998) and a higher conformational flexibility

(Krajewski and Becker, 1998). Higher order chromatin compaction is dependent on contacts between the H4 tail with the H2A-H2B acidic patch on the adjacent nucleosome, and this compaction is inhibited upon acetylation of H4K16 (Dorigo et al., 2003; Shogren-Knaak et al., 2006).

Histone modifications also act as platforms to recruit effectors proteins (or readers) that create an environment permissive to transcription. The -1 and +1 nucleosomes flanking the TSS are highly positioned with a fast rate of turnover (Struhl and Segal, 2013). The turnover of these histones is mediated by high levels of histone acetylation, H3K4me3, and incorporation of the H2A.Z variant that promotes nucleosome instability. At promoters, histone acetylation of H3/H4 tails and H3K56, located in the globular histone folds, recruit the SWR complex which exchanges H2A-H2B for the variant H2A.Z-H2B (Durant and Pugh, 2007; Watanabe et al., 2013). Furthermore, H2A.Z acetylation stimulates SWR to promote its own retention at +1 nucleosomes (Altaf et al., 2010). H3K4me3, also enriched at active promoters is a docking site for the transcription initiation factor TAF3 (Lauberth et al., 2013) an important factor in the assembly of the transcription preinitiation complex (PIC). H3K4me3 is recognised by the chromatin remodeller CHD1 (Flanagan et al., 2005; Sims et al., 2005) which regulates chromatin accessibility at promoters (Siggens et al., 2015). CHD1 associates with early elongation factors to release stalled Pol II from the promoter (Park et al., 2014; Skene et al., 2014). A PHD finger in the BPTF subunit of the chromatin

remodeller NURF, binds H3K4me3 and recruits NURF to Hox gene promoters where its remodelling activity is important for Hox gene expression (Li et al., 2006; Wysocka et al., 2006). These marks at the promoter work together to promote the recruitment and formation of the PIC at promoters and help Pol II initiate productive early elongation by overcoming the physical barriers imposed by nucleosomes.

Interestingly, another set of histone modifications, H2B ubiquitylation (H2BK120u1 in humans, H2BK123u1 in yeast), H3K79me3, and H3K36me3 mark the bodies of active genes and regulate transcription elongation and other events over coding regions (Bannister et al., 2005; Kouskouti and Talianidis, 2005; Minsky et al., 2008). H2B ubiquitylation stimulates H2A-H2B eviction by FACT and promotes passage of the elongating Pol II (Fleming et al., 2008; Pavri et al., 2006). DOT1L, the enzyme that catalyses H3K79me3, associates with the elongating Pol II and has been identified as a component in multiple elongation factor complexes (reviewed in (Nguyen and Zhang, 2011)). However the mechanisms through which DOT1L and H3K79me3 regulate transcription elongation remain to be elucidated.

The C-terminal domain of RNA Polymerase II can undergo phosphorylation at different serine residues. SETD2 binds the elongating form of Ser2 phosphorylated Pol II and catalyses H3K36me3 in the wake of

transcription (Edmunds et al., 2008). Accordingly, H3K36me3 is restricted to active gene bodies and peaks towards the 3' end of genes (Bannister et al., 2005; Edmunds et al., 2008). In budding yeast, H3K36me3 maintains the integrity of chromatin following histone eviction after Pol II progression to prevent spurious transcription initiation from within gene bodies (Venkatesh et al., 2015). H3K36me3 recruits the histone chaperone FACT, and the chromatin remodellers Isw1 and Chd1 which regulate histone disassembly in front of the elongating Pol II and reassembly following its progression (Carvalho et al., 2013; Smolle et al., 2012). In yeast, H3K36me3 suppresses spurious transcription initiation from gene bodies by recruiting the histone deacetylase Rpd3S, which in turn maintains a hypo-acetylated state that is non-permissive to transcription initiation (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Lickwar et al., 2009). Additionally, H3K36me3 also supresses the incorporation of hyperacetylated H3/H4 from the soluble pool into chromatin by Asf1. Accordingly, loss of H3K36me3 causes increased histone exchange and accumulation of H3K56ac and H4ac over gene bodies (Venkatesh et al., 2012).

Thus it appears that some histone marks in active genes contribute to generating an open environment permissive to transcription at the promoter, while others promote elongation while counteracting erroneous transcription initiation from coding regions of the gene body. Together, these histone marks

regulate both the transcriptional status and accuracy of the genes over which they are deposited.

The function of histone modifications in gene repression

Another set of modifications H3K9me3, H3K27me3, and H2AK119u1 have been associated with gene repression. H3K9me3 is associated principally with constitutive heterochromatin although also with a small number of gene promoters. H3K27me3 and H2AK119u1 are catalysed by the Polycomb complexes PRC2 and PRC1 respectively and are associated with facultative heterochromatin (recently reviewed in (Saksouk et al., 2015)).

Constitutive heterochromatin is localised at the pericentromeric and telomeric regions of chromosomes; these areas are gene-poor and contain many tandem repeats or satellites (Eymery et al., 2009). Constitutive heterochromatin is relatively static and almost always stably silenced independent of cell type. Heterochromatin silencing and spreading is a result of the concerted actions of DNA methyltransferases, chromatin modifiers, and chromatin readers, notably HP1 – the family of proteins that bind to H3K9me3. HDACs and H3K9 histone methyltransferases from the SUV39 family are recruited to constitutive heterochromatin to maintain a hypoacetylated and H3K9me3 rich chromatin environment (Peters et al., 2003; Schotta et al., 2004; Taddei et al., 2001). This in turn recruits the HP1

proteins and the *de novo* DNA methyltransferases DNMT3A/B (Fuks et al., 2003; Lehnertz et al., 2003).

Facultative heterochromatin by contrast, forms at developmentally regulated genes which are activated or silenced depending on developmental cues. The formation of facultative chromatin is associated with the Polycomb complexes PRC2 which catalyses H3K27me3 and PRC1 which catalyses H2AK119u1, that together regulate the spatio-temporal silencing of key developmental genes (Simon and Kingston, 2009). First identified in *Drosophila*, loss or mutation of Polycomb genes results in failure of Hox gene silencing leading to defects in body-patterning (Lewis, 1978; Moazed and O'Farrell, 1992; Struhl, 1981). The mechanistic function of H3K27me3 in gene silencing is poorly understood. H3K27me3 may function as a docking site for PRC2 and PRC1 recruitment, consequently promoting Polycomb spreading at target genes. This function has been proposed to be important for the maintenance of gene repression through cell division, where small amounts of H3K27me3 and the catalytic subunit of PRC2 EZH2 remain associated with mitotic chromatin, and may serve as a site of initiation for the re-establishment of Polycomb domains following DNA replication (Francis et al., 2009; Lanzuolo et al., 2011). H3K27 methylation also blocks H3K27ac by the histone acetyltransferase CBP, as the two marks are by nature mutually exclusive. Loss of H3K27me3 through SUZ12 deletion

results in the accumulation of H3K27ac and de-repression of Polycomb target genes in mESCs (Pasini et al., 2010).

Several mechanisms have been suggested for PRC1 and H2AK119u1 function in gene silencing. Ser-5 phosphorylated Pol II co-occupies Polycomb target promoters genome-wide with PRC1 and PRC2 complexes, but fails to produce mature transcripts (Brookes et al., 2012). H2AK119u1 has been suggested to block Pol II transcriptional elongation at promoters (Stock et al., 2007; Zhou et al., 2008). Another mechanism for Polycomb silencing involves PRC1's ability to compact chromatin independent of its enzymatic activity. Multiple studies have found that PRC1 components induce compaction in nucleosome arrays *in vitro* (Francis et al., 2004; Grau et al., 2011). In *Drosophila* PRC1 mediates the compact chromatin state at Hox loci *in vivo* in a non-enzymatic manner independent of H2AK119u1 (Eskeland et al., 2010). Recent work by Schoenfelder et al. propose a silencing mechanism in which PRC1 physically maintains promoter-promoter contacts between Polycomb target genes in a repressive chromatin network in mESCs (Schoenfelder et al., 2015). Presumably this three-dimensional chromatin architecture is important for excluding factors that would promote gene activation.

1.2 Establishment of the chromatin landscape through epigenetic crosstalk

Polycomb and Trithorax group proteins are two important families of transcriptional regulators. The Polycomb group consists of proteins that form the two chromatin modifying complexes PRC1 and PRC2. The Trithorax group consist of many more members with different functions. These include chromatin remodellers from the SWI/SNF, CHD and NURF families, H3K4 methyltransferases from the SET1 and MLL families, H3K36 methyltransferases, and histone acetyltransferases (Schuettengruber et al., 2011). Importantly, Polycomb and Trithorax act antagonistically to maintain gene repression and activation respectively.

Conventionally, the recruitment of chromatin modifiers to target genes has been attributed to DNA sequence-specific modes via association with transcription factors or non-coding nuclear RNAs. In *Drosophila,* Polycomb and Trithorax proteins are recruited by DNA sequences named Polycomb/Trithorax group response elements (PRE/TREs) (Ringrose and Paro, 2007), yet despite the high degree of conservation of Polycomb/Trithorax proteins, PRE/TRE elements are curiously absent in mammalian genomes (Bauer et al., 2016). Additionally, given the thousands of target sites marked by these epigenetic marks, it appears unlikely that direct interactions with

sequence-specific TFs can mediate Polycomb/Trithorax recruitment to all target sites.

An alternative model towards understanding Polycomb and Trithorax recruitment to target genes has been offered by Klose et al. The authors propose that Polycomb and Trithorax proteins sample chromatin genome wide and are stabilised or repelled by feedback mechanisms dependent on the local chromatin environment they encounter (Klose et al., 2013). In fact, this model of regulation through the interplay between epigenetic marks is not only limited to Polycomb/Trithorax but can be extended to other chromatin modifying enzymes. The vast majority of chromatin modifiers possess domains which can bind different histone and DNA modifications, and in many cases the modification that they catalyse, and their recruitment and activity are often regulated by histone crosstalk. In this section, I will discuss multiple examples of histone crosstalk and how the interplay between the various types of chromatin modifications contribute to the distinct epigenetic signature seen at active and repressed genes.

Crosstalk at active genes

The pattern of H3K4me3 is highly coupled to CpG islands (CpGI), regions of GC and CpG dense DNA devoid of 5mC methylation and enriched at vertebrate promoters (Deaton and Bird, 2011). CpGI promoters are bound

by Zn-finger CxxC domain containing proteins, which include the H3K4 methyltransferase complexes MLL1 and MLL2, and CFP1, a subunit of the SET1A and SET1B complexes. CFP1 and the SET1A/B complexes are enriched at active promoters with the highest levels of H3K4me3 (Clouaire et al., 2012), while MLL2 is responsible for the H3K4me3 observed at bivalent promoters with low or no expression (Denissov et al., 2014; Hu et al., 2013). Bivalent promoters are Polycomb target genes that harbour both the active H3K4me3 mark and the repressive H2AK119u1 and H3K27me3. The presence of H3K4me3 and H3K4 writers at these Polycomb repressed genes may poise these sites for rapid activation upon developmental cues.

H2BK120u1 in mammals (or H2BK123u1 in yeast) positively regulates H3K4me3 and H3K79me3 at target genes. In yeast, depletion of the E2/E3 enzymes Rad6/Bre1 which catalyse H2BK123u1 severely reduces levels of H3K4me3 (Sun and Allis, 2002). This crosstalk pathway is conserved in mammals, as H3K4me3 levels reduce globally upon knockdown of the Bre1 homologs, RNF20/40 in human cells (Kim et al., 2005). *In vitro* studies have observed that ubiquitylation of H2BK120 and H2BK34 allosterically stimulates MLL activity through an interaction between the MLL subunit ASH2L with the ubiquitin (Wu et al., 2013a). These two lysine residues H2BK120 and H2BK34 lie at the exposed surface of the nucleosome and their coupling to ubiquitin may provide a substrate favourable for MLL binding or activity.

H3K4me3 writers also positively regulate the deposition of their own marks through a feed-forward mechanism. CFP1 contains a PHD finger domain that is able to bind H3K4me3, thereby targeting or reinforcing SET1's accumulation at sites already containing its mark (Eberl et al., 2013; Murton et al., 2010; Shi et al., 2007). A PHD finger domain in the MLL1 has been shown to bind H3K4me3 and recruit MLL1 to Hox gene targets (Wang et al., 2010). Other PHD finger domains within SET1 and MLL complexes have also been found to interact with H3K4me3 *in vitro* (reviewed in (Ali et al., 2014)), and may contribute to the self-propagation of H3K4me3.

Several pathways of histone crosstalk exist between H3K4me3 and histone acetylation. In budding yeast, H3K4me3 recruits the NuA3 histone acetyltransferase complex to target sites where it deposits H3K14ac (Taverna et al., 2006). Similarly, a tandem tudor domain in the SAGA HAT complex links H3K4me3 binding to H3K9 acetylation at gene targets (Bian et al., 2011; Vermeulen et al., 2010). The budding yeast NuA4 and the mammalian HBO1 are two other HAT complexes that contain PHD fingers able to bind H3K4me3 (Doyon et al., 2004; Hung et al., 2009). Similar to HATs, histone deacetylase complexes may also be recruited to sites of H3K4me3 through direct recognition of the mark. ING2, a component of the SIN3-HDAC1 deacetylase complex binds specifically to H3K4me2/3 (Shi et al., 2006). H3K4me3 containing histones show high levels of acetylation turnover (Crump et al., 2011), and HAT and HDAC complexes colocalise with

H3K4me3 at active gene promoters genome wide, as seen by ChIP-sequencing (Wang et al., 2009b). H3K4me3 may be one mechanism responsible for the recruitment or reinforcement of HAT and HDAC activity at active genes.

In summary, the colocalisation of histone acetylation, H3K4me3, and H2Bub at active gene promoters synergistically regulate their own accumulation through multiple positive feedback mechanisms. The maintenance of these marks at promoter regions help govern productive transcription initiation.

At gene body regions, the crosstalk between histone deacetylation and H3K36me3 prevents spurious transcription initiation. In budding yeast, H3K36me2/3 is recognised by the chromodomain of Eaf3, a component of the Rpd3S HDAC complex (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2009). Rpd3S recruitment to active gene bodies ensures a hypoacetylated state over coding regions that inhibits aberrant transcription initiation events. Upon loss of Set2 and H3K36me3, subsequent increases in histone acetylation over coding regions leads to aberrant transcription initiation resulting in the production of short intragenic transcripts (Carrozza et al., 2005).

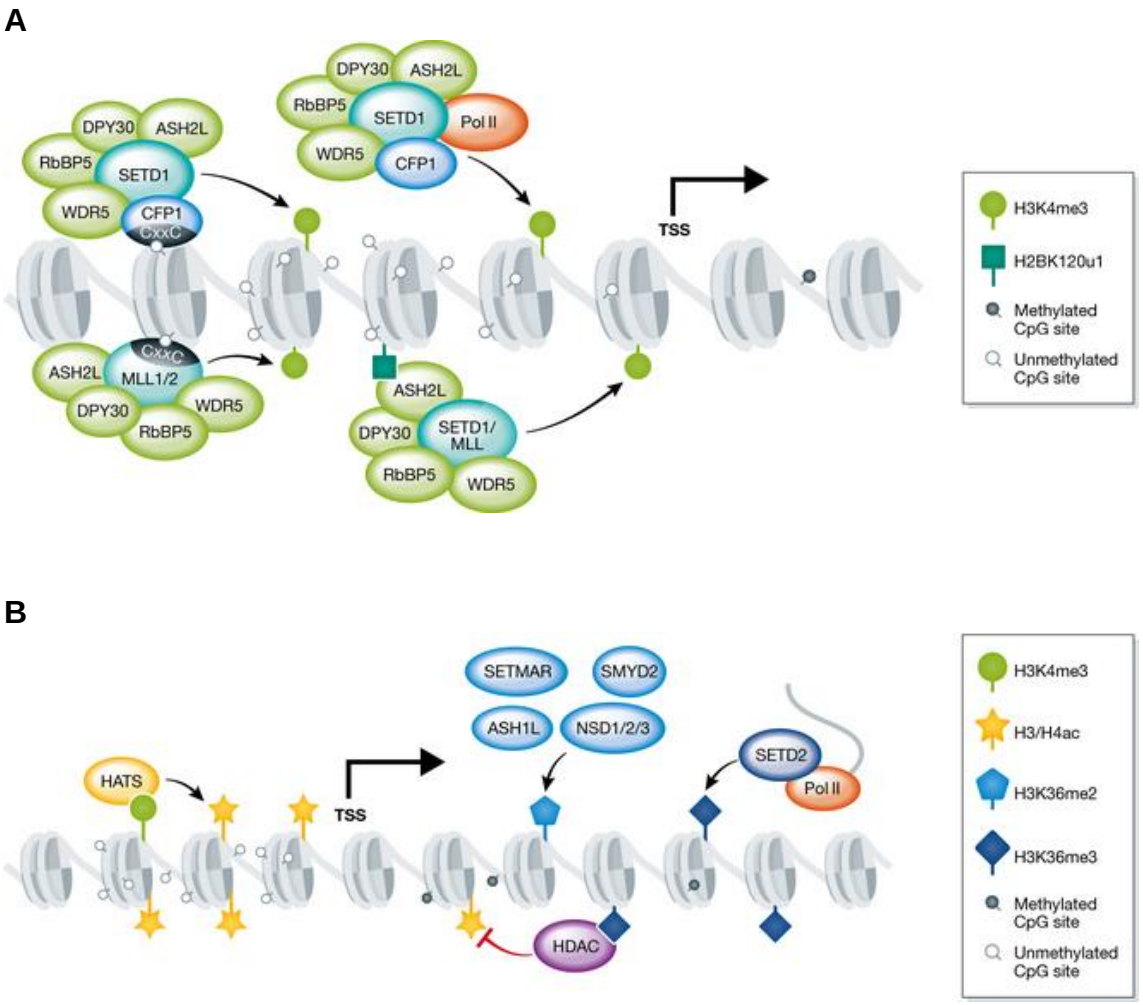


Figure 1. Chromatin crosstalk pathways at active genes

(adapted from Zhang et al. 2015)

A At active gene promoters, H3K4 methyltransferases SETD1 and MLL1/2 are recruited by binding to unmethylated DNA sequences called CpG islands, and through prior deposition of H2BK120u1.

B At active gene promoters, H3K4me3 recruits histone acetyltransferases. On gene bodies, multiple H3K36 methyltransferases catalyse H3K36me1 and H3K36me2. SETD2 associates with the elongating RNA Pol II and catalyses H3K36me3. In budding yeast, both H3K36me2 and H3K36me3 can recruit an HDAC complex to remove histone acetylation from gene bodies.

While both are hallmarks of active transcription H3K4me3 and H3K36me3 occupy spatially distinct regions. Several mechanisms regulate the spatial separation of H3K4me3 and H3K36me3. The H3K4 writer SET1 associates with Ser-5 phosphorylated initiating form of Pol II, while the H3K36 writer SETD2 associates with the Ser-2 phosphorylated elongating form of Pol II, leading to H3K4me3 enrichment at promoters and H3K36me3 over gene bodies. Negative crosstalk exists between DNA methylation and H3K4me3. DNA methylation seems to repress the establishment of H3K4me3 (Okitsu and Hsieh, 2007), and confines it to nonmethylated CpG islands at promoters and occludes it from highly methylated gene body regions. Meanwhile, H3K36me2/3 is restricted to gene bodies in part through the active removal of H3K36me2 from the promoter by the CpGI binding H3K36 demethylases of the KDM2 family (Blackledge et al., 2010; Farcas et al., 2012). These crosstalk pathways ensure the establishment of a chromatin landscape permissible to transcription at the TSS and refractory to transcription initiation from coding regions.

Crosstalk at Polycomb repressed genes

The Polycomb complexes PRC1 and PRC2 target several hundred promoters genome wide. Many positive feedback loops regulate the establishment and accumulation of the Polycomb marks H2AK119u1 and H3K27me3. PRC2 binds to H3K27me3, the modification it deposits through

its subunit EED a WD40 repeat containing protein. The binding of H3K27me3 by the aromatic cage at the centre of the EED's beta-propeller allosterically stimulates the activity of EZH2, the catalytic subunit of PRC2 (Margueron et al., 2009). Binding of PRC2 to H3K27me3 is required for the proper formation of H3K27me3 domains at target sites. PRC1 also recognises its own mark H2AK119u1 via its subunit RYBP and its homolog YAF2. RYBP contains a ubiquitin binding domain that is able to bind H2AK119u1 *in vivo* (Arrigoni et al., 2006)

Both PRC1 and PRC2 and their respective marks also reinforce one another's activity. Canonical PRC1 complexes bind the PRC2 mark H3K37me3 through its chromodomain-containing CBX subunit (Ren et al., 2008). CBX2/4/6/7/8 have all been shown to bind H3K27me3 peptides *in vitro* (Kaustov et al., 2011). It was initially believed that PRC2 initiated Polycomb silencing at target genes and deposition of H3K27me3 would recruit PRC1 through CBX. However, occupancy of PRC1 at Polycomb target sites in the absence of PRC2 pointed to another mechanism of PRC1 recruitment that was independent of PRC2 (Schoeftner et al., 2006; Tavares et al., 2012). Discovery of a biochemical interaction between a variant PCGF1-containing PRC1 and the CxxC domain containing KDM2B provided a mechanism through which variant PRC1 could be recruited *de novo* to CpG islands genome wide (Farcas et al., 2012; Wu et al., 2013b). Recent studies have found that H2AK119u1, deposited principally by variant PRC1 complexes, can

recruit PRC2 complexes to chromatin (Blackledge et al., 2014; Cooper et al., 2014). H2AK119u1 recruits and stimulates the activity of the PRC2 complex containing the accessory factors JARID2 and AEBP2 (Kalb et al., 2014).

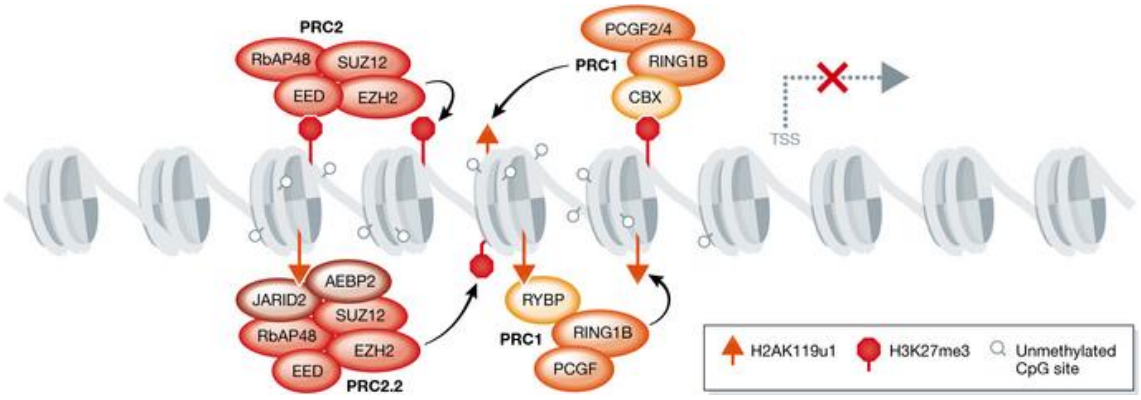


Figure 2. Chromatin crosstalk pathways at Polycomb repressed genes

(adapted from Zhang et al. 2015)

Polycomb complexes PRC1 and PRC2 positively reinforces the accumulation of their own and one another's recruitment. PRC2's binding to its own mark H3K27me3 via EED stimulates its activity. Variant RYBP-PRC1 can also bind its own mark H2AK119u1 through RYBP. PRC2 can recruit canonical CBX-PRC1 complexes through the ability of CBX to bind H3K27me3. Similarly, JARID2/AEBP2-PRC2 can bind the PRC1 mark H2AK119u1.

These examples illustrate how many different mechanisms and positive feedback pathways lead to the recruitment and accumulation of PRC1 and PRC2 at Polycomb targets. Notably, these studies also highlight the importance of accessory factors and the unique functions of variant complexes in regulating PRC1 and PRC2 function.

Crosstalk between active and repressive chromatin modifications

Histone crosstalk regulates the transcriptional state of genes at active and repressed genes, whereby marks associated with either the on or off state positively reinforces the accumulation of marks associated with the same transcription state. Moreover crosstalk between writers and marks of opposite transcriptional states can also antagonise each other's function. The dynamic interplay between epigenetic marks and the plethora of feedback systems may play a role in creating a bistable state primed for the rapid switching of gene expression during development and differentiation.

The crosstalk between Polycomb and Trithorax complexes have been well characterised. In *Drosophila*, Polycomb and Trithorax proteins regulate the spatio-temporal expression of Hox genes during development through their opposing functions of gene repression and activation respectively. Histone crosstalk governs the molecular mechanisms responsible for this mutual antagonism. Schmitges et al. showed that the presence of Trithorax modifications H3K4me2/3 and H3K36me2/3 on recombinant nucleosomes directly inhibits the activity of PRC2 (Schmitges et al., 2011). The VEFS domain of SUZ12 has been proposed to allosterically inhibit the catalytic subunit EZH2 upon binding to H3K4me3 or H3K36me3 *in cis* (Schmitges et al., 2011). The ability of H3K36me3 to inhibit catalysis of H3K27me3 by PRC2 is congruent with mass spectrometry data from several cell types

showing that H3K36me3 and H3K27me3 never or rarely exist on the same H3 tail in multiple cell types (Jung et al., 2010; Yuan et al., 2011; Zheng et al., 2012). SETD2 deletion and loss of H3K36me3 leads to an increase in H3K27me3 over active genes coupled with a reduction in expression (Yoh et al., 2008). Importantly this inhibition only occurs *in cis*, and PRC2 is able to catalyse H3K27me3 on H3 in a nucleosome where the other H3 harbours H3K4me3 or H3K36me3 (Voigt et al., 2012).

Polycomb in turn, has been shown to inhibit Trithorax activity. Several H3K36 methyltransferases are inhibited by the PRC1 mark H2AK119u1 *in vitro*. H3K4me3 writers MLL1 and possibly MLL3 may also be inhibited by H2AK119u1 *in vitro* (Nakagawa et al., 2008; Wu et al., 2013a). There is some *in vivo* work that supports this model. For example Endoh et al. see, upon deletion of RING1A/B the catalytic subunit of PRC1 and loss of H2AK119u1, levels of H3K4me3 increase at Polycomb target genes (Endoh et al., 2012).

Together, these crosstalk processes may explain the absence of PRC2 and H3K27me3 at actively transcribed genes with high levels of promoter H3K4me3 and gene body H3K36me3. While, at bivalent Polycomb target genes, low levels of H3K4me3 coexist with H3K27me3 and H2AK119u1 resulting in low or no transcriptional output. Regardless, many parts of the complex network of interactions between Polycomb and Trithorax remain to

be explained. An intriguing series of studies found that the PRC2-associated PCL1-3 proteins contain tudor domains that bind H3K36me3 with high specificity (Ballare et al., 2012; Brien et al., 2012; Cai et al., 2013; Musselman et al., 2012). However, neither PCL nor any other PRC2 components colocalise with H3K36me3 by ChIP-sequencing (Ballare et al., 2012; Cai et al., 2013). The presence of an H3K36me3 reader within the PRC2 complex may be interpreted in several ways. Firstly, while active gene bodies lack H3K27me3, they do contain H3K27me1/2, thus PCL-PRC2 may be important for the maintenance of these marks throughout the genome. Another interpretation is that PCL mediated recognition of H3K36me3 may aid the switching of transcriptional states by *de novo* recruitment of PRC2 to active genes upon specific developmental cues. In support of this view the recruitment of the H3K36 demethylase NO66 has been linked to PCL3; prior removal of H3K36me2/3 would allow the subsequent deposition of H3K27me3 by PRC2 (Brien et al., 2012).

The interplay between acetylation and methylation at H3K27 is another example of antagonism between Polycomb and Trithorax proteins. H3K27ac is deposited by the Trithorax histone acetyltransferases p300 and CBP in mammals (Pasini et al., 2010). H3K27ac and any degree of methylation (me1/2/3) at H3K27 are mutually exclusive. As previously mentioned, knockout of the PRC2 subunit SUZ12 depletes H3K27me3 and leads to an increase in H3K27ac deposition by p300 and CBP at Polycomb

target genes in mammalian cells (Jung et al., 2010; Pasini et al., 2010). Similarly, knockdown of E(Z) the catalytic component of PRC2 in *Drosophila*, results in a global increase in H3K27ac which negatively affects Polycomb silencing (Tie et al., 2009). In flies CBP physically interacts with the H3K4 methyltransferase Trithorax (Petruk et al., 2001; Tie et al., 2009), and is also recruited by H3K4me3 to promoters (Tie et al., 2009). In addition, H3K4me1 has been shown to stimulate CBP-dependent catalysis of H3K27ac; these two marks are both enriched at enhancers and may function together to prevent the spreading of H3K27me3 into enhancer regions (Tie et al., 2014).

Phosphorylation of H3S28 is known to promote H3K27ac by inhibiting H3K27me3 (Gehani et al., 2010; Lau and Cheung, 2011). Consequently removal of H3K27ac is required for Polycomb silencing. The nucleosome remodelling and deacetylase complex NuRD has been implicated in facilitating Polycomb silencing through the removal of H3K27ac. The NuRD complex is required for the recruitment of PRC2 and accumulation of H3K27me3 at Polycomb gene targets and for the silencing of pericentric heterochromatin upon loss of DNA methylation (Reynolds et al., 2012; Saksouk et al., 2014). In ESC differentiation, a transcriptional corepressor CTBP2 recruits NuRD for the removal of H3K27ac at active genes allowing for PRC2 mediated H3K27me3 to occur (Kim et al., 2015). Thus NuRD mediated PRC2 recruitment may also facilitate the switching of transcriptional states at a subset of active genes during ESC differentiation.

In summary chromatin modifications seem to have several roles in transcriptional regulation. Modifications can affect the accessibility of chromatin by altering the physical interactions within and between nucleosomes. They further act as docking sites for effector proteins that regulate the various aspects of transcription from formation of the preinitiation complex, to transcriptional elongation, to maintaining the accuracy of transcription initiation events. Chromatin crosstalk may also be important for transcriptional memory, as marks associated with a transcriptional state positively reinforce the same transcriptional state while inhibiting writers and marks associated with the opposing state. These feedback systems may be important for re-establishing gene expression patterns following DNA replication and cell division. X-chromosome inactivation in females is an important example of transcriptional memory; the lncRNA Xist triggers this process early in development but then the silenced state is maintained throughout cell differentiation through the concerted action of a complex network of factors including chromatin modifiers. Importantly, chromatin modifications generate the dynamic bistability that allows the rapid switching of gene expression states upon developmental or environmental triggers.

1.3 Biology of the NuRD complex

NuRD is a highly conserved protein complex that has been purified from multiple cell types and organisms (Tong et al., 1998; Wade et al., 1998; Xue et al., 1998; Zhang et al., 1998). When the NuRD complex was first discovered, it was unique in that it contained two enzymatic components, one with chromatin remodelling function and the other with histone deacetylase function. NuRD is composed of six basic subunits, the chromatin helicase-DNA-binding protein CHD3/4, histone deacetylases HDAC1/2, the metastasis-associated MTA1/2/3, the retinoblastoma-binding protein RBBP4/7, the methyl-CpG-binding domain MBD2/3, and the gata zinc finger domain containing GATAD2A/B.

CHD3/4 belongs to the CHD family of chromatin remodellers that use the energy of ATP to mobilise nucleosomes. CHD3 and CHD4 are large ~200 kDa proteins that harbour two Zn finger PHD domains, a tandem chromodomain, and a SNF2-like ATPase domain. The PHD domains, chromodomains and ATPase domain are all required for association with chromatin and NuRD-mediated repression (Ramirez et al., 2012; Watson et al., 2012). Genome-wide mapping places CHD4 at the promoters of active and bivalent genes; CHD4 is enriched at the -1 and +1 nucleosomes flanking the DNAseI hypersensitivity site and is particularly enriched at the -1 nucleosome (de Dieuleveult et al., 2016). CHD4 is critical in early

development, with knockout embryos failing to implant at E3.5-4.0 due to defects in trophectoderm formation (O'Shaughnessy-Kirwan et al., 2015). Accumulating evidence suggest that CHD3/4 also functions outside of the NuRD complex, and has been implicated in NuRD-independent transcriptional regulation, DNA repair and cell cycle progression (O'Shaughnessy and Hendrich, 2013).

HDAC1 and HDAC2 belong to the Class I HDACs which are homologous to Rpd3 in yeast. HDAC1 and HDAC2 are highly similar, with an N-terminal dimerisation domain, a large deacetylase domain spanning most of the protein, and a more divergent C-terminal domain. HDAC1 and HDAC2 exist as homodimers or heterodimers *in vivo*, and this dimerisation is important for deacetylase activity (Khan et al., 2013; Luo et al., 2009). HDAC1 and HDAC2 associate with a variety of proteins to form three different classes of deacetylase complexes – NuRD, SIN3, and CoREST. By chromatin immunoprecipitation, HDAC1/2 are enriched at the promoters and proximal gene bodies of active genes (Kidder and Palmer, 2012; Wang et al., 2009b) and correlate well with histone acetylation and gene activity in ES and TS cells (Kidder and Palmer, 2012). While these two proteins seem to have redundant functions in many cellular and tissue models, germ-line deletion of HDAC1 and HDAC2 have different consequences. Deletion of HDAC1 in mice results in embryonic lethality by day E9.5, while HDAC2 knockout mice die shortly after birth (Montgomery et al., 2007).

The MTA proteins have been well studied in cancer research, where MTA is aberrantly expressed in many cancer lines. MTA are highly conserved among vertebrates and invertebrates. Three MTA proteins MTA1/2/3 exist in vertebrates. They are highly similar in the N-terminus in the BAH, ELM2, and SANT domains but diverged in the C-terminal end of the protein. MTA1-3 do not have any enzymatic activity, instead serving as a scaffold protein that binds and assembles the other components of the NuRD complex.

RBBP4 and RBBP7 are WD-repeat domain containing proteins found in many chromatin complexes including NuRD, SIN3, PRC2, and the CAF-1 complex. They have histone-binding ability and may function to target or stabilise binding of these complexes to chromatin. RBBP4 and RBBP7 maintain the deacetylated state of histones in oocyte maturation (Balboula et al., 2014, 2015) and depletion of the RBBP4/7 homolog in *Drosophila* results in homeotic transformations similar to the Polycomb mutant phenotype (Anderson et al., 2011).

MBD2/3 are an integral part of the NuRD complex. MBD2 and MBD3 are highly homologous in mammals. MBD2 contains a canonical methyl DNA binding domain and recruits NuRD to methylated DNA, but MBD3 lacks four important residues in the MBD domain and does not bind methylated DNA (Zhang et al., 1999). From biochemical studies, MBD2 and MBD3 are

mutually exclusive within NuRD (Le Guezennec et al., 2006). NuRD complexes with MBD2 and MBD3 may be targeted to very different parts of the genome. In accordance with this, MBD2 and MBD3 appear to have distinct functions as MBD3 null mice are embryonic lethal (E8.5) while MBD2 mice are viable (Hendrich et al., 2001). Recruitment of NuRD to methylated gene promoters is dependent on MBD2 (Ramirez et al., 2012). MBD3 maps to the promoters of active and bivalent genes, and its localisation is severely affected by loss of TET1 the enzyme which converts 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) (Yildirim et al., 2011).

GATAD2A and GATAD2B (also p66 and p68) were first identified as part of the NuRD complex. Further characterisation revealed that GATAD2A/B harbours an N-terminal coiled coil region that bind MBD2/3, and C-terminal GATA-Zn finger (Feng et al., 2002). It has been shown that GATAD2A/B represses transcription from a reporter gene (Brackertz et al., 2006; Feng et al., 2002), and this activity is in part through the deacetylase activity of NuRD (Feng et al., 2002). GATAD2A and GATAD2B may function together to repress transcription as knockdown of one affected the other's ability to repress transcription from a reporter (Brackertz et al., 2006). The nuclear localisation of GATAD2A/B is also dependent on its binding to MBD2 (Brackertz et al., 2006). GATAD2A/B has been reported to play an important role in organ development in organisms from *Drosophila* to Humans (Kim et al., 2004; Vasquez-Doorman and Petersen, 2016; Willemsen et al., 2013).

GATAD2A is critical in development, as knockout mice show defects in multiple tissues and die at embryonic day E9.5 (Marino and Nusse, 2007).

1.4 Structural organisation of the NuRD complex

Structural studies have provided much insight into the structural organisation of NuRD. The crystal structure of HDAC1 with the ELM2-SANT domains of MTA1 shows that MTA1 acts as a scaffold around HDAC1 (Millard et al., 2013). The interaction surface between the MTA1 SANT domain and HDAC1 forms a pocket for inositol phosphate binding. Inositol phosphate activates HDAC activity and these residues are highly conserved in HDAC1, 2, and 3 complexes. The ELM2 domain wraps around HDAC1 and also dimerises bringing two copies of MTA1 and HDAC1 into one complex (Millard et al., 2013).

Interaction of RBBP4/7 with MTA1 is required for its inclusion in the NuRD complex (Alqarni et al., 2014). Each MTA1 proteins appears to bind two RBBP4/7 proteins through two regions (464-546 and 670-691) in the C-terminal half of the protein (Alqarni et al., 2014; Millard et al., 2016). The two RBBP4/7 interaction motifs in MTA1 occupy the same surface on RBBP4/7 as histone H4 (Alqarni et al., 2014; Millard et al., 2016). Thus RBBP4/7 in the NuRD complex may only have H3 binding activity. A negative stain EM model of this core deacetylase complex of NuRD shows two MTA1,

two HDAC1 and four RBBP4 molecules arranged in a zig-zag pattern centred around the MTA1 dimer (Millard et al., 2016). RBBP4/7 has also been found to be the most dynamic subunits of NuRD that dissociate more readily from the complex than other components, possibly owing to their presence in many nuclear complexes (Kloet et al., 2015).

There is no structural information for the interaction between the deacetylase core of NuRD with MBD2/3, CHD3/4, and GATAD2A/B, but crosslinking studies have mapped crosslinks between MTA and all components of the NuRD complex (Kloet et al., 2015). Biochemical studies show that methylation of MTA1 residue K562 is required for its association with CHD4, though the actual interaction motif remains to be characterised (Nair et al., 2013).

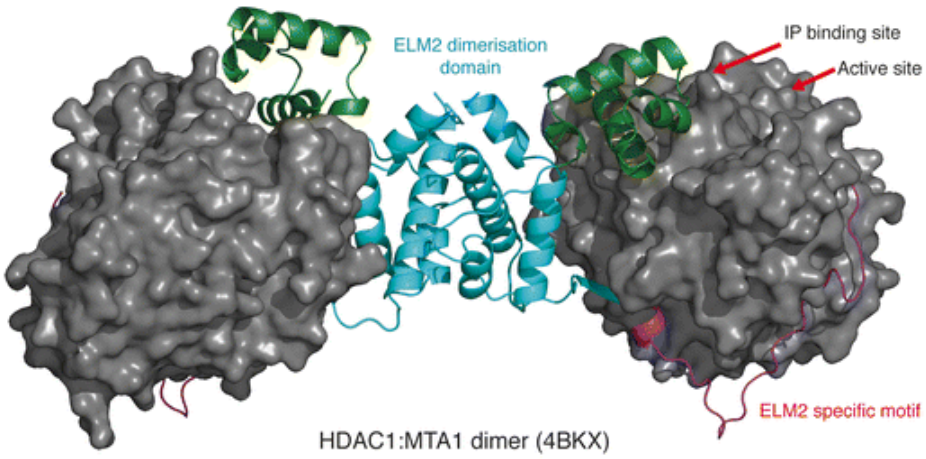
Multiple interactions also exist between MBD2/3, CHD3/4, and GATAD2A/B. The coiled-coil domains in GATAD2A and MBD2 heterodimerises to form the site of interaction between these proteins (Gnanapragasam et al., 2011). Interactions between CHD3/4 and GATAD2A/B have also been captured by crosslinking (Kloet et al., 2015).

A large degree of heterogeneity exists within the NuRD complex due to the number of subunits and their paralogs. At this point, it is unclear if all

paralogous subunits are interchangeable or if they form paralog-specific interactions that confer different biological functions to the complex. Furthermore the stoichiometry of NuRD may also be variable. As previously mentioned RBBP4/7 subunits dynamically associate with the complex (Kloet et al., 2015), and CHD4 is also observed to be substoichiometric within NuRD (Low et al., 2016; Zhang et al., 2016).

A study of NuRD complex assembly in *Drosophila* which only has one gene for each subunit showed that the *Drosophila* homologs of MTA1-3, HDAC1/2, RBBP4/7, and MBD2/3 associate in a stable subcomplex that transiently associates with CHD3/4 and GATAD2A/B (Zhang et al., 2016). This subcomplex has a stoichiometry of 4:2:2:1 for RBB4/7: MTA1-3: HDAC1/2: MBD2/3 and displays robust deacetylase activity. CHD3/4 is also functional as a chromatin remodeller outside of the NuRD complex, and has many NuRD-independent roles in chromatin biology (O'Shaughnessy-Kirwan et al., 2015; O'Shaughnessy and Hendrich, 2013). Therefore it appears that several subcomplexes of NuRD co-exist stably within the cell and the formation of the holo-NuRD complex may occur only transiently at certain sites on chromatin to perform specific functions.

A



B

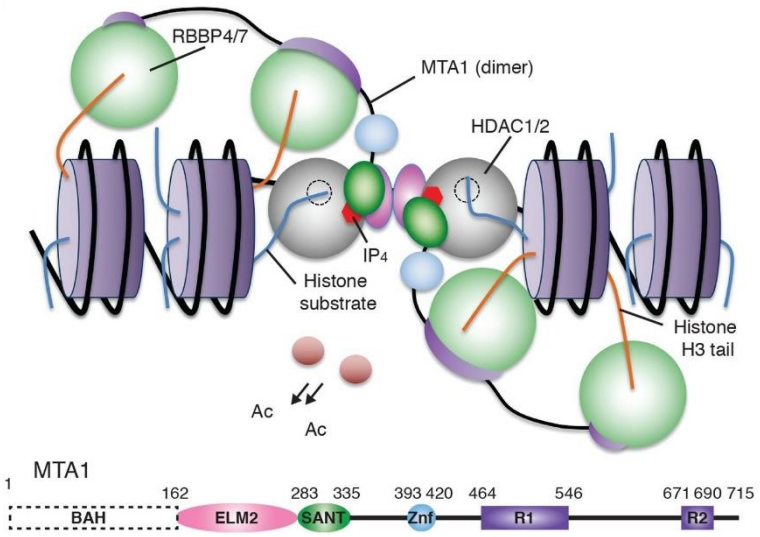


Figure 3. Crystal structure of HDAC1 in complex with ELM2-SANT domains from MTA1.

A (adapted from Millard et al. 2014) MTA1 homodimerises through the ELM2 dimerisation domain (cyan). The ELM2 specific motif wraps around HDAC1 (grey). An inositol phosphate binding pocket is formed at the interaction surface between HDAC1 and the SANT domain (green) of MTA1.

B (adapted from Millard et al. 2016) The deacetylase core of NuRD is form by the dimerization of MTA1 which acts as a scaffolding protein. Each MTA1 protein can bind one HDAC1/2 via its ELM2-SANT domains and up to two RBBP4/7 through the R1 and R2 domains. In this model, RBBP4/7 binding to histone tails recruits the complex to chromatin and mediates histone deacetylation through the HDAC1/2 subunits.

1.5 Regulation of NuRD by accessory factors

Multiple core subunits within NuRD such as MBD2, MBD3, and CHD4 have been found to be important for NuRD's localisation as described earlier. Moreover, NuRD is known to interact with many chromatin binding proteins that may recruit the complex to different targets genome wide. The H3K4 demethylase LSD1 biochemically co-purifies all NuRD complex subunits in mammalian cells and co-occupies many target promoters with MTA1-3 (Wang et al., 2009a). BEND3, which colocalises with HP1 and H3K9me3 at pericentric heterochromatin, is another protein that has been found to interact with NuRD and SALL4 and may recruit NuRD to heterochromatin (Saksouk et al., 2014; Sathyan et al., 2011). In *Xenopus* the interaction of NuRD with the transcriptional factor FOG is essential for primitive erythropoiesis (Mimoto and Christian, 2012).

Multiple other binding partners including PHF6, CDK2AP1, and APPL1 have been identified (Banach-Orlowska et al., 2009; Spruijt et al., 2010; Todd and Picketts, 2012). A plethora of transcription factors have also been found to associate with NuRD. The vast array of interactors may mediate the recruit of NuRD to genomic targets where its deacetylase and chromatin remodelling activities are necessary for gene regulation during differentiation and development.

PROJECT AIMS

1. Studying the molecular mechanisms governing the mutual antagonism between Polycomb and Trithorax proteins

While it is well established in the field that Polycomb and Trithorax proteins have opposing functions in the regulation of gene expression, the molecular mechanisms governing their mutual antagonism were not well characterised. Trithorax modifications H3K4me3 and H3K36me3 are known to inhibit the activity of PRC2, but their effect on PRC1 activity had not been studied. We hypothesised that histone crosstalk exists in both directions and that Polycomb modifications would negatively regulate the activity of Trithorax enzymes as well. Using recombinant nucleosomes, I performed *in vitro* enzyme assays to test these ideas.

2. Identification of novel readers of the Polycomb modification H3K27me3 and the Trithorax modification H3K36me3

Identification of epigenetic readers can give great insight into the functions of histone modifications. Multiple readers of H3K27me3 were identified in a previous screen by Bartke et al. 2010, but readers of H3K36me3 in a nucleosome template had not been performed. I set out to find novel readers of H3K27me3 and H3K36me3 using a pulldown from SILAC labelled nuclear extract with recombinant H3K27me3 and H3K36me3 modified nucleosomes.

RESULTS CHAPTER 1

Histone crosstalk between Polycomb and Trithorax

Introduction

Polycomb group and Trithorax group proteins are evolutionarily conserved transcriptional regulators that control the expression of developmental genes during embryogenesis. First discovered as regulators of Hox gene expression critical in body pattern development in *Drosophila*, Polycomb and Trithorax proteins have since been found to co-occupy thousands of target sites genome-wide in flies and vertebrate organisms. Polycomb and Trithorax are chromatin modifying enzymes that epigenetically regulate the expression of key developmental genes through their antagonistic functions of gene repression and activation respectively. The co-occupancy of Polycomb and Trithorax proteins with opposing functions at key developmental genes point to the importance of bistability in embryonic development, when the interplay between multiple factors can mediate the rapid turning on or off of genes during differentiation. How Polycomb and Trithorax are recruited to these shared target genes and their antagonistic interplay in controlling gene expression is a longstanding area of interest in the field of epigenetics.

There is emerging evidence that histone crosstalk plays an important role in the establishment and maintenance of Polycomb and Trithorax modifications over genic regions. Apart from their chromatin modifying ability, Polycomb and Trithorax proteins possess multiple chromatin binding domains capable of reading the epigenetic landscape. In an elegant *in vitro* study by Schmitges et al., the activity of the Polycomb complex PRC2 towards H3K27 was inhibited by the presence of H3K4me3 and H3K36me3, two Trithorax marks associated with active transcription. This work suggested that establishment of Trithorax at active genes could negatively regulate the recruitment/activity of the opposing Polycomb proteins. Furthermore, mass spectrometry of H3 tail peptides showed that H3K27me3 and H3K36me3 are mutually exclusive. We hypothesised that multiple histone crosstalk pathways must exist between these two groups of proteins. Using recombinant enzymes and modified nucleosome substrates, I tested the effect of different histone modifications on the activity the Polycomb complexes PRC1 and PRC2 which catalyse H2AK119u1 and H3K27me3 respectively, and the Trithorax H3K4me3 methyltransferases SET1B, MLL1, and MLL4 complexes and the H3K36me1/2 methyltransferase NSD2.

Results

Purification of NSD2 and NSD2$_{SET}$

The expression and purification of full length NSD2 was first attempted using large scale transient transfection in HEK293T mammalian

cells. Figure 1.1A shows the expression and purification of three different NSD2 constructs. NSD2 expression (band above 150kDa as seen by anti-Flag western) 48h post transfection was achieved in all three constructs, however only Flag purification of the C-terminally tagged NSD2-Flag yielded un-degraded protein but at very low amounts. I also tried in parallel baculovirus expression of Flag-NSD2 in Sf9 insect cells (Figure 1.1B). After optimising for several factors such cell density, multiplicity of infection, infection duration, and protein purification strategy, pure full length Flag-NSD2 was achieved. However NSD2 expression in the baculovirus/Sf9 system was also very low, and it was difficult to obtain large quantities of the pure protein for downstream biochemical applications at the scale I was using for expression and purification. Thus for long term use, baculovirus/insect cell derived full length NSD2 was purchased from the company Reaction Biology Corporation, of which a coomassied stained gel can be seen in Figure 1.2B. The catalytic domain of the NSD2 protein, $NSD2_{SET}$, was expressed as 6xHis-$NSD2_{SET}$ in the *E. coli* strain BL21 DE3 pLysS and affinity purified using a His-trap purification followed by gel filtration (Figure 1.1C and D). Milligram quantities of the protein was obtained.

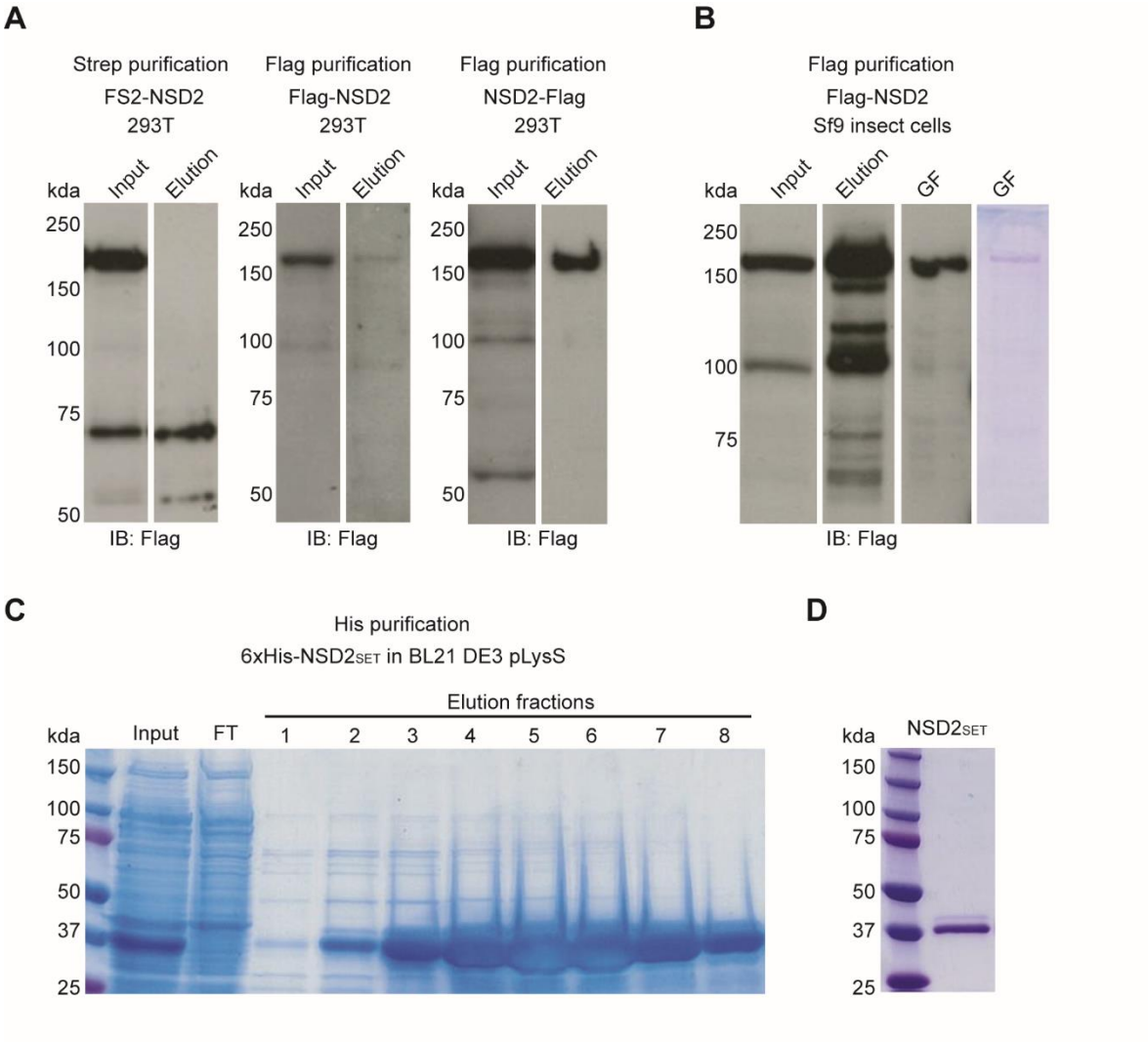


Figure 1.1 Purification of full length NSD2 and NSD2SET

A Expression and purification of different full length NSD2 constructs from mammalian HEK 293T cells. The input and elution samples were probed with an antibody against Flag.

B Baculovirus expression and purification of Flag-NSD2 from Sf9 insect cells. Input, elution, and the purified sample after gel filtration (GF) were probed with an antibody against Flag. A coomassie of the purified sample is also shown.

C His-trap purification of $NSD2_{SET}$ catalytic domain from *E. coli*. $NSD2_{SET}$ runs at 37 kda.

D Coomassie stained gel of purified and concentrated $NSD2_{SET}$.

Enzymes for biochemical assays

Figure 1.2 shows all the enzymes used for biochemical assays, 1-5 μg of protein was loaded for each sample. Many of these enzymes were provided by collaborators. The PRC2 complex was a gift from the lab of Steve Gamblin, the PRC1 complexes from the lab of Sarah Elderkin, and the MLL1 and MLL4 complexes from Jon Wilson and Ying Zhang. The SET1B complex and full length NSD2 were purchased from Reaction Biology Corporation, while $NSD2_{SET}$ was purified in house as described. As visualised by coomassie, the complexes were all relatively pure and stoichiometric and the individual subunits can be easily identified by size. The PRC2 complex used is the core complex comprised of the four subunits EZH2, EED, SUZ12, and RBBP4/7. The canonical PRC1 complex we used contains the two ring finger proteins RING1B and MEL18 which form the E3 ligase and the chromodomain containing CBX protein, but lacks the subunit PHC which is normally associated with canonical PRC1 *in vivo*. The variant PRC1 complex used in this project contains RING1B, MEL18, and RYBP. All the H3K4 histone methyltransferase complexes SET1B, MLL1, and MLL4 are formed by the catalytic domains $SET1B_{SET}$, $MLL1_{SET}$, and $MLL4_{SET}$ along with the other core subunits critical for activity, WDR5, RBBP5, ASH2L, and DPY30. NSD2 activity is not dependent on binding to accessory proteins and the $NSD2_{SET}$ catalytic domain alone is also active *in vitro*.

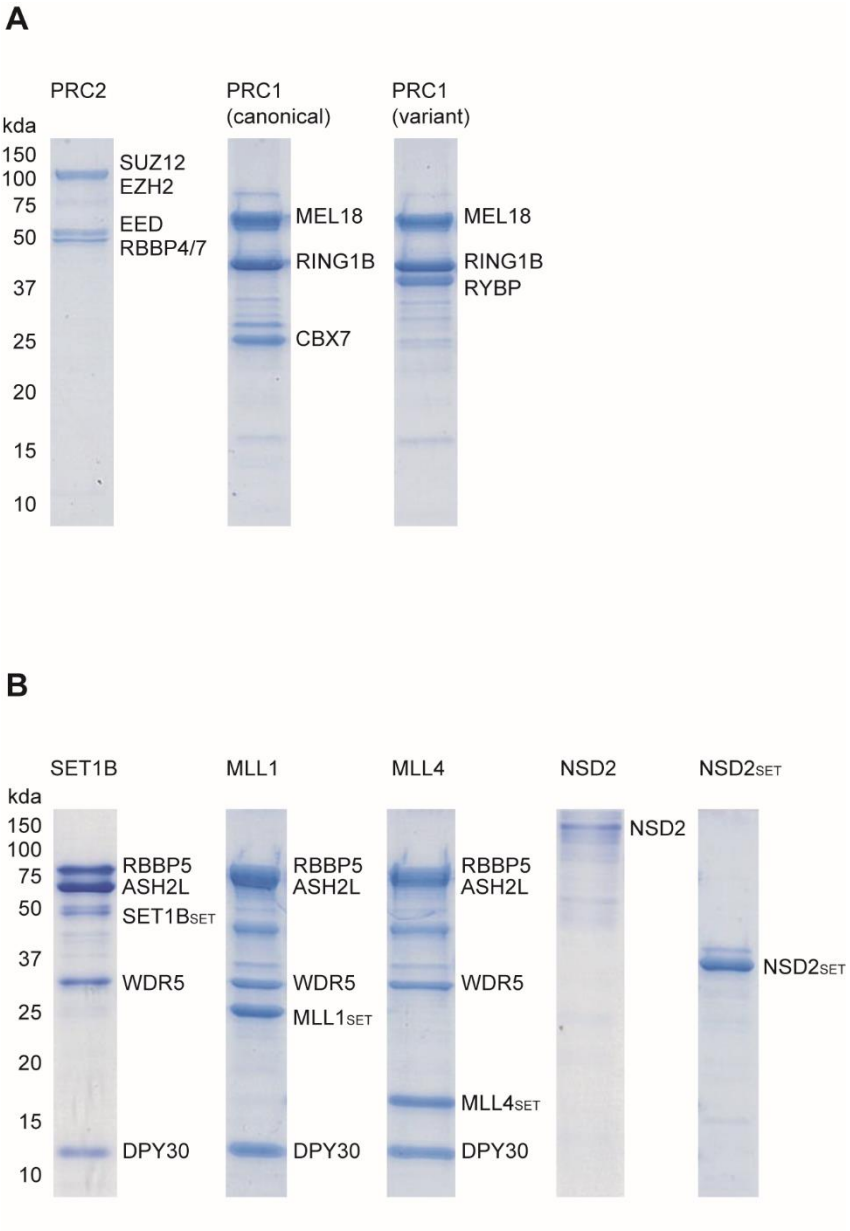


Figure 1.2 Purified enzymes used in *in vitro* activity assays

A PRC2 complex (SUZ12, EZH2, EED, RBBP4/7), canonical PRC1 complex (RING1B, MEL18, CBX7), variant PRC1 complex (RING1B, MEL18, RYBP) generated from multibac expression in insect cells.

B SET1B complex (SET1B$_{SET}$, WDR5, RBBP5, ASH2L, DPY30), MLL1 complex (MLL1$_{SET}$, WDR5, RBBP5, ASH2L, DPY30), MLL4 complex (MLL4$_{SET}$, WDR5, RBBP5, ASH2L, DPY30) generated by expression purification of individual subunits from bacteria. Full length NSD2 purified from baculovirus expression in insect cells, and NSD2$_{SET}$ purified from bacteria.

Generation of nucleosome substrates

Histones were expressed in bacteria, purified under denaturing conditions, and then refolded through gradual dialysis into water. H3 with site specific lysine to cysteine mutations on the histone tail were alkylated to produce methyllysine analogues of H3K4me3, H3K9me3, H3K27me3 and H3K36me3 (Figure 1.3A and B). The histones H2A, H2B, H3 (or a modified variant), and H4 were mixed in equimolar ratios, denatured and refolded together in high salt buffer, then purified through gel filtration to produce pure histone octamers (Figure 1.3C). Octamers were incubated with equimolar amounts of the high affinity 2 × 601 Widom nucleosome positioning DNA sequence and serially diluted into low salt buffer to form dinucleosomes. Complete dinucleosome formation can be seen from the shift from the free DNA band around 400 bp to a size just under 1000 bp in a non-denaturing TB gel (Figure 1.3D). The recombinant reconstituted dinucleosomes retain the correct stoichiometry as seen by SDS-PAGE (Figure 1.3E). Fresh dinucleosomes were generated for each assay.

The effect of H3 lysine methylations on PRC2 and PRC1 activity

The activity assays for the polycomb complexes PRC2 and PRC1 on these five different dinucleosome substrates are presented in Figure 1.4A. PRC2 activity was measured by the incorporation of ^{3}H-methyl on histone H3 by autoradiography. PRC2 has robust activity on unmodified and H3K9me3

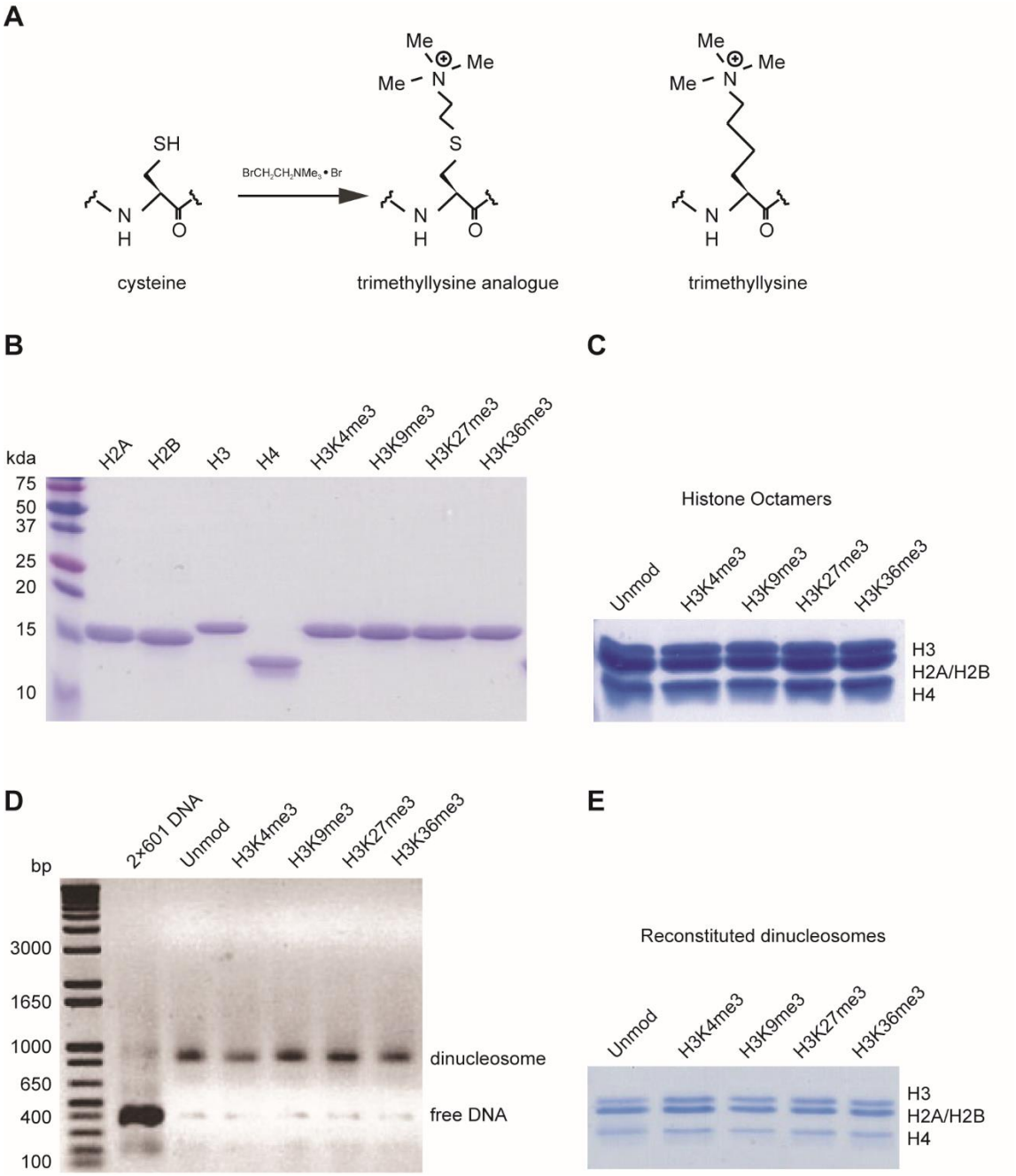


Figure 1.3 Purification and generation of nucleosome substrates

A Alkylation of cysteine to produce the trimethyllysine analogue.

B Purified canonical histones H2A, H2B, H3, H4, and H3 with site-specific trimethyllysine modifications.

C Purified unmodified and modified histone octamers.

D Reconstituted nucleosomes seen by gel shift – free 2×601 DNA and fully reconstituted dinucleosomes were run on a denaturing agarose gel then stained with EtBr.

E Reconstituted nucleosomes run on a SDS-PAGE

containing nucleosomes. Interestingly, H3K9me3 and H3K27me3 have been shown to exist in separate over-lapping domains *in vivo* by immunofluorescence (Cooper et al., 2014). This separation of H3K9me3 and H3K27me3 *in vivo* must be due to other features of the chromatin such as DNA methylation or HP1 binding, because PRC2 activity is not affected by the presence of H3K9me3. In agreement with previous studies, PRC2 activity is inhibited by the presence of H3K4me3 and H3K36me3 on nucleosome substrates, modifications associated with active chromatin. Reassuringly, the PRC2 complex used for these assays shows very specific activity for H3K27 as it has no activity on nucleosomes already modified with H3K27me3.

The ubiquitin ligase activity of PRC1 towards H2AK119 was measured by western blot with an H2AK119u1-specific antibody. The RYBP-containing variant PRC1 complex had similar activity on unmodified and all modified nucleosomes (Figure 1.4B). The CBX-containing canonical PRC1 complex displayed minor increase in activity on nucleosomes modified with the PRC2 mark H3K27me3 (Figure 1.4C), though this needs to be verified at a broader range of concentrations. Various CBX proteins can bind methylated histone tail peptides including H3K27me3. Increased canonical PRC1 activity on H3K27me3 nucleosomes would support the model that canonical PRC1 can be recruited to polycomb target sites in the genome through binding of the chromodomain of CBX to H3K27me3, which is deposited by PRC2.

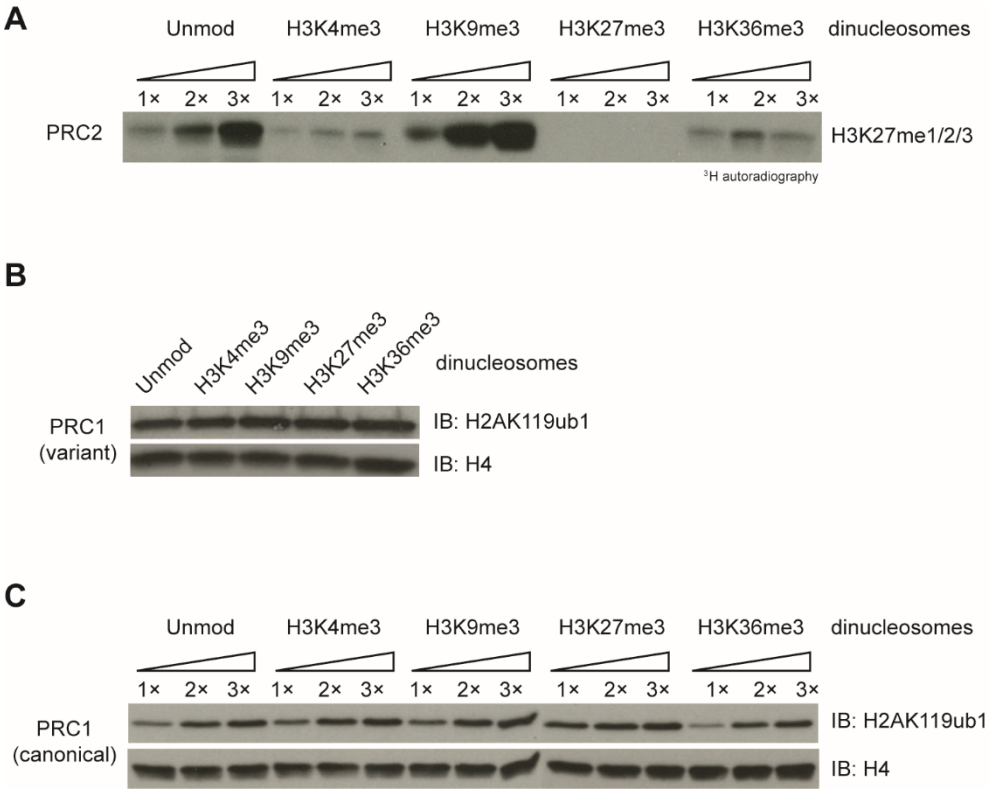


Figure 1.4 Effect of H3 methylation on Polycomb complex activity

A The activity of PRC2 on unmodified and modified nucleosomes. Total activity is measured by the transfer of ^{3}H-methyl group from the methyl-donor S-adenosylmethionine to the target site H3K27. PRC2 is inhibited by the active chromatin marks H3K4me3 and H3K36me3 as first reported by Schmitges et al. (2011).

B The activity of the variant PRC1 complex is unaffected by methylations on H3K4, K9, K27, or K36, as measured by H2AK119u1 western.

C The canonical PRC1 complex is unaffected by methylations on H3K4me3, H3K9me3, H3K36me3, and stimulated by methylation of H3K27me3, as measured by H2AK119u1 western.

The effect of H3 tail lysine methylations on Trithorax activity

The H3K36 methyltransferase NSD2 contains six PHD fingers and two PWWP domains. These chromatin binding domain may regulate the recruitment or activity of the enzyme by reading the epigenetic landscape. Much evidence in the literature points to the mutual antagonism between methylation at H3K27 and H3K36. Mass spectrometry of H3 tail peptides found that H3K36me3 and H3K27me3 rarely coexist on the same histone tail (Jung et al., 2010; Yuan et al., 2011; Zheng et al., 2012). Thus it was expected that the H3K36 methyltransferase NSD2 would be inhibited by H3K27me3 similar to the inhibition of PRC2 by H3K36me3. However, NSD2 was not inhibited by H3K27me3 or any of the other methylations on the H3 tail as measured by autoradiography or westerns against H3K36me1/2 (Figure 1.5 A). Notably, only H3K36me1 activity was observed under these assay conditions so it remains unclear if the activity of NSD2 in the conversion between H3K36me1 to H3K36me2 could be affected by any of the H3 modifications tested.

Next I assayed the activity of the H3K4 methyltransferase complexes SET1B, MLL1 and MLL4 (Figure 1.5 B &C). SET1B, MLL1 and MLL4 were active on all the nucleosome substrates tested. All three forms of H3K4 methylation, mono, di, and tri were observed by western. None of the H3K4

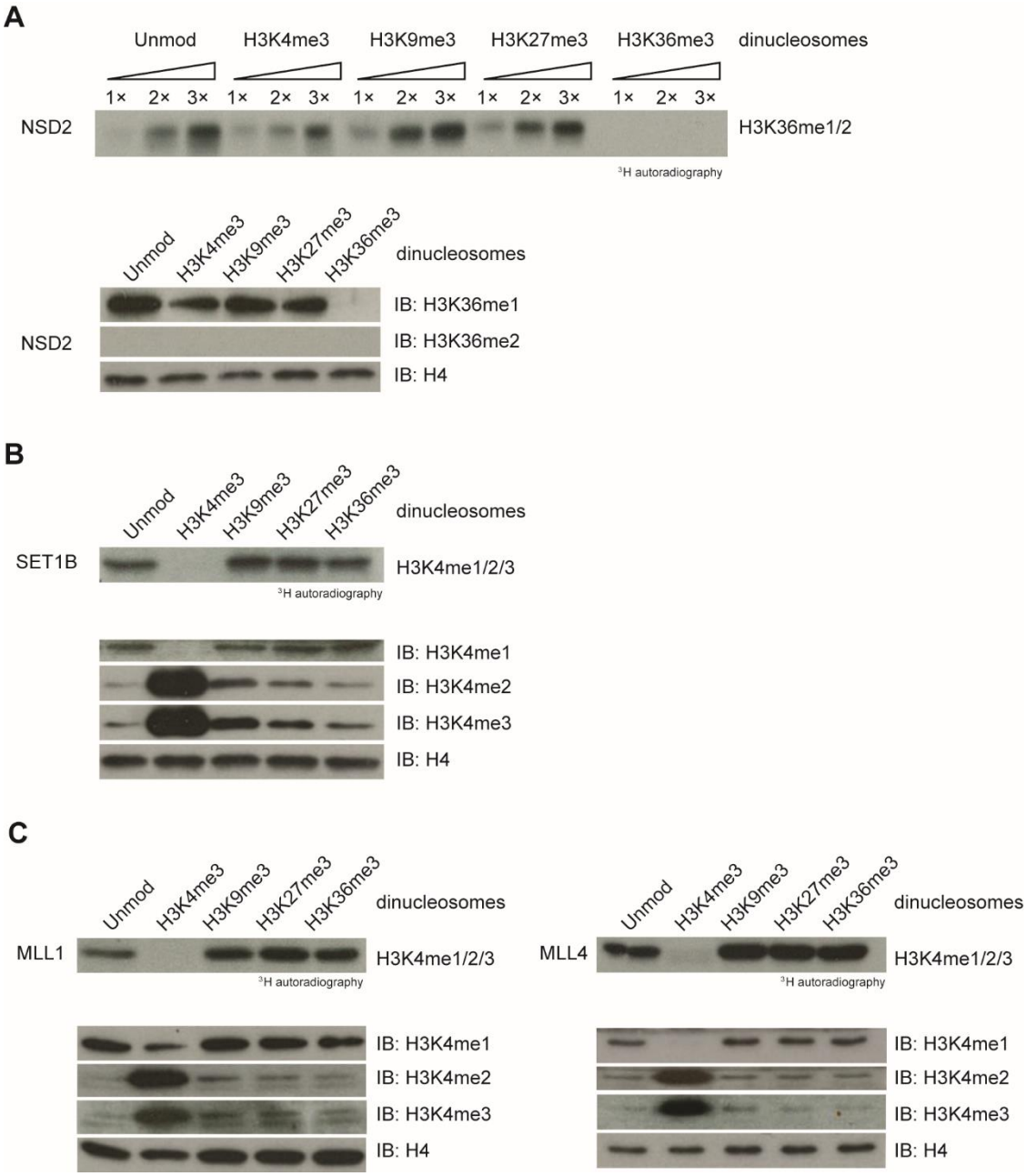


Figure 1.5 Effect of H3 tail methylation on Trithorax activity

A NSD2 activity was not affected by any of the modifications tested as measured by both total activity, and H3K36me1. H3K36me2 activity was not observed.

B SET1B is not inhibited by any of the modifications tested. H3K4me2/3 activity may be slightly stimulated by H3K9me3 and H3K27me3. Note: the H3K4me2 antibody crossreacts with H3K4me3.

C MLL1 and MLL4 are not inhibited by any of the modifications tested as measured by total activity or H3K4me1/2/3 by western.

methyltransferases were inhibited by the repressive modifications H3K9me3 or H3K27me3, and SET1B seemed to display slightly increased H3K4me2/3 activity on H3K9me3 containing nucleosomes which was surprising as *in vivo* H3K4me3 is enriched at actively transcribed genes while H3K9me3 is a marker of constitutive heterochromatin. Overall, it may be concluded that SET1B, MLL1 and MLL4 are not inhibited by the presence of other methylated lysines on the H3 tail. Due to the homology between the SET1 and MLL complexes especially in their catalytic domains and the shared accessory subunits, SET1A, MLL2 and MLL3 by extension are also unlikely to be inhibited by the modifications tested.

H2AK119u1 stimulates PRC2/JARID2 and inhibits NSD2 and MLL1/4

Nucleosomes containing the PRC1 modification H2AK119u1 were generated in collaboration with a postdoc in our lab Lizzi Underwood. The core PRC2 complex displayed similar activity on unmodified and H2AK119u1 (uH2A) nucleosomes, as seen by autoradiography and quantified by liquid scintillation counting. However in the presence of JARID2, a substoichiometric component of the mammalian PRC2 complex, this JARID2-PRC2 was stimulated by H2AK119u1 (Figure 1.6A).

While not inhibited by the PRC2 modification H3K27me3, NSD2 activity was inhibited by the PRC1 mark H2AK119u1. This was true for the

full length protein as well as for the catalytic SET domain which was inhibited to the same degree (Figure 1.6B). The target site of NSD2, the residue H3K36, lies near where the H2A tail exits from the nucleosome core, therefore ubiquitylation of the H2A tail at H2AK119 may sterically hinder the binding of the catalytic site of NSD2 with its target lysine resulting in a loss of activity on H2AK119u1 nucleosomes. The inhibition of NSD2 by H2AK119u1 was recently reported by Yuan et al. 2015; furthermore they found that many H3K36 methyltransferases were similarly inhibited by H2AK119u1.

MLL1 and MLL4 were less active on nucleosomes containing H2AK119u1 (Figure 1.6C), though the degree of inhibition was less than what was observed for NSD2. Trans-histone ubiquitin signalling between ubiquitylation of H2B and H3K4me3 has been well studied. H2B ubiquitylation promotes H3K4me3 in yeast and mammalian organisms. *In vitro* activity assays have shown that both H2BK120u1 and H2BK34ub1 enhances MLL1 complex activity on mononucleosomes (Wu et al., 2013a). Moreover, this study reported that the winged-helix motif in the subunit ASH2L mediates this trans-ubiquitin stimulation on H3K4me3. To test if ASH2L was involved in the inhibition of MLL activity by H2AK119u1, the assay was repeated with MLL complexes lacking the ASH2L and DPY30 subunits. In the absence of ASH2L and DPY30, the inhibition of H3K4 methyltransferase activity was lost (Figure 1.6D). The ability of H2B

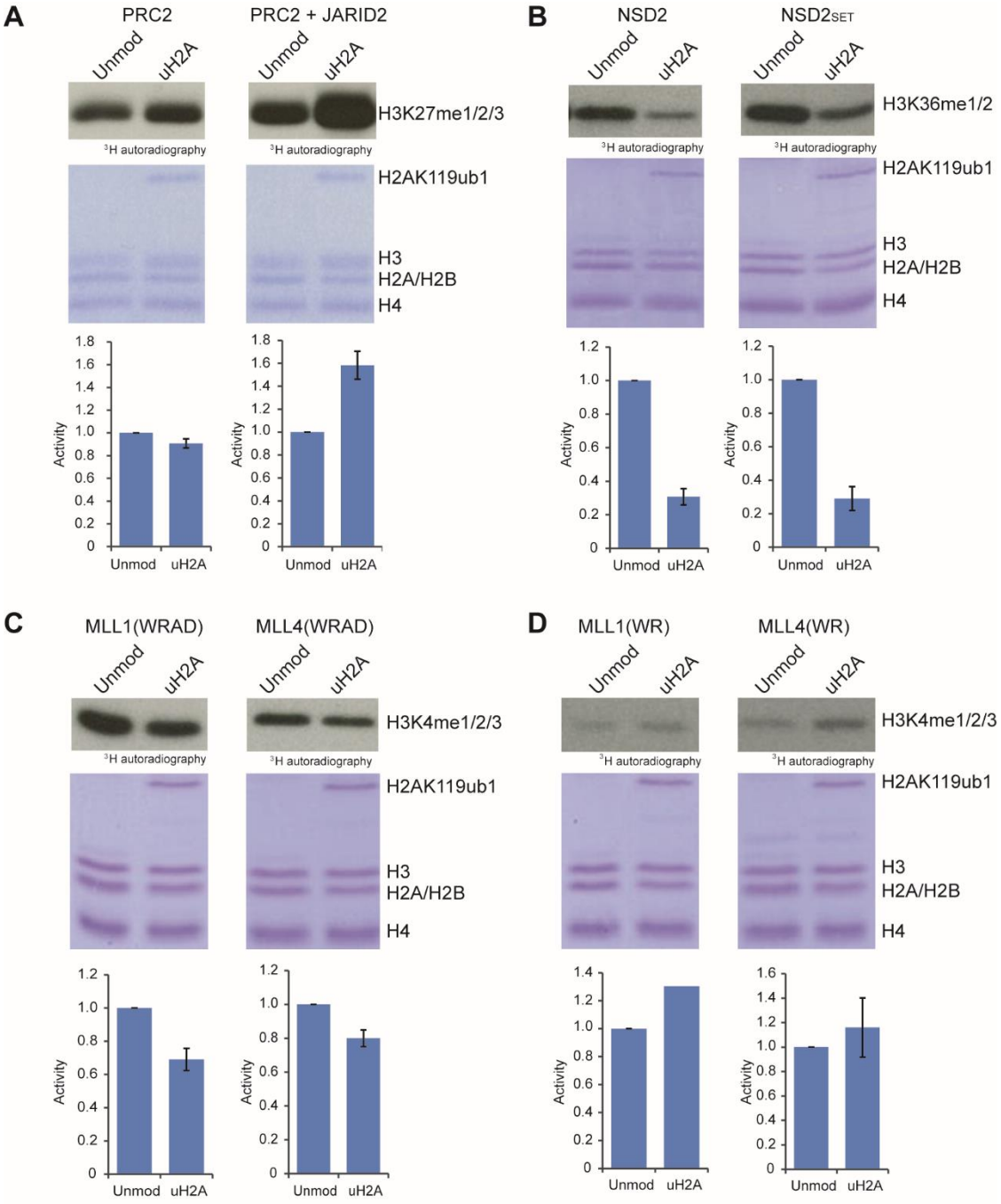


Figure 1.6 the PRC1 mark H2AK119u1 stimulates PRC2-JARID2 and inhibits Trithorax proteins NSD2, MLL1, and MLL4

Total activity is measured by autoradiography and quantified by liquid scintillation counting. Three biological replicates were performed.

A PRC2 is stimulated by uH2A only in the presence of JARID2.

B Both full length and the catalytic domain of NSD2 are inhibited by uH2A

C The MLL1 and MLL4 complexes are inhibited by uH2A. This inhibition is lost in the absence of ASH2L and DPY30 (**D**)

ubiquitylation to stimulate H3K4 methylation and H2A ubiquitylation to inhibit H3K4 methylation implies that this trans-histone crosstalk is highly position specific. Ubiquitin sensing may be mediated by ASH2L in both cases but the effect on activity seems to depend on the site of ubiquitylation.

Effect of chromatin density and linker length

In (Yuan et al., 2012) PRC2 was assayed against chromatin substrates with different linker lengths, and PRC2 was found to have a preference for dense oligonucleosome substrates over dispersed oligonucleosome substrates. The authors suggested that the preference of PRC2 for dense chromatin was tied to its repressive function and involvement in sensing silent compact chromatin. However when performing the same experiment, I found that NSD2 the H3K36 methyltransferase associated with actively transcribed chromatin also showed a preference for dense oligonucleosomes. For this study, histone octamers were incubated with varying amounts of a long DNA sequence corresponding to 25 repeats of the Widom 601 sequence separated by 48bp linkers. Four pools of oligonucleosomes with different densities but the same amount of total histones were generated (Figure 1.7A &B). PRC2 as previously shown, was most active on the densest chromatin substrate (Figure 1.7C), however so was NSD2 (Figure 1.7D) which displayed less activity with increasing DNA to octamer ratio. Clearly the most favourable

substrate for both enzymes was the saturated substrate with 48bp linkers between each nucleosome.

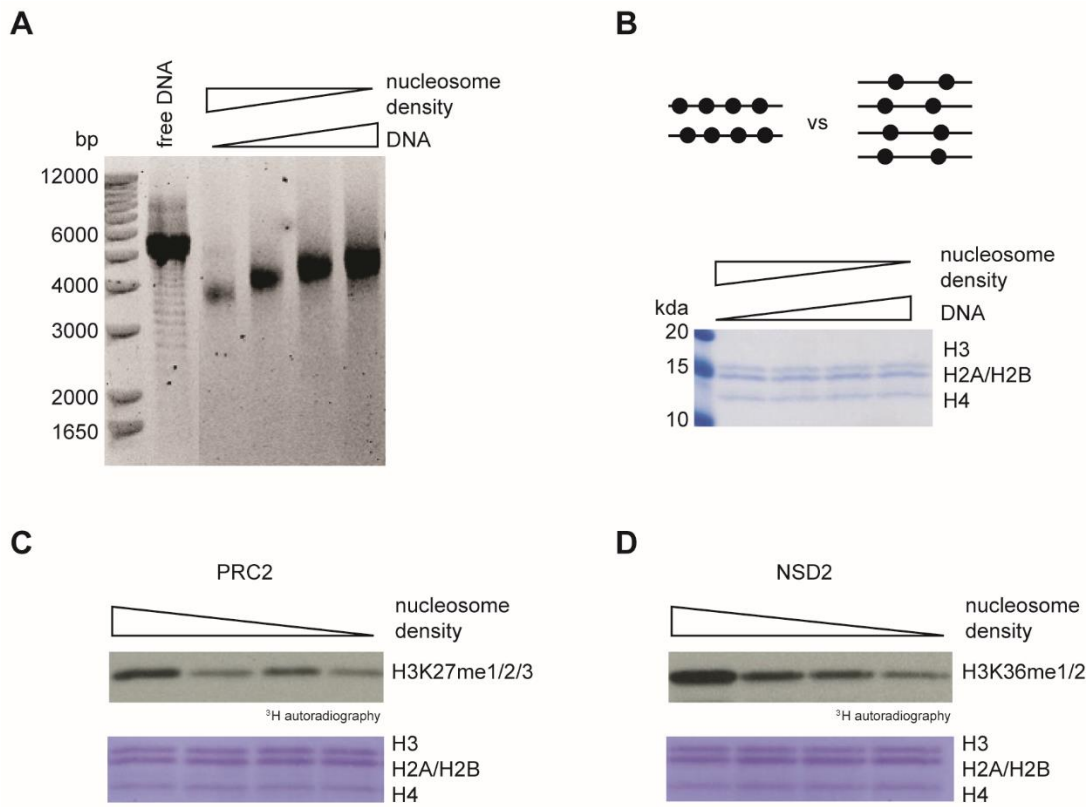


Figure 1.7 PRC2 and NSD2 prefer dense over sparse chromatin

A Four samples of oligonucleosomes with decreasing nucleosome density (from left to right) as seen by gel shift

B These oligonucleosomes differ only in their density and have the same amount of histone proteins as seen by SDS-PAGE

C and **D** both PRC2 and NSD2 are most active on the densest oligonucleosome substrates.

To determine if linker DNA was a requirement for activity, I assayed PRC2 and NSD2 against dinucleosomes with a 48bp linker and compact dinucleosomes with no linker DNA. There was a strong preference for both PRC2 and NSD2 for dinucleosomes with a 48bp linker. It appears that while dense chromatin is favoured over very sparse chromatin, there is a requirement for some spacing between nucleosomes to achieve optimal activity. Considering that both H3K27 and H3K36 residues are located close to where the H3 tail protrudes from the nucleosome core, complete lack of spacing between nucleosomes may block access of PRC2 and NSD2 to the target site on the histone tail. Multiple studies have suggested that PRC2 may dock on linker DNA and interact with tails from two adjacent nucleosomes. EED allosterically stimulates PRC2 activity through binding to H3K27me3 *in trans* (Margueron et al., 2009). Surprisingly, MLL4 was just as active on compact dinucleosomes with no linker DNA (Figure 1.8C). This suggests a mode of nucleosome engagement that differs from that of PRC2 and NSD2. The H3K4me3 tail is further removed from the nucleosome core particle and may be more accessible to binding by the MLL4 complex. Alternatively, the MLL complex may also interact with the exposed surface of the nucleosome and thus remain unaffected by linker length or chromatin density.

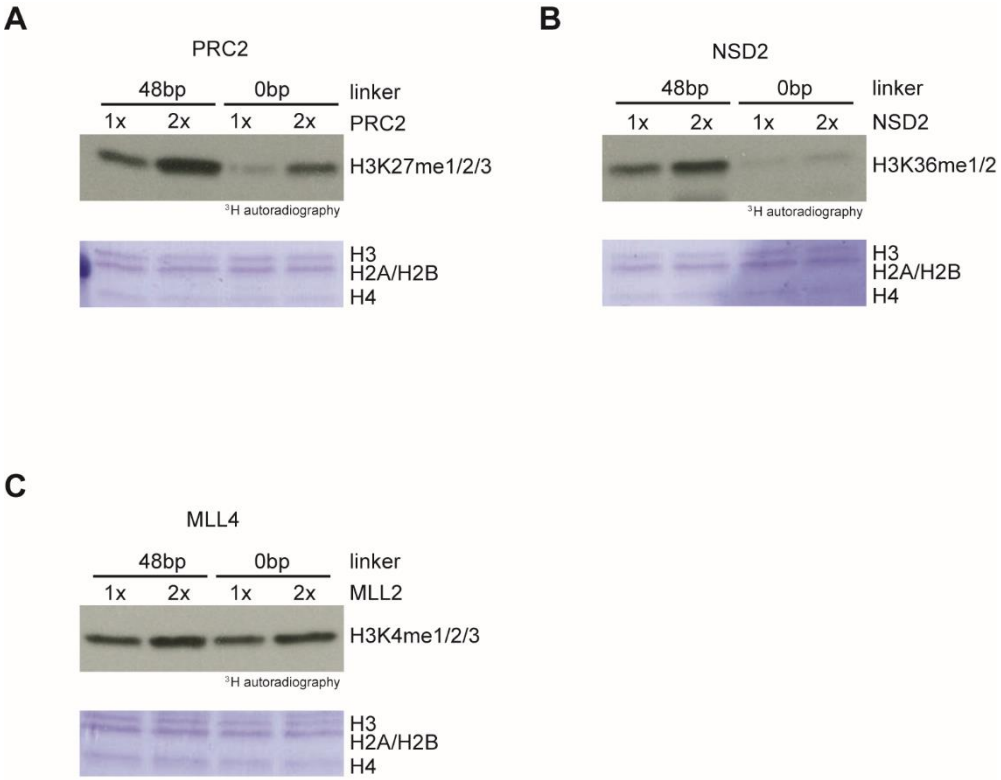


Figure 1.8 PRC2 and NSD2 but not MLL4 requires linker DNA for optimal activity

A PRC2 was assayed on dinucleosomes containing a 48bp linker, or no linker. PRC2 is more active on dinucleosomes with a 48bp linker.

B NSD2 was assayed on dinucleosomes containing a 48bp linker, or no linker. NSD2 is more active on dinucleosomes with a 48bp linker.

C The MLL4 complex is equally active on dinucleosomes with a 48bp linker or no linker.

Discussion

To summarise, histone crosstalk may regulate the binding or activity of Polycomb and Trithorax proteins on chromatin. Crosstalk between PRC1 and PRC2 positively enforces Polycomb activity – whereby the PRC2 mark H3K27me3 stimulates the activity of CBX-containing PRC1 complexes, and the PRC1 mark H2AK119u1 stimulates the activity of JARID2-containing PRC2 complexes. In the Polycomb field, the classical model of Polycomb recruitment was based on a hierarchical model in which PRC2 was recruited first to target genes and deposition of H3K27me3 would serve as a docking site for CBX containing canonical PRC1. Recently, *in vivo* studies by Cooper et al. (2014) and Blackledge et al. (2014) saw that PRC2 can be recruited to specific sites by H2AK119u1, showing that PRC2 can also follow PRC1 activity. A biochemical link was established when Kalb et al. (2014) identified that the PRC2 complex containing JARID2 and another Polycomb protein AEBP2 bound preferentially to H2AK119u1 containing nucleosomes. Through *in vitro* activity assays, Kalb et al. (2014) found that AEBP2 stimulates PRC2 activity on H2AK119u1 containing nucleosomes, while we saw that JARID2 can do the same. Further work by Lizzi Underwood in our lab identified an ubiquitin interaction motif in JARID2 that binds H2AK119u1 nucleosomes *in vitro* is required for the JARID2-dependent recruitment of PRC2 to sites containing H2AK119u1 *in vivo* (data not published).

Additionally, histone crosstalk between Polycomb and Trithorax marks can also negatively regulate the opposing group – seen in the ability of Trithorax marks H3K4me3/H3K36me3 to inhibit PRC2 activity, and the ability of the PRC1 mark H2AK119u1 to inhibit MLL1/4 and NSD2 activity. The inhibition of PRC2 by the active modifications H3K4me3 and H3K36me3 has been often proposed as a mechanism to prevent the spreading of repressive H3K27me3 into active chromatin domains. This inhibition of PRC2 activity occurs only *in cis*; PRC2 is active on nucleosomes where only one of the two H3 tails contains H3K4me3/H3K36me3. H3K27me3 and H3K4me3 do coexist at bivalent domains, which have lower levels of H3K4me3 and have low or no transcriptional activity. H3K27me3 did not inhibit the H3K36 methyltransferase NSD2, which was an unexpected result given the mutual exclusivity of H3K27me3 and H3K36me3 as seen by multiple mass spectrometry studies. However, this may be explained by the coexistence of H3K27me3 with H2AK119u1 at Polycomb target sites and the ability of H2AK119u1 to inhibit NSD2 and other H3K36 methyltransferases.

Notably, PRC1 was not inhibited by any of the Trithorax associated marks tested. This data suggests that PRC1 may be able to bind and modify nucleosomes in active chromatin environments non-permissive to PRC2. Through Chip-sequencing of PRC1 components or the PRC1 mark H2AK119u1, studies have found that PRC1 can co-occupy the promoters of active genes with H3K4me3 and gene expression independent of H3K27me3

(Frangini et al., 2013; Pemberton et al., 2014; van den Boom et al., 2016). It would also be interesting to test the effects of other marks associated with active chromatin such as H3/H4Ac or H2BK123u1 on PRC1 activity.

Together these crosstalk pathways support a model in which Polycomb and Trithorax proteins propagate their own marks while inhibiting those of the opposing group. Genome-wide mapping of chromatin modifications revealed the existence of bivalent genes harbouring both H3K4me3 and H3K27me3 marks associated with active and repressed chromatin. This state of equilibrium between these two groups of epigenetic regulators at key developmental genes may shift upon specific developmental cues to favour gene activation or gene repression through the action of other factors such as transcription factors and the transcriptional machinery.

While not extensively analysed, the preference of different methyltransferases for dense or dispersed chromatin and linker length seems dependent not on whether they are associated with an activation or repression function but on their specific modes of nucleosome binding. Only structural studies can provide the molecular details of the interaction between these enzymes and chromatin.

Native chromatin is decorated by such a vast array of DNA and histone modifications, the molecular mechanisms behind the function of chromatin modifying complexes are often difficult to study *in vivo*. *In vitro* studies allow us to simplify and disentangle these multitude of factors. A good example of this is that H3K9me3 and H3K27me3 are often found to localise in distinct non-overlapping domains by immunofluorescence (Cooper et al., 2014). However, this is not due to any biochemical antagonism between the two marks as PRC2 is highly active on H3K9me3 nucleosomes, suggesting that factors associated with H3K9me3 rich heterochromatin such as DNA methylation or heterochromatin associated proteins may prevent PRC2 binding or activity.

The limitation of these *in vitro* assays lies in the fact that it can be difficult to produce functional complexes with all subunits present. The presence of transiently interacting or substoichiometric subunits may confer additional functions to the core complex. For instance, the addition of the subunit JARID2 to the PRC2 complex confers stimulation by H2AK119u1, which is not observed on the core PRC2 alone. Clearly *in vitro* assays cannot recapitulate the full complexity of chromatin biology, but are a valuable way to study the molecular mechanisms behind the regulation of these epigenetic processes. Directed cellular and biochemical *in vivo* experiments are needed to verify these observations made by myself and others.

RESULTS CHAPTER 2

Identification of readers of H3K27me3 and H3K36me3

Introduction

The initial focus of my work was to understand how patterns of histone modifications are established and regulated through histone crosstalk, with great focus on the interplay between the repressive Polycomb and active Trithorax complexes. I wanted to complement this work by determining the nuclear interactome that associated with repressed and active histone modifications. I therefore chose to perform an unbiased screen to identify readers of H3K27me3 and H3K36me3 using quantitative mass spectrometry with SILAC labelling.

A previous screen identified readers of H3K4me3, H3K9me3, and H3K27me3, but did not include H3K36me3 (Bartke et al., 2010). In this study, their method for generating modified H3 was produced by ligation of a modified peptide corresponding to H3 1 - 31 to the rest of the protein and thus could not be extended to H3K36me3. Thus screening for factors that bind preferentially to H3K36me3 modified nucleosomes has not been reported although a screen for readers of H3K36me3 has been performed using H3 tail peptides (Vermeulen et al., 2010). Making use of methods to generate site-specific methyllysine analogs on H3K27 and H3K36 followed by nucleosome

reconstitution allowed us to probe and compare the interactomes of these marks in a nucleosome-specific context. Methyllysine analogs behave similarly to their native counterparts; they can be recognised by antibodies and known readers of the native modification (Simon et al., 2007). Binding affinities could be different due to the sulfur (instead of the γ carbon) in the side chain of the methyllysine analog. Overall we anticipated that this approach would allow us to identify biologically relevant hits.

These two lysine modifications at residues K27 and K36 on the H3 tail are in close proximity but are associated with very distinct biological functions as described in the introduction. H3K27me3 is catalysed by the PRC2 complex and is localised at the promoters of repressed and lowly expressed bivalent genes, while H3K36me3 is deposited co-transcriptionally by the Pol II-associated SETD2 and is found the gene bodies of active genes in a pattern that increases toward the 3' end of genes. Bidirectional antagonism between these two marks has been observed. In vitro studies have shown that PRC2 activity is inhibited on nucleosomes containing H3K36me3 (Schmitges et al., 2011). Furthermore, comprehensive mass spec studies of H3 tail peptides isolated from different cell types have observed that while many combinations of methylation at H3K27 and H3K36 exist, histone H3 never or rarely harbours both H3K27me3 and H3K36me3 on the same tail (Jung et al., 2010; Yuan et al., 2011; Zheng et al., 2012). I therefore

expected to identify very different groups of interactors for H3K27me3 and H3K36me3 that would reveal more about their distinct biological functions.

Results

Establishment of a nucleosome pulldown assay using symmetrically methylated H3K27me3 and H3K36me3 nucleosomes

Freshly purified histone octamers were reconstituted into dinucleosomes with a 2 × Widom 601 DNA sequence biotinylated at the 5' end. 24 μg of nucleosomes was generated for each pulldown, and 2 μg of each sample from unmodified, H3K27me3 modified, and H3K36me3 modified nucleosome was run on an SDS-PAGE and agarose gel to check the quality of the histones and the nucleosome formation (Figure 2.1A).

HeLa cells were grown for nine passages in heavy R10K8 (L-Arg $^{13}C6$ $^{15}N4$, L-Lys $^{13}C6$ $^{15}N2$) media, or light R0K0 (L-Arg $^{12}C6$ $^{14}N4$, L-Lys $^{12}C6$ $^{14}N2$). This number of passages was sufficient for the full incorporation of the heavy amino acids into the proteome (as checked by mass spectrometry of whole cell lysate, data not shown). Cells were then expanded and harvested for the preparation of nuclear extract. The nuclear extracts for the heavy and light labelled cells were prepared in parallel. The nuclear extract was enriched for nuclear proteins (EZH2, CBX7), and uncontaminated by cytoplasmic components (tubulin) (Figure 2.1B).

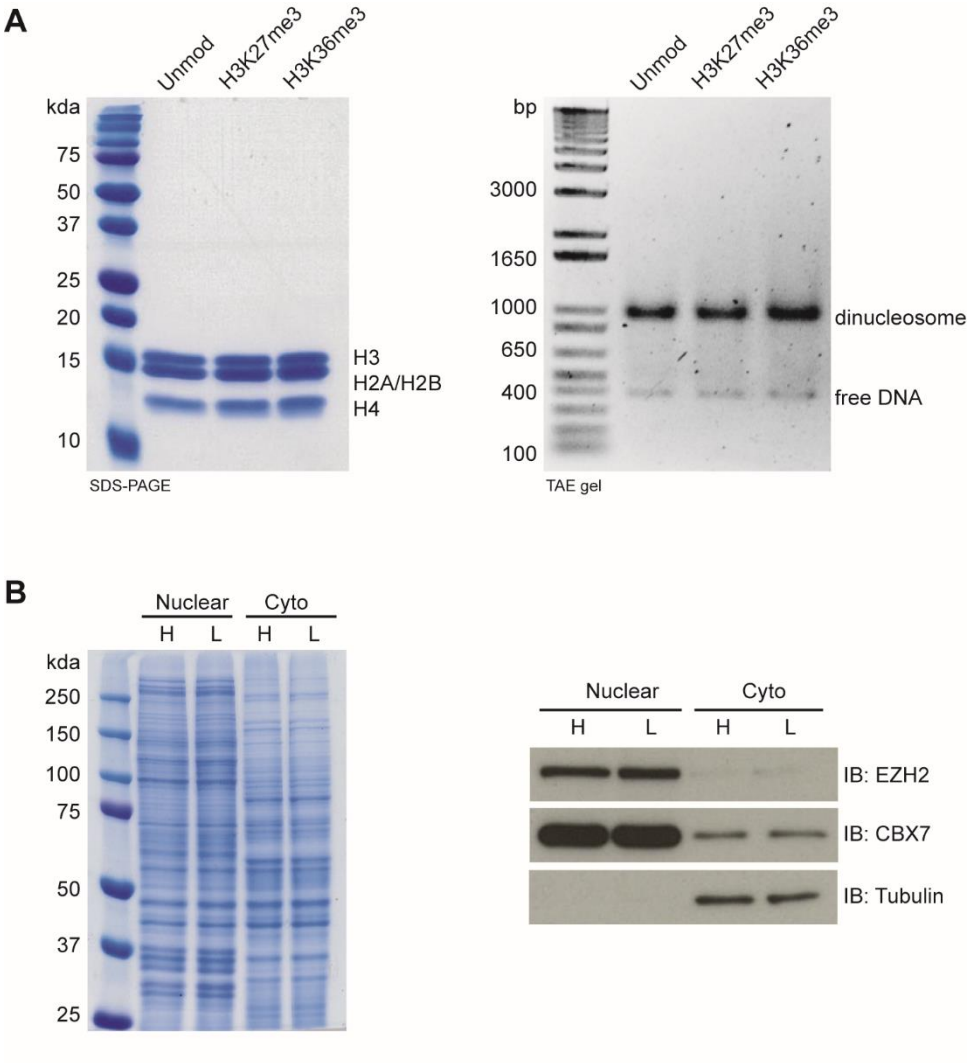


Figure 2.1 Bait nucleosomes and SILAC labelled nuclear extract

A Unmodified, H3K27me3, and H3K36me3 dinucleosomes (2×601 with 48bp linker) run on SDS-PAGE and non-denaturing agarose gel

B Nuclear extract prepared from SILAC heavy (R10K8) and light (R0K0) labelled HeLa cells was run on SDS-PAGE, and immunoblotted for nuclear protein markers (EZH2, CBX7) and cytoplasmic protein markers (Tubulin).

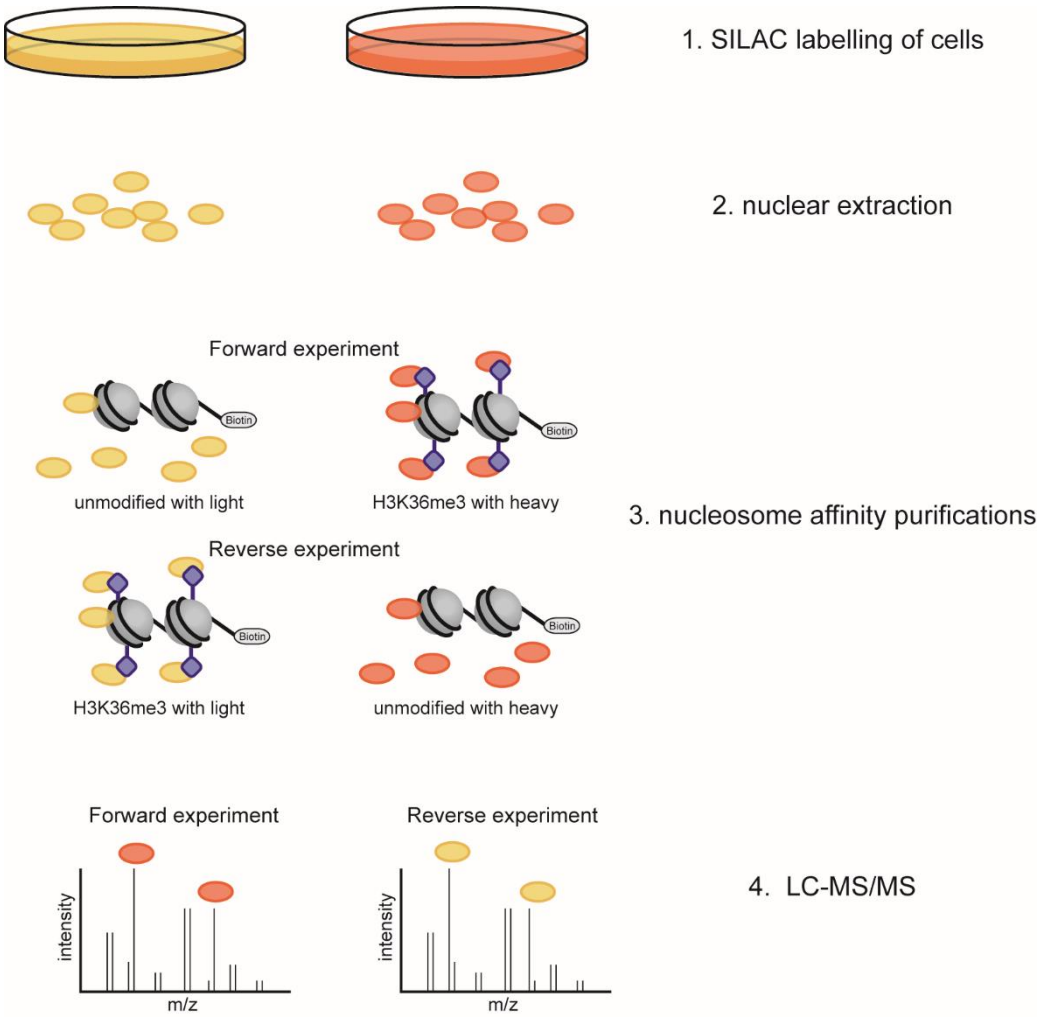


Figure 2.2 Experimental setup for the SILAC nucleosome affinity purifications

HeLa cells were labelled with light or heavy media (1). Nuclear extract was prepared (2). Nucleosomes were conjugated to magnetic streptavidin beads via 5' Biotin on the DNA and incubated with light or heavy nuclear extract (NE). In the Forward experiment, unmodified nucleosomes were mixed with light NE and modified nucleosomes with heavy NE, and the elutions were pooled. And vice versa for the Reverse experiment (3). Samples were trypsin digested and sent for mass spectrometry.

SILAC Nucleosome Affinity Purification

The experimental setup for the nucleosome affinity pulldowns is shown in Figure 2.2. Prior to pooling together the paired pulldown reactions (unmodified nucleosomes with light nuclear extract and modified nucleosomes with heavy nuclear extract, and vice versa), samples were run on SDS-PAGE for quality control. Importantly, the structural integrity of the bait nucleosomes was preserved as seen in the coomassie gel and H4 western (Figure 2.3A). Had the histones dissociated from the biotinylated DNA which was conjugated to beads, they would have been lost during the incubation or wash steps. The pattern of proteins pulled down is different than in the input, and is clearly enriched compared to the beads only control (Figure 2.3A).

Paired samples were pooled for mass spectrometry. The data collection was performed by MS collaborators Rebecca Konietzny and Roman Fischer from the lab of Benedikt Kessler, and I performed the downstream analysis. Peptides were called using Mascot and enriched values were calculated using MaxQuant. For each pulldown, a Heavy/Light ratio was assigned for each protein. A protein that preferentially binds to H3K27me3 over unmodified nucleosomes has an H/L ratio >1 in the Forward experiment, and an H/L ratio <1 in the Reverse experiment. Likewise, a protein that preferentially binds unmodified nucleosomes over H3K27me3 will have an H/L ratio <1 in the Forward experiment, and an H/L ratio >1 in the Reverse experiment.

In Figure 2.3C the log2(H/L Forward) is plotted against the log2(H/L Reverse) for the H3K27me3 and the H3K36me3 pulldowns. Proteins enriched on the modified nucleosome are predominantly found in quadrant 3, while proteins depleted from the modified nucleosome are found in quadrant 1. Contaminants, and chromatin readers which bind both unmodified and modified equally are found primarily in quadrants 2 and 4. Although the Forward and Reverse experiments can be considered biological replicates, it is important to consider both results in conjunction to find true interactors. Due to slight differences in cell growth conditions or sample preparation, a protein could be more enriched in the heavy labelled nuclear extract compared to the light labelled nuclear extract, and will have an H/L ratio > 1 in both the Forward and Reverse experiment. Therefore, only proteins with H/L ratio that differ between the Forward and Reverse experiment can be considered true candidates.

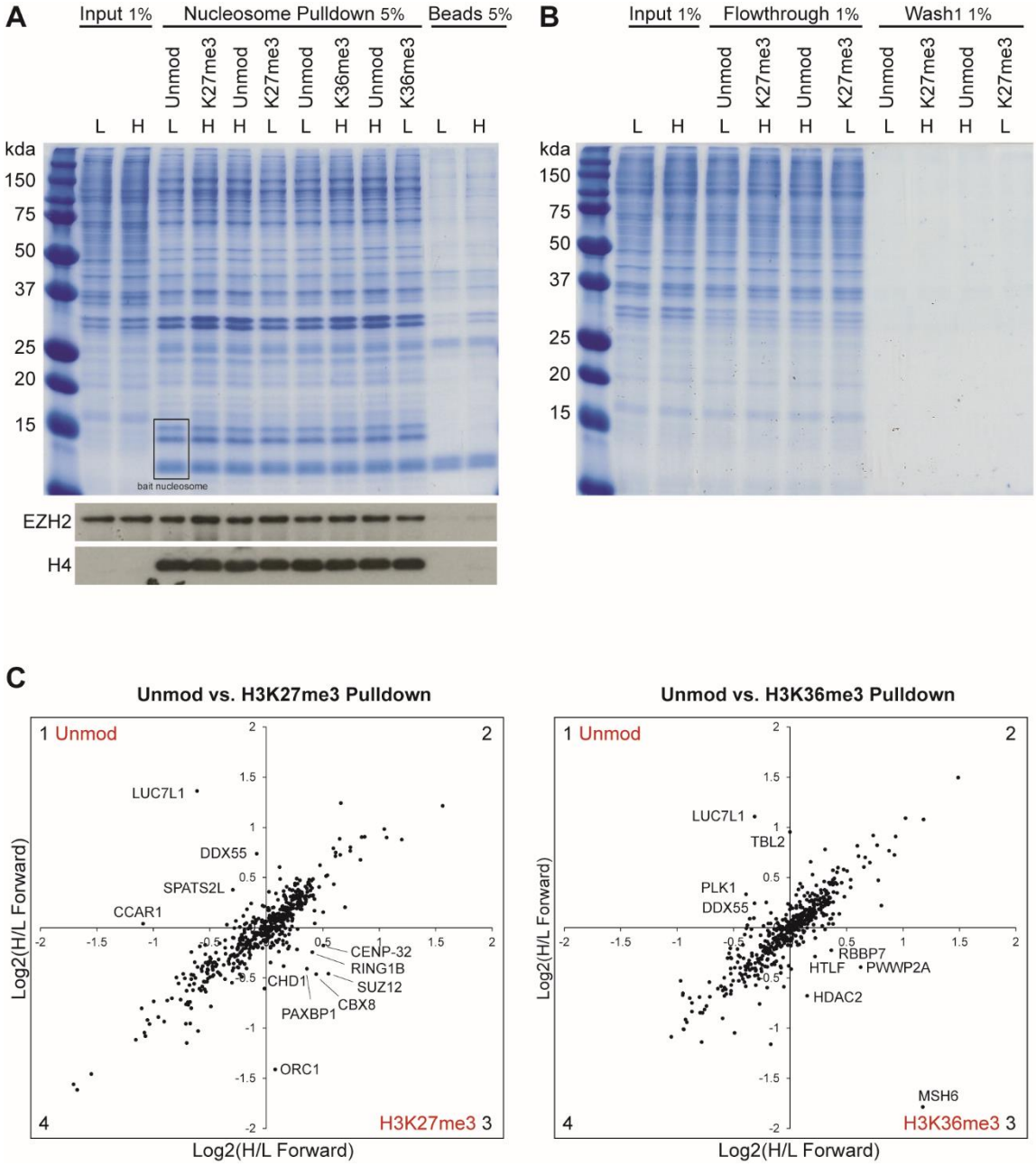


Figure 2.3 Proteins enriched on H3K27me3 and H3K36me3

A Light (L) and heavy (H) nuclear extract input samples, nucleosome pulldowns (Unmodified nucleosomes with L, H3K27me3 nucleosomes with H, unmodified nucleosomes with H, H3K27me3 nucleosomes with L, unmodified nucleosomes with L, H3K36me3 nucleosomes with H, unmodified nucleosomes with H, H3K36me3 nucleosomes with L), and beads only pulldowns with L and H were run on SDS-PAGE. Integrity of bait nucleosomes were checked by H4 western.

B Input, flowthrough, and washes were run on SDS-PAGE for quality control.

C Identification of proteins enriched (quadrant 3) or repelled (quadrant1) by H3K27me3 and H3K36me3. The log of heavy/light ratio for each protein was plotted in the Forward and Reverse experiments.

The majority of proteins identified have similar H/L Forward and H/L Reverse values, suggesting that most chromatin interactors bind to chromatin irrespective of the presence or absence of H3K27me3 and H3K36me3. Figure 2.4 presents a list of outliers found by interquartile range analysis, which have statistically different H/L values between the Forward and Reverse experiments. These outliers are likely candidates that preferentially bind to or are excluded by H3K27me3 or H3K36me3.

The screen identified several H3K27me3 interactors, many of which have been previously characterised. H3K27me3 is catalysed by the polycomb repressive complex PRC2, and PRC2 is known to bind its own mark through an aromatic cage in the WD40 beta propeller domain of its subunit EED. Binding of EED to H3K27me3 allosterically stimulates PRC2 activity, and may lead to the selective recruitment and propagation of PRC2 activity at silenced chromatin (Margueron et al., 2009). Similarly, the PRC1 subunit CBX2/4/6/7/8 binds H3K27me3 through its chromodomain (Kaustov et al., 2011). The recruitment of these CBX-containing canonical PRC1 complexes to Polycomb target sites may be dependent on H3K27me3 and PRC2. Figure 2.4C by western confirms the enrichment of various Polycomb subunits (CBX7, EED, EZH2) on H3K27me3 modified nucleosomes.

A

Enriched on H3K27me3 nucleosomes			
		H/L Forward	H/L Reverse
ORC1	Origin recognition complex	1.0565	0.37581
SUZ12	Polycomb PRC2 subunit	1.4647	0.72763
RBBP7	Polycomb PRC2 subunit	1.1433	0.86459
CBX8	Polycomb PRC1 subunit	1.3596	0.72545
Ring1B	Polycomb PRC1 subunit	1.3264	0.8439
CENP-32	Centromere/kinetochore associated protein	1.4201	0.88442
PAXBP1	PAX3 and PAX7 binding protein	1.2838	0.75254
Depleted from H3K27me3 nucleosomes			
		H/L Forward	H/L Reverse
LUC7L1	Splicing factor associated with U1 snRNP	0.65397	2.5701
DDX55	DEAD-Box RNA helicase	0.94295	1.6668
SPATS2L	Spermatogenesis associated serine rich 2	0.81503	1.2996
CCAR1	Transcriptional coactivator	0.4696	1.0277
SMARCA5	SWI/SNF chromatin remodelling protein	0.83007	1.0473
BAZ1B	Tyrosine kinase involved in remodelling	0.88833	1.1352
KAT7	Histone H4 acetyltransferase	0.86892	1.084

B

Enriched on H3K36me3 nucleosomes			
		H/L Forward	H/L Reverse
MSH6	DNA mismatch repair protein	2.2596	0.29004
MSH2	DNA mismatch repair protein	0.94696	0.75802
PWWP2A	PWWP domain containing protein	1.5425	0.76096
HDAC2	Histone deacetylase complex	1.1107	0.62482
RBBP7	Histone binding protein	1.2886	0.85542
ADARB1	Adenosine to inosine RNA editase	1.7518	1.1646
HLTF	Helicase like transcription factor	1.1656	0.81928
PRP31	U4/U6 snRNP spliceosome factor	0.8882	0.44707
RPA1	DNA replication protein A1	1.2401	0.98759
Depleted from H3K36me3 nucleosomes			
		H/L Forward	H/L Reverse
Luc7L1	Splicing factor associated with U1 snRNP	0.80443	2.1544
PLK1	Polo-like kinase 1	0.76328	1.2597
DDX55	DEAD-Box RNA helicase	0.80377	1.1829

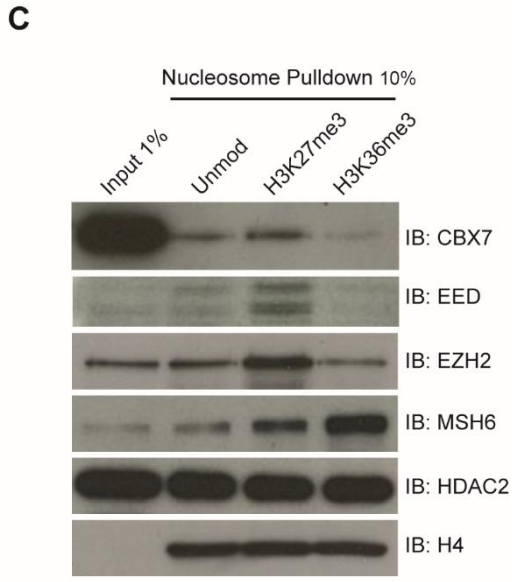


Figure 2.4 List of proteins enriched or depleted from H3K27me3 and H3K36me3 and validation

A and **B** Proteins enriched or depleted from H3K27me3 and H3K36me3 nucleosomes and their enrichment values

C Validation of selected hits by western. Polycomb proteins CBX7, EED, and EZH2 are enriched on H3K27me3 nucleosomes. MSH6 is enriched on H3K36me3 nucleosomes. H4 from bait nucleosomes was used as a loading control.

ORC1 was also identified in the previous screen for H3K27me3 interactors (Bartke et al., 2010). In addition to its function in DNA replication, the ORC complex has been implicated in transcriptional silencing. The ORC-associated protein LRWD1 recognises the heterochromatic H3K9me3, H3K27me3 and H3K20me3, but not marks associated with active chromatin (Bartke et al., 2010; Shen et al., 2010; Vermeulen et al., 2010). It has been demonstrated that the WD40 domain of LRWD1 is necessary and sufficient to bind these heterochromatic marks *in vitro*. The localisation of LRWD1 and ORC2 to the pericentric

heterochromatin is H3K9me3 dependent, and loss of LRWD1 and ORC2 leads to increased transcription from the major satellites (Chan and Zhang, 2012). More recent work suggests that LRWD1 and the ORC complex may interact and stabilise H3K9 histone methyltransferase complexes at heterochromatin (Giri et al., 2015). ORC function at H3K27me3-containing chromatin is poorly understood.

Other identified candidates CENP-32 and PAXBP1 have not been observed previously as H3K27me3 interactors. CENP-32 is a centromere protein that is important for centrosome association to mitotic spindles (Ohta et al., 2015). PAXBP1 is a PAX3/PAX7 binding protein which contains a C-terminal GC-rich DNA binding domain. PAXBP1 has been suggested to link the transcription factor PAX7 to the H3K4 methyltransferase MLL through interacting directly with the MLL subunit WDR5 (Diao et al., 2012).

The H3K36me3 nucleosome affinity pulldowns also captured several candidate interactors. MSH6, verified by western blot (Figure 2.4B), the factor most enriched on H3K36me3 compared to unmodified nucleosomes, is a known reader of H3K36me3. The hMutSα complex formed by MSH6 and MSH2 (also identified in the screen) functions in the DNA Mismatch Repair (MMR) pathway. MSH6 binds to H3K36me3 through an aromatic cage in its PWWP domain, this interaction is essential for the loading of MSH6 on

chromatin in early S phase. Li et al. (2013) observe that the highest levels of H3K36me3 accumulate just prior to S phase or in early S phase, and knockdown of the H3K36me3 methyltransferase SETD2 abrogates MSH6 recruitment to chromatin. From this data, it appears that the MMR machinery is recruited to chromatin in a cell cycle dependent manner during early S phase in preparation for DNA replication where it is required correct small base insertion/deletion errors during replication (Li et al., 2013). Like MSH6 mutants, depletion of SETD2/H3K36me3 in HeLa cells also cause a mutator phenotype. Furthermore, SETD2 has been identified as a tumour suppressor in clear cell renal cell carcinoma (Duns et al., 2010), suggesting that H3K36me3 plays an important role in maintenance of DNA integrity through recruiting MSH6 and the MMR machinery.

The second most enriched protein on H3K36me3 nucleosomes was PWWP2A, an uncharacterised nuclear protein containing a C-terminal PWWP domain. PWWP2A had been identified as an interactor of HDAC1 and HDAC2 in human T cells previously (Joshi et al., 2013). Two members of HDAC core complex HDAC2 and RBBP7 were also enriched on H3K36me3 containing nucleosomes in my screen. A PWWP2A antibody referenced in the literature (Joshi et al., 2013) failed to work for western blot, while HDAC2 bound at such high levels to both unmodified and modified nucleosomes that a preference for H3K36me3 was not observable by western blot (Figure 2.4B). The interaction between PWWP2A and H3K36me3 has nevertheless been

confirmed through ChIP studies, which will be discussed in Chapter 4. We predict that PWWP2A is the component of this putative HDAC complex that mediates the interaction with H3K36me3 due to the presence of its C-terminal PWWP domain.

Other proteins enriched on H3K36me3 modified nucleosomes were HLTF, RPA1, ADARB1, and PRP31. HLTF and RPA1 have some role in DNA repair while ADARB1 and PRP31 are both involved in RNA processing. None of these factors have been previously linked to H3K36me3. HLTF, the Helicase-like Transcription factor is a SWI/SNF protein implicated in chromatin remodelling, transcription, and post replication DNA repair (reviewed in (Dhont et al., 2016)). It is the mammalian ortholog of RAD5 in yeast. Like RAD5, HLTF is a double stranded DNA translocase that can resolve stalled or damaged replication forks (Achar et al., 2015; Blastyak et al., 2010). Following UV induced DNA damage, HLTF promotes ubiquitylation of PCNA and recruitment of the DNA damage associated DNA Polymerase η (Lin et al., 2011).

RPA1 forms a heterotrimer with RPA2 and RPA3 and binds ssDNA in replication and in DNA repair. RPA1 has been found to physically interact with HLTF through affinity purification in human HEK293T cells but the functional significance of this interaction remains unclear (MacKay et al.,

2009). Unfortunately this type of screen cannot identify the direct readers of the histone modification. Neither HLTF nor RPA1 contain any domains that are known to bind chromatin or histone modifications, further work is necessary to confirm this as a real interactor of H3K36me3. H3K36me3 has been linked to DNA repair in multiple studies. It is involved in the mismatch repair pathway as described previously through the recruitment of MSH6 in S phase. It is known to play a role in DNA double strand break repair pathway choice in yeast, by promoting non-homologous end joining (NHEJ) over homologous recombination (HR) in G1 (Pai et al., 2014). A contrasting study in mammalian cells finds H3K36me3 to be critical for the recruitment of HR factors, RPA1 and RAD51, to sites of DNA damage. Loss of RPA1 and RAD51 recruitment and HR efficiency upon knockdown of the H3K36me3 methyltransferase SETD2 could only be rescued by expression of wildtype SETD2 but not a catalytic null mutant (Pfister et al., 2014). Clearly much work remains to verify the roles of these candidate interactors and understand fully the function of H3K36me3 in DNA damage repair.

As a histone mark associated with elongating Pol II, it is possible that H3K36me3 could be involved in many nascent RNA processing pathways. ADARB1 belongs to the adenosine deaminase acting on RNA (ADAR) family of proteins that catalyse adenosine to inosine RNA editing on nascent nuclear RNAs. The function of A-to-I RNA editing remains obscure, it known that the inosine is interpreted by the splicing and translation machineries as a

guanosine. This may represent one mode of alternative splicing regulation as another ADAR protein ADAR2 has been shown to regulate alternative splicing through A-to-I RNA editing (Rueter et al., 1999). Another splicing factor that was captured from the H3K36me3 affinity purification screen was PRP31 – a component of the U4/U6 snRNP complex which forms the spliceosome that mediates pre-mRNA splicing (Du et al., 2015; Liu et al., 2007; Weidenhammer et al., 1996). H3K36me3 has previously been implicated in the regulation of alternative splicing through recruitment of the splice regulators PTB (Luco et al., 2010) and PSIP1 (Pradeepa et al., 2012). Further work is required before drawing any concrete conclusions about the ability of H3K36me3 to recruit these various RNA processing proteins.

Lastly as our focus was to identify readers of H3K27me3 and H3K36me3, we did not concentrate on factors that were depleted from these modified nucleosomes. However the exclusion of certain factors by histone modifications may also represent a way of regulating chromatin processes in the nucleus, and is an interesting area of investigation.

Discussion

This screen succeeded in identifying many candidates which were known and others that were novel H3K27me3 and H3K36me3 interactors, but it was not without technical weaknesses. Here, I will aim to highlight some of the technical successes and limitations of the screen and the resulting data.

In the H3K27me3 pulldown, most interactors were Polycomb or ORC related proteins – canonical PRC1 (RING1B, CBX8), PRC2 (SUZ12, RBBP7), and ORC (ORC1). The comparable screen performed by Bartke et al (2010) identified similar proteins from canonical PRC1 (CBX8), PRC2 (SUZ12, EED, EZH2, MTF2, C17orf96), and ORC (ORC2/3/4/5, LRWD1). There were also some differences. I observed CENP-32 and PAXBP1, which were not seen by Bartke et al. (2010), while they identified a PPIB a protein isomerase which did not feature on my list. Furthermore, I found several proteins (Figure 2.4A) which were depleted from H3K27me3 containing nucleosomes while Bartke et al. (2010) did not observe any proteins that were depleted. Both our studies found that the vast majority of H3K27me3 binding proteins fall into these two groups – Polycomb and ORC. Functionally it appears that a significant role of H3K27me3 is to recruit Polycomb proteins and mediate the propagation of Polycomb marks. The role of ORC in repression is not well

understood, though as previously discussed ORC has affinity for both H3K9me3 and H3K27me3 which are both associated with gene repression.

Similarly a recent study by Kalb et al. (2014) found that the PRC1 modification H2AK119u1 recruits both Polycomb complexes as well. Specifically variant RYBP-containing PRC1, and JARID2/AEBP2-containing PRC2. Canonical PRC1 subunits CBX and PHC were not observed, neither were the PCL1/2/3 proteins from the alternative form of PRC2 observed. H2AK119u1 containing nucleosomes were also enriched for several factors involved in ubiquitin metabolism. Multiple ubiquitin E3 (BRCA1, BARD1) enzymes and ubiquitin proteases (USP7, USP48) were observed, it is unknown currently if these proteins are recruited to Polycomb target genes *in vivo* or have any role in gene repression. Looking at these studies together, H3K27me3 and H2AK119u1 in combination are able to recruit all known forms of Polycomb complexes – canonical CBX-containing PRC1, variant RYBP-containing PRC1, PCL1/2/3-containing PRC2, and JARID2/AEBP2-containing PRC2. The ability of Polycomb marks to positively reinforce Polycomb propagation must be an important aspect of Polycomb biology and gene repression.

H3K36me3 interactors on the other hand, seem more abundant and fall into a wider range of functions. In the H3K36me3 pulldown, the top hits

were MSH6 and PWWP2A. Both these proteins contain a PWWP domain which is known to bind H3K36me3 in many proteins. A study by Vermeulen et al. (2010) using H3K36me3 peptides as bait also identified several PWWP domain containing proteins – N-PAC, MSH6, and NSD1/2. Neither of these screens purified a comprehensive list of all PWWP-containing proteins. Notably, the *de novo* DNA methyltransferases DNMT3A and DNMT3B were absent from both screens. Both of these have been identified as readers of H3K36me3 through their PWWP domains (Baubec et al., 2015; Dhayalan et al., 2010). Other proteins involved in DNA damage repair and RNA processing were also found in my screen but absent in Vermeulen et al. (2010). It is possible that use of different baits (nucleosomes vs. peptide) may explain these differences.

The method we employed used dinucleosomes as bait rather than histone tail peptides, providing a more specific context for recognition of the modified histone residue. The H3K36 residue lies near where the DNA wraps around the histone octamer core, the recognition and binding of H3K36me3 by the PWWP domain is enhanced by the presence of two positively charged DNA binding surfaces adjacent to the H3K36me3 binding pocket (Qiu et al., 2002). The downside to using nucleosomal substrates substrates as bait is that the presence of nucleosomal DNA and linker DNA increases the background by attracting unspecific DNA binding proteins.

This screen allowed the unbiased identification of not only direct binders but also other proteins in complex with the reader. On one hand, this provides a broader range of information about all the components which are enriched on a specific modification but it can also make identifying the direct binder difficult especially if the candidates are poorly characterised in the literature.

The majority of proteins identified also show equal preference for the unmodified and modified bait nucleosomes, thus the majority of the peptides identified in the mass spec are from unspecific chromatin binding proteins. Due to the complexity of the sample, the trypsinised peptides were separated by charge over a column prior to entering the mass spectrometer. Not all peptides entering the mass spectrometer are detected, at any one point only the 15 most abundant peptides are sequenced. In the cell, the dynamic range of protein expression is huge. As a result of this, the detection of peptides in the mass spectrometer is skewed towards more abundant proteins present in the sample, thus many known readers such as EED, the PRC2 subunit that binds H3K27me3, were not detected altogether or only detected in either the Forward or Reverse pulldown. Proteins with more sequenced peptides have more reliable H/L values, thus the enrichment values of the candidate interactors could be improved if more peptides were sequenced.

From the SDS-PAGE gels in Figure 2.3B, the majority of the unbound proteins are found in the flowthrough, however very little is observed in the Wash 1. Increasing the wash stringency could aid in removing more weakly bound proteins and reduce sample complexity. Additionally, more sophisticated methods of peptide separation could also be tested to achieve greater resolution and reduce the sample complexity of peptides entering the mass spectrometer at once.

Overall, I was satisfied with the number of known and novel candidates identified in the screen. The most promising candidate was PWWP2A, which we decided to investigate in detail.

RESULTS CHAPTER 3

Biochemical characterisation of the PWWP2A-NuRD complex

Introduction

For many proteins identified in the screen, it is unclear from their domain structure if the H3K36me3 interaction is direct or mediated through another protein. The interaction of PWWP2A with H3K36me3 on the other hand is likely to be direct as it contains a PWWP domain, which is found in many H3K36me3 readers.

PWWP2A has never been biochemically characterised but has appeared in proteomic studies as an HDAC interactor. Joshi et al. (2013) mapped the interaction network for HDACs 1-11 in human T cells, and found PWWP2A associated exclusively with HDAC1 and HDAC2, but not HDAC3-11. Thus PWWP2A may represent a new mode of HDAC1/2 targeting to chromatin containing H3K36me3 in mammalian cells. This may be an analogous pathway to the recruitment of the HDAC Rpd3S in budding yeast which is targeted to active gene bodies through H3K36me3 binding. In mammalian organisms, the function of HDACs have primarily been studied in the context of their functions at gene promoters or enhancers, but HDAC activity over gene bodies has not been well studied.

HDAC1/2 form the deacetylase core of three Class I HDAC complexes, SIN3, CoREST, and NuRD. It is unknown which of these complexes PWWP2A associates with or if the interaction between PWWP2A and HDAC1/2 is conserved in all cell types. We therefore decided to characterise PWWP2A's interaction with HDAC1/2 in HeLa cells and E14 mESCs.

Results

Introduction to PWWP2A

Human PWWP2A is an 82 kDa 755 amino acid long nuclear protein. There are several splice isoforms and alternative polyadenylation sites generating different protein isoforms of PWWP2A. The long protein isoform is 755 amino acids and is the only isoform that includes the C-terminal PWWP domain, while the short isoforms lack the last ~ 230 amino acids of the protein (Figure 3.1A). In vertebrates there exists a paralog called PWWP2B, the human protein is 590 amino acid long and is most similar to amino acids 160 – 755 of human PWWP2A. Through BLAST analysis, homologs of PWWP2A exist in all chordates (dorsal nerve chord and notochord containing organisms) and echinoderms such as sea urchins (Figure 3.1B). The sea squirt *Ciona intestinalis* and the lancet *Branchiostoma floridae* are invertebrate chordates which have only one ortholog of

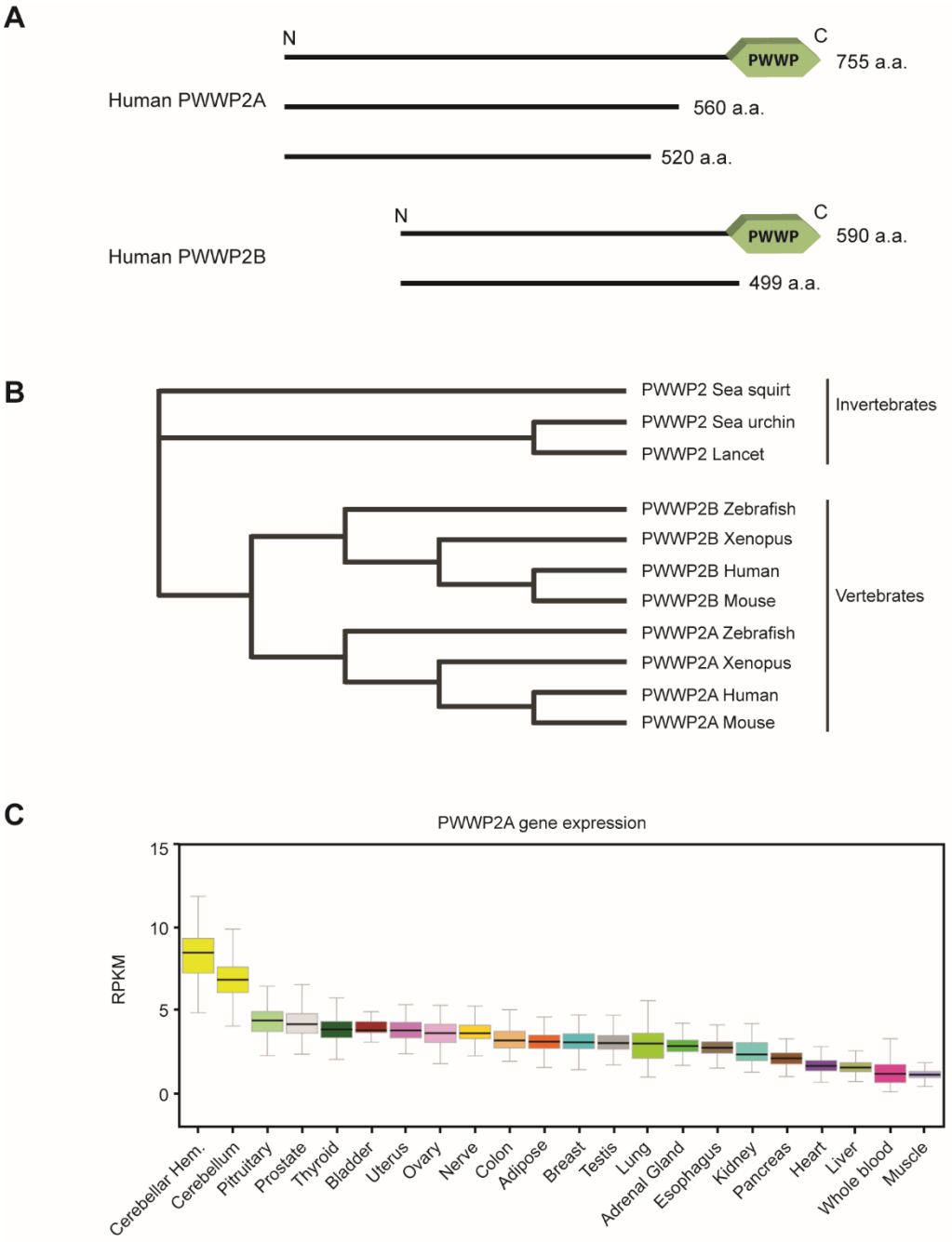


Figure 3.1 PWWP2A is an evolutionary conserved protein expressed in multiple tissues

A Scaled representation of human PWWP2A and its paralog PWWP2B. The protein is predicted to be highly disordered, only the C-terminal PWWP domain is annotated.

B Phylogenetic tree of PWWP2A/B in various chordates. Invertebrates have only one ortholog of PWWP2A, while vertebrates have PWWP2A and PWWP2B. The PWWP domain is the most conserved part of the protein. See appendix for alignment of PWWP2A and PWWP2B.

C PWWP2A is broadly expressed in a range of tissues

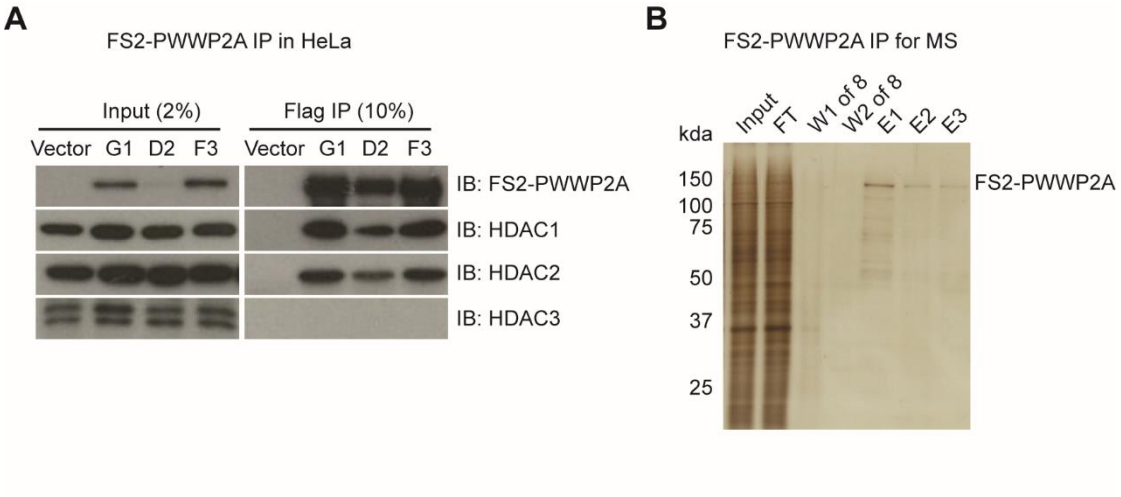
PWWP2A/B, suggesting that the gene duplication event that generated the paralogs PWWP2A and PWWP2B occurred in an early vertebrate evolution.

Consulting human gene expression databases, several large studies such as Genotype-Tissue Expression project and the RIKEN Fantom5 Project found PWWP2A to be expressed from all human tissues tested, with a higher level of expression observed in the brain (Figure 3.1C). In mESCs, both isoforms of PWWP2A are expressed at medium levels.

PWWP2A associates with NuRD subunits in HeLa cells

As previously mentioned, PWWP2A was found in a proteomic screen to associate with HDAC1 and HDAC2 in human T cells (Joshi et al., 2013). However it was unknown if this interaction was conserved in other cell types, or if HDAC1/2 was a main interactor or one of many nuclear complexes PWWP2A associates with.

I established a cell line stably expressing N-terminal tagged Flag-2×Strep human PWWP2A (henceforth shortened to FS2-PWWP2A) in HeLa cells, the cell line used in the SILAC screen. First I confirmed that FS2-PWWP2A co-immunoprecipitates endogenous HDAC1 and HDAC2 but not HDAC3 in all three clones tested (Figure 3.2A).



C

	Protein	Score	Significant matches	Significant sequences
1	PWWP2A	14324	650	63
2	MTA1.3	5346	208	63
3	MTA1	5004	194	59
4	MTA2	2307	85	41
5	MTA3	1765	72	30
6	RBBP7	2074	85	21
7	RBBP4	1489	67	19
8	HDAC2	1650	86	25
9	HDAC1	1562	84	26
10	BEND3	986	26	26
11	PWWP2B	455	20	18

Figure 3.2 PWWP2A interacts with NuRD components in HeLa cells

A PWWP2A immunoprecipitates endogenous HDAC1 and HDAC2 but not HDAC3 in HeLa cells. Three clones (G1, D2, F3) with different expression levels of FS2-PWWP2A showed the same result.

B Silver stain of the PWWP2A immunoprecipitation Input, flowthrough (FT), washes (W1 and W2), and elutions (E1, E2, and E3) for mass spectrometry.

C PWWP2A immunoprecipitates NuRD subunits MTA1-3, HDAC1/2, and RBBP4/7, a NuRD associated chromatin binding protein BEND3, and its paralog PWWP2B in HeLa cells.

Next to determine all interactors of PWWP2A by LC-MS/MS, 15 mg of nuclear extract was prepared from the FS2-PWWP2A expressing HeLa cell line for streptactin purification. Highly stringent conditions were used for the washes (eight washes in high salt and detergent 500mM NaCl 0.2% NP40) so that only the most stably associated proteins were eluted (see silver stain gel Figure 3.2B). The material from the streptactin purification was sent for LC-MS/MS.

The main PWWP2A interactors identified from the affinity purification were members of the NuRD complex (Figure 3.2C). The core deacetylase complex HDAC1, HDAC2, RBBP4, and RBBP7 were observed, while HDAC3-11 were absent. The highest hits were the metastasis-associated family of proteins. All three homologs present in humans MTA1 (two isoforms), MTA2, and MTA3 were co-isolated with PWWP2A. MTA1-3 are specific to the NuRD complex and provide the scaffolding function in the complex by bridging many of the other subunits. None of the other proteins associated with the other two Class I HDAC complexes SIN3 (SIN3, SAP18, SAP30) or the CoREST (CoREST, LSD1) were detected, indicating that PWWP2A specifically interacts with the NuRD complex. The MTA proteins appeared as the most abundant proteins co-isolated with PWWP2A, making it likely that MTA1-3 bind PWWP2A directly and mediates its association with the rest of the complex. Notably, other known subunits of the NuRD complex such as CHD3/4, GATAD2A/B, and MBD3 were not observed.

Only two other proteins of significance were found, the paralog PWWP2B and another chromatin binding protein BEND3. PWWP2B shares a 33% sequence identity with PWWP2A, and may also interact with MTA1-3 or another member of the NuRD complex. MTA1 dimerizes through its N-terminal ELM2 domain (Millard et al., 2013), therefore each subunit that binds MTA1-3 (HDAC1/2, RBBP4/7, and potentially PWWP2A/B) is present in at least two copies in the NuRD complex. BEND3 has been shown to pulldown all NuRD components in mouse ESCs (Saksouk et al., 2014).

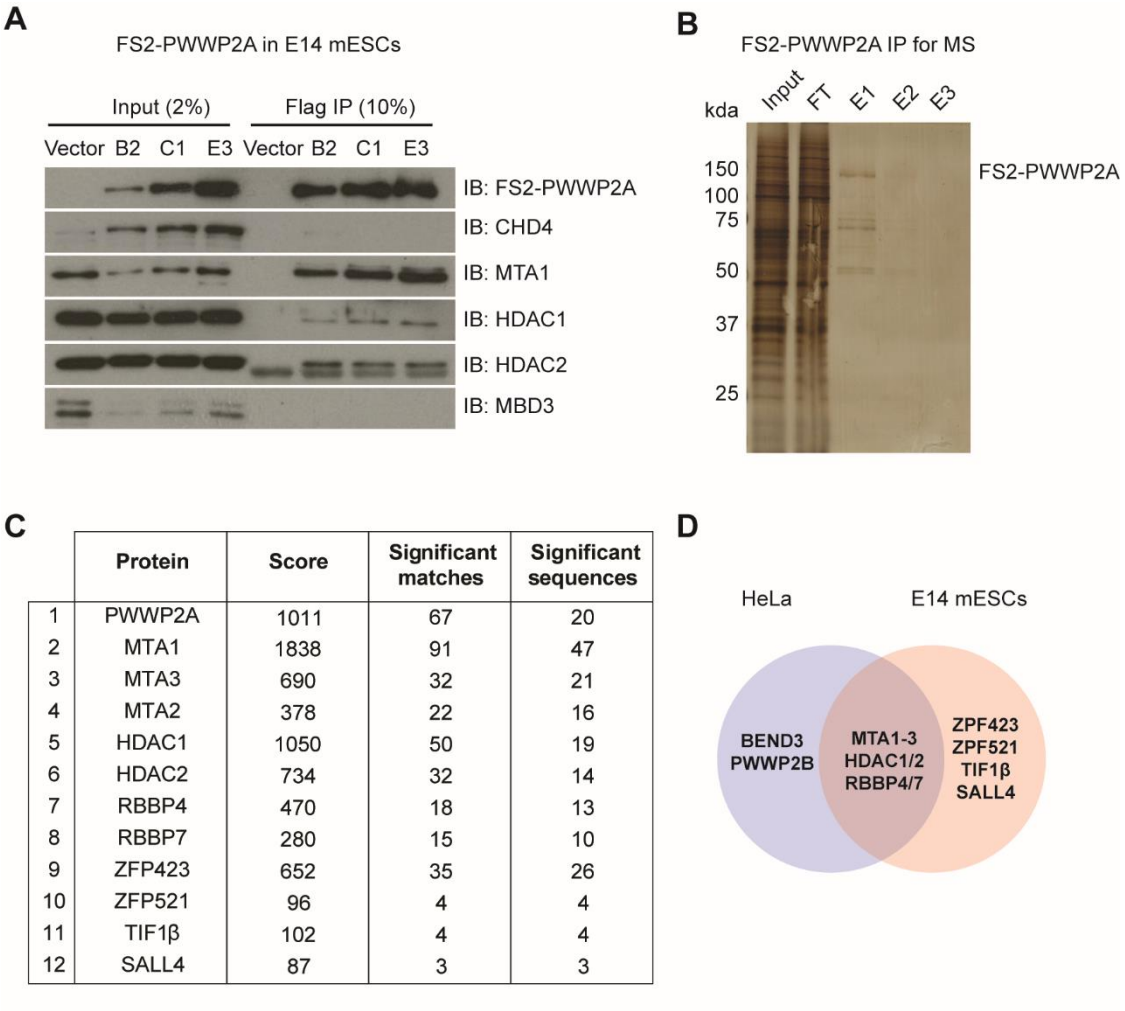
From this mass spectrometry experiment, there was no indication that PWWP2A associates stably with any other nuclear proteins or chromatin interacting complexes in HeLa cells. Therefore it seems reasonable to conclude that PWWP2A functions either as a component of the NuRD complex or alone *in vivo*.

Having established the interaction of PWWP2A with the NuRD complex in HeLa cells, I concluded that the observed HDAC2/RBBP7 enrichment on H3K36me3 nucleosomes seen in the SILAC screen presented in Chapter 2 is likely to be mediated by PWWP2A.

PWWP2A associates with NuRD subunits in mESCs

The same construct FS2-PWWP2A was stably expressed in the mESCs E14 line, and three clones with different expression levels were chosen for further analysis. In all three clones, endogenous mouse HDAC1, HDAC2, and MTA1 co-immunoprecipitated with FS2-PWWP2A while CHD4 and MBD3 did not as seen by western (Figure 3.3A).

This was corroborated by mass spectrometry data. Similar to HeLa cells, all three MTA proteins MTA1-3, HDAC1, HDAC2, RBBP4, and RBBP7 were observed under stringent IP conditions (Figure 3.3C). There were also some differences in the PWWP2A interactome in mESCs compared to HeLa (Figure 3.3D). Several peptides from PWWP2B and BEND3 were detected in HeLa but none were picked up in the FS2-PWWP2A IP from mESC nuclear extract. This is most likely due to experimental rather than biological differences as the BEND3-NuRD interaction has been observed in mESCs before (Saksouk et al., 2014). The HeLa clone had a higher level of FS2-PWWP2A expression than the mESC clone, more material was obtained from the HeLa IP compared to the mESC IP, and the sample preparation and mass spectrometry runs were performed on different days. More replicates of the mass spec or verification by immunoblotting against PWWP2B or BEND3 in mESCs could be used to resolve this issue.



	Protein	Score	Significant matches	Significant sequences
1	PWWP2A	1011	67	20
2	MTA1	1838	91	47
3	MTA3	690	32	21
4	MTA2	378	22	16
5	HDAC1	1050	50	19
6	HDAC2	734	32	14
7	RBBP4	470	18	13
8	RBBP7	280	15	10
9	ZFP423	652	35	26
10	ZFP521	96	4	4
11	TIF1β	102	4	4
12	SALL4	87	3	3

Figure 3.3 PWWP2A interacts with NuRD components in E14 mESCs

A PWWP2A immunoprecipitates a subset of NuRD subunits MTA1, HDAC1 and HDAC2 but not CHD4 or MBD3. Three clones with different expression levels of FS2-PWWP2A showed the same result.

B Silver stain of the PWWP2A immunoprecipitation for mass spectrometry

C PWWP2A immunoprecipitates NuRD subunits MTA1-3, HDAC1/2, RBBP4/7, and four transcription factors ZFP423, ZFP521, TIF1ß, and SALL4 in E14 mouse embryonic stem cells.

D Core NuRD subunits MTA1-3, HDAC1/2, and RBBP4/7 were identified in both MS experiments.

Interestingly, four transcription factors were present the mESC dataset that were not present in the affinity purification from HeLa. These were ZFP423, ZFP521, SALL4, and TIF1ß. All four have previously been linked to NuRD though their possible role in NuRD recruitment or function requires further investigation.

The N-terminal half of PWWP2A mediates NuRD interaction

To further investigate the physical interaction of PWWP2A with the NuRD complex, three truncated constructs were generated – the PWWP2A N-terminal half FS2-PWWP2A$_{1-373}$, PWWP2A lacking the PWWP domain FS2-PWWP2A$_{\Delta PWWP}$, and the C-terminal half FS2-PWWP2A$_{382-755}$ (Figure 3.4A). Unfortunately the C-terminal construct FS2-PWWP2A$_{382-755}$ failed to localise to the nucleus despite containing a predicted NLS. As seen in Figure 3.4B, both the N-terminal FS2-PWWP2A$_{1-373}$, and FS2-PWWP2A$_{\Delta PWWP}$ can co-purify MTA1 and HDAC1. Thus it appears that the PWWP domain (a.a 651 – 755) is entirely dispensable and the N-terminal half of PWWP2A (a.a. 1 – 373) is sufficient for the NuRD interaction. While sufficient for the NuRD interaction, it does not mean that the N-terminal domain is necessary, as the a.a 374 – 651 could also mediate an interaction with NuRD. Ideally a C-terminal PWWP2A containing a functional NLS should be tested for NuRD binding capabilities. Nevertheless, the N-terminal half of PWWP2A looks likely to confer the NuRD interaction.

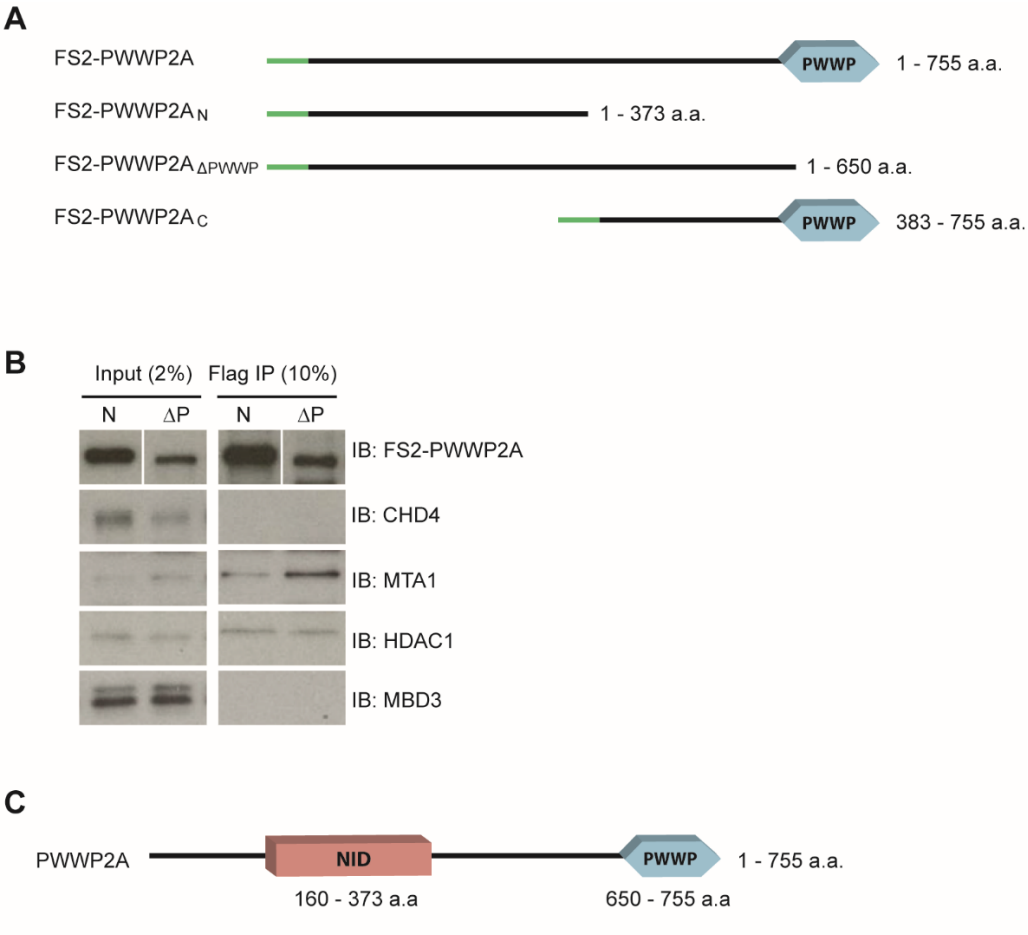


Figure 3.4 A putative NuRD interaction domain exists within 160 – 373 a.a. of PWWP2A

A Truncated FS2-PWWP2A constructs expressed in E14 mESCs.

B The N-terminal half of PWWP2A is sufficient to immunoprecipitate MTA1 and HDAC1. The PWWP domain is not required for NuRD interaction.

C The N-terminal half of PWWP2A (1 – 373 a.a.) must contain a Nurd interaction domain. The most conserved part of PWWP2A starts from 160 a.a. onwards, therefore the NID is likely within 160 – 373 a.a.

From clustalomega sequence alignments, the very N-terminus is the most diverged part of the protein, while amino acid 160 – 755 show a high degree of conservation between frog, zebrafish, mouse, and human, leading us to believe that the NuRD interaction domain of PWWP2A could reside within amino acids 160 – 373. Of note, the two annotated short protein isoforms of PWWP2A which result from alternative splicing and polyadenylation share amino acids 1 – 517 and 1 – 518 with the long isoform, meaning that they also have NuRD binding capacity. This putative NuRD binding region in PWWP2A 160 – 373 is also found in the paralog PWWP2B.

PWWP2A co-elutes with NuRD in a high molecular weight complex

To study the PWWP2A-NuRD interaction further, I examine the co-migration of PWWP2A with other NuRD components over size exclusion (Figure 3.5A). Nuclear extract from an FS2-PWWP2A expressing mESC clone was run through a Superose6 size exclusion column, and 0.25 ml fractions were collected. Every other fraction was loaded on a denaturing gel for western blot and probed with antibodies against CHD4, Flag (to see FS2-PWWP2A), MTA1, HDAC1, and MBD3. The majority of FS2-PWWP2A elutes at a much higher molecular weight than expected for a monomer of 88 kDa, indicating that it co-migrates with other interaction partners within a larger complex. The bulk of FS2-PWWP2A is found in fraction B8 – C3 and overlaps with MTA1 and HDAC1 over several fractions. Fractions A7 – B8 contain all

the NuRD components tested and most likely represents the principal NuRD complex. A large proportion of CHD4 appears to also exist outside of the NuRD complex, and can be seen as the lower molecular weight species in fractions B4 – C3. MBD3 co-elutes with other NuRD components over fractions A7 – B8. HDAC1 and MTA1 elute in broader peaks over many fractions, likely highlighting the multiplicity of NuRD and sub-NuRD complexes they can form.

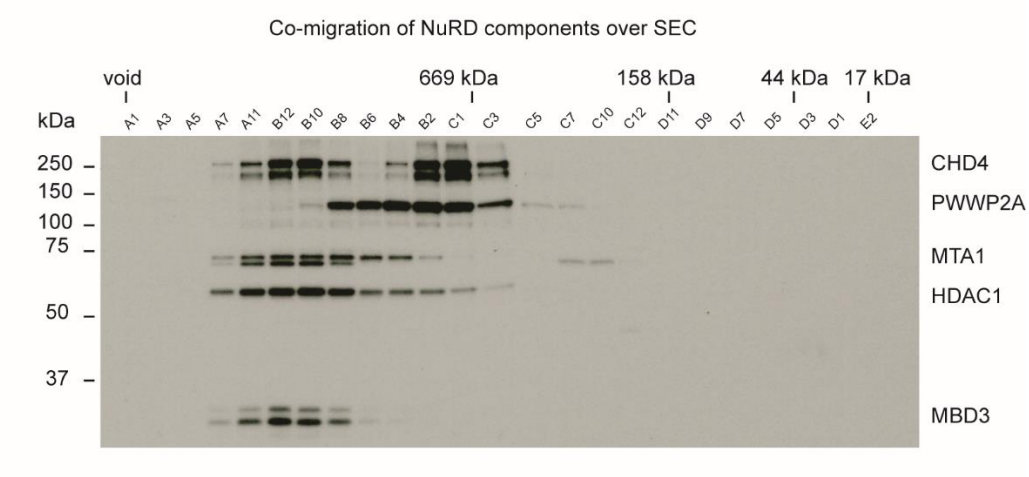


Figure 3.5 PWWP2A comigrates with NuRD subunits in a high molecular weight complex

Separation of nuclear extract under high salt (300mM NaCl) through size exclusion chromatography. Elution profiles of the NuRD subunits CHD4, FS2-PWWP2A, MTA1, HDAC1, and MBD3 were probed by western.

Discussion

In summary, I found that PWWP2A stably associates with several members of the NuRD complex in Hela cells and in mouse ESCs. The interaction with MTA proteins and the core deacetylase complex HDAC1/2 and RBBP4/7 is the strongest. My results suggest that PWWP2A associates directly with one or more of these proteins though a conserved region 160-373 in the N-terminal half of the protein. MTA1-3 is a scaffolding protein upon which the rest of the NuRD complex assembles. From the mass spectrometry of the PWWP2A affinity purification, the MTA1-3 proteins are consistently the most enriched, leading me to hypothesise that PWWP2A binds to MTA1-3 directly, and through MTA1-3 interacts with the deacetylase core components. To confirm this idea definitively, one would also need to identify the domain in MTA1-3 that mediates the interaction with PWWP2A. Interactors that were less enriched, and presumably less stably associated with PWWP2A included PWWP2B, BEND3, and the four transcription factors ZFP423, ZFP521, SALL4, and TIF1B. All of these with the exception of PWWP2B have been previously known to associate with NuRD.

BEND3 is able to co-IP all NuRD components in mESCs. Saksouk et al. (2014) found that BEND3-NuRD mediates an alternative mechanism of chromatin silencing at the pericentric heterochromatin in the absence of DNA methylation or H3K9me3. Constitutive pericentric heterochromatin is

normally enriched in DNA methylation and H3K9me3, while facultative heterochromatin is marked by the polycomb PRC2 modification H3K27me3. In the absence of either DNA methylation or H3K9me3, BEND3 mediates an alternative mode of repression of pericentric elements by recruiting the PRC2 complex. The *de novo* targeting of PRC2 to these novel sites is not through an interaction between BEND3 and PRC2, but through BEND3's interaction with NuRD. While the authors did not dissect the mechanism by which NuRD facilities PRC2 recruitment, other studies have found that the NuRD-mediated deacetylation of H3K27ac is a prerequisite for PRC2 recruitment to shared NuRD/polycomb CpG island target promoters in mESCs (Kim et al., 2015; Reynolds et al., 2012). The mechanism through which BEND3 recruits NuRD to chromatin targets is also not understood.

ZFP423 and its homolog ZFP521 are a class of large multifunctional transcription factors containing up to 30 zinc finger domains (Harder et al., 2014). These zinc fingers are clustered into several distinct domains with both DNA-binding and protein-binding capacity. ZFP423 plays an important role in neurogenesis; loss of ZFP423 in mice leads to abnormal development in the cerebellum, forebrain, and olfactory bulb resulting in death shortly after birth (Cheng et al., 2007; Warming et al., 2006). ZFP423 binds and regulates the activity of EBF1, a transcription factor expressed in neuronal and olfactory tissues (Mesuraca et al., 2015; Tsai and Reed, 1998). Many developmental pathways in embryogenesis including neuronal development have been

linked to BMP signalling. ZFP423 is involved in BMP signalling through its interaction with the SMAD1-SMAD4 complex and recognition of BRE sequences (BMP responsive elements) via its DNA binding domain (Gupta et al., 2010; Hata et al., 2000). Additionally, it has been described to play a role in retinoic acid signalling, Notch signalling, and the DNA damage response. Both ZPF423 and ZPF521 contain an N-terminal NuRD interaction domain, a 12 amino acid motif RRKQ*X*K-P*XX*F first discovered in the SALL family of proteins (Lauberth and Rauchman, 2006). It has been postulated that through its deacetylase and remodelling activities NuRD could enable some of the regulatory roles of ZPF423 in these diverse pathways, though the molecular mechanisms still need to be elucidated.

TIF1ß is a transcription corepressor associated with the KRAB zinc finger first studied in the context of heterochromatin silencing (Groner et al., 2010; Iyengar et al., 2011; Schultz et al., 2002). Recently it was shown that TIF1ß also occupies a large number of gene promoters in ESCs (Iyengar et al., 2011; O'Geen et al., 2007), and may play a role in regulating Pol II pausing (Bunch et al., 2014). TIF1ß is essential in early embryonic development, and null mice die at E5.5 (Cammas et al., 2004). TIF1ß coimmunoprecipitates NuRD subunits CHD3, HDAC2, and RBBP4/7, and this interaction is dependent on its PHD and bromodomain (Schultz et al., 2002). Beyond this, the association and function of TIF1ß within NuRD has not been investigated.

SALL4 also appeared in the affinity purification of FS2-PWWP2A from mESCs. SALL4 has an important role in early embryonic development, mice with homozygous deletion of SALL4 die early during the peri-implantation stage E6.5 (Elling et al., 2006; Sakaki-Yumoto et al., 2006). A homolog of SALL1, SALL4 also contains this 12 amino acid motif; a peptide of this NuRD interaction motif in SALL4 can co-immunoprecipitate HDAC1 and HDAC2 (Gao et al., 2013). In two cancerous cell lines, SALL4 and NuRD have been shown to co-occupy and repress the tumour suppressor *PTEN* gene (Gao et al., 2013; Lu et al., 2009). These studies have led many in the field to attribute an important role of NuRD in facilitating SALL4's function in development. However, conflicting data is presented in a recent study looking specifically to define the role of SALL4 and NuRD in mESCs development. Miller et al. (2016) find that *SALL4* null ESCs and *MBD3* null ESCs show no correlation in expression changes, and that MBD3 binding by ChIP is not strongly affected in *SALL4* null cells. The authors conclude that NuRD recruitment to target genes is not dependent on SALL4 (Miller et al., 2016). While only about 7% of SALL4 associates with NuRD, a large portion of NuRD contains SALL4. Although the authors find SALL4 is not required for MBD3 targeting to chromatin, the presence of SALL4 within the NuRD complex may have other functions.

From the mass spectrometry and nuclear separation studies, it appears that under high salt conditions PWWP2A does not associate with CHD3/4, MBD2/3, or GATAD2A/B subunits of NuRD. This could mean one of two things – either PWWP2A binding occludes the binding of these three subunits and obstructs the formation of the full NuRD complex, or that PWWP2A stably associates with MTA and the deacetylase core and transiently associates with CHD3/4, MBD2/3, and GATAD2A/B as part of a full NuRD complex.

In the introduction, I discussed several papers analysing the structural organisation and stoichiometry of the NuRD complex. These studies agree that the deacetylase portion of NuRD is formed around a dimer of the MTA1-3 protein which binds HDAC1/2 and RBBP4/7 (Millard et al., 2016; Millard et al., 2013; Zhang et al., 2016). Zhang et al. (2016) also found that MBD3 can associate with this subcomplex without affecting its overall structure. The stoichiometry of this complex was 4 RBBP: 2 MTA: 2 HDAC: 1 MBD3. Crosslinking studies by Kloet et al. (2014) have put MTA and MBD3 at the centre of the NuRD complex. Multiple crosslinks exist between MTA1-3, HDAC1/2 and RBBP4/7 that correlate well with structural data. CHD3/4 and GATAD2A/B only share crosslinks with MBD3 and MTA1-3, and do not make any contacts with the deacetylase subunits, once again supporting the idea of stable sub-complexes within NuRD. This sub-NuRD complex with functional deacetylase activity may likely represent a stable functional unit *in vivo* that

transiently interacts with CHD4 and GATAD2A through MBD3 to form the holo-NuRD.

A greater structural or biochemical understanding of how PWWP2A interacts with the sub-NuRD complex formed by MTA1-3, HDAC1/2, and RBBP4/7 is required. Mapping of the MBD3-MTA1 and PWWP2A-MTA1 interaction domains may help determine if MBD3 and PWWP2A are mutually exclusive, or if PWWP2A can associate as part of the holo-NuRD complex. Henceforth, I will refer to this PWWP2A associated NuRD complex as PWWP2A-NuRD.

RESULTS CHAPTER 4

Genome-wide mapping of PWWP2A and HDAC1 targets

Introduction

Having established that PWWP2A interacts with subunits of the NuRD complex, I went on to characterise its localisation to chromatin by ChIP-sequencing to determine if it colocalised to regions occupied by H3K36me3 and NuRD components.

ChIP-sequencing data for multiple components of the NuRD complex is available. In a seminal work on the genome-wide mapping of HAT and HDAC targets, Wang et al. showed that despite being a transcriptional co-repressor complex, HDACs 1, 2, 3, and 6 are enriched on active genes and positively correlate with gene expression, histone acetylation, and Pol II occupancy in $CD4^+$ T cells (Wang et al., 2009b). At these active gene targets, HDAC1 and HDAC3 display predominantly promoter enrichment, while HDAC2 and HDAC6 also show enrichment in the gene bodies. In ESCs, HDAC1 shows enrichment at active promoters and enhancers, with some binding at bivalent promoters and gene bodies (Lee et al., 2015).

MBD3 was shown to be enriched at the enhancers, promoters, and gene bodies of active genes in breast cancer cell lines (Shimbo et al., 2013). In ESCs, MBD3 displays mainly promoter binding and a clear preference for bivalent genes co-occupied by H3K27me3 and Polycomb (Yildirim et al., 2011). CHD4 mainly occupies the promoters and enhancers of active genes, but is also present at gene bodies and bivalent promoters in ESCs. In the Luo et al. (2015) study, the authors identified 11,058 and 7,421 target sites bound by CHD4 and MBD3 in ESCs of which only 3773 are shared; they postulate that CHD4 and MBD3 could function independently of each other or outside of the NuRD complex at many genomic loci, or that multiple variants of the NuRD complex exist (Luo et al., 2015). From these studies, NuRD seems to be associated with the enhancers, promoters and gene bodies of active genes, the promoters of bivalent genes, and not enriched at all on repressed or heterochromatin. It was expected that PWWP2A would show a preference for active genes given its putative binding to H3K36me3 containing nucleosomes, a histone modification that is co-transcriptionally deposited.

Results

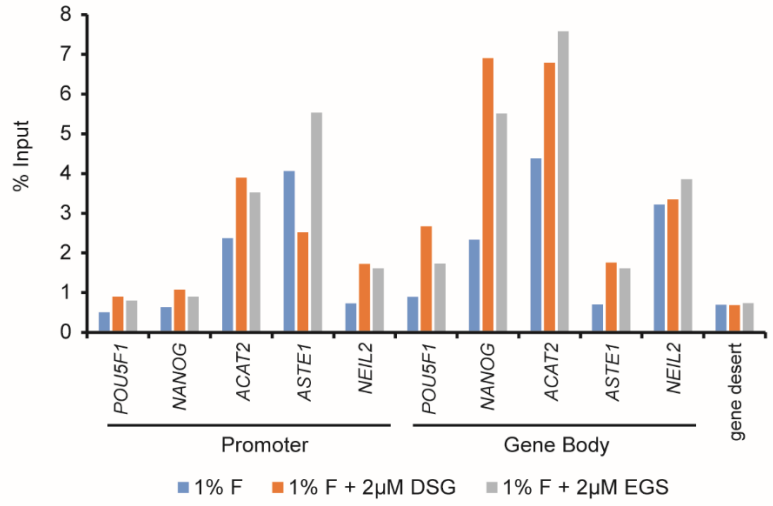
Optimisation of ChIP conditions

I set out to perform ChIP-seq to determine the genomic targets of PWWP2A. I hypothesised that due to PWWP2A's interaction with HDAC1, they would share common target genes. Moreover, as a putative H3K36me3

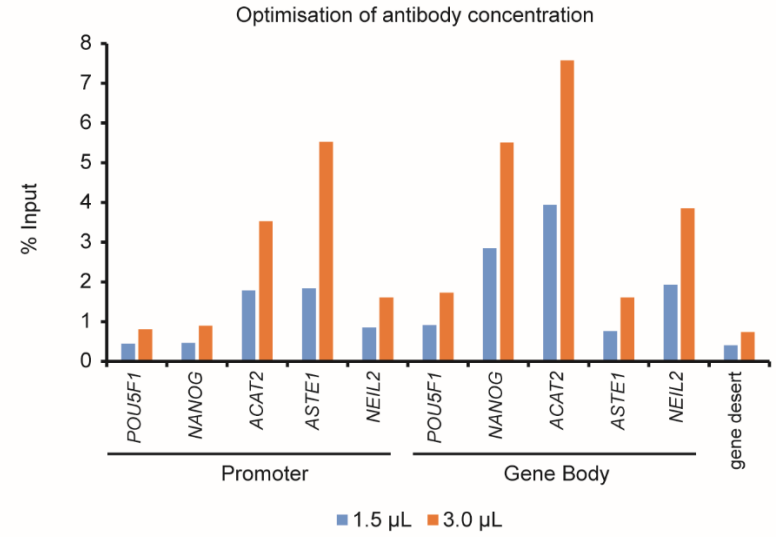
reader, I postulated that PWWP2A may be found at gene bodies of HDAC1 target genes. I optimised the ChIP conditions by selecting five genes occupied by HDAC1 in ESCs ranging from high to low expression – *Pou5f1*, *Nanog*, *Acat2*, *Aste1*, and *Neil2*. Primer pairs were designed for the promoters of these genes (which are occupied by HDAC1), and a gene body region far downstream of the promoter. A gene desert region on Chromosome 19 devoid of any genic elements and histones modifications related to active or heterochromatin was used to estimate the background level of the FS2-ChIP.

Fig 4.1 shows the optimisations for crosslinking conditions, the antibody concentration, and the beads for the affinity purification. Single crosslinking using 1% formaldehyde was tested against double crosslinking with 1% formaldehyde in combination with 2 μM of crosslinkers DSG (disuccinimidyl glutarate at 7.7 Å) or EGS (ethylene glycol bis(succinimidyl succinate) at 16.1 Å). For large chromatin complexes, using long range crosslinkers can improve the immunoprecipitation and recovery of chromatin material. Reassuringly, FS2-PWWP2A was enriched at these HDAC1 target loci. The pattern of enrichment of the genes tested was the same between all three conditions, but under double crosslinking conditions with formaldehyde and DSG or EGS FS2-PWWP2A was clearly more enriched relative to the gene desert background compared to the single crosslinking with 1% formaldehyde. There was minimal difference between DSG and EGS, so EGS was chosen. Figure 4.1B compares two different antibody concentrations used

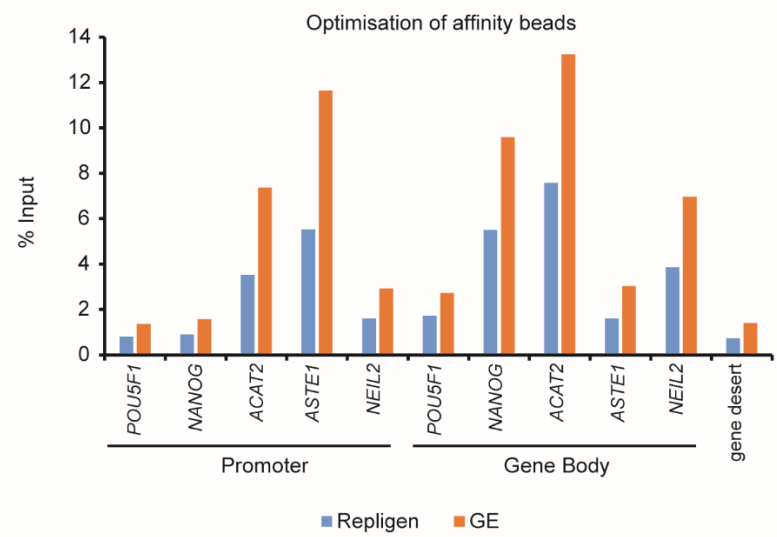
A


Optimisation of crosslinking conditions
% Input
0
1
2
3
4
5
6
7
8
POU5F1
NANOG
ACAT2
ASTE1
NEIL2
POU5F1
NANOG
ACAT2
ASTE1
NEIL2
gene desert
Promoter
Gene Body
1% F
1% F + 2μM DSG
1% F + 2μM EGS

B


Optimisation of antibody concentration
% Input
0
1
2
3
4
5
6
7
8
POU5F1
NANOG
ACAT2
ASTE1
NEIL2
POU5F1
NANOG
ACAT2
ASTE1
NEIL2
gene desert
Promoter
Gene Body
1.5 μL
3.0 μL

C


Optimisation of affinity beads
% Input
0
2
4
6
8
10
12
14
POU5F1
NANOG
ACAT2
ASTE1
NEIL2
POU5F1
NANOG
ACAT2
ASTE1
NEIL2
gene desert
Promoter
Gene Body
Repligen
GE

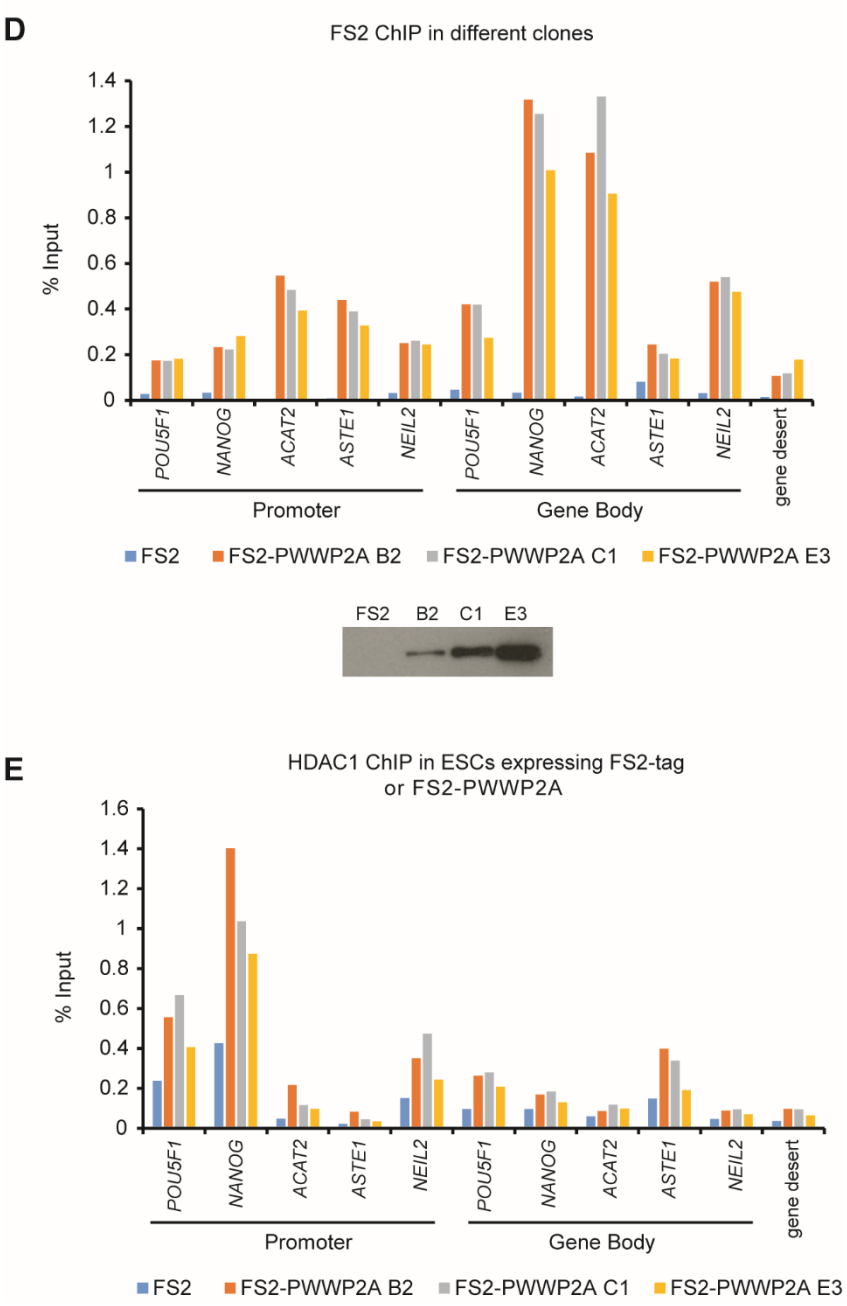


Figure 4.1 Optimisation of FS2-PWWP2A ChIP

A Comparison of single formaldehyde crosslinking vs double crosslinking with formaldehyde and long range crosslinkers EGS or DSG

B Comparison of 1.5 µL vs 3.0 µL of anti-FS2 antibody

C Comparison of protein A affinity beads from two companies

D FS2-PWWP2A ChIP in clones with three different expression levels and in a clone expressing the FS2-tag empty vector.

E ChIP for endogenous HDAC1 the three clones overexpressing FS2-PWWP2A and in the clone expressing the FS2-tag empty vector.

in the ChIP. The antibody used was a rabbit polyclonal derived against a synthetic Flag-StrepII-StrepII peptide. The ChIP conditions using 3 µl per IP were more favourable than with 1.5 µl. Finally two different Protein A affinity resins were tested – IPA300 from Repligen and Protein A Sepharose from GE. ChIP using both beads achieved the same enrichment over background, but the Protein A resin from Repligen had a lower level of background overall and was therefore used for ChIP-sequencing.

Once the general ChIP conditions were optimised, different sonication conditions were also tested to achieve the optimal DNA fragment length for sequencing. Sonication of 40 million cells in 1 mL for 27 minutes (30 sec on 30 sec off) was determined to be the best condition that consistently gave a majority of sheared DNA fragments between 100 – 300 bp in size. Next, three ESC lines expressing different levels of FS2-PWWP2A were tested against a line expressing the FS2 tag alone. ChIP for the FS2 tag yielded a very low signal. Encouragingly, the three different clones with different expression levels of FS2-PWWP2A (B2 – low, C1 – medium, E3 – high) showed a very similar pattern and level of enrichment at the tested loci. (Figure 4.1D) All three clones showed greater gene body enrichment compared to at the promoter, with the greatest enrichment consistently seen at the gene bodies of *Nanog* and *Acat2*. This reassured us that the pattern FS2-PWWP2A localisation observed at these HDAC1 targets was physiological and not due to clonal differences or differences in expression levels. At this point, the

optimisation for FS2-PWWP2A ChIP was considered sufficient to proceed to ChIP-seq.

While HDAC1 data already exists for mouse ESCs, I also decided to perform my own HDAC1 ChIP-seq in the FS2-tag and FS2-PWWP2A expressing lines to account for changes that overexpression of the FS2-tag and PWWP2A may cause to HDAC1 localisation. From Figure 4.1E HDAC1 localisation was only observed at the promoters of three of the five HDAC1 target genes. This could be due to slight differences in cell types, or the primers designed for qPCR did not overlap exactly over the HDAC1 peak. Overall, the HDAC1 ChIP appeared to work well under the conditions optimised for the FS2-PWWP2A ChIP. Interestingly, HDAC1 showed greater enrichment in the ESCs expressing FS2-PWWP2A than the tag, suggesting that increased levels of FS2-PWWP2A could be recruiting more HDAC1 to their shared genomic targets.

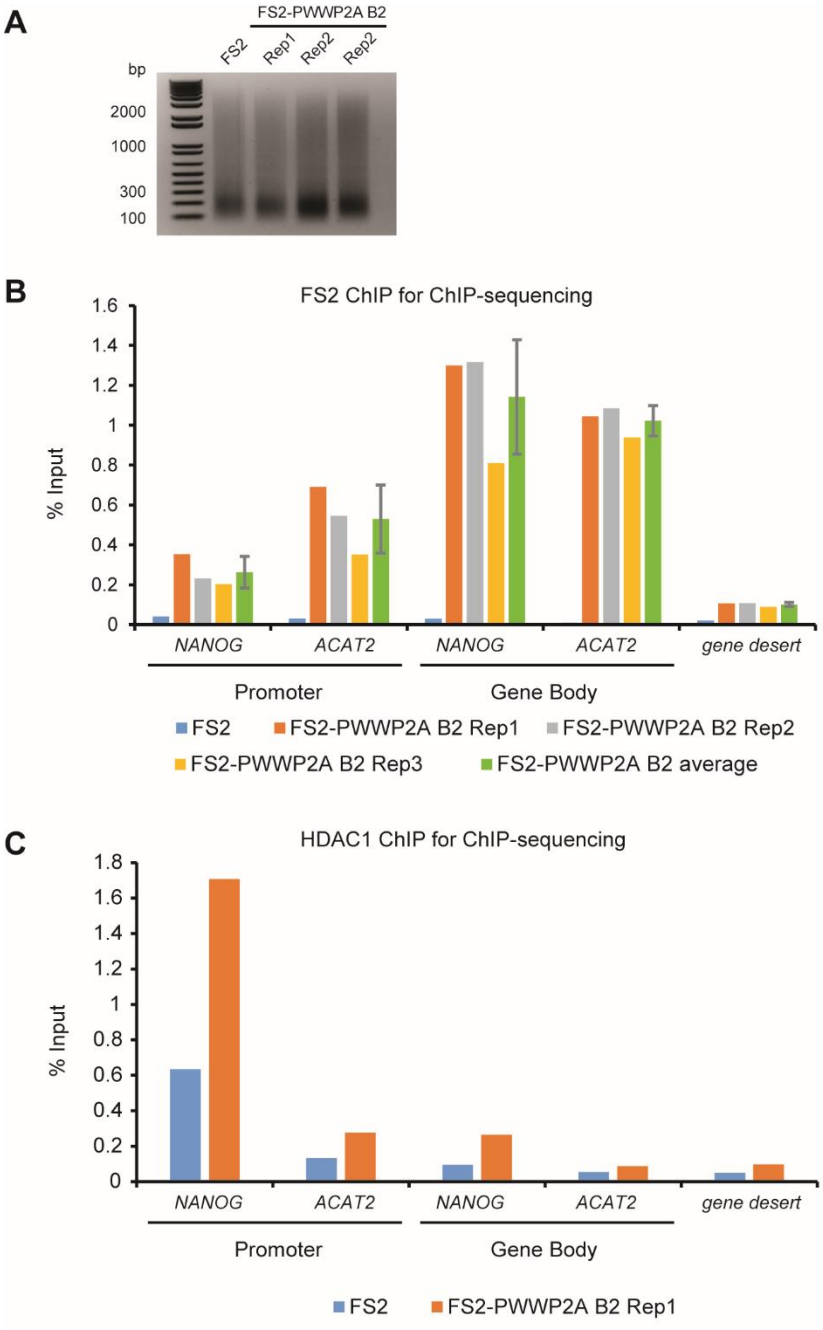


Figure 4.2 Samples for FS2-PWWP2A and HDAC1 ChIP sequencing

A Sonicated chromatin for ChIP-seq. FS2-tag only (FS2), and three biological replicates (FS2-PWWP2A B2 Rep1/2/3).

B FS2-PWWP2A ChIP in three biological replicates of clone B2. One ChIP was performed in the FS2-tag clone to measure background.

C HDAC1 ChIP was performed in one of the biological replicates of the clone overexpressing FS2-PWWP2A, and in the FS2-tag clone which should represent the native condition.

ChIP-sequencing analysis of PWWP2A in mESCs

The clone FS2-PWWP2A B2 with the lowest expression was chosen for ChIP-seq and cells were grown and harvested in three biological replicates. A clone expressing the FS2 tag alone was also used for the background control. Figure 4.2A shows the chromatin was well sonicated with most of the DNA between 100 – 300 bp in size and a light smear above with no large DNA fragments trapped in the wells of the agarose gel. HDAC1 ChIP-seq was performed in the FS2-tag clone and FS2-PWWP2A B2 replicate 1, while FS2 ChIP was performed in all four chromatin samples. To verify the ChIP had worked, a small amount of each sample was used for qPCR (Figure 4.2B and C).

** Bioinformatic analysis and presentation of these data were performed in collaboration with a postdoc in the lab Guifeng Wei.

PWWP2A colocalisation with H3K36me3 is PWWP-domain dependent

The ChIP-seq yielded around 70 million reads per sample, where approximately 60% of the reads mapped to the mm10 mouse reference genome. The three biological replicates for the FS2-PWWP2A ChIP showed very strong correlation and were combined for downstream analyses. Figure 4.3A displays the Chip-sequencing tracks over a 90 kb window for FS2-PWWP2A, HDAC1, the epigenetic modifications H3K4me3, H3K27me3, and

H3K27ac, as well as other genomic features such as CpG islands and DNaseI accessibility. The peaks for FS2-PWWP2A are broad and lie across the entire gene bodies of the genes in this region. The pattern of FS2-PWWP2A localisation strongly resembles the pattern of H3K36me3 over gene bodies, and is very different from the other epigenetic marks H3K4me3, H3K27me3, and H3K27Ac which are found as narrow and sharp peaks at the CpG island promoters. Statistically, all FS2-PWWP2A target sites in mESCs show a very high correlation with H3K36me3 (R = 0.95) (Figure 4.3B).

This Chip-sequencing data supported the idea of PWWP2A as a putative H3K36me3 binder, as speculated based on the initial H3K36me3 SILAC affinity purification. We then tested if the localisation of FS2-PWWP2A at gene bodies was dependent on its H3K36me3-binding C-terminal PWWP domain by comparing the enrichment of full length PWWP2A with a construct lacking the PWWP domain at three genes *Nanog*, *Acat2*, and *Neil2*. The PWWP2A signal at the gene bodies of *Nanog*, *Acat2*, and *Neil2* is lost in the ΔPWWP construct, but not significantly affected at the promoter (Figure 4.3C). This suggests that the PWWP domain of PWWP2A mediates its interaction with the histone modification H3K36me3 and is responsible for the recruitment and localisation of PWWP2A at gene body regions.

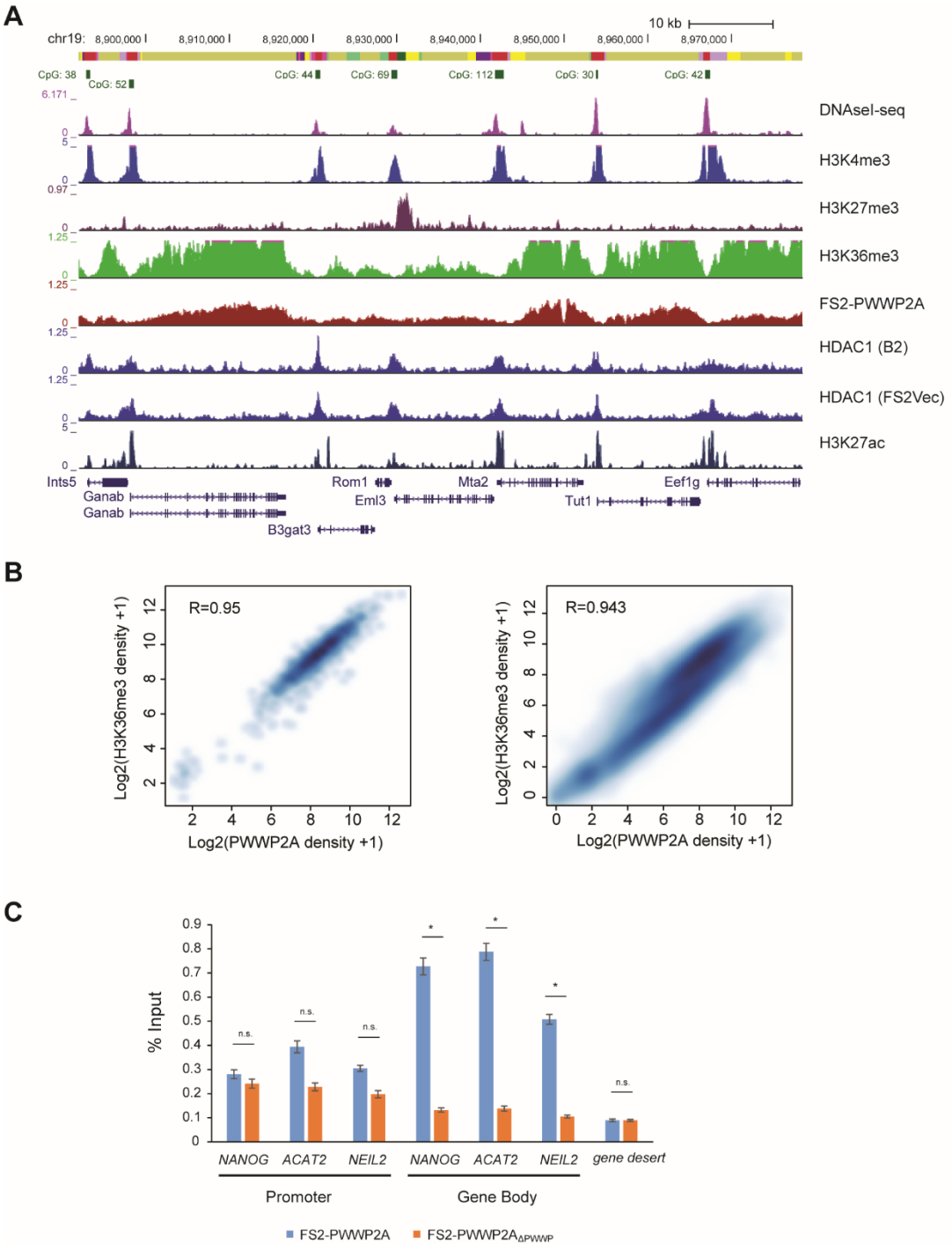


Figure 4.3 PWWP2A colocalises with H3K36me3 at gene bodies

A A snapshot of PWWP2A localisation over a region on chromosome 19. FS2-PWWP2A follows the pattern of H3K36me3 enrichment.

B FS2-PWWP2A ChIP signal highly correlates with that of H3K36me3 in all PWWP2A target genes as well as at all genes.

C ChIP-qPCR of full length FS2-PWWP2A vs. FS2-PWWP2A$_{\Delta PWWP}$. Localisation of PWWP2A to the gene bodies of target genes is dependent on the H3K36me3-binding PWWP domain.

Localisation of PWWP2A at gene promoters is present at lower levels and is not dependent on the presence of the PWWP domain. As mentioned previously, the short isoforms of PWWP2A which are also expressed in mESCs lack the PWWP domain and thus may only bind gene promoters. The presence of both long and short isoforms of the protein with potentially different chromatin targets suggest that PWWP2A may have divergent functions at promoters and gene bodies of target genes.

PWWP2A colocalises with HDAC1 at highly expressed genes

Meta-analysis of ChIP-seq data shows the characteristic binding signature for FS2-PWWP2A, HDAC1, and H3K36me3 over target genes (Figure 4.4A). H3K36me3 and FS2-PWWP2A are both enriched over gene bodies with increasing occupancy towards the 3' end of genes. This characteristic of localisation of H3K36me3 has been observed previously and well noted; FS2-PWWP2A likely shows the same pattern due to its direct binding and recruitment to H3K36me3 containing nucleosomes. HDAC1 on the other hand is highly enriched at gene promoters.

In Figure 4.4B, genes were divided evenly by expression level using previously published RNA-seq data from E14 mESCs. 5985 HDAC1 targets were identified, over 80% of HDAC1 targets were intermediately or highly

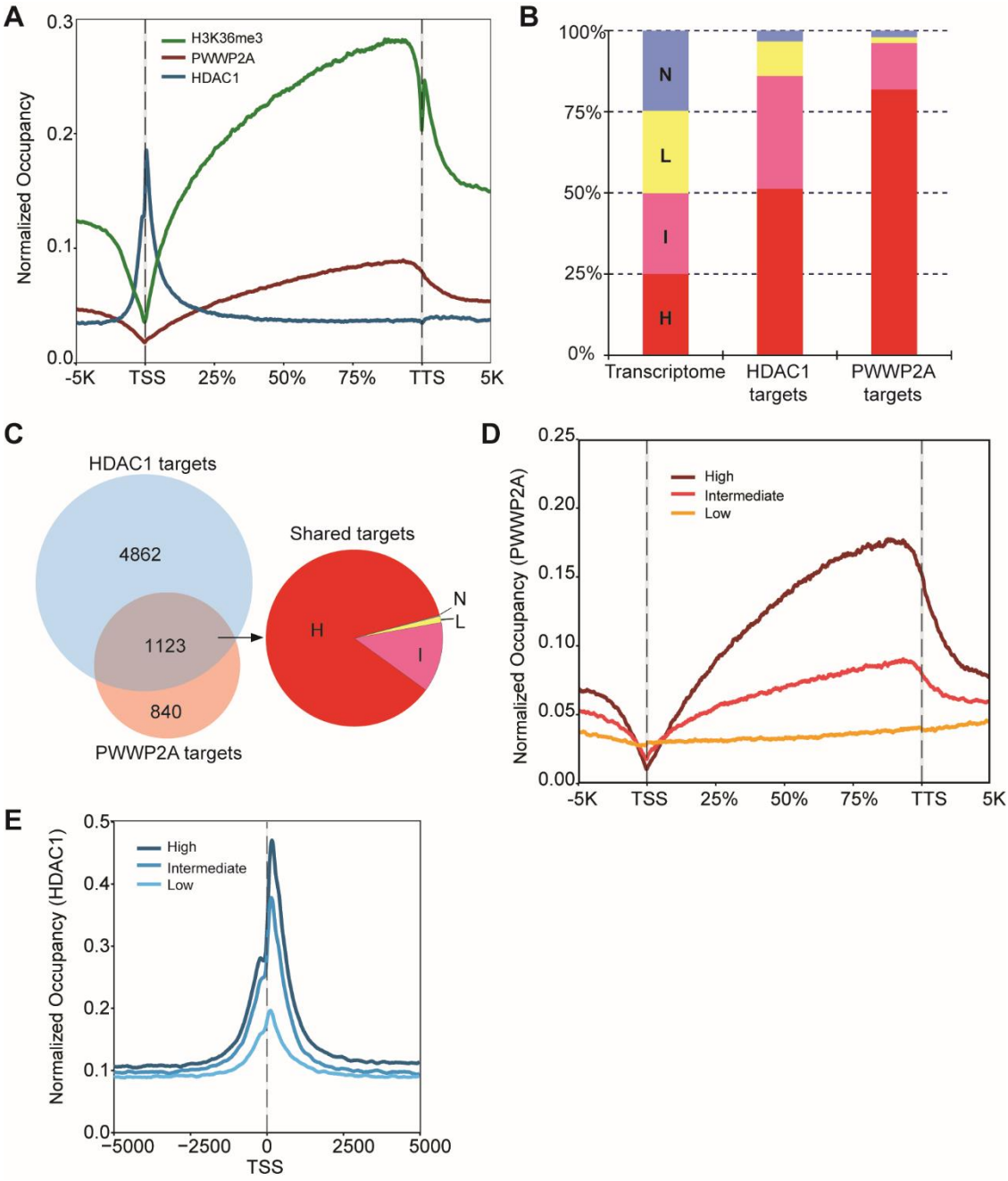


Figure 4.4 PWWP2A and HDAC1 target genes are highly expressed in mESCs

A Meta-analysis of the profiles for FS2-PWWP2A, HDAC1, and H3K36me3

B HDAC1 and PWWP2A target genes are categorised by expression level

C HDAC1 and PWWP2A share over 1000 target genes, most of which are highly or intermediately expressed

D and E PWWP2A and HDAC1 enrichment are both positively correlated with gene expression level

expressed in mESCs. This agrees with previously published data that point to HDAC1's role in maintaining a dynamic turnover of acetylation at active genes. FS2-PWWP2A target genes numbered at 1963, and over 90% of FS2-PWWP2A targets were of intermediate or high expression. Peak calling requires setting specific thresholds for what can be considered enough enrichment over background, and due to the broad nature of FS2-PWWP2A ChIP peaks, this may be an under-representation of PWWP2A targets. Of the 5985 HDAC1 targets and 1963 FS2-PWWP2A targets, 1123 are shared. Of these shared gene targets, the majority are once again highly or intermediately expressed (Figure 4.4B). Unsurprisingly, both FS2-PWWP2A and HDAC1 occupancy increase with gene expression level Figure 4.4 D and E. This correlation between PWWP2A and gene expression may be due to the fact that the H3K36 methyltransferase SETD2 associates with the elongating RNA Pol II, and increased Pol II occupancy leads to greater H3K36me3.

The chromatin environment at PWWP2A target genes

To understand the chromatin environment at PWWP2A target genes, ChIP-seq data from epigenetic modifications (H3K4me3, H3K27me3, H3K36me3, H3K9ac, H3K27ac, H2A.Zac), chromatin remodellers (CHD2, CHD4), pluripotency factors (NANOG, OCT4), and the different forms of POL II (S5P, S2P) were divided and plotted in meta-analysis as PWWP2A targets and PWWP2A non-targets. PWWP2A target genes correlated well with all

markers of active genes, with higher levels of promoter H3K4me3, H3K9ac, H3K27ac, and H2A.Zac and gene body H3K36me3 than PWWP2A non-target genes. Conversely PWWP2A non-target genes had much higher levels of the repressive histone mark H3K27me3. Both chromatin remodellers CHD2, CHD4 were also more enriched by ChIP at PWWP2A targets genes, as were the transcription factors NANOG and OCT4. Likewise both the initiating Ser-5 and the elongating Ser-2 forms of RNA Pol II are more enriched at PWWP2A targets compared to non-targets. PWWP2A targets are genes with all the conventional markers of active transcription.

Through Gene ontology analysis of the PWWP2A target genes identified through ChIP-seq, these genes are involved in a wide range of nuclear processes from RNA processing, to cell cycle, to transcription, and etc (Figure 4.5B).

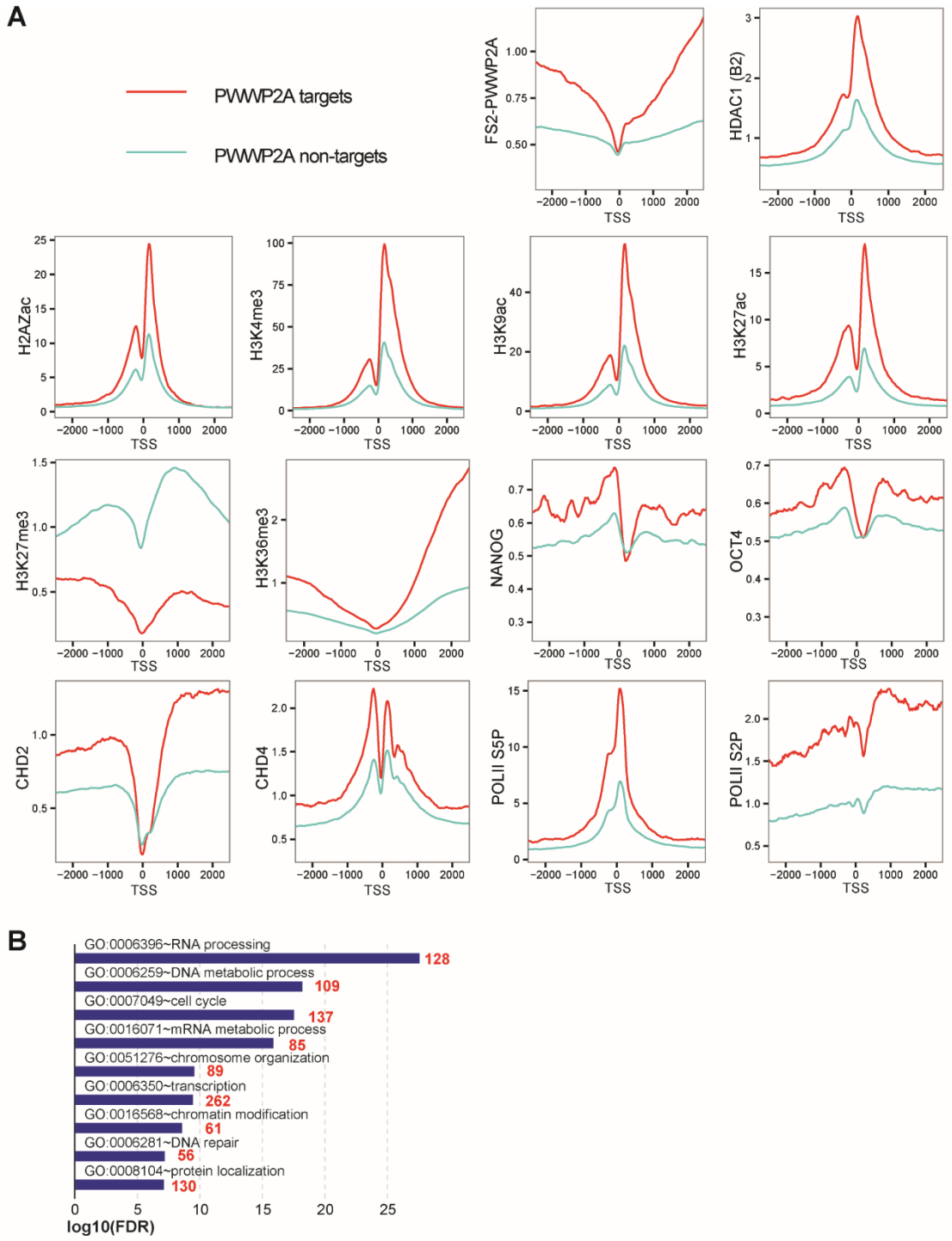


Figure 4.5 Comparison of the chromatin signature for PWWP2A targets and non-targets

A The ChIP profile of PWWP2A targets and non-target genes are compared for different epigenetic marks, binding of transcription factors NANOG and OCT4, chromatin remodellers, and POL II.

B Gene ontology analysis of PWWP2A targets.

Discussion

While FS2-PWWP2A and HDAC1 are observed at a number of the same genes, they appear to be distributed differently with FS2-PWWP2A occupying mostly the gene bodies, and HDAC1 localising principally at the promoter. The biochemical studies presented in Chapter 3 strongly support a physical interaction between FS2-PWWP2A and HDAC1, therefore it was initially surprising to see such distinct non-overlapping ChIP profiles for the two proteins. However, affinity purification of FS2-PWWP2A only recovers a fraction HDAC1 meaning that only a portion of HDAC1 is associated with PWWP2A-NuRD in the nucleus. Thus given that only a small proportion of HDAC1 may be recruited to gene body regions by FS2-PWWP2A, this would be masked by the strong promoter signal.

There is some evidence that PWWP2A recruits HDAC1 to shared target sites. Overexpression of FS2-PWWP2A in mESCs increases HDAC1 occupancy at promoter target sites. By ChIP-qPCR HDAC1 enrichment was greater in clones expressing FS2-PWWP2A than FS2 tag alone (Figure 4.1E and 4.2C). This effect is also observed in the ChIP-seq data where by meta-analysis the combined HDAC1 signal in the FS2-PWWP2A line is slightly greater than the HDAC1 signal in the FS2-vector line (data not shown). This increased HDAC1 occupancy must result from greater recruitment to shared target sites upon overexpression of PWWP2A. While higher levels of HDAC1

is observed at the promoters in cells overexpressing FS2-PWWP2A, increases in HDAC1 occupancies at gene bodies is not easily observable.

We suspect that these increases of HDAC1 at the gene bodies are not being captured by ChIP for several reasons. Gene promoters are much smaller in size compared to gene bodies. HDAC1 enrichment at promoters appears as a sharp peak by ChIP-seq whereas HDAC1signal over gene bodies is broad, flat, and difficult to distinguish over the background. HAT and HDAC are highly enriched at promoters of active gene and the turnover of promoter acetylation is greater than the turnover of acetylation at gene bodies. Accordingly, the dynamics of HDAC1 binding at promoters could be very different to its binding over gene bodies, gene body HDACs may not be as efficiently not captured by the ChIP crosslinking. At promoters, HDAC1/NuRD may interact more with the DNA or promoter associated histone modifications or histone variants, while at gene bodies PWWP2A binding to H3K36me3 may be one of few modes of NuRD recruitment and the interaction with chromatin over gene bodies may be weaker or more transient and not be captured by the crosslinking.

While the majority of PWWP2A localises to the H3K36me3 containing gene bodies, PWWP2A promoter enrichment can be seen by ChIP-qPCR. Upon deletion of the PWWP domain, FS2-PWWP$_{\Delta PWWP}$ was lost from gene

bodies but not greatly affected at the promoters. PWWP2A may also be recruited to promoters by another mechanism. The existence of short isoforms of PWWP2A and PWWP2B which lack the PWWP domain point to a possible function of PWWP2A/B outside of H3K36me3 targeting. It would be interesting to perform ChIP-seq on the short protein isoform of PWWP2A study its genomic localisation independent of H3K36me3 binding.

The role of PWWP2A at target sites genome wide remains unclear, but due to its H3K36me3-dependent targeting, it must have a general function at the gene bodies of actively transcribed genes. Also, its expression in a wide range of tissues from embryonic tissue to differentiated somatic tissue allude to a general biological function at active genes. As a protein with strong gene body localisation at active genes, PWWP2A's function could very well be related to transcriptional regulation at gene bodies, possible prevention of aberrant transcription initiation or regulation of transcriptional elongation.

RESULTS CHAPTER 5

Generation of PWWP2A and PWWP2A/B knockout mESCs

Introduction

So far I have established that PWWP2A is an H3K36me3 reader that forms a stable complex with NuRD subunits MTA1-3, HDAC1/2, and RBBP4/7 and is enriched principally at the gene bodies of actively transcribed genes. To begin to address the functional roles of PWWP2A in ESCs, PWWP2A single knockout and PWWP2A/PWWP2B double knockout lines were generated in mESCs using CRISPR/Cas9.

This chapter will describe how these knockout cell lines were generated and characterised. Then I will outline the future experiments we will employ in these cells to address the biological roles of PWWP2A-NuRD.

Results

Generation of *Pwwp2a* knockout cells

The gene annotation for *Pwwp2a* can be seen in Figure 5.1A, the two major transcripts are shown. Transcript variant 1 contains 4 exons and codes for the short protein isoform of PWWP2A, while the transcript variant 2 contains 2 exons and codes for the long protein isoform that harbours the

PWWP domain. Due to lack of an antibody to screen for loss of the PWWP2A protein, my strategy was to generate large deletions within the *Pwwp2a* gene to prevent transcription from starting at alternative transcription initiation sites.

I designed three guide sequences for CRISPR/Cas9 nuclease mediated cleavage within the *Pwwp2a* gene (Figure 5.1B). These were transfected into E14 mESCs. Transfection of plasmids containing the guides 1 and 3 in combination with the Cas9 nuclease yielded a near-complete deletion of the *Pwwp2a* gene. PWWP2A null clones B8 and G12 both had homozygous deletions of around ~ 23 kb that removed most of exon 1 and exon 2 (Figure 5.2). These deletions also generated an early stop codon so that if the transcript was stable, only a peptide of 40 – 50 AA long is produced. Transfection of plasmids containing guides 2 and 3 deleted the majority of exon 2 to produce a transcript that codes for only the first ~ 200 AA of the protein. Two clones E9 and F3 were selected (Figure 5.2).

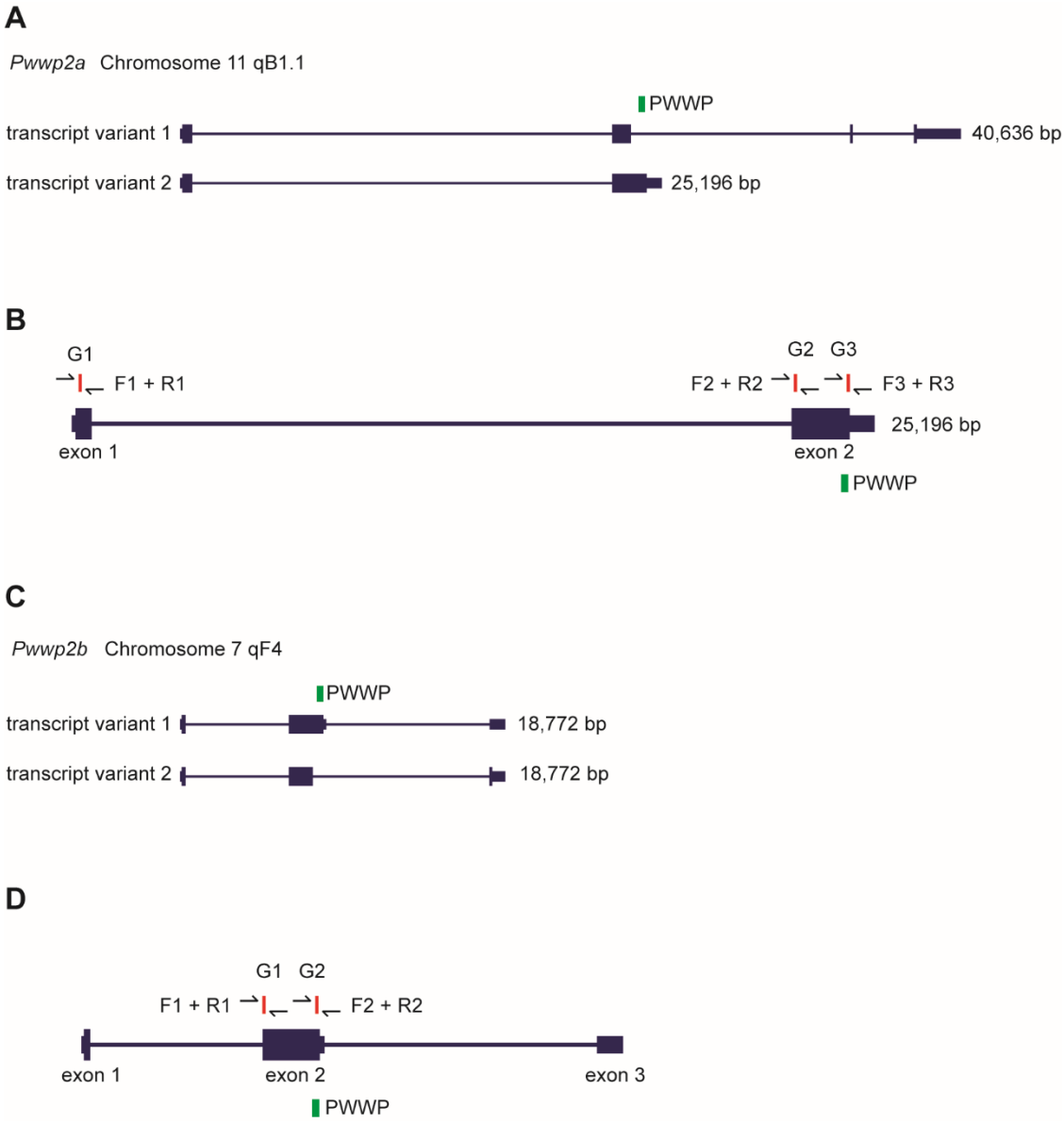


Figure 5.1 Design of guides for CRISPR/Cas9 mediated deletion of PWWP2A and PWWP2B

A *Pwwp2a* has two major transcripts resulting from alternative splicing. Only transcript variant 2 contains the exon coding for the PWWP domain (indicated in green).

B Three guide sequences were designed, G1 in exon1, and G2 and G3 in exon2. Primer pairs were designed around each guide to aid in screening for deletions by PCR.

C *Pwwp2b* also has two major transcripts. Transcript variant 1 codes for the long protein isoform containing the PWWP domain (green).

D Two guide sequences were designed to delete exon2 of P*wwp2b* which codes for the majority of the protein.

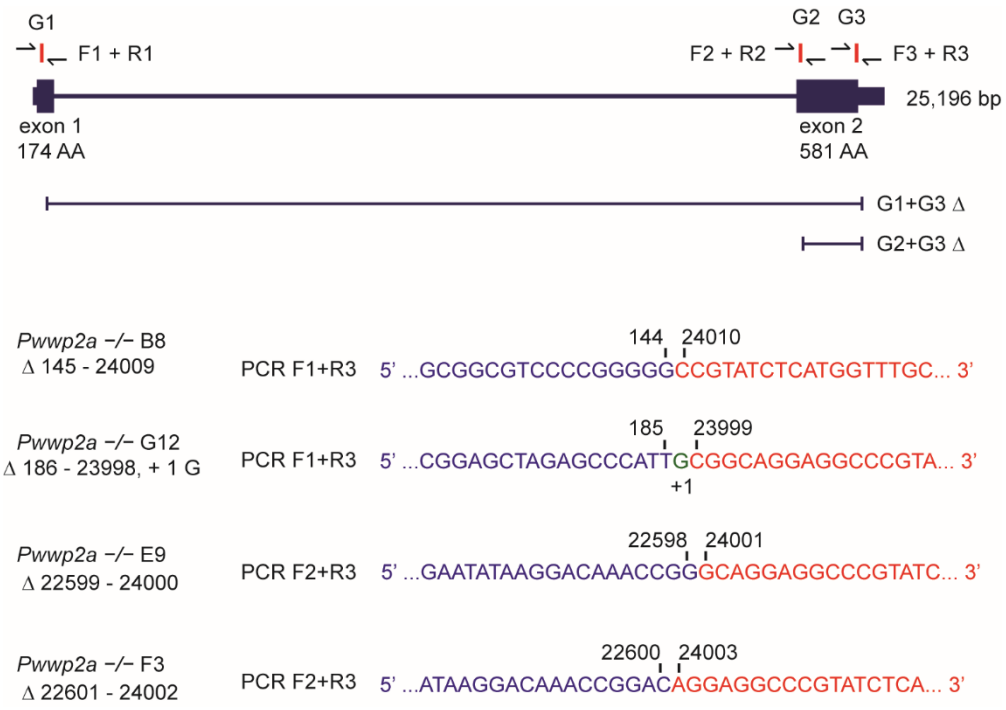


Figure 5.2 Generation of *Pwwp2a* -/- mESCs

Transfection of guides G1 + G3 generated large deletions spanning most of the gene as indicated. Screening was performed by PCR using primers F1 + R3. *Pwwp2a* -/- clones B8 and G12 contain homozygous deletions and the frame shift also introduces early stop codons into the sequence.

Transfection of guides G2 + G3 generated deletions of exon2. Clones were screened using primers F2 + R3. *Pwwp2a* -/- clones E9 and F3 contain homozygous deletions of exon2 which codes for the majority of the protein and introduces an early stop codon into the sequence.

Characterisation of PWWP2A null mESCs

Due to PWWP2A's association with NuRD, I wanted to test if levels of NuRD subunits would be affected upon deletion of PWWP2A. Nuclear extract was prepared from wildtype E14 and the four PWWP2A null lines were tested for global levels of CHD4, MTA1, HDAC1, and HDAC2. Loss of PWWP2A did not appear to affect the levels of any NuRD complex components in mESCs (Figure 5.3A).

All knockout lines grew in colonies and appeared ES-like upon visual inspection, and none had any observable defects in growth rate. PWWP2A null clones B8 and F3 stained positive for alkaline phosphatase similar to the E14 WT mESCs (Figure 5.3B).

Generation of *Pwwp2a/b* double knockout mESCs

Due to the high degree of homology between PWWP2A and PWWP2B, they might have redundant functions in the cell. PWWP2B is likely a reader of H3K36me3 as the C-terminal PWWP domain is particularly conserved. Furthermore, PWWP2B may also interact with the NuRD complex, as it was purified from the screen for PWWP2A interactors (see Chapter 3). Therefore, I decided to generate a *Pwwp2b* deletion in the *Pwwp2a -/-* background.

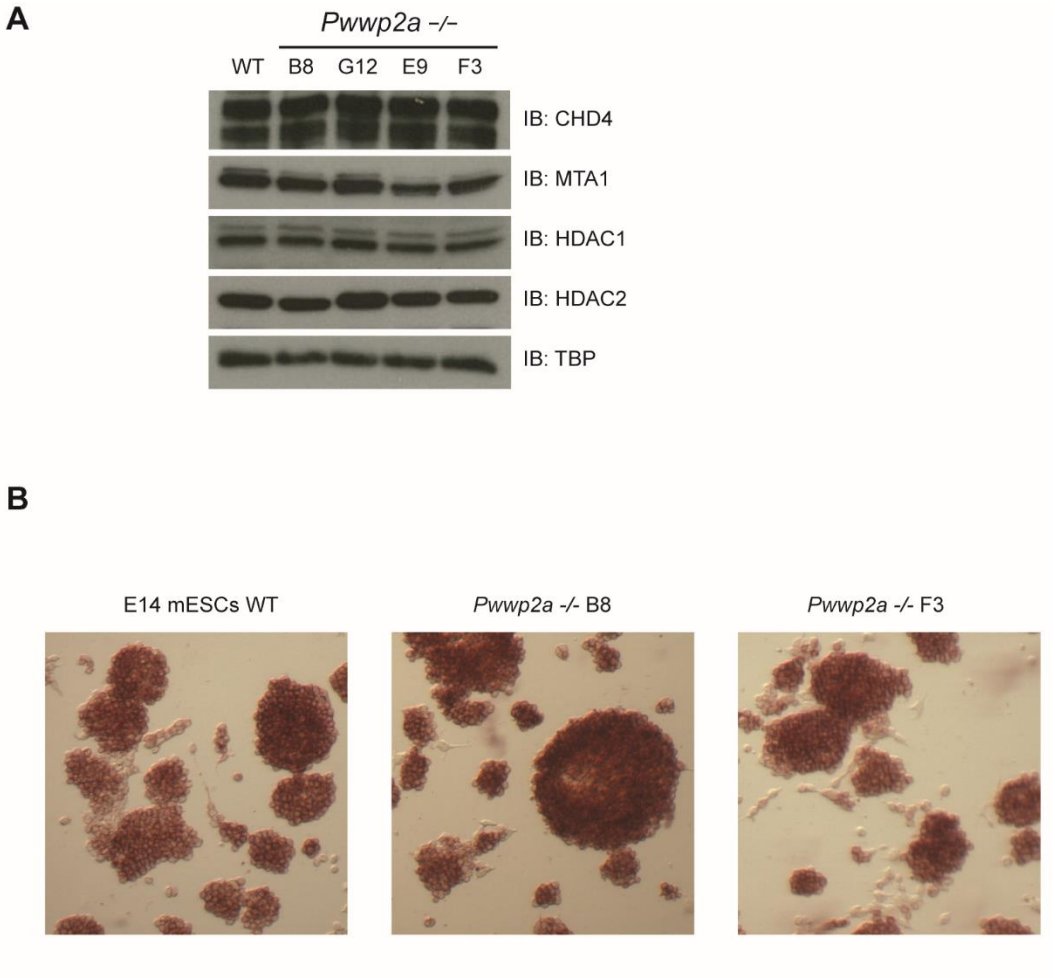


Figure 5.3 Characterisation of PWWP2A null mESCs

A Nuclear extract was prepared for E14 WT mESCs and PWWP2A null mESCs, and probed for NuRD subunits CHD4, MTA1, HDAC1, HDAC2, and the loading control TBP.

B E14 WT mESCs and PWWP2A null clones B8 and F3 were stained for alkaline phosphatase.

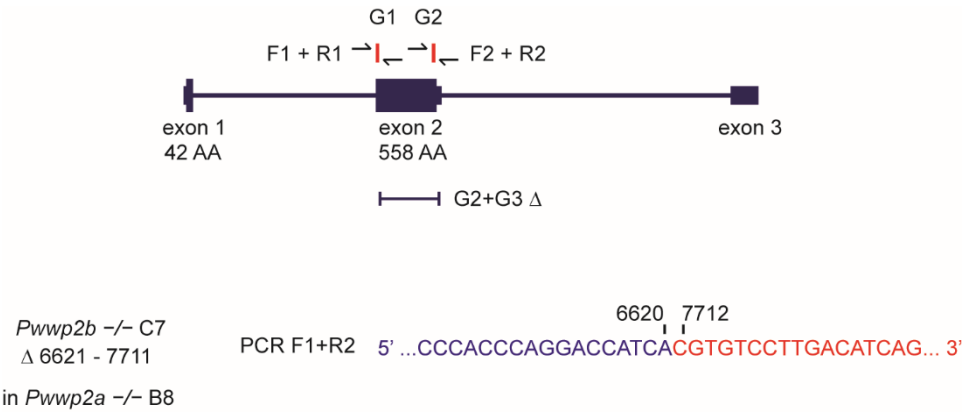


Figure 5.4 Generation of PWWP2A/B double knockout mESCs

Guides G1 and G2 were used for the deletion of *Pwwp2b* in the background of *Pwwp2a* -/- clone B8. Clones were screened using primers F1 + R2. Clone C7 has homozygous deletion of exon2 of *Pwwp2b*, where an early stop codon is also introduced in the sequence.

The clone *Pwwp2a -/-* B8 was used for subsequent deletion of *Pwwp2b* by CRISPR/Cas9. The *Pwwp2b* gene transcripts can be seen in Figure 5.1C. Both transcripts contain 3 exons but due to alternative splicing, only transcript variant 1 contains the region coding for the PWWP domain. Guides were designed to delete exon 2 which codes for the majority of the protein including the PWWP domain (Figure 5.3D). The best clone was determined to be C7 in which the deletion of exon 2 produced a frame shift resulting in an early stop codon. The *Pwwp2a/b -/-* lines were also ES-like and indiscernible from WT mESCs in appearance.

PWWP2A null and PWWP2A/B null mESCs are able to differentiate

To determine if deletion of PWWP2A and PWWP2B had any effects on the ability of the mESCs to differentiate, E14 WT cells, PWWP2A null lines B8 and F3, and the PWWP2A/B null line C7 were differentiated into embryoid bodies.

To initiate differentiation, cells were seeded at low density (1 million) in the absence of LIF. Upon LIF withdrawal ES cells spontaneously aggregate to form three dimensional spheres called embryoid bodies. Embryoid bodies can give rise to cells from all three germ layers (endoderm, ectoderm, mesoderm) which recapitulates many of the early steps in embryogenesis (Kurosawa, 2007). On Day 3 of LIF withdrawal, cells were blown into suspension and passaged into bacterial-grade dishes to prevent cell adhesion and to initiate embryoid body formation. All three deletion lines were able to form embryoid bodies. The size, number and appearance of the embryoid bodies were very similar to the WT mESCs line (Figure 5.5). On Day 7 the embryoid bodies were passaged back into regular tissue culture dishes and allowed to attach. By Day 10, many differentiated cells could be seen on the plates in all four cell lines. There were no obvious differences visually in the knockouts compared to the wildtype. However expression of pluripotency markers and differentiated lineage-specific markers still need to be compared

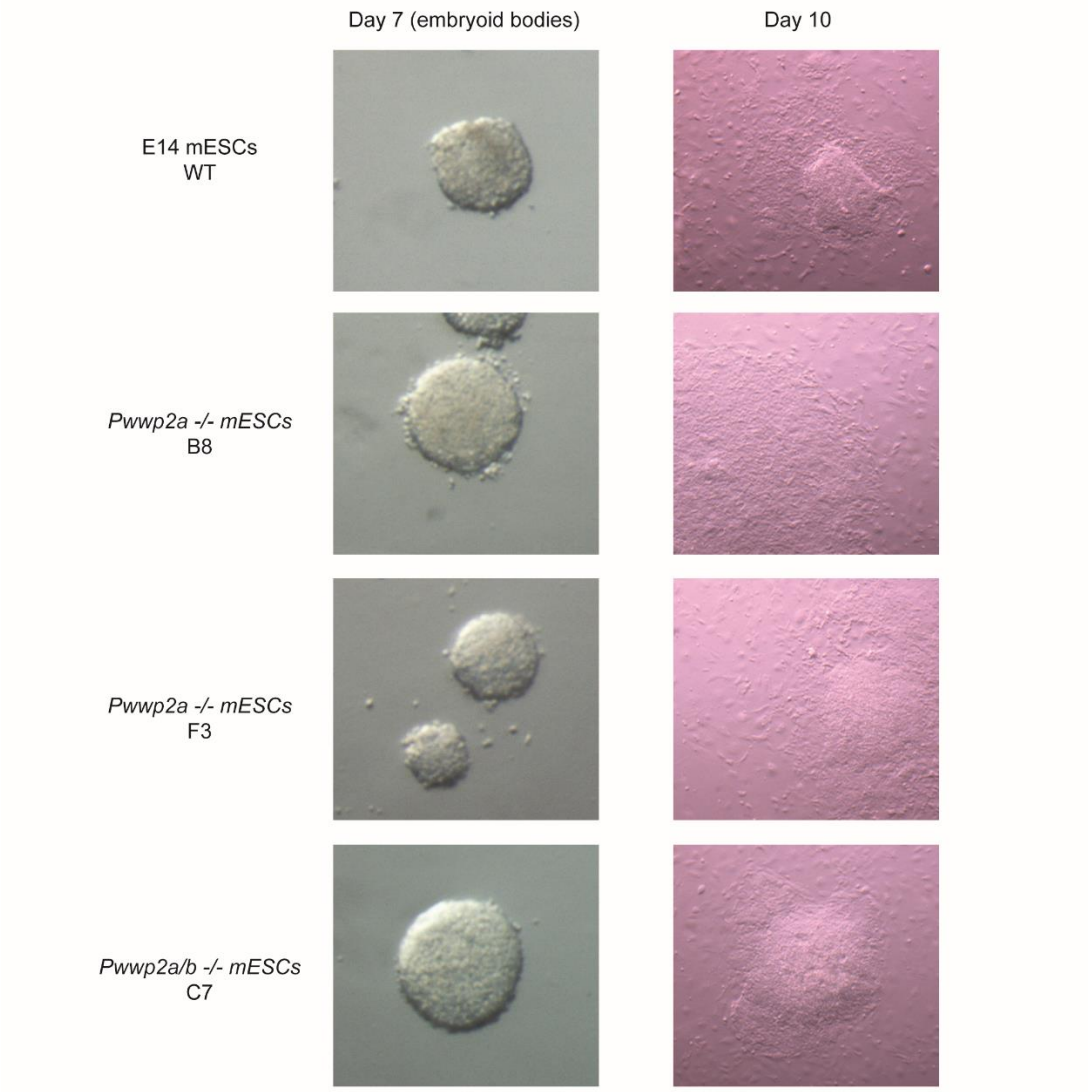


Figure 5.5 Differentiation of WT and PWWP2A and PWWP2A/B null mESCs into embryoid bodies

E14 WT and PWWP2A and PWWP2A/B null mESCs were differentiated into embryoid bodies via LIF withdrawal. Embryoid bodies was observed for all four cell lines (Day 7). Embryoid bodies were maintained in suspension for 4 days, then plated onto gelatin-coated tissue culture dishes and allowed to reattach. Differentiated fibroblast-like cells can be seen by Day 10 in all four cell lines.

for WT and mutant cells by RT-PCR to test if the PWWP2A and PWWP2A/B null mESCs can differentiate normally.

Global acetylation levels on H3K36me3 histones are unaffected by PWWP2A and PWWP2B knockout

Based on my hypothesis that PWWP2A-NuRD may be recruited by H3K36me3 to gene bodies to mediate histone deacetylation at these regions, I set out to test if global levels of acetylation were affected in PWWP2A and PWWP2A/B null cells in collaboration with Dr. Nicholas Crump.

The acetylation levels of histones can be easily visually when denatured histones are run in an acid urea gel. Because each acetylation neutralises the charge of one lysine residue, H3 histones with different numbers of acetylations migrate at different speeds resulting in a clear banding pattern (Figure 5.6). Immunoblotting with antibodies against modifications on H3, such as H3K4me3, H3K36me3, and H3K79me3 allows the visualisation of the co-occurrence of these marks with histone acetylation.

From Figure 5.6, most H3K4me3 containing histones in WT cells typically carry between one to three acetylations. Upon inhibition of deacetylation by treatment with the HDAC inhibitor TSA, H3K4me3

containing histones become hyperacetylated, shifting to four to five acetylated residues. H3K36me3 containing histones are generally less acetylated carrying between zero to two acetylated residues, and upon treatment with TSA progressively shift to becoming hyperacetylated.

Of note, H3K36me3 containing histones in mESCs are more sensitive to TSA treatment than H3K36me3 containing histones in fibroblasts. In fibroblasts, H3K36me3 containing histones do not seem to be substrates for HDACs, as there is no shift in acetylation levels upon HDAC inhibition by TSA treatment. This suggest that there may be fundamental differences in the chromatin organisation in mESCs compared to differentiated somatic cells. H3K36me3 containing nucleosomes may experience higher turnover of acetylation in mESCs.

If PWWP2A and PWWP2B were responsible for regulation of global deacetylation, we would expect to see a shift to the hyperacetylated state in PWWPA and PWWP2A/B null cells, similar to TSA treatment. However no changes in the global levels of acetylation of H3K4me3, H3K36me3, and H3K79me3 containing histones is observed in the null lines. H3K36me3 containing histones are substrates for deacetylation because they respond to TSA treatment, but there are likely multiple pathways that recruit HDAC

activity to H3K36me3 containing histones, with PWWP2A-NuRD being just one.

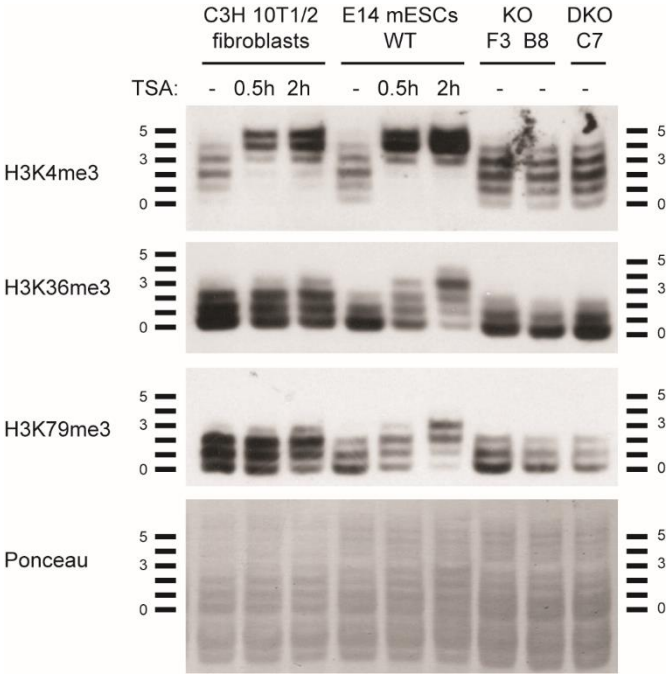


Figure 5.6 PWWP2A and PWWP2A/B null mESCs do not show global increases in acetylation on H3K4me3, H3K36me3, or H3K79me3 containing histones

This experiment was done in collaboration with Dr. Nicholas Crump. I grew, TSA-treated, and harvested the cells. Nicholas Crump extracted the nuclei, and processed the histones for acid urea gel analysis.

Histones from C3H 10T1/2 fibroblasts, E14 WT and PWWP2A null and PWW2A/B null mESCs were run on an acid urea gel, and immunoblotted for H3K4me3, H3K36me3, and H3K79me3. The levels of acetylation on histone H3 are marked from 0 to 5. C3H 10T1/2 fibroblasts and E14 mESCs were treated with the HDAC inhibitor TSA for 0.5h and 2h. Histone become hyperacetylated upon HDAC inhibition. Loss of PWWP2A alone or in combination with PWWP2B did not result in global increases in acetylation on H3K4me3, H3K36me3, or H3K79me3 containing histones.

Discussion

PWWP2A and PWWP2A/B null mESCs were generated using a CRISPR/Cas9 mediated strategy. Loss of PWWP2A from mESCs did not appear to affect cell growth, pluripotency, or the protein levels of NuRD subunits.

Deletion of MBD3 has been observed to affect the ability of mESCs to differentiate. MBD3 null cells can initiate the formation of embryoid bodies but fail to express lineage-specific markers and unable to commit to larger stages of differentiation (Kaji et al., 2006). PWWP2A and PWWP2A/B null mESCs seem to differentiate to be able to undergo differentiation into embryoid bodies, and form fibroblast-like cells when allowed to reattach. Analysis of the expression profile of lineage-specific markers in these cells will need to be performed to verify that these cells can undergo differentiation to cells from all three germ layers.

PWWP2A and PWWP2B deletion did not appear to affect global levels of acetylation associated with histones containing the active marks H3K4me3, H3K36me3, and H3K79me3. As previously mentioned, there may be redundant mechanisms that mediate deacetylation at gene bodies, thus global levels of acetylation would not be affected. Loss of PWWP2A may affect NuRD recruitment or acetylation only at a subset of genes. To look in detail,

I am planning to perform calibrated ChIP-seq for HDAC1 and MTA2 in WT, PWWP2A null, and PWWP2A/B null lines, to determine if loss of PWWP2A alone or in combination with PWWP2B affects the recruitment of NuRD subunits to target sites in mESCs.

Additionally, as a factor that binds gene bodies, PWWP2A may have a role in transcriptional elongation. Calibrated ChIP-seq for Ser-5 and Ser-2 phosphorylated Pol II will also be performed to compare Pol II recruitment in WT and PWWP2A and PWWP2A/B null lines. In conjunction, 4-sU nascent RNA sequencing will be used to address differences in transcription upon PWWP2A or PWWP2A/B deletion. Together, these experiments should give some mechanistic insight into the role of PWWP2A-NuRD in mESCs.

DISCUSSION

In the first part of my project I studied the effect of pre-existing histone modifications on the activity of different Polycomb and Trithorax chromatin modifiers. In brief, I found that histone crosstalk positively regulated the activity of chromatin modifiers from the same group while negatively regulating the activity of the opposing group. The PRC1 modification H2AK119u1 stimulated the activity of JARID2-containing PRC2 complexes, and the PRC2 modification H3K27me3 stimulated the activity of canonical CBX-containing PRC1 complexes.

While H3K4me3 and H3K36me3 negatively regulate PRC2 activity, surprisingly H3K27me3 did not inhibit any of the Trithorax H3K4 and H3K36 methyltransferases tested. We hypothesised that the absence of H3 molecules carrying both H3K27me3 and H3K36me3 on the same tail was a result of mutual inhibition, namely inhibition of PRC2 by H3K36me3 and inhibition of H3K36 methyltransferases by H3K27me3. Contrary to our expectation, this trans-histone crosstalk only occurs one way, and H3K27me3 did not inhibit the H3K36 methyltransferase NSD2 as expected. However, the Polycomb marks H3K27me3 and H2AK119u1 strongly colocalise and as previously mentioned reinforce one another's accumulation, meaning that nucleosomes at Polycomb target genes are likely carry both modifications.

Thus, the *in vivo* data which the mutual exclusion between H3K27me3 and H3K36me3 may be explained by the ability of the PRC1 mark H2AK119u1 to inhibit H3K36 methyltransferases.

Polycomb and Trithorax proteins are essential for proper development, and their mutation or misexpression is linked to a plethora of developmental conditions, and diseases such as cancer (Mills, 2010; Morishita and di Luccio, 2011; Richly et al., 2011; Steffen and Ringrose, 2014). Studying the molecular mechanisms of their interaction with chromatin is important part of understanding their biological function. Native chromatin is decorated with a myriad of different modifications. *In vitro* studies using recombinant chromatin templates allows us to test very specific questions and disentangle the functions of different modifications. These effects like the crosstalk pathways summarised earlier can then be tested *in vivo*.

These histone crosstalk pathways may regulate the recruitment and accumulation of many chromatin modifiers at target sites. The ability of chromatin modifiers to read pre-existing modifications may allow the maintenance of the transcriptional state even when the initial stimulus is gone, or following cell division when most nuclear proteins dissociate from chromatin. Furthermore the interplay between marks and chromatin modifiers of opposing functions may be important for the establishment of

bistable transcriptional states amenable to dynamic changes in gene expression during development.

The functionally of histone modifications is conferred in a large part by their ability to recruit effectors. To investigate the biological functions of Polycomb and Trithorax modifications, I performed a screen to identify readers of H3K27me3 and H3K36me3. Briefly, I found that H3K27me3 readers consisted mostly of proteins from the Polycomb family. Again, this suggests that self-propagation of Polycomb activity is an important aspect of Polycomb function. H3K36me3 readers consisted of proteins of various functions, many associated with DNA repair, and RNA-processing. Notably, I found a novel reader of H3K36me3, PWWP2A. PWWP2A may link H3K36me3 to NuRD mediated deacetylation at gene bodies in mammalian cells, an important pathway of histone crosstalk critical to transcription regulation well studied in budding yeast.

PWWP2A physically associates with NuRD subunits MTA1-3, HDAC1/2, RBBP4/7 under high salt conditions. It also co-purified to a lesser degree with other known NuRD binding proteins such as BEND3 and transcription factors ZFP423, ZFP521, TIF1ß, and SALL4. A putative NuRD interaction region in the N-terminal half of PWWP2A was identified but requires further characterisation.

The absence of MBD3, GATAD2A/B, CHD3/4 from the FS2-PWWP2A affinity purifications in HeLa and mESCs is very peculiar. Previous biochemical studies have shown that affinity purification of any core subunit of NuRD is sufficient to pull down all other core subunits; this has been observed for MTA2 (Zhang et al., 1999), HDAC1 (Zhang et al., 1999), MBD3 (Kloet et al., 2015), and the *Drosophila* homolog of RBBP4/7 (Zhang et al., 2016). Moreover, immunoprecipitation of accessory NuRD interactors like FOG1 (Low et al., 2016) and BEND3 (Saksouk et al., 2014) could also recover all core subunits of the NuRD complex. It was therefore surprising that no peptides of MBD2/3, GATAD2A/B, and CHD3/4 were identified from the FS2-PWWP2A immunoprecipitation. It is believed that MBD2/3 bridges the association between MTA1-3 and the deacetylase core with GATAD2A/B and CHD3/4. Binding of PWWP2A to MTA1/2/3-HDAC1/2-RRBP4/7 may occlude the association of this subcomplex with the rest of the NuRD subunits, though this is just speculative. Further biochemical or structural characterisation of the interactions within PWWP2A-NuRD are needed in order to establish this as a distinct type of complex.

PWWP2A-NuRD may represent a distinct NuRD complex that is specifically targeted to active gene bodies through PWWP2A's recognition of H3K36me3. From our genome-wide mapping of PWWP2A targets, we show

that PWWP2A is enriched at the gene bodies of highly transcribed genes. The gene body localisation of PWWP2A was dependent its PWWP domain, which is known to recognise H3K36me3 in other proteins. From our ChIP-seq analysis, we found that PWWP2A co-occupies many gene targets with HDAC1. NuRD targets genes can be difficult to identify because many components of the complex have independent functions or associate with other complexes. Ideally I would also like to obtain a ChIP-seq dataset for other NuRD subunits such as MTA1-3 and from the overlap, identify PWWP2A-NuRD specific targets.

PWWP2A may also have targets beyond H3K36me3 modified chromatin. Short isoforms of PWWP2A and PWWP2B lacking the PWWP domain are also expressed in multiple tissues similar to the long isoforms. The PWWP domain is the only annotated domain in the protein, but it's likely that other chromatin binding domains exist as well. PWWP2A was identified in a proteomics screen as an interactor of H2A.Z (Draker et al., 2012). H2A.Z is enriched at the promoters of active genes, and by ChIP-qPCR low levels of PWWP2A can be found at promoters. ChIP-seq of the PWWP2A short isoform should allow us to see H3K36me3-independent localisation on chromatin. From truncation studies in Chapter 3, the N-terminal half of PWWP2A was sufficient to mediate the interaction with NuRD subunits MTA1 and HDAC1. Consequently, the short isoform of PWWP2A should also be able to interact

with NuRD. This suggests that PWWP2A may have roles in NuRD function independent of H3K36me3 binding.

PWWP2A and PWWP2B deletions were generated in mESCs to study the biological functions of these proteins. So far, only simple characterisation of differentiation ability and effects on global acetylation levels have been performed. Loss of PWWP2A and PWWP2A/B does not appear to affect the pluripotency of mESCs, or their ability to undergo differentiation into embryoid bodies. I am now looking into the effect of PWWP2A and PWWP2B deletion on the localisation of NuRD subunits HDAC1 and MTA2 in mESCs.

The difficulty is that by ChIP the majority of HDAC1 and MTA2 peaks are found at promoter regions where they are most enriched. Thus differences in their gene body localisation may be difficult to discern given their low levels of binding at these regions. I have considered looking into the effect of PWWP2A/B deletion on gene body acetylation levels by ChIP as well. Global levels of histone acetylation do not increase on H3K36me3 containing nucleosomes in the absence of PWWP2A and PWWP2A/B, possibly pointing to other mechanisms of HDAC recruitment to these sites. Chromatin may also be accessible to HDAC activity generally, as the chromatin structure in mESCs has been reported to be less condensed than in differentiated cells (Gaspar-Maia et al., 2011). However, we hope calibrated ChIP-seq will enable

us to quantify differences in HDAC1, MTA2, or acetylation at a subset of PWWP2A target genes.

I would like to further examine the role of PWWP2A in transcriptional regulation. As discussed in the Introduction, H3K36me3 recruits many proteins that regulate the progression of Pol II by mediating histone disassembly and reassembly. As a protein that is enriched over active gene bodies, it is conceivable that PWWP2A may also function in some aspect of transcriptional elongation regulation. I am planning to perform ChIP-seq for Ser-5 and Ser-2 phosphorylated Pol II to see if Pol II dynamics on chromatin are effected by PWWP2A or PWWP2A/B deletion. Additionally, nascent 4-sU RNA-seq will be used to examine effects of PWWP2A and PWWP2A/B deletion on transcription directly. If loss of PWWP2A/B leads to more spurious transcription initiation events, we should see a greater number of non-coding transcripts and antisense transcripts from the RNA-seq in PWWP2A or PWWP2A/B null cells (Murray et al., 2015).

Although histone deacetylation has traditionally been associated with gene repression, the enrichment of HDAC1/2 proteins and other NuRD subunits at active genes has suggested a more dynamic role for these proteins in gene regulation. At the promoters of active genes, the turnover of acetylation is high. Acetylation removed by HDAC1/2 is quickly replaced by

HATs through the multiple pathways that reinforce HAT recruitment at active genes as discussed previously in the introduction. During the development and transition to new cell states, a subset of these genes will lose factors promoting their activation, and HDAC activity may mediate a shift to a more hypoacetylated state that promotes the accumulation of repressive factors such H3K27me3 and Polycomb. The interplay between NuRD and other chromatin modulators like Polycomb is an interesting area of study as NuRD has been found to work in concert with PRC2 to co-regulate the expression of a subset of developmentally important genes through mediating an H3K27 acetylation to methylation switch that promotes gene repression (Reynolds et al., 2012).

How NuRD is recruited to all its target sites genome-wide remains an important question in the field. Targeting of NuRD can be attributed in part to the diversity of complexes it is able to form. In mammalian organisms, there are two to three paralogs for all core NuRD subunits, which may confer different biological functions to the complex and target the complex to different sites on chromatin. Additionally, the association of NuRD with an array of transcriptional factors and chromatin binding proteins is also important for its recruitment to genomic targets in mESCs and differentiated cells. We believe that PWWP2A may be one of these factors.

In this work, I have identified a new interactor PWWP2A that may recruit NuRD to gene bodies enriched in H3K36me3. Elucidating the mechanisms behind PWWP2A-NuRD recruitment and its role at active gene bodies will contribute a small part to our understanding of NuRD and HDAC1/2 function in cellular biology.

MATERIALS AND METHODS

General Biochemistry

Nuclear extraction

Nuclear extraction was performed as described in (Dignam et al., 1983). Briefly, cells were harvested by using either trypsin or scraping. The pellet was resuspended in hypotonic buffer (10 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF, complete protease inhibitors) to allow cells to swell prior to lysis. Lysis of the cytoplasmic membrane was achieved using hypotnic buffer buffer containing 0.1% NP-40. Nuclei were washed and pelleted then extracted for 1 h on ice using high salt (5 mM Hepes pH 7.9, 26% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 300 – 350 NaCl/KCl). The nuclei was centrifuged at maximum speed at 4 ^{o}C, and the supernatant was taken as the nuclear extract. After quantification by bradford, the nuclear extract was snap frozen.

Flag immunoprecipitation

Flag affinity purifications for co-IP experiments were carried out using the anti-Flag M2 affinity gel. All co-IP experiments described were performed using nuclear extract. Nuclear extract was treated with benzonase prior to use to eliminate DNA-facilitated protein-protein interactions. 0.5 mg of

nuclear extract was incubated with 50 ul of beads for 3 h at 4 ºC. This was followed by 8 – 10 stringent washes in buffer containing (25 mM Tris pH 7.5, 300 mM NaCl, 0.2% NP-40, complete protease inhibitors). Samples were boiled in SDS-PAGE loading buffer for downstream analysis by western.

For the purification of Flag-tagged proteins, 10 ml of cellular lysate was incubated with 100 – 200 µl of anti-Flag M2 beads for 4 – 5 h at 4 degrees C. Samples were washed 3 – 5 times in buffer containing 25 mM Tris pH 7.5, 150 mM NaCl, 0.1 % NP40, 1 mM DTT, 1mM PMSF, complete protease inhibitors and eluted three times using 3×Flag peptides. The elutions were pooled and concentrated for clean-up over gel filtration. Everything was performed on ice.

Streptactin purification for Mass spec

For each mass spectrometry experiment, cells were collected from forty 15 cm diameter tissue culture dishes, this corresponded to around 2 - 3 x 10^9 cells. Nuclear extract was prepared in advance and treated with benzonase for 1 h just prior to use. 15 mg of nuclear extract was incubated with 50 µl of streptactin agarose beads for 4 h. The samples were then washed for 8 × 15 minutes in high salt wash buffer (20 mM Tris pH 8.0, 500 mM NaCl, 0.2% NP-40, 5% glycerol, 1 mM DTT, complete protease inhibitors), and finally eluted with 10 mM Desthiobiotin in 20 mM Tris pH 8.0, 150 mM NaCl, 0.2%

NP-40, 5% glycerol, 1 mM DTT, complete protease inhibitors. Samples were snap frozen and stored at -80.

Preparation of samples for mass spectrometry

IP samples were checked for quality by silverstain and western prior to processing for mass spectrometry. The sample was first reduced with 5 mM DTT for 30 minutes at room temperature then all cysteines were alkylated with 20 mM iodoacetamide for 30 minutes at room temperature. Proteins were precipitated by methanol/chloroform extraction, and the pellet was solubilised in denaturing 6 M urea buffer to disrupt protein folding. In-solution digestion was carried out using mass spec grade trypsin overnight shaking at 37 ºC. Trypsin was added in 1:50 trypsin to protein ratio. The digestion was stopped by the addition of formic acid, and the peptides were purified using SEP-PAK C18 cartridges. Briefly, the SEP-PAK C18 column was eliquibrated with 5 ml of Solution B (65% acetonitrile, 35 % milliQ water, 0.1% TFA), followed by 10 ml of Solution A (98% milliQ water, 2% acetonitrile, 0.1% TFA). The sample was then applied in to the column, and washed with 10 ml of solution A. Samples were eluted in Solution B, dried down in a speed-vac, then stored at – 80 ºC.

Western blot

Samples were transferred onto PVDF membranes, blocked in PBS + 5% milk. Primary antibodies were prepared in PBS usually at 1:1000 and incubated with the membrane overnight. Washing was done in PBS, followed by incubation with secondary antibodies at 1:2000 prepared in PBS + 5% milk for 1.5 h. Detection was achieved with the enhanced chemiluminescence method.

Separation of nuclear extract by size exclusion chromatography

To determine the co-migration of PWWP2A with NuRD components, nuclear extract was prepared from the E14 cell line B2 expressing FS2-PWWP2A. 0.3 mg nuclear extract was treated with benzonase for 1 h then centrifuged at maximum speed for 10 min at 4 ºC, then concentrated to 100 µl before loading on a Superose 6 10/300 column equilibriated with 20 mM Tris pH 7.5, 300mM NaCl, 2 mM $MgCl_2$, 5% glycerol, 1 mM PMSF, and complete protease inhibitors. 0.25 ml fractions were collected and precipitated with TCA. The protein pellets were resuspended and boiled in SDS-PAGE loading buffer, and every other fraction was loaded for western analysis.

Generation of Recombinant Nucleosomes

(adapted from (Dyer et al., 2004))

Histone purification

The plasmids for the expression of canonical *Xenopus laevis* histones H2A, H2B, H3, H4 were obtained from the group of Timothy Richmond. The pET-histone expression plasmids were expressed in BL21 DE3 pLysS *E. coli* cells, grown in 2×TY medium and induced at 0.5 mM IPTG. Histones were purified from inclusion bodies under denaturing conditions (7 M urea, 20 mM sodium acetate pH 5.2, 1 M NaCl, 5 mM 2-mercaptoethanol, 1 mM EDTA), over a HiPrep 26/60 Sephacryl S-200 HR column. Finally the purified histones were gradually refolded by dialyses against 3 × 4 L of water with 2 mM 2-mercaptoethanol at 4 ºC.

Generation of site specific methyllysine analogs

This was performed as in (Simon et al., 2007). Briefly, to introduce site-specific methyllysine analogs into H3, the lysine of interest was mutated to a cysteine. This method relies on the reactivity of cysteines, thus the only native cysteine in H3 at residue 110 is mutated to an alanine, leaving only one cysteine residue at the site of interest. The mutant H3 was expressed and purified as described above. Next, under reducing conditions, the mutant H3 was incubated with excess (2-bromoethyl) trimethylammonium bromide, resulting in the alkylation of the cysteine at the site of interest resulting in a trimethyllysine (Kme3) mimic. The final product differs from the native Kme3 only by the sulphur in the side chain. The reaction was monitored by

mass spectrometry, the efficiency of the reaction was usually near completion with total conversion of the reactive cysteine to the trimethyllysine analog.

Histone octamer assembly and purification

3mg of each histone H2A, H2B, H3 (unmodified or modified), H4 was unfolded in denaturing buffer (7 M guanidine HCl, 20 mM Tris pH 7.5, 10 mM DTT), then mixed together at equal ratios. The histones were then dialysed into 3 × 4 L refolding buffer containing high salt (2M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, 5 mM 2-mercaptoethanol) to allow the histones to slowly fold together. The dialysed sample was run over a HiLoad Superdex 200 16/60 column to separate the histone octamer from H3/H4 tetramers, H2A/H2B dimers, and individual histones. Octamers were stored at 4 ºC for short term use, or in 50% glycerol at - 20ºC for long term storage.

DNA purification

The plasmid containing the 2×601 Widom nucleosome positioning sequence was obtained from the lab of Jonathan Widom. The 2×601 sequence for nucleosome reconstitution was PCR amplified from the plasmid pBlu2SKP 2×601+48, which contains two 601 nucleosome positioning sequences separated by a 48 bp linker. Another version of the plasmid contained the 2×601 sequence without any linker. The PCR product was

purified by anion exchange in a 1ml Resource Q column, ethanol precipitated, and finally dissolved into milliQ water.

The 25×601 construct was obtained from the lab of Till Bartke. This pWM530 plasmid contained 25×601 sequences separated by 48bp linkers, and flanked by EcoRI restriction sites cutting at regular intervals along the backbone. Large amounts of plasmid was obtained through maxiprep, then digested using EcoRI. The 25×601 was separated from the small DNA fragments using PEG precipitation. 5% of PEG6000 was used to precipitate the 25×601 sequence while leaving the small digested products in solution. The pellet containing the desired fragment was washed and concentrated using ethanol precipitation before being finally resuspended in milliQ water.

Nucleosome reconstitution

Octamers and 2×601 DNA were mixed at 2:1 molar ratio to obtain dinucleosomes at high salt (2 M NaCl), and step wise dialysed into slow salt buffer (100 mM NaCl, 10 mM Tris pH 7.5) to allow octamers to fold and position correctly against the DNA. Proper reconstitution was visualised by running the free DNA and nucleosomes in a nondenaturing 0.8% TB gel, then staining with EtBr. The 2×601 runs at ~400 bp, and the dinucleosome runs between 800 – 1000 bp. Partially formed nucleosomes results in a smear of multiple bands from 400 – 800 bp.

To generate oligonucleosomes, octamers were mixed with the 25×601 sequence at varying amounts to determine the optimal ratio. The efficiency of nucleosome formation could be measured by running the samples containing different octamer:DNA ratios over a non-denaturing 0.8% TB gel. Too much octamer leads to DNA precipitation and loss of the DNA band entirely, thus the optimal ratio was taken to be the sample with highest octamer:DNA that did not precipitate.

Expression and purification of NSD2

Expression in Bacteria

$NSD2_{SET}$ was cloned into the pet14b vector containing an N-terminal 6×Histidine tag. $NSD2_{SET}$ was induced with 0.5 mM IPTG in BL21 DE3 pLysS *E. coli* cells for 3 h at 37 ºC. Cells were lysed and run over a His-trap column and the protein was eluted over a 10 mM – 500 mM imidazole gradient. $NSD2_{SET}$ was dialysed and concentrated into physiological buffer (20mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT).

Expression in Mammalian Cells

Three NSD2 constructs with different tags were cloned into a mammalian expression vector under a CAG promoter. HEK293T cells were

transiently transfected and harvested after 48 h. Nuclear extract was prepared and NSD2 protein was purified by pulling on either the Flag or Strep tags, and the crude eluate was further purified by gel filtration over a Superose 6 10/300 column.

Expression in Insect Cells

The Bac-to-Bac Baculovirus Expression system from Thermo Fisher Scientific was used to produce NSD2 from Sf9 insect cells. NSD2 was cloned into the pFastBac donor vector, that when transformed into the *E. coli* strain DH10BAC, introduced the recombinant protein into the recombinant bacmid containing the genes for baculovirus production. The bacmid was transfected into Sf9 insect cells and after 5 – 7 days, a small titre of virus designated as P1 was obtained. After two round of viral amplification, the high titre P3 virus was used for large-scale expression. Cells were infected at an MOI of around 1-2, and harvested 48 – 72 h post-infection. A Flag purification was performed followed by gel filtration on a Superose6 10/300 column. The purification was checked by western blotting against the Flag tag and by coomassie.

Enzyme assays

Histone methyltransferase assays

Reactions were set up in 25 µl, with 5 µg of nucleosomes, 0.5 – 1 µg of enzyme, and 2 µl ^{3}H labelled S-adenosyl methionine (final concentration 1 µM). PRC2 assays were performed in buffer containing 20 mM Tris pH 8.0, 4 mM EDTA, 1 mM PMSF, 0.5 mM DTT, while MLL, SET, and NSD2 assays were performed in 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM PMSF, 0.5 mM DTT. Reactions were allowed to proceed for 1-2 hours at 30 ºC, then boiled in SDS-PAGE loading buffer. To visualise the histones, the reaction was run on an 18% SDS-PAGE gel, coomassie stained, then incubated in an autoradiography enhancer solution En3hance (Perkin Elmer) for 40 minutes. The gel was dried under, and exposed on Kodak Biomax XAR film at – 80 ºC before developing. For quantification by liquid scintillation counting, histone bands were excised from the gel, and counted in Ultima Gold F liquid scintillation solution.

Histone ubiquitylation assays

Reactions were set up in 25 µl, with 5 µg of nucleosomes, 0.2 µg of E1 enzyme (UBE1 from Boston Biochem), 0.2 µg of E2 (UbcH5c from Enzo Life Sciences), 200 µM ATP (GE Healthcare), 1 µg of methylated ubiquitin, and 0.5 – 1 µg of the E3 ligase PRC1 complex, in 50 mM Tris pH 7.5, 5 mM $MgCl_2$, 1 mM DTT. The reaction was incubated for 1 h at 37 ºC, then stopped by

boiling in SDS-PAGE loading buffer. Samples were run on an 18% SDS-PAGE gel, and wet transferred onto PVDF. PRC1 activity was detected by immunoblot against H2AK119u1.

SILAC Nucleosome Affinity Purification

Labelling of HeLa cells

HeLa cells were cultured in RPMI medium minus arginine and lysine. This was supplemented with 10% dialysed fetal calf serum, and with 76 μg/ml arginine and 133 μg/ml lysine (either light R0K0 or heavy R10K8). Trypsin-free cell dissociation buffer from Invitrogen was used for passaging cells. Cells were passaged nine times to ensure full incorporation of heavy amino acids which was checked by mass spectrometry to be >98% heavy label.

Pulldowns

(Adapted from (Bartke et al., 2010))

Biotinylated nucleosomes were prepared fresh and checked for purify and correct histone stoichiometry. 20 μg of nucleosomes was conjugated to 75 μl Pierce streptavidin magnetic beads for 1 h, then incubated for 4 h with 0.5 mg of nuclear extract pretreated treated with benzonase. The paired pulldowns were performed as illustrated in Figure 2.2. For the H3K27me3 Forward Experiment, unmodified nucleosomes were incubated with light

extract, and H3K27me3 nucleosomes with heavy extract. In the H3K27me3 Reverse Experiment, unmodified nucleosomes were incubated with heavy extract and H3K27me3 nuclesomes with light extract. The same was done for the H3K36me3 Forward and Reverse Experiments. After incubation, samples were washed five times in the binding buffer 20 mM Hepes pH 7.5, 150 mM NaCl, 0.2 mM EDTA, 20% glycerol, 0.1% NP-40, 1 mM, and complete protease inhibitors, then boiled in SDS-PAGE loading buffer.

Mass spectrometry

Samples were SDS-PAGE gels to check the efficiency and quality of the pulldown. The paired samples were then mixed together for in solution trypsin digestion and mass spectrometry analysis. Samples were electrosprayed into an LTQ Orbitrap mass spectrometer, The spectra was acquired with 100 000 resolution, with the 15 most intense precursor ions fragmented for MS/MS analysis. The proteins were identified using the Mascot and MaxQuant computational platform for the quantification of the SILAC datasets.

Chromatin Immunoprecipitation

Preparation of Chromatin

For the preparation of chromatin for ChIP, 40 million E14 mESCs was used for each sample. Cells were crosslinked in 10 ml of PBS with 2 μM EGS or DSG for 45 minutes, followed by 15 minutes of 1% formaldehyde at 25 ºC rotating. The reaction was quenched with the addition of 1.5 ml of 1M glycine. The cells were lysed in 50 mM Hepes pH 7.9, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton × 100, complete protease inhibitors, and the nuclei was washed in 10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, complete protease inhibitors. The nuclei were then resuspended in 1 ml of 10 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % sodium deoxycholate, 0.5% N-lauryolsarcosine, complete protease inhibitors. Samples were sonicated at 4 ºC between 20 – 27 minutes 30 sec on, 30 sec off using the Diogenode Biorupter Pico. Soluble chromatin was isolated after centrifuging at maximum speed for 10 min at 4 ºC. 30 ul of sample was taken every time, reverse crosslinked, purified, and run on a gel to check for the sonication efficiency. The chromatin was then snap frozen and stored at – 80 ºC.

Chromatin immunoprecipitation

For each ChIP reaction, 100 μl chromatin (corresponding to 4 million cells) was used. The chromatin was first diluted into ChIP dilution buffer (1%

triton × 100, 1 mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl) then precleared with 20 µl protein A agarose beads blocked in yeast 1 mg/ml tRNA and 1 mg/ml BSA for 30 minutes. Following preclearing, 3 – 5 µg of antibody was added to the chromatin for overnight incubation at 4 ºC. The following day, samples were incubated with 20 µl of preblocked protein A agarose beads for 1 h at 4 ºC. The beads were then washed with 1 × 1ml of low salt buffer (0.1% SDS, 1% triton × 100, 2 mM EDTA, 20 mM Tris, 150 mM NaCl), high salt buffer (0.1 % SDS, 1% triton × 100, 2 mM EDTA, 20 mM Tris, 500 mM NaCl), lithium chloride buffer (250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris pH 8.0), and 2 × 1ml TE buffer. The immunoprecipitated chromatin was then eluted off the beads with 200 µl of elution buffer (1% SDS, 0.1 M $NaHCO_3$), vortexing, and shaking for 30 minutes at room temperature. The DNA was then reverse crosslinked overnight shaking at 65 ºC in 200 mM NaCl. The next day, the samples were treated with RNAse and proteinase K, and the DNA was purified using the Chip Clean and Concentrator kit from Zymo Research.

Library preparation and Sequencing

Samples for sequencing were quantified on a bioanalyzer and Qubit. 10 ng of each sample was used for the preparation of DNA libraries. Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina, and quantified using the Kapa Biosystems DNA library

quantification kit. Sequencing was performed on the Illumina NextSeq machine. 600 million 40 bp paired end reads were achieved.

Data analysis

ChIP-seq data for H3K36me3, H3K4me3, H3K27me3, H3K4me1, H3K27ac, H3K9ac, H3K9me3, Nanog, Oct4, and CTCF were obtained from the ENCODE Project, ChIP-seq data for chromatin remodellers were retrieved from GSE64825, ChIP-seq data for histone variants H2A.Z, H2A.Zac) were from GSE34483, RNA Pol II ChIO-seq data (Ser2P, Ser5P, Ser7P) were from GSE34520.

Sequencing reads were mapped to the mm10 reference genome with Bowtie (Langmead et al., 2009), and only unique read pairs were kept for further analysis. Both the ChIP-seq data generated and downloaded were listed in supplementary table SX. MACS2 (2.0.10) (Zhang et al., 2008) was employed to call the ChIP-seq peaks with the default parameters except (-f BAMPE -g mm –q 0.01). We defined the PWWP2A targets by overlapping the called peaks with the annotated gene body regions, resulting 1963 genes (2955 transcripts), while the HDAC1 targets were further defined if the peaks were fully within the 10 kb regions surrounding the transcriptional start site (TSS) in two cell lines used (i.e. 5985 genes in total). The metagene binding patterns and heatmaps were generated using DANPOS (Chen et al., 2013)

with the key parameters (--flank_up 5000 --flank_dn 5000 --region_size 20000 --bin_size 20 --excludeP 0.01). The bigwig tracks shown in UCSC Genome Browser were generated by script xbam2wig.py from https://github.com/guifengwei/glib, normalizing to one million mapped reads. HOMER was employed to analyse the occupancy differences for various histone modifications and transcription factors between PWWP2A targets and non-targets. Chromatin state segregation was performed with ChromHMM software (v1.11) (Ernst and Kellis, 2010, 2012) , consisting of 12 candidate chromatin state.

Cell culture techniques

Growing cells

E14 cells were grown feederless on 0.1% gelatin coated tissue culture dishes in DMEM high glucose media supplemented with 10% FCS, 1 × minimum essential media non-essential amino acids from Gibco, 1.8 mM L-glutamate, 1Pen/Strep, 55 μM 2-mercaptoethanol, and LIF. Cells were fed every 1-2 days and passaged once 70-80% confluency was reached.

HEK293 and HeLa cells were grown in DMEM high glucose media supplemented with 10% FCS, 1 × minimum essential media non-essential amino acids from Gibco, 1.8 mM L-glutamate, 1Pen/Strep, and 55 μM 2-

mercaptoethanol. Medium was changed every 2 days and cells were passaged at 90% confluency.

Differentiation into Embryoid Bodies

Wildtype and knockout E14 mESCs cells were trypsinised and counted. 1 million cells were seeded onto non-gelatinised 9 cm tissue culture plates in media without LIF. On the third day, colonies were gently blown off the plate with a pipette and transferred to bacterial Petri dishes. At this point, embryoid bodies started to form in suspension. The medium was changed every day until the 7th day. On Day 7, the embryoid bodies were plated back onto 9 cm tissue culture dishes.

Generating knockouts using CRISPR/Cas9

CRISPR guides were designed using CRISPR Design from crispr.mit.edu, and cloned into pSpCas9(BB)-2A-Puro (the Feng Zhang lab's vector used in Ran Nature Prot 2013). This is a mammalian expression vector containing Cas9 nuclease from *S. pyogenes*, sites for insertion of the guide DNA sequences, and puromycin selectable marker. For knockout of PWWP2A, three guides were chosen and were used in pairs to delete large sections of the genes as illustrated in Figure _____. For the knockout of PWWP2B, two guides were used to delete exon 2 which codes for the majority of the protein.

E14 mESCs in 6 well plates were transfected with 1.5 μg of each vector, and passaged into 9 cm dishes the following day. The cells were then grown in ESC media with 2 μg/ml puromycin for 48 h to select for transfected colonies. After a further 4 – 6 days of growth in ESC media without puromycin, colonies were picked and screen by PCR for the expected deletions.

REFERENCES

Achar, Y.J., Balogh, D., Neculai, D., Juhasz, S., Morocz, M., Gali, H., Dhe-Paganon, S., Venclovas, C., and Haracska, L. (2015). Human HLTF mediates postreplication repair by its HIRAN domain-dependent replication fork remodelling. Nucleic Acids Res *43*, 10277-10291.

Ali, M., Hom, R.A., Blakeslee, W., Ikenouye, L., and Kutateladze, T.G. (2014). Diverse functions of PHD fingers of the MLL/KMT2 subfamily. Biochim Biophys Acta *1843*, 366-371.

Alqarni, S.S., Murthy, A., Zhang, W., Przewloka, M.R., Silva, A.P., Watson, A.A., Lejon, S., Pei, X.Y., Smits, A.H., Kloet, S.L., *et al.* (2014). Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. J Biol Chem *289*, 21844-21855.

Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M., Bouchard, N., Lacoste, N., Utley, R.T., Gaudreau, L., *et al.* (2010). NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. J Biol Chem *285*, 15966-15977.

Anderson, A.E., Karandikar, U.C., Pepple, K.L., Chen, Z., Bergmann, A., and Mardon, G. (2011). The enhancer of trithorax and polycomb gene Caf1/p55 is essential for cell survival and patterning in Drosophila development. Development *138*, 1957-1966.

Araki, Y., Wang, Z., Zang, C., Wood, W.H., 3rd, Schones, D., Cui, K., Roh, T.Y., Lhotsky, B., Wersto, R.P., Peng, W., *et al.* (2009). Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity *30*, 912-925.

Arrigoni, R., Alam, S.L., Wamstad, J.A., Bardwell, V.J., Sundquist, W.I., and Schreiber-Agus, N. (2006). The Polycomb-associated protein Rybp is a ubiquitin binding protein. FEBS Lett *580*, 6233-6241.

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

Balboula, A.Z., Stein, P., Schultz, R.M., and Schindler, K. (2014). Knockdown of RBBP7 unveils a requirement of histone deacetylation for CPC function in mouse oocytes. Cell Cycle *13*, 600-611.

Balboula, A.Z., Stein, P., Schultz, R.M., and Schindler, K. (2015). RBBP4 regulates histone deacetylation and bipolar spindle assembly during oocyte maturation in the mouse. Biol Reprod *92*, 105.

Ballare, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., *et al.* (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat Struct Mol Biol *19*, 1257-1265.

Banach-Orlowska, M., Pilecka, I., Torun, A., Pyrzynska, B., and Miaczynska, M. (2009). Functional characterization of the interactions between endosomal adaptor protein APPL1 and the NuRD co-repressor complex. Biochem J *423*, 389-400.

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

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J Biol Chem *280*, 17732-17736.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470-484.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243-247.

Bauer, M., Trupke, J., and Ringrose, L. (2016). The quest for mammalian Polycomb response elements: are we there yet? Chromosoma *125*, 471-496.

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.

Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O.*, et al.* (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. EMBO J *30*, 2829-2842.

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.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell *38*, 179-190.

Blastyak, A., Hajdu, I., Unk, I., and Haracska, L. (2010). Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. Mol Cell Biol *30*, 684-693.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczynski, B., Riddell, A., and Furlong, E.E. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet *44*, 148-156.

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

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L.*, et al.* (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol *19*, 1273-1281.

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.

Bunch, H., Zheng, X., Burkholder, A., Dillon, S.T., Motola, S., Birrane, G., Ebmeier, C.C., Levine, S., Fargo, D., Hu, G.*, et al.* (2014). TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nat Struct Mol Biol *21*, 876-883.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D.*, et al.* (2013). An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Mol Cell *49*, 571-582.

Cammas, F., Herzog, M., Lerouge, T., Chambon, P., and Losson, R. (2004). Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. Genes Dev *18*, 2147-2160.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Carvalho, S., Raposo, A.C., Martins, F.B., Grosso, A.R., Sridhara, S.C., Rino, J., Carmo-Fonseca, M., and de Almeida, S.F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Res *41*, 2881-2893.

Chan, K.M., and Zhang, Z. (2012). Leucine-rich repeat and WD repeat-containing protein 1 is recruited to pericentric heterochromatin by trimethylated lysine 9 of histone H3 and maintains heterochromatin silencing. J Biol Chem *287*, 15024-15033.

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., Dent, S., He, X., and Li, W. (2013). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res *23*, 341-351.

Cheng, C., Yan, K.K., Yip, K.Y., Rozowsky, J., Alexander, R., Shou, C., and Gerstein, M. (2011). A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. Genome Biol *12*, R15.

Cheng, L.E., Zhang, J., and Reed, R.R. (2007). The transcription factor Zfp423/OAZ is required for cerebellar development and CNS midline patterning. Dev Biol *307*, 43-52.

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

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.

Crump, N.T., Hazzalin, C.A., Bowers, E.M., Alani, R.M., Cole, P.A., and Mahadevan, L.C. (2011). Dynamic acetylation of all lysine-4 trimethylated histone H3 is evolutionarily conserved and mediated by p300/CBP. Proc Natl Acad Sci U S A *108*, 7814-7819.

de Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Bou Dargham, D., Jounier, S., Humbertclaude, H., Ribierre, F., Baulard, C.*, et al.* (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. Nature *530*, 113-116.

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

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

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem *285*, 26114-26120.

Dhont, L., Mascaux, C., and Belayew, A. (2016). The helicase-like transcription factor (HLTF) in cancer: loss of function or oncomorphic conversion of a tumor suppressor? Cell Mol Life Sci *73*, 129-147.

Diao, Y., Guo, X., Li, Y., Sun, K., Lu, L., Jiang, L., Fu, X., Zhu, H., Sun, H., Wang, H., *et al.* (2012). Pax3/7BP is a Pax7- and Pax3-binding protein that regulates the proliferation of muscle precursor cells by an epigenetic mechanism. Cell Stem Cell *11*, 231-241.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Dong, X., Greven, M.C., Kundaje, A., Djebali, S., Brown, J.B., Cheng, C., Gingeras, T.R., Gerstein, M., Guigo, R., Birney, E., *et al.* (2012). Modeling gene expression using chromatin features in various cellular contexts. Genome Biol *13*, R53.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J Mol Biol *327*, 85-96.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Cote, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol *24*, 1884-1896.

Draker, R., Ng, M.K., Sarcinella, E., Ignatchenko, V., Kislinger, T., and Cheung, P. (2012). A combination of H2A.Z and H4 acetylation recruits Brd2 to chromatin during transcriptional activation. PLoS Genet *8*, e1003047.

Du, J.L., Zhang, S.W., Huang, H.W., Cai, T., Li, L., Chen, S., and He, X.J. (2015). The Splicing Factor PRP31 Is Involved in Transcriptional Gene Silencing and Stress Response in Arabidopsis. Mol Plant *8*, 1053-1068.

Duns, G., van den Berg, E., van Duivenbode, I., Osinga, J., Hollema, H., Hofstra, R.M., and Kok, K. (2010). Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. Cancer Res *70*, 4287-4291.

Durant, M., and Pugh, B.F. (2007). NuA4-directed chromatin transactions throughout the Saccharomyces cerevisiae genome. Mol Cell Biol *27*, 5327-5335.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol *375*, 23-44.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Mol Cell *49*, 368-378.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J *27*, 406-420.

Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. Proc Natl Acad Sci U S A *103*, 16319-16324.

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.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology *28*, 817-825.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

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.

Eymery, A., Callanan, M., and Vourc'h, C. (2009). The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. Int J Dev Biol *53*, 259-268.

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

Feng, Q., Cao, R., Xia, L., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2002). Identification and functional characterization of the p66/p68 components of the MeCP1 complex. Mol Cell Biol *22*, 536-546.

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

Fleming, A.B., Kao, C.F., Hillyer, C., Pikaart, M., and Osley, M.A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. Mol Cell *31*, 57-66.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. Cell *137*, 110-122.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574-1577.

Frangini, A., Sjoberg, M., Roman-Trufero, M., Dharmalingam, G., Haberle, V., Bartke, T., Lenhard, B., Malumbres, M., Vidal, M., and Dillon, N. (2013). The aurora B kinase and the

polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. Mol Cell *51*, 647-661.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res *31*, 2305-2312.

Gao, C., Dimitrov, T., Yong, K.J., Tatetsu, H., Jeong, H.W., Luo, H.R., Bradner, J.E., Tenen, D.G., and Chai, L. (2013). Targeting transcription factor SALL4 in acute myeloid leukemia by interrupting its interaction with an epigenetic complex. Blood *121*, 1413-1421.

Garcia-Ramirez, M., Rocchini, C., and Ausio, J. (1995). Modulation of chromatin folding by histone acetylation. J Biol Chem *270*, 17923-17928.

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

Gehani, S.S., Agrawal-Singh, S., Dietrich, N., Christophersen, N.S., Helin, K., and Hansen, K. (2010). Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation. Mol Cell *39*, 886-900.

Giri, S., Aggarwal, V., Pontis, J., Shen, Z., Chakraborty, A., Khan, A., Mizzen, C., Prasanth, K.V., Ait-Si-Ali, S., Ha, T.*, et al.* (2015). The preRC protein ORCA organizes heterochromatin by assembling histone H3 lysine 9 methyltransferases on chromatin. Elife *4*.

Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M., Wang, S.Z., Zu Zhu, S., Ginder, G.D., and Williams, D.C., Jr. (2011). p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. Proc Natl Acad Sci U S A *108*, 7487-7492.

Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev *25*, 2210-2221.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet *6*, e1000869.

Gupta, R.K., Arany, Z., Seale, P., Mepani, R.J., Ye, L., Conroe, H.M., Roby, Y.A., Kulaga, H., Reed, R.R., and Spiegelman, B.M. (2010). Transcriptional control of preadipocyte determination by Zfp423. Nature *464*, 619-623.

Harder, L., Puller, A.C., and Horstmann, M.A. (2014). ZNF423: Transcriptional modulation in development and cancer. Mol Cell Oncol *1*, e969655.

Hata, A., Seoane, J., Lagna, G., Montalvo, E., Hemmati-Brivanlou, A., and Massague, J. (2000). OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. Cell *100*, 229-240.

Hebbes, T.R., Clayton, A.L., Thorne, A.W., and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO J *13*, 1823-1830.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

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

Ho, J.W., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.A., Minoda, A., Tolstorukov, M.Y., Appert, A., *et al.* (2014). Comparative analysis of metazoan chromatin organization. Nature *512*, 449-452.

Hon, G.C., Hawkins, R.D., and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. Hum Mol Genet *18*, R195-201.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat Struct Mol Biol *20*, 1093-1097.

Hung, T., Binda, O., Champagne, K.S., Kuo, A.J., Johnson, K., Chang, H.Y., Simon, M.D., Kutateladze, T.G., and Gozani, O. (2009). ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. Mol Cell *33*, 248-256.

Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J., 3rd, and Farnham, P.J. (2011). Functional analysis of KAP1 genomic recruitment. Mol Cell Biol *31*, 1833-1847.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol Cell *20*, 971-978.

Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., and Cristea, I.M. (2013). The functional interactome landscape of the human histone deacetylase family. Mol Syst Biol *9*, 672.

Jung, H.R., Pasini, D., Helin, K., and Jensen, O.N. (2010). Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. Mol Cell Proteomics *9*, 838-850.

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

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., 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.

Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A *107*, 2926-2931.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G.A., Li, Z., Vedadi, M., Schapira, M., *et al.* (2011). Recognition and specificity determinants of the human cbx chromodomains. J Biol Chem *286*, 521-529.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., *et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

Khan, D.H., He, S., Yu, J., Winter, S., Cao, W., Seiser, C., and Davie, J.R. (2013). Protein kinase CK2 regulates the dimerization of histone deacetylase 1 (HDAC1) and HDAC2 during mitosis. J Biol Chem *288*, 16518-16528.

Kidder, B.L., and Palmer, S. (2012). HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. Nucleic Acids Res *40*, 2925-2939.

Kim, J., Hake, S.B., and Roeder, R.G. (2005). The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. Mol Cell *20*, 759-770.

Kim, T.W., Kang, B.H., Jang, H., Kwak, S., Shin, J., Kim, H., Lee, S.E., Lee, S.M., Lee, J.H., Kim, J.H., *et al.* (2015). Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active Embryonic Stem Cell Genes During Exit from Pluripotency. Stem Cells *33*, 2442-2455.

Kim, Y.O., Park, S.J., Balaban, R.S., Nirenberg, M., and Kim, Y. (2004). A functional genomic screen for cardiogenic genes using RNA interference in developing Drosophila embryos. Proc Natl Acad Sci U S A *101*, 159-164.

Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W., Berendsen, M., Niazi, H., Kops, G.J., and Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. FEBS J *282*, 1774-1785.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. PLoS Genet *9*, e1003717.

Kouskouti, A., and Talianidis, I. (2005). Histone modifications defining active genes persist after transcriptional and mitotic inactivation. EMBO J *24*, 347-357.

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

Krajewski, W.A., and Becker, P.B. (1998). Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. Proc Natl Acad Sci U S A *95*, 1540-1545.

Kurosawa, H. (2007). Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. J Biosci Bioeng *103*, 389-398.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lanzuolo, C., Lo Sardo, F., Diamantini, A., and Orlando, V. (2011). PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication. PLoS Genet *7*, e1002370.

Lau, P.N., and Cheung, P. (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. Proc Natl Acad Sci U S A *108*, 2801-2806.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021-1036.

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

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

Lee, B.K., Shen, W., Lee, J., Rhee, C., Chung, H., Kim, K.Y., Park, I.H., and Kim, J. (2015). Tgif1 Counterbalances the Activity of Core Pluripotency Factors in Mouse Embryonic Stem Cells. Cell Rep *13*, 52-60.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol *13*, 1192-1200.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. J Biol Chem *284*, 7970-7976.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell *153*, 590-600.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91-95.

Lickwar, C.R., Rao, B., Shabalin, A.A., Nobel, A.B., Strahl, B.D., and Lieb, J.D. (2009). The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency. PLoS One *4*, e4886.

Lin, J.R., Zeman, M.K., Chen, J.Y., Yee, M.C., and Cimprich, K.A. (2011). SHPRH and HLTF act in a damage-specific manner to coordinate different forms of postreplication repair and prevent mutagenesis. Mol Cell *42*, 237-249.

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R., Carlomagno, T., and Wahl, M.C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. Science *316*, 115-120.

Low, J.K., Webb, S.R., Silva, A.P., Saathoff, H., Ryan, D.P., Torrado, M., Brofelth, M., Parker, B.L., Shepherd, N.E., and Mackay, J.P. (2016). CHD4 Is a Peripheral Component of the Nucleosome Remodeling and Deacetylase Complex. J Biol Chem *291*, 15853-15866.

Lu, J., Jeong, H.W., Kong, N., Yang, Y., Carroll, J., Luo, H.R., Silberstein, L.E., Yupoma, and Chai, L. (2009). Stem cell factor SALL4 represses the transcriptions of PTEN and SALL1 through an epigenetic repressor complex. PLoS One *4*, e5577.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996-1000.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luo, Y., Jian, W., Stavreva, D., Fu, X., Hager, G., Bungert, J., Huang, S., and Qiu, Y. (2009). Trans-regulation of histone deacetylase activities through acetylation. J Biol Chem *284*, 34901-34910.

Luo, Z., Gao, X., Lin, C., Smith, E.R., Marshall, S.A., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. Mol Cell *57*, 685-694.

MacKay, C., Toth, R., and Rouse, J. (2009). Biochemical characterisation of the SWI/SNF family member HLTF. Biochem Biophys Res Commun *390*, 187-191.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Marino, S., and Nusse, R. (2007). Mutants in the mouse NuRD/Mi2 component P66alpha are embryonic lethal. PLoS One *2*, e519.

Mesuraca, M., Chiarella, E., Scicchitano, S., Codispoti, B., Giordano, M., Nappo, G., Bond, H.M., and Morrone, G. (2015). ZNF423 and ZNF521: EBF1 Antagonists of Potential Relevance in B-Lymphoid Malignancies. Biomed Res Int *2015*, 165238.

Millard, C.J., Varma, N., Saleh, A., Morris, K., Watson, P.J., Bottrill, A.R., Fairall, L., Smith, C.J., and Schwabe, J.W. (2016). The structure of the core NuRD repression complex provides insights into its interaction with chromatin. Elife *5*, e13941.

Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C.V., Fairall, L., and Schwabe, J.W. (2013). Class I HDACs share a common mechanism of regulation by inositol phosphates. Mol Cell *51*, 57-67.

Miller, A., Ralser, M., Kloet, S.L., Loos, R., Nishinakamura, R., Bertone, P., Vermeulen, M., and Hendrich, B. (2016). Sall4 controls differentiation of pluripotent cells independently of the Nucleosome Remodelling and Deacetylation (NuRD) complex. Development *143*, 3074-3084.

Mills, A.A. (2010). Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins. Nat Rev Cancer *10*, 669-682.

Mimoto, M.S., and Christian, J.L. (2012). Friend of GATA (FOG) interacts with the nucleosome remodeling and deacetylase complex (NuRD) to support primitive erythropoiesis in Xenopus laevis. PLoS One *7*, e29882.

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol *10*, 483-488.

Moazed, D., and O'Farrell, P.H. (1992). Maintenance of the engrailed expression pattern by Polycomb group genes in Drosophila. Development *116*, 805-810.

Montgomery, R.L., Davis, C.A., Potthoff, M.J., Haberland, M., Fielitz, J., Qi, X., Hill, J.A., Richardson, J.A., and Olson, E.N. (2007). Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. Genes Dev *21*, 1790-1802.

Morishita, M., and di Luccio, E. (2011). Cancers and the NSD family of histone lysine methyltransferases. Biochim Biophys Acta *1816*, 158-163.

Murray, S.C., Haenni, S., Howe, F.S., Fischl, H., Chocian, K., Nair, A., and Mellor, J. (2015). Sense and antisense transcription are associated with distinct chromatin architectures across genes. Nucleic Acids Res *43*, 7823-7837.

Murton, B.L., Chin, W.L., Ponting, C.P., and Itzhaki, L.S. (2010). Characterising the binding specificities of the subunits associated with the KMT2/Set1 histone lysine methyltransferase. J Mol Biol *398*, 481-488.

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.E., Hong, Z., Allen, C., Roy, S., Nunez, J.K., Nickoloff, J.*, et al.* (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol *19*, 1266-1272.

Nair, S.S., Li, D.Q., and Kumar, R. (2013). A core chromatin remodeling factor instructs global chromatin signaling through multivalent reading of nucleosome codes. Mol Cell *49*, 704-718.

Nakagawa, T., Kajitani, T., Togo, S., Masuko, N., Ohdan, H., Hishikawa, Y., Koji, T., Matsuyama, T., Ikura, T., Muramatsu, M.*, et al.* (2008). Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. Genes Dev *22*, 37-49.

Nguyen, A.T., and Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. Genes Dev *25*, 1345-1358.

O'Geen, H., Squazzo, S.L., Iyengar, S., Blahnik, K., Rinn, J.L., Chang, H.Y., Green, R., and Farnham, P.J. (2007). Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. PLoS Genet *3*, e89.

O'Shaughnessy-Kirwan, A., Signolet, J., Costello, I., Gharbi, S., and Hendrich, B. (2015). Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. Development *142*, 2586-2597.

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

Ohta, S., Wood, L., Toramoto, I., Yagyu, K., Fukagawa, T., and Earnshaw, W.C. (2015). CENP-32 is required to maintain centrosomal dominance in bipolar spindle assembly. Mol Biol Cell *26*, 1225-1237.

Okitsu, C.Y., and Hsieh, C.L. (2007). DNA methylation dictates histone H3K4 methylation. Mol Cell Biol *27*, 2746-2757.

Pai, C.C., Deegan, R.S., Subramanian, L., Gal, C., Sarkar, S., Blaikley, E.J., Walker, C., Hulme, L., Bernhard, E., Codlin, S., *et al.* (2014). A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. Nat Commun *5*, 4091.

Park, D., Shivram, H., and Iyer, V.R. (2014). Chd1 co-localizes with early transcription elongation factors independently of H3K36 methylation and releases stalled RNA polymerase II at introns. Epigenetics Chromatin *7*, 32.

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., *et al.* (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Res *38*, 4958-4969.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.

Pemberton, H., Anderton, E., Patel, H., Brookes, S., Chandler, H., Palermo, R., Stock, J., Rodriguez-Niedenfuhr, M., Racek, T., de Breed, L., *et al.* (2014). Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. Genome Biology *15*.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., *et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell *12*, 1577-1589.

Petruk, S., Sedkov, Y., Smith, S., Tillib, S., Kraevski, V., Nakamura, T., Canaani, E., Croce, C.M., and Mazo, A. (2001). Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. Science *294*, 1331-1334.

Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., La Thangue, N.B., Porter, A.C., *et al.* (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. Cell Rep *7*, 2006-2018.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genet *8*, e1002717.

Qiu, C., Sawada, K., Zhang, X., and Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol *9*, 217-224.

Ramirez, J., Dege, C., Kutateladze, T.G., and Hagman, J. (2012). MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes. Mol Cell Biol *32*, 5078-5088.

Ren, X., Vincenz, C., and Kerppola, T.K. (2008). Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. Mol Cell Biol *28*, 2884-2895.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012). NuRD-mediated deacetylation of H3K27

facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J *31*, 593-605.

Richly, H., Aloia, L., and Di Croce, L. (2011). Roles of the Polycomb group proteins in stem cells and cancer. Cell Death Dis *2*, e204.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223-232.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75-80.

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

Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.E.*, et al.* (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol Cell *56*, 580-594.

Saksouk, N., Simboeck, E., and Dejardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*, 3.

Sathyan, K.M., Shen, Z., Tripathi, V., Prasanth, K.V., and Prasanth, S.G. (2011). A BEN-domain-containing protein associates with heterochromatin and represses transcription. J Cell Sci *124*, 3149-3163.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

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.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev *18*, 1251-1262.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol *12*, 799-814.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev *16*, 919-932.

Shen, Z., Sathyan, K.M., Geng, Y., Zheng, R., Chakraborty, A., Freeman, B., Wang, F., Prasanth, K.V., and Prasanth, S.G. (2010). A WD-repeat protein stabilizes ORC binding to chromatin. Mol Cell *40*, 99-111.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., *et al.* (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96-99.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G., Fingerman, I.M., Briggs, S.D., *et al.* (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J Biol Chem *282*, 2450-2455.

Shimbo, T., Du, Y., Grimm, S.A., Dhasarathy, A., Mav, D., Shah, R.R., Shi, H., and Wade, P.A. (2013). MBD3 localizes at promoters, gene bodies and enhancers of active genes. PLoS Genet *9*, e1004028.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Siggens, L., Cordeddu, L., Ronnerblad, M., Lennartsson, A., and Ekwall, K. (2015). Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions. Epigenetics Chromatin *8*, 4.

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, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. Cell *128*, 1003-1012.

Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem *280*, 41789-41792.

Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. Elife *3*, e02042.

Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. Nat Struct Mol Biol *19*, 884-892.

Spruijt, C.G., Bartels, S.J., Brinkman, A.B., Tjeertes, J.V., Poser, I., Stunnenberg, H.G., and Vermeulen, M. (2010). CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. Mol Biosyst *6*, 1700-1706.

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.

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.

Struhl, G. (1981). A gene product required for correct initiation of segmental determination in Drosophila. Nature *293*, 36-41.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol *20*, 267-273.

Sun, Z.W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Taddei, A., Maison, C., Roche, D., and Almouzni, G. (2001). Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. Nat Cell Biol *3*, 114-120.

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.

Taverna, S.D., Ilin, S., Rogers, R.S., Tanny, J.C., Lavender, H., Li, H., Baker, L., Boyle, J., Blair, L.P., Chait, B.T.*, et al.* (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. Mol Cell *24*, 785-796.

Tie, F., Banerjee, R., Saiakhova, A.R., Howard, B., Monteith, K.E., Scacheri, P.C., Cosgrove, M.S., and Harte, P.J. (2014). Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. Development *141*, 1129-1139.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131-3141.

Todd, M.A., and Picketts, D.J. (2012). PHF6 interacts with the nucleosome remodeling and deacetylation (NuRD) complex. J Proteome Res *11*, 4326-4337.

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

Tsai, R.Y., and Reed, R.R. (1998). Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. Mol Cell Biol *18*, 6447-6456.

Tse, C., Sera, T., Wolffe, A.P., and Hansen, J.C. (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Mol Cell Biol *18*, 4629-4638.

van den Boom, V., Maat, H., Geugien, M., Rodriguez Lopez, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W., Yuan, H.*, et al.* (2016). Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep *14*, 332-346.

Vasquez-Doorman, C., and Petersen, C.P. (2016). The NuRD complex component p66 suppresses photoreceptor neuron regeneration in planarians. Regeneration (Oxf) *3*, 168-178.

Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., and Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature *489*, 452-455.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., *et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

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

Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., *et al.* (2009a). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell *138*, 660-672.

Wang, Z., Song, J., Milne, T.A., Wang, G.G., Li, H., Allis, C.D., and Patel, D.J. (2010). Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. Cell *141*, 1183-1194.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009b). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell *138*, 1019-1031.

Warming, S., Rachel, R.A., Jenkins, N.A., and Copeland, N.G. (2006). Zfp423 is required for normal cerebellar development. Mol Cell Biol *26*, 6913-6922.

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science *340*, 195-199.

Watson, A.A., Mahajan, P., Mertens, H.D., Deery, M.J., Zhang, W., Pham, P., Du, X., Bartke, T., Zhang, W., Edlich, C., *et al.* (2012). The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. J Mol Biol *422*, 3-17.

Weidenhammer, E.M., Singh, M., Ruiz-Noriega, M., and Woolford, J.L., Jr. (1996). The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccharomyces cerevisiae. Nucleic Acids Res *24*, 1164-1170.

Willemsen, M.H., Nijhof, B., Fenckova, M., Nillesen, W.M., Bongers, E.M., Castells-Nobau, A., Asztalos, L., Viragh, E., van Bon, B.W., Tezel, E., *et al.* (2013). GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. J Med Genet *50*, 507-514.

Wu, L., Lee, S.Y., Zhou, B., Nguyen, U.T., Muir, T.W., Tan, S., and Dou, Y. (2013a). ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. Mol Cell *49*, 1108-1120.

Wu, X., Johansen, J.V., and Helin, K. (2013b). Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol Cell *49*, 1134-1146.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

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

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498-1510.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T.*, et al.* (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971-975.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem *286*, 7983-7989.

Zhang, W., Aubert, A., Gomez de Segura, J.M., Karuppasamy, M., Basu, S., Murthy, A.S., Diamante, A., Drury, T.A., Balmer, J., Cramard, J.*, et al.* (2016). The Nucleosome Remodeling and Deacetylase Complex NuRD Is Built from Preformed Catalytically Active Sub-modules. J Mol Biol *428*, 2931-2942.

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

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

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

Zheng, Y., Sweet, S.M., Popovic, R., Martinez-Garcia, E., Tipton, J.D., Thomas, P.M., Licht, J.D., and Kelleher, N.L. (2012). Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. Proc Natl Acad Sci U S A *109*, 13549-13554.

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.

APPENDIX

Sequence alignment between human protein PWWP2A and PWWP2B

```
CLUSTAL O(1.2.4) multiple sequence alignment

sp|Q96N64|PWP2A_HUMAN      MAAVAAEAAATAASPGEGGAGEAEPEMEPIPGSEAGTDPLPVTATEASVPDGETDGQQSA
sp|Q6NUJ5|PWP2B_HUMAN      ------------------------------------------------------------

sp|Q96N64|PWP2A_HUMAN      PQADEPPLPPPPPPPGELARSPEAVGPELEAEEKLSVRVAESAAAAPQGGPELPPSPASP
sp|Q6NUJ5|PWP2B_HUMAN      ------------------------------------------------------------

sp|Q96N64|PWP2A_HUMAN      PEQPPAPEEREEPPLPQPVAPALVPPAGGDSTVSQLIPGSEVRVTLDHIIEDALVVSFRF
sp|Q6NUJ5|PWP2B_HUMAN      ---------------------------------MEPRAGCRLPVRVEQVVNGALVVTVSC
                                                              :   *..: * ::::::.****:.

sp|Q96N64|PWP2A_HUMAN      GEKLFSGVLMDLSKRFGPHGIPVTVF-PKREY---KDKPEAMPL--QSNTFQEGTEVKCE
sp|Q6NUJ5|PWP2B_HUMAN      GERSFAGILLDCTKKSGLFGLPPLAPLPQVDESPVNDSHGRAPEEGDAEVMQLGSSSPPP
                           **: *:*:*:* :*: * .*:*  .  *: :    :*.    *   :::.:* *:.

sp|Q96N64|PWP2A_HUMAN      ANGAVP-DDPSPVPHPELSLAESLWTSKPPPLFHEGAPYPPPLFIRDTYNQSIPQPPPRK
sp|Q6NUJ5|PWP2B_HUMAN      ARGVQPPETTRPEPPPPLVPPLPAGSLPPYPPYFEGAPFPHPLWLRDTYKLWVPQPPPRT
                           *.*. * :   * * * *       :  * * :.****:* **::****:  :******.

sp|Q96N64|PWP2A_HUMAN      IKRPKRKMYREE--PTSIMNAIKLRPRQVLCDKCKNSVVAEKKEIRK----------GSS
sp|Q6NUJ5|PWP2B_HUMAN      IKRTRRRLSRNRDPGRLILSTIRLRPRQVLCEKCKSTLSPPEASPGPPAAPRARRRLGSG
                           *** :*:: *:.     *:.:*:********:***.::   : .            **.

sp|Q96N64|PWP2A_HUMAN      ATDSSKYEDKKRRNE-SVTTVNKKLKTDHKVDGKNQNESQKRNAVVKVSN-IAHSRGRVV
sp|Q6NUJ5|PWP2B_HUMAN      PDRELRKPEEPENGEPTAAATARRSKRERREEDRAPAEQVPRSPVIKISYSTPQGKGEVV
                              . :  :: ...* :.::. :: * ::: :.:   *.  *. *:*:*    :.:*.**

sp|Q96N64|PWP2A_HUMAN      KVSAQANTSKAQLSTKKVLQSKNMDHAKAREVLKIAKEKAQKKQNETSTSKNAHSKVHFT
sp|Q6NUJ5|PWP2B_HUMAN      KIPSRVHGSLEPFRPQQAPQ---DDGSQDPEVLDRES--RD---RPSCAPSASIPKLKLT
                           *: ::.: *   :  ::. *    * ::  ***.  .   :   . :.: . :  *:::*

sp|Q96N64|PWP2A_HUMAN      RRYQNPSSGSLPPRVRLKPQRYRNEENDSSLKTGLEKMRSGKMAPKPQSRCTSTRSAGEA
sp|Q6NUJ5|PWP2B_HUMAN      RP-VPAGADLPPPKIRLKPHRLGDSEHEPVYRAELVGELNGYLRDSSPAPCADGPAGGLA
                           *     .:.  **::****:*  :.*::   :: *    .* :  .  : *:.  :.* *

sp|Q96N64|PWP2A_HUMAN      PSENQSPSKGPEEASSEVQDTNEVHVPGDQDEP--QTLGKKGSKNNISVYMTLNQKKSDS
sp|Q6NUJ5|PWP2B_HUMAN      DL------------SSGS--------SGEDDDFKSCPQGPQ-GREGLAFLVSCPEGRADC
                                        **          *::*:      * : .::.::. ::  : ::*.

sp|Q96N64|PWP2A_HUMAN      SSASVCSIDSTDDLKSSNSECSSSESFDFPPGSMHAPSTSSTSSSSKEEKKLSNSLKMKV
sp|Q6NUJ5|PWP2B_HUMAN      ASESACSSDSLDEARSSGSEGTPADTGDLSPGHGA-----SAPSVSREARQTVPPLTVRL
                           :* *.** ** *: :**.** : ::: *: **        *: * *:* ::    *.:::

sp|Q96N64|PWP2A_HUMAN      FSKNVSKCVTPDGRTICVGDIVWAKIYGFPWWPARILTITVSRKDNGLLVRQEARISWFG
sp|Q6NUJ5|PWP2B_HUMAN      HTQSVSECITEDGRTVAVGDIVWGKIHGFPWWPARVLDISLGQKEDGEPSWREAKVSWFG
                           .::.**:*:* ****:.******.**:********:* *::.:*::*    :**::****

sp|Q96N64|PWP2A_HUMAN      SPTTSFLALSQLSPFLENFQSRFNKKRKGLYRKAITEAAKAAKQLTPEVRALLTQFET
sp|Q6NUJ5|PWP2B_HUMAN      SPTTSFLSISKLSPFSEFFKLRFNRKKKGMYRKAITEAANAARHVAPEIRELLTQFET
                           *******::*:**** * *: ***:*:**:**********:**::::**:* *******
```