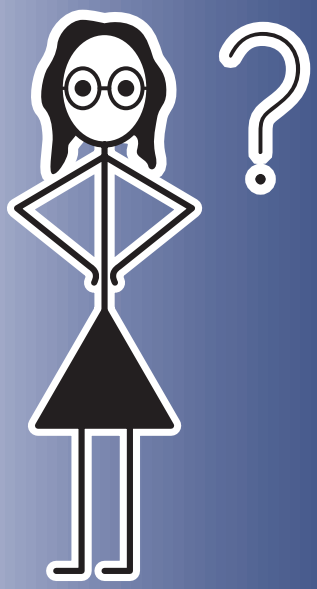


Chromatin Regulatory Complexes from an Interaction Proteomics Perspective

H. İrem Baymaz

Chromatin Regulatory Complexes from an Interaction Proteomics Perspective

H. İrem Baymaz

printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

Chromatin Regulatory Complexes from an Interaction Proteomics Perspective

Chromatine-Regulerende Complexen
vanuit een Interactie-Proteomics Perspectief

Proefschrift

ter verkrijging van de graad van doctor aan
de Radboud Universiteit Nijmegen op gezag
van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen in het openbaar te verdedigen
op donderdag 26 november 2015 om 12.30 uur precies

door

Hamdiye İrem Baymaz

geboren op 8 maart 1986 te Ankara, Turkije

Promoter: Prof. dr. M. Vermeulen

Manuscriptcommissie:
Prof. dr. J.H.L.M. van Bokhoven
Prof. dr. H.T.M. Timmers (UMCU)
Dr. L.M. Kamminga

Anne ve Babama.

Table of Contents

List of Abbreviations

ATP: Adenosine Triphosphosphate
cDNA: complementary DNA
CpGI: Cytosine-Guanine Island
DDR: DNA Damage Response
DNA: Deoxyribonucleic Acid
GTF: General Transcription Factor
ICR: Imprint Control Region
KDM: Lysine-specific Demethylase
LMR: Low Methylated Region
MBD: Methyl-CpG-binding domain
meCpG: methylated Cytosine-Guanine linear dinucleotide
NuRD: Nucleosome Remodeling and Histone Deacetylation
O-GlcNAc: O-linked beta-N-acetylglucosamine
PcG: Polycomb Group
PGC: Primordial Germ Cell
PIC: Preinitiation Complex
PPI: Protein-Protein Interaction
PR-Dub: Polycomb Repressive Deubiquitinase
PRC: Polycomb Repressive Complex
PRE: Polycomb Response Element
PTM: Post-Translational Modification
RNA pol II: RNA polymerase II
RNA: Ribonucleic Acid
TF: Transcription Factor
TrxG: Trithorax Group
TSS: Transcription Start Site
WGBS: Whole Genome Bisulphite Sequencing

Chapter 1

Introduction

"'But for what purpose was the earth formed?' asked Candide.
'To drive us mad.' replied Martin."

Voltaire, *Candide* (1759)

Basic concepts in molecular biology in relation to epigenetics

Deoxyribonucleic acid (DNA) is the hereditary material in all domains of life except for the cases of some obligate parasites (**figure 1**)[1]. DNA is made up of nucleotides that consist of a phosphate group, a sugar (deoxyribose), and either of the 4 bases; adenine, guanine, cytosine or thymine. The double helix structure consists of the phosphate backbone facing outward while the bases form hydrogen bonds (A basepairing with T, and G with C) in-between the 2 antiparallel strands. DNA is able to inherit all the information it carries by replication where it serves as the template.

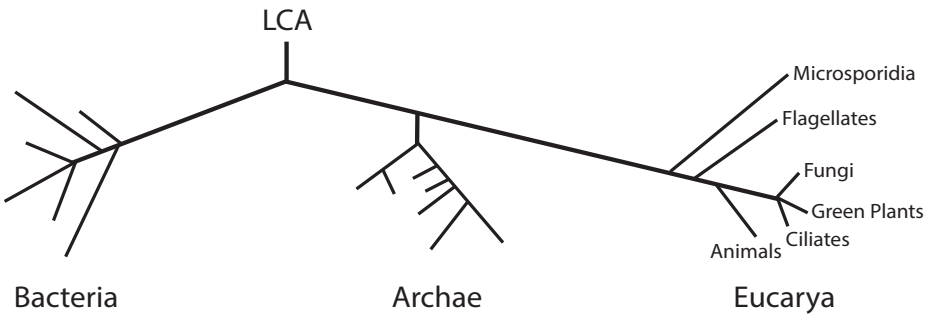


Figure 1: Representative sketch of the tree of life. The 3 domains of life are depicted as Bacteria, Archaea and Eucarya, originating from the last common ancestor (LCA). It is still a matter of debate whether and where the obligate parasites with an RNA genome belong in this picture[2], especially considering the predominant view in the field that the LCA had DNA as its genetic material.[3] Figure adapted from Woese et al, 1990;[4] however the branch lengths do not represent any phylogenetic information. All model organisms mentioned in this thesis belong to the Eucarya domain: *Saccharomyces cerevisiaea* in Fungi; *Arabidopsis thaliana* in Plants; *Caernohabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* in Animals (metazoans).

The central dogma of molecular biology posits that information stored on DNA is passed on to ribonucleic acid (RNA) molecules in a process called transcription[5]. RNA molecules are further 'decoded' into functional agents called proteins via a process called translation. Although the flow of information was originally postulated to be unidirectional, nature has presented exceptions to the rule[6,7]. A gene is the structural unit encoding for an RNA molecule[8]. Messenger RNAs are translated into proteins, while ribosomal RNAs and transfer RNAs are not; but they are required for the process of translation. More recently discovered types of RNA, such as long non-coding RNAs and microRNAs, expand the group of untranslated and functional RNAs, playing important roles in modulating transcription or translation processes[9–11]. The transcriptome of a cell refers to all the genes that are transcribed (active, expressed) and the proteome refers to all the proteins that are translated and folded correctly into a functionally capable structure.

A cell is the most basic unit of life and, all organisms, however complex they may be in the end, arise from a single cell[1]. More specifically, focusing on the mammalian development, the fertilized egg (zygote) transitions into the blastocyst with an inner cell mass (ICM). The cells within the ICM then give rise to the three germ layers that form all the subsequent organs and tissues of the organism. These germ layers are the endoderm (forming e.g. lungs, liver, gut), the mesoderm (muscular system, heart, blood) and the ectoderm (nervous

system, skin)[12]. A cell with the potential to give rise to a whole organism is called totipotent. A cell with the potential to give rise to any cell type of an organism is called pluripotent. From pluripotent cells, progenitors of different tissues are derived which have the potential to both self-renew (i.e. proliferate) and further specialize (i.e. differentiate) into the different cell types of specific tissues (**figure 2**). In the end, there are around 250 different types of cells in a human[12]. Starting from a single cell, this level of differentiation is impressive, especially considering the fact that all the different cells have the same DNA. However, it is well-established that besides some common genes (the housekeeping genes) that are required for the basic cellular structures and functions, the transcriptomes and hence the proteomes of different cells types vary remarkably[12].

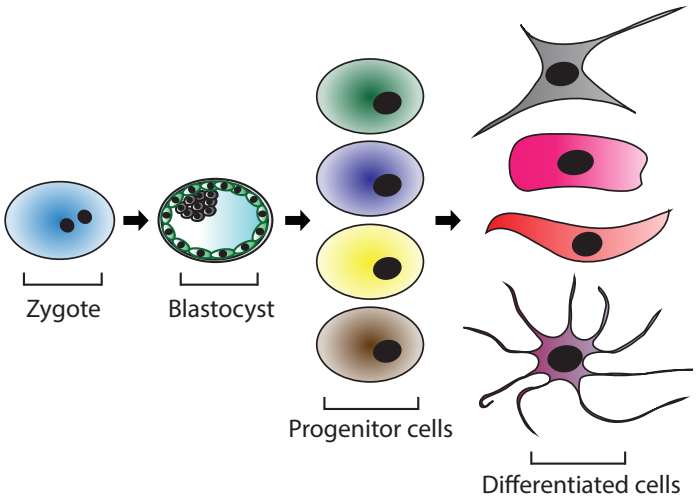


Figure 2: A simplified sketch of cellular differentiation. The zygote transitions into the blastocyst. Cells within the inner cell mass are pluripotent. Progenitors capable of self-renewal and differentiation further differentiate into tissue-specific cells.

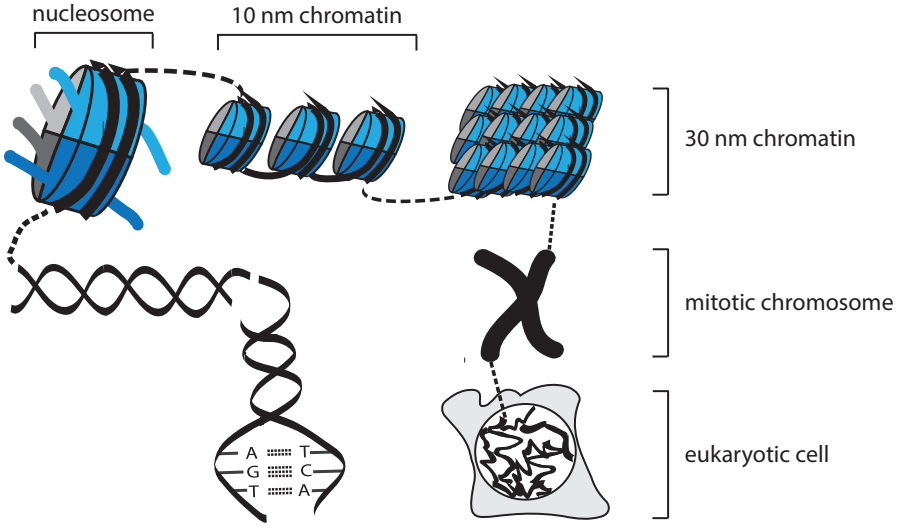
A defining property of the eukaryotic family of organisms is that DNA is not free-floating in the cell, but confined to the nucleus. Due to the sheer size of DNA (e.g. 2 meters when all human chromosomes are aligned end to end), it's compacted thoroughly by being wrapped around 4 types of histone proteins that, in pairs, make up an octamer. DNA wrapped around this octamer of histone proteins is called the nucleosome (**box 1** and **figure 3**). The strands of DNA that enter and exit the nucleosome structure are stabilized by another histone protein, Histone H1, also called the linker histone. Nucleosomes, nucleosome-associated proteins and long noncoding RNAs are collectively called chromatin[1].

Early visualization studies yielded the 'beads-on-a-string' image of DNA wrapped around nucleosomes which represents the 10 nanometer (nm) chromatin fiber. Images with a 30 nm diameter show stacks of polynucleosomes in a higher order chromatin structure, although the exact organization of nucleosomes at this stage and beyond need to be resolved[13–15]. It is also known that chromosomes are not in a state of uniform compaction in interphase (the longest phase of the cell cycle). Highly dense regions of chromatin are stained more strongly by basic dyes and called heterochromatin while the less dense, less intensely

stained regions are called euchromatin. These classifications predate the discovery of the structure of DNA and nucleosomes. Amazingly, the functional implications articulated at the time hold true to this day, i.e. the euchromatic regions are more transcriptionally active than the heterochromatic regions[16].

Initially considered a structural requirement for the compaction of DNA, the functional involvement of nucleosomes in all DNA-templated processes became clear soon enough. First of all, chromatin structure is refractory to all processes that require access to DNA such as DNA replication, DNA repair and transcription, hence chromatin manipulation prior to any of these processes is essential[15]. Secondly, in addition to acting as a structural barrier between DNA and DNA-binding proteins, the components of nucleosomes and various modifications they hold affect DNA accessibility and recruitment of factors that alter accessibility[17,18].

Figure 3: DNA is found in the context of chromatin in eukaryotic cells. The double helix is held together by the hydrogen bonds between adenine-thymine, and guanine-cytosine. DNA is wrapped around histone octamers (beads-on-a-string vision in 10 nm) and nucleosomes are further compacted (30 nm) in the nucleus. Mitotic chromosomes represent the most compact form of chromatin. Humans have 46 chromosomes.



The proteins that can manipulate chromatin are divided into 2 groups: Chromatin remodelers and chromatin modifiers. The first group of proteins can slide or evict nucleosomes using ATP, hence changing the nucleosome density and directly affecting DNA accessibility[19,20]. The second group of proteins catalyzes covalent modifications on DNA or the histones, capable of affecting the electrostatic interaction between DNA and histones as well as presenting binding platforms for other proteins[21]. The histone tails (N- and C-terminal) that protrude from the nucleosome provide an ideal platform for chromatin modifier activity. The modifiers that add a functional group to chromatin are called writers, those that remove the mark are erasers; and the proteins that bind to the marks are called readers. There are a vast number of chromatin modifications present in a eukaryotic cell, some of the most

common ones being DNA methylation, histone acetylation, phosphorylation, methylation and ubiquitination[22](**figure 4**). Collectively, these modifications can be referred to as the histone post-translational modifications (PTMs); or to be more comprehensive and include the modifications on DNA as well, the term 'epigenetic/chromatin marks' could be used. Epigenetics refers to all the information passed onto the next generation of cells or organisms that is independent of the DNA sequence (i.e. genetic information)[23]. The variations in the definition of epigenetics and the hereditary aspect of the epigenetic marks are discussed in detail in chapter 7.

Box 1: Histone variants and some well-known histone marks

The 4 histones mentioned above that make up the nucleosome are Histone H3, H4, H2A and H2B. Histone H1 serves as the linker histone between nucleosomes. In addition to these canonical histones, there are histone variants that display slight differences in sequence and structure[24]. The variant histones seem to have different functional implications as well. For example, Histone CenH3 (also known as CENP-A) is exclusively found in the centromeric regions of chromosomes and play a major role in kinetochore formation and sister chromatid separation. Histone H2A.Z is another variant enriched at transcription start sites. Depending on the Histone H3 it is coupled with (H3.1 or H3.3), it increases or decreases the stability of the nucleosome respectively[24].
Histone marks are typically denoted with the histone name, the amino acid and its position and an acronym for the modification. For example, H3K4me3 represents Histone H3 modified with 3 methyl groups on lysine that is the fourth amino acid in the protein sequence. Some well-known histone marks are Histone H3K4me3 found at transcription start sites, H3K27me3 found in repressed regions and H3K36me3 found at actively transcribed gene bodies. Gamma-H2A.X (which stands for phosphorylated H2A.X in an exceptional denotation) is enriched at DNA damage sites.

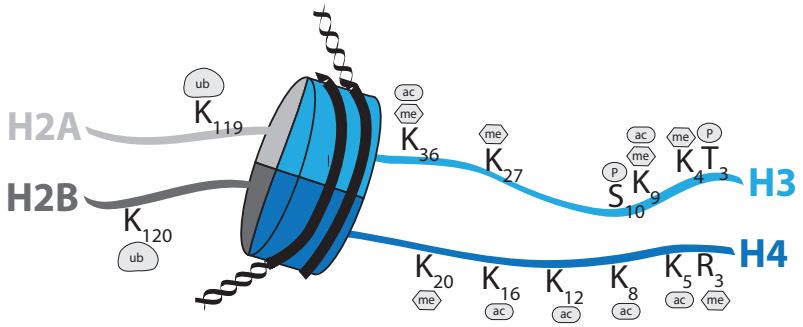


Figure 4: Histone tails carry multiple post-translational modifications. Here, a small selection of modifications is shown: Methylation, phosphorylation, acetylation and ubiquitination. Note that N-terminal tails of Histone H3 and H4, and C-terminal tails of Histone H2A and H2B are depicted.

Eukaryotic transcription machinery

The transcription machinery is assembled on promoter regions of genes. Promoters, around 50-100 basepairs in size spanning the transcription start site (TSS), have sequences

within that are recognized by general and specific transcription factors (TFs)[25]. TFs are specialized proteins with a sequence-specific DNA-binding domain and an activation domain for interacting with other regulators[26]. The sequences recognized by general transcription factors (GTFs) are called core promoter elements (e.g. TATA box, Initiator, Downstream Promoter Element (DPE)). The assembly of GTFs, namely, TFIID, TFIIB, TFIIA followed by TFIIE, TFIIF and TFIIH, result in the melting of the double strand, providing the transcriptional machinery with a single strand of DNA[25]. GTFs and co-factors, such as the Mediator, recruit RNA polymerase II (RNA Pol II), which is the polymerase that transcribes all messenger RNAs as well as most noncoding RNAs in eukaryotes[27](**figure 5**). The assembly of all these complexes on core promoter elements is called the Preinitiation Complex (PIC) formation. In simpler eukaryotes such as *Saccharomyces cerevisiae*, PIC formation is followed by transcriptional initiation and elongation[28]. However, in metazoans, another level of regulation takes place in the form of RNA pol II pausing at promoter proximal sites: After PIC formation, transcription may start, but usually is aborted after a few bases (30-60 nucleotides) are transcribed (abortive initiation). It is only with the help of positive elongation factors (and the dissociation of negative elongation factors) that productive elongation of transcription is possible[28–30]. Evidently, it is not only in the proximity of TSS that RNA Pol II pausing occurs, but the best studied instance of pausing so far is indeed the proximal pausing prior to productive elongation.

Core promoter elements, the GTFs and the related co-factors mentioned above are sufficient for a basal level of transcription to take place *in vitro*[25]. Regulated (activated) transcription requires the binding of specific transcription factors to their cognate sites (motifs) in the proximal promoter region[31]. Co-factors such as the Mediator seems to communicate the signals between the complexes bound to the proximal and core promoter elements and regulate the initiation of transcription[27].

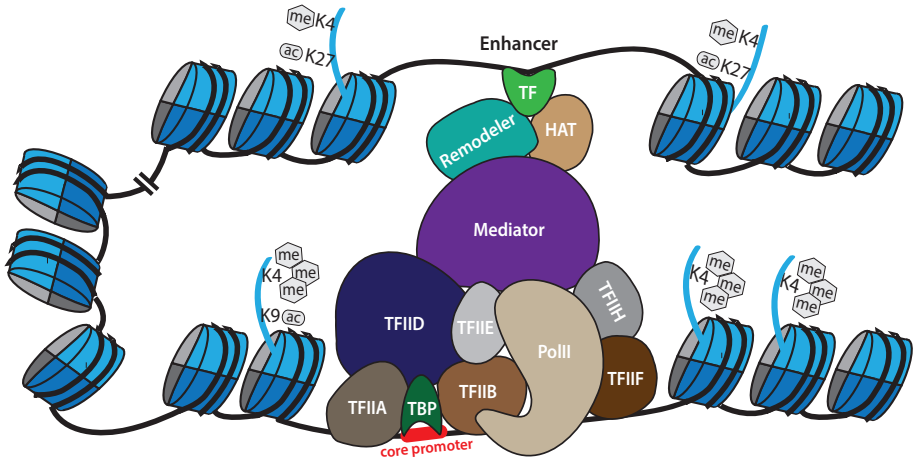


Figure 5: Basic transcription machinery at a promoter and its corresponding enhancer. Multisubunit general transcription factors assemble on the core promoter. Chromatin remodeling and histone acetylation take place at promoter and enhancer regions. Chromatin looping brings the enhancer into the proximity of the promoter. The Mediator complex is involved in the regulation of transcription initiation. Figure adapted from Koster et al., 2015[36].

Distal elements, such as enhancers and locus control regions (LCRs) are other gene regulatory elements that affect transcription[25]. Enhancers, like proximal promoters, harbor TF-binding motifs and also have an 'open' chromatin structure (low nucleosome density and high histone acetylation levels)[32]. Enhancers and LCRs are brought to the proximity of promoters that they regulate via chromatin looping events and activate/increase transcription of those genes[33,34](**figure 5**). A negative regulatory element that is called the insulator, disrupts the enhancer-promoter interaction and inhibits/decreases transcription[35].

Transcription termination is mediated by the recognition of the A-rich sequences at the 3' end of transcripts by termination factors. These factors cause the dissociation of RNA pol II from the transcribed strand of DNA[37].

The chromatin environment plays an important part in transcriptional regulation by allowing or hindering the formation of transcriptional machinery on promoters as well as affecting the progression of RNA pol II on gene bodies. Nucleosome depletion, i.e. chromatin remodeling, is required at the TSS[19]. Certain histone PTMs, such as acetylation of certain residues on Histone H3 and H4 in addition to H3K4 methylation, are strongly correlated with actively transcribed gene promoters[18]. On the other hand, modifications on DNA, such as methylation and hydroxymethylation, seem to affect transcription context-dependently[38]. Chromatin remodelers and modifiers, proteins that lack specificity, are recruited by general or specific TFs and various chromatin marks. The rest of this introduction will elaborate on specific aspects of the chromatin environment and the proteins involved in the processes of controlling those aspects.

DNA methylation

One of the best studied epigenetic marks is DNA methylation. Notably, it is also the epigenetic mark of which the inheritence is well-known with functional consequences[39]. DNA methylation is widely present in many organisms from bacteria to plants, showing mosaic or global patterns[40]. In vertebrates, global methylation of Cytosine in a CpG context is observed along with some low abundant nonCpG (CpA, CpT) methylation[41]. Because only CpG methylation (meCpG) is symmetrical on the double-stranded DNA, it is thought to be the only type of methylation that is faithfully propagated during DNA replication[42]. All instances of DNA methylation refer to CpG methylation in the rest of this thesis unless otherwise stated.

Repetitive regions of chromosomes such as centromeres and telomeres are heavily methylated in addition to transposable elements within the genome. There is extensive evidence suggesting DNA methylation-mediated repression at these particular sequences[40]. In placental mammals, promoters of genes in the inactive X chromosome and some promoters of imprinted genes along with non-promoter regulatory sequences (i.e. imprinting control regions (ICR)) are also methylated[39]. Moreover, mammalian genomes, in total slightly depleted of CpG (~40%), have long stretches of DNA with high and unmethylated CpG content called CpG islands (CGIs). CGIs overlap with ~60% of gene promoters and CpG methylation at these sites is tightly regulated during development[43,44]. All these instances show DNA methylation effecting repression and support its role as the model epigenetic mark: heritable across cell divisons and having a functional consequence[23]. Whether (and how) it is heritable across generations is a major question the epigenetics field is trying to answer[45,46].

For mammals, transgenerational inheritance of DNA methylation has only been shown in the context of imprinting: Imprinting is the phenomenon of allele-specific gene silencing

dependent on the parent of origin[47]. In spite of the genome-wide demethylation the primordial germ cells (PGCs) undergo, during maturation, alleles that should be silenced and imprinting control regions are correctly methylated in a sex-specific manner[48]. These loci are protected against the further wave of demethylation that takes place during pre-implantation[49].

More recent studies, with insight provided by high resolution whole genome bisulphite sequencing (WGBS, see **box 2**), have challenged the DNA methylation-repression association. Many genes harbor DNA methylation on the gene-body (genic methylation) which seems not to interfere with transcriptional activity[40]. Further WGBS data sets across different human tissue and cell types revealed dynamic regulation of DNA methylation in around 22% of autosomal CpGs[50]. These regions, (referred to as differentially/dynamically methylated regions (DMRs)) coincide with enhancers and TF-binding sequences. Other studies in parallel reported that at regions with low CpG density (low methylated regions (LMR)), it is mainly TF-binding dynamics that determine DNA methylation patterns[51,52]. These studies are important for 1. narrowing down the sites to investigate the methylation status in as yet uncharted genomes from different/diseased tissues for comparison purposes 2. narrowing down the sites to investigate the mechanistic details of DNA methylation and gene regulation. Importantly, the observation in which the presence or absence of TFs seems to regulate DNA methylation state is not an instance of DNA methylation 'instructing' transcription[53]. On the contrary, it shows the effect of the transcriptional regulatory elements on the methylation state. This is in opposition to the promoter methylation examples discussed in the previous paragraphs in which DNA methylation instructs repression.

In summary, the DNA methylation field is ever expanding with surprising findings. The functional consequences of DNA methylation need to be dissected in the right context: e.g. differentially methylated regions in development or disease and the nature of the DNA element that is methylated (promoter, gene-body, intergenic, etc.) should be taken into account when a functional link is investigated.

Box 2: Mammalian DNA methyltransferases, DNA demethylation and WGBS

There are three DNA methyltransferases in mammals: the maintenance methyltransferase DNMT1 active at replication forks to propagate the methyl-CpG sites to the daughter strand, and de novo methyltransferases DNMT3A and 3B, capable of putting down methylation to 'new' target CpGs. This functional demarcation, while valid for the most part, has been shown to be flexible to a certain extent[54]. A related protein, DNMTL, is expressed only in germ cells. It lacks the methyltransferase activity, but is nonetheless essential for the maintenance of imprinting[42].

Methyl-CpG is oxidized to hydroxymethyl-CpG by the TET enzymes (and further oxidized to carboxyl- and formyl-Cytosine). This pathway is thought to constitute the active DNA demethylation process. Whether the intermediate products on the genome serve another, specific regulatory purpose is not yet known[55,56].

Whole genome bisulphite sequencing (WGBS) has allowed the mapping of methylated CpGs with basepair resolution[57]. Bisulphite-treated cytosines convert to uracil while meC is resistant to this conversion. Subsequent high-throughput sequencing reveals the U:G mismatches, i.e. unmethylated CpGs. Of note, bisulphite sequencing does not distinguish between meC and hydroxymeC. Other methods for sequencing both meC and hydroxymeC have been developed.[55,58]

Proteins that recognize methylated DNA

Canonically, there are three groups of proteins that interact with methyl-CpG in mammals: Methyl-CpG Binding Domain (MBD) family of proteins (discussed in detail below), Kaiso and Kaiso-like proteins (Kaiso, ZBTB4, ZBTB38) and the Set and Ring finger associated (SRA) domain proteins (UHRF1 and UHRF2)[59]. Recent unbiased interaction screens have revealed more proteins with an affinity for Methyl-CpG[60–62].

MBD Family of proteins

The Methyl-CpG binding domain was first discovered in the MeCP2 protein due to its strong interaction with methylated DNA sequences in band shift assays[63]. Other proteins with a highly homologous MBD are MBD1, MBD2, MBD3 and MBD4. MBD5 and 6 proteins were discovered later than the canonical MBDs in a cDNA screen from brain tissue in 2000[64]. Sequence homology studies revealed more proteins with an MBD: BAZ2A, BAZ2B, SETDB1 and SETDB2 (see **figure 1** in chapter 3 (MBD family of proteins)). The canonical MBDs have been better studied, probably for the reason of having been discovered earlier. Importantly, none of the MBDs of the 'non-canonical' members of the family, in addition to MBD3 of the canonical ones, have retained an affinity for methylated CpG in mammals[59]. Nonetheless, they all interact with chromatin.

Mutations in the MeCP2 gene, located on the X chromosome, have been linked to the Rett syndrome where neurodevelopmental defects are observed in females (males are also affected in some cases, however the frequency is far lower)[65]. MeCP2 is also associated with other neurological disorders from autism spectrum disorder to schizophrenia. Very abundant in the brain, it's thought to compete with the linker Histone H1 in neurons, implying that MeCP2 may function as a chromosomal architecture protein as well. Interestingly, both knock-down and overexpression of MeCP2 result in neurological phenotypes, suggesting the level of this protein requires a tight regulation. While MeCP2 has been proposed to interact with some chromatin complexes (e.g. Sin3A, N-CoR) to exert repression, the exact molecular mechanisms and the specific targets of this protein remain under intense investigation.

MBD1 is unique within the MBD family because it contains three CXXC domains (variable among different isoforms) in addition to its MBD[59]. CXXC is a zinc finger domain known for strongly interacting with unmethylated CpGs. Functionally, MBD1 has been linked with heterochromatin maintenance via interactions with HP1 and H3K9 methyltransferase SUV39H1[66]. It was also shown to target some genes involved in neuronal differentiation[67–69]. In addition, it has been implicated in propagating histone marks during DNA replication due to its reported interaction with a member of the chromatin assembly complex, CAF-1, and Histone H3K9 methyltransferase SETDB1 in the S phase of the cell cycle[70].

MBD2 and MBD3 are members of the chromatin complex NuRD that was originally identified as a transcriptionally repressive complex (more on NuRD complex later). These proteins are found mutually-exclusively within the NuRD complex[71]. Their function is expected to be non-redundant as MBD2 can interact with methylated CpGs whereas MBD3 cannot. In mammals, MBD3 is highly expressed at the embryonic stage while the expression of MBD2 increases upon differentiation.[72] This may partly explain why knock-out phenotypes are so different: Embryonic lethality in the case of MBD3; whereas in the case of MBD2, some defects in maternal behaviour are observed in female knock-out mice[59].

MBD4 is involved in base excision repair via its glycosylase domain[73]. It acts on G:T mismatches caused by the deamination of methyl-Cytosine, providing an additional link to the process of DNA demethylation. In addition to DNA repair and demethylation processes,

MBD4 has been associated with apoptosis[73].

For the non-canonical MBD-bearing proteins, SETDB1 and SETDB2 both have a SET domain that methylates Histone H3K9. This mark, mostly linked to repression, is enriched on centromeres as well as some euchromatic regions[74]. Additionally, SETDB1 is implicated in the imprinting process and in maintaining the silence of various retrotransposable elements as all these sequences also harbor H3K9me3[75,76]. SETDB2 has been reported to regulate left-right asymmetry in addition to mediating anterior-posterior axis elongation during zebrafish embryogenesis[77].

BAZ2A and BAZ2B proteins, on the other hand, have a PHD-type zinc finger and a bromo domain that interacts with acetylated histones. BAZ2A is the better characterized protein as it is part of the Nucleolar Remodeling Complex (NoRC) that regulates the expression of ribosomal DNA[78]. More data on the function and interactors of BAZ2B are yet to be published.

MBD5 and MBD6

MBD5 and MBD6, as mentioned above, were first discovered in a cDNA screen in 2000. A thorough biochemical characterization of these proteins was published in 2010[79]. In this paper, the proteins were shown not to have a preference for methylated CpG in band shift assays. In addition, the chromocentric localization of the proteins were mostly unaffected in DNMT knock-out mouse embryonic fibroblasts (MEFs): In wild-type cells, chromocenters are hypermethylated, but in DNMT knock-out MEFs the methyl mark is lost[80], indicating that MBD5 and 6 localization to chromocenters is indepedent of the methylation status.

MBD5 has two isoforms. The long isoform (isoform 1) has a PWWP domain, a known chromatin interaction domain[81]. Expression profiling in various mouse tissues showed that MBD5 isoform 1 and MBD6 are highly expressed in brain and testes. The second isoform of MBD5 is highly expressed in oocytes. The observed high expression of MBD5 in the brain is expected as it is one of the genes that is either deleted or duplicated in the 2q23.1 microdeletion/duplication syndrome[82–87]. This locus is within the Autism Spectrum Disorder-associated loci. In addition to the drastic change that would be the deletion or duplication of a locus, several mutations in the Mbd5 gene are also reported in patients with various neurological disorders[88]. In one such study, genetic interaction screens performed in *Drosophila melanogaster* showed a synergistic relationship between the MBD5 homologue, sba, and the EHMT1 homologue, EHMT, which is a histone H3K9 methyltransferase[89]. The mechanistic details of this genetic interaction, however, are not known. In 2012, a paper describing Mbd5 knock-out mice was published: The mice displayed perturbed glucose metabolism in addition to neurological retardation[90]. For MBD6, functional data is still lacking.

Polycomb Repression

Among the known chromatin regulatory complexes (excluding remodelers), Polycomb group complexes are possibly the best-studied repressors. Initially characterized during *Drosophila melanogaster* development, a series of mutants were investigated that resulted in the repression or derepression of the homeotic genes. The mutated genes that caused derepression were classified as the Polycomb group genes (PcG)[91]. Further studies revealed that many other genes are also targeted by PcG proteins[92–94].

The PcG proteins form two major complexes: Polycomb Repressor Complex 1 (PRC1) and PRC2[94]. Conserved from Drosophila to mammals, both complexes mediate repression of targets, driving differentiation and development. Both have chromatin-modifying

subunits, with the histone marks they catalyze instrumental in mediating repression: Ring1b (Sce in Drosophila) in PRC1 lays down monoubiquitin on Histone H2AK119Ub in mammals (H2AK118Ub in fly)[95,96] and Ezh2 (E(z) in Drosophila) in PRC2 catalyzes Histone H3K27 methylation[97,98]. PRC1 was also shown to repress transcription independently of Histone H2A ubiquitination[99,100]. In addition, there are several variant PRC1 complexes with remarkably different subunit compositions[101]. For a summary of the PcG complexes and their subunits in Drosophila and in human, see **table 1**. For variant complex subunits and additional interactors of the PcG complexes in human, see **table 2**.

Drosophila melanogaster	**Characteristic domains**	***Homo sapiens***
Polycomb repressive complex 1 (PRC1)		
E3 ubiquitin-protein ligase dRING (Sce)	RING	RING2 (RING1B) and RING1 (RING1A)
Posterior sex combs (Psc) and Suppressor of Zeste (Su(z)2)	RING	PCGF4 (BMI1) and PCGF2 (MEL18)
Polyhomeotic-proximal (Ph-p) and Polyhomeotic-distal (Ph-d)	Sterile α-motif (SAM) and Zinc finger	PHC1 (EDR1), PHC2 (EDR2), PHC3 (EDR3)
Polycomb (Pc)	Chromodomain	CBX2,CBX4, CBX6, CBX7, CBX8
Sex combs on midleg (Scm)	Malignant brain tumor (MBT), SAM and zinc finger	SCMH1 and SCML2
Polycomb repressive complex 2 (PRC2)		
Enhancer of Zeste (E(z))	SANT, CXC and SET	EZH2 and EZH1
Extra sex combs (Esc) and Extra sex combs-like	WD40	EED
Suppressor of zeste 12 (Su(z)12)	Zinc finger and VEFS box	SUZ12
Chromatin assembly factor 1 subunit	WD40	RBBP4 and 7
Jing	Zinc finger	AEBP2
PR-Dub complex		
Calypso	UCH	BAP1
Asx	atypical PHD type zinc finger and LXXLL motif 1 and 2	ASXL1/2/3

Table 1: PRC1, PRC2 and PR-Dub complex (core) subunits in *Drosophila melanogaster* and humans with critical domains denoted. Adapted from Schwartz and Pirrotta, 2013[94].

Since none of the PcG proteins carries a DNA-binding domain, how recruitment to target genes is mediated is an intense area of investigation[94,102,103]. In Drosophila, the targeting of the complexes is mediated by specific sequences called the Polycomb Response Elements (PREs)[104]. In mammals, the targeting mechanisms seem to be more complex, involving specific recruiter proteins and long non-coding RNAs. The histone marks catalyzed by both complexes are mostly found together on chromatin, forming 'Polycomb domains'[102]. Even the PRC1 variant complexes that do not catalyze H2AK118/9Ub localize to Polycomb domains to mediate repression in an H2AUb-independent manner[99]. A key finding providing insight as to how this cooperation was regulated emerged in 2003: The CBX proteins found in PRC1 complexes were shown to interact with the chromatin mark laid down by the PRC2

complex, H3K27me3[105,106]. This finding developed into the hierarchical model in which the initial recruitment of PRC2 to target sites catalyzes the methylation of H3K27, which acts as a recruiter for both more PRC2 complexes[107,108] and PRC1 complexes, which then catalyze H2AK119Ub[109]. The challenge to this model comes from sites on the genome in which PRC1 is found without the presence of H3K27me2/3[110–113]. In addition, in mESCs that have diminished PRC2 activity (and hence no H3K27me3 mark), PRC1 targeting and H2AUb levels are not completely disrupted. Therefore the hierarchical model of PRC1 and PRC2 recruitment seems to be valid for a subset of targets where both marks are present.

In addition, in the past year, consecutive publication of 3 papers (one of which is chapter 4 of this thesis) showed a re-shuffling in the hierarchical model of polycomb recruitment: The biochemical characterization H2AK119Ub interactors *in vitro*, as well as induced PRC1 targeting studies *in vivo* showed a role of H2AK119Ub in recruiting the PRC2 complex[114,115]. These studies form the basis of a model where PRC1 recruitment feeds forward to PRC2 recruitment which feeds forward to more PRC1 and PRC2 recruitment, and hence, Polycomb domain formation.

Subunit / Interactor	**Characteristic domain**	**Complex**
RING1 and YY1-binding protein (RYBP)	Zinc finger	PRC1 variant (multiple)
YY1-associated Factor (YAF2)	Zinc finger	PRC1 variant (multiple)
BCL6 Corepressor (BcoR)	ANK Repeat 1-3	PRC1 variant (1.1)
BCL6 Corepressor-like protein 1 (BcoRL1)	ANK Repeat 1-3	PRC1 variant (1.1)
Lysine specific demethylase 2B (KDM2B)	CXXC-type zinc finger and Jumonji	PRC1 variant (1.1)
Ubiquitin carboxyl-terminal hydrolase (USP7)	USP	PRC1 variant (1.1)
Lethal(3)malignant brain tumor-like protein 2 (L3MBTL2)	Zinc finger	PRC1 variant (1.6)
Max protein (MAX)	Basic helix loop helix	PRC1 variant (1.6)
Max gene-associated protein (MGA)	Basic helix loop helix	PRC1 variant (1.6)
WD repeat-containing protein 5 (WDR5)	WD 1-7	PRC1 variant (1.6)
E2F6	DEF box and Leucine zipper	PRC1 variant (1.6)
Polycomb group RING finger protein 1 (PCGF 1)	RING	PRC1 variant (1.1)
Polycomb group RING finger protein 3 (PCGF3)	RING	PRC1 variant (1.3)
Polycomb group RING finger protein 5 (PCGF5)	RING	PRC1 variant (1.5)
Polycomb group RING finger protein 6 (PCGF6)	RING	PRC1 variant (1.6)
Autism susceptibility gene 2 protein (AUTS2)		PRC1 variant (1.3 and 1.5)
Protein Jumonji (JARID2)	Jumonji	PRC2
PHD finger protein 1 (PHF1)	PHD type zinc finger 1-2	PRC2
PHD finger protein 19 (PHF19/PCL3)	PHD type zinc finger 1-2	PRC2
Metal-response element-binding TF2 **(**MTF2**/**PCL2)	PHD type zinc finger 1-2	PRC2
Lysine specific demethylase 1B (KDM1B)	CW-type Zinc finger, SWIRM, FAD	PR-Dub
Forkhead box protein 1 (FOXK1)	FHA	PR-Dub
Forkhead box protein 2 (FOXK2)	FHA	PR-Dub
MBD5	MBD*	PR-Dub
MBD6	MBD*	PR-Dub
Host cell factor 1 (HCFC1)	Kelch 1-5	PR-Dub
O-GlcNAc transferase subunit p110 (OGT)	TPR 1-13	PR-Dub

Table 2: Additional complex subunits/interactors for the three Polycomb complexes in humans. Note that a selection of reported interactors is listed. PRC1 variant complex definitions are based on Gao et al., 2012[101]. (*) indicates loss of meCpG-binding.

Although the repressive function of all PcG complexes is well-established, there are recent reports emerging with exceptional observations where PcGs seem to mediate gene activation[116–119]. Whether these exceptional cases are due to the composition of the PcG complexes in question or the chromatin environment that they act on (or both) remains to be seen.

Minibox 1: Trithorax Group proteins

The family of proteins in which mutations resulted in the repression of genes as opposed to derepression (PcG) were named the Trithorax Group (TrxG). PcGs and TrxGs work in an antagonistic manner. Just like PcG, TrxG proteins also catalyze certain chromatin marks: methylation of Histone H3K4 and H3K36, marks associated with active transcription. There are 3 TrxG complexes in Drosophila melanogaster (Set1, Trx, Trr) and 6 in mammals (Set1A, Set1B, MLL1-4). Interestingly, although there is no yeast homologue for PcG complexes, there is one TrxG complex (Set1/COMPASS). The corresponding marks, i.e. H3K4me1/2/3 and H3K36me1/2/3 are also present in *Saccharomyces cerevisiae*[120].

A Novel PcG complex: Polycomb Repressive Deubiquitinase

In 2010, a novel repressive complex that was associated with PREs was characterized in *Drosophila melanogaster*[121]. The two proteins that make up the complex, CALYPSO and ASX, catalyze the deubiquitination of Histone H2AK118Ub. CALYPSO is the catalytic subunit, and ASX enhances the activity of CALYPSO. The mammalian homologues, BAP1 (CALYPSO) and ASXL1 (ASX) were also shown to catalyze the deubiquitination of H2AK119Ub on *in vitro* reconstituted nucleosomes. This novel complex was categorized as a PcG complex because the knock-out of Calypso and Asx resulted in the derepression of target genes and defects in development. In addition, both Calypso and Asx were found to associate with PRE as mentioned above, hence the complex was named Polycomb Repressive Deubiquitinase (PR-Dub).

The observation that removal of monoubiquitin on Histone H2AK118 resulted in derepression was surprising, considering the fact that Histone H2AK118 is monoubiquitinated by PRC1 and this mark has been long linked with repression.

Although the PR-Dub complex was identified and mostly characterized in Drosophila, the mammalian homologue of CALYPSO, BAP1, was already a well-known and well-studied protein due to its alleged interaction with BRCA1[122]. That is the reason why BAP1 stands for **B**RCA1 **A**ssociated **P**rotein **1**. Brca1 is a tumor suppressor gene mutated in breast and ovarian cancers and mutation carriers face high risk of developing these tumors[123]. Subsequent studies showed that BAP1 deubiquitinates the BRCA1 interactor, BARD1, rather than interact with BRCA1[124]. Since BARD1 and BRCA1 form a heterodimer with an E3 ubiquitin ligase activity[125], BAP1 indirectly affects BRCA1 activity via BARD1. However, within the BAP1 interactome that was characterized in HeLa cell lines, BARD1 was not detected as an interactor as shown in chapter 3 of this thesis.

In addition to indirectly modulating BRCA1 activity, BAP1 was linked to several types of cancer as mutations in the Bap1 gene were detected in patients with uveal and cuteanous melanoma, mesothelioma, renal carcinoma, breast and lung tumors[126–129]. Strikingly, multiple patients from 2 families with germline mutations in Bap1 developed uveal melanoma *and* mesothelioma, an occurrence that would happen 36 times in a trillion if only coincidental.

Based on this observation, it was proposed that germline mutations in Bap1 cause a 'hereditary cancer syndrome' that is characterized by patients developing multiple (and particular) malignancies[130,131]. Now a bona fide tumor suppressor, Bap1 was also classified as one of the 'driver' genes mutated in malignancy according to a tumorigenesis catalogue put together by Vogelstein et al[132].

Knocking out Bap1 results in embryonic lethality[133]. In 2012, a study reported the effects of conditional knock-out of Bap1 in mice[133]. These mice displayed splenomegaly, monocytosis and neutrophilia. This myeloproliferative/myelodysplastic syndrome (MDS) observed in mice resembles chronic myelomonocytic leukemia (CMML) in humans. The knock-out mice do not display any of the cancer types known to be associated with BAP1 deregulation in humans.

While the cancer research field established BAP1 as a tumor suppressor, the exact role of BAP1 in cell cycle regulation, based on studies performed in various cell lines, remains contradictory. Several studies reported that BAP1 knock-down enhances cell proliferation, others showed that it slows down the cell cycle[134–137]. The mammalian homologue is interesting in that it has gained a protein interaction domain that is missing in the Drosophila CALYPSO: An HCFC1-binding domain (HBM)[134–136]. HCFC1 is a rather intriguing regulatory factor involved in progression of the cell cycle via its interaction with the E2F family of transcription factors[138,139]. HCFC1 undergoes proteolysis for maturation and the resulting subunits, HCFC1-N and HCFC1-C, seem to drive different phases of the cell cycle (G1 and M progression, respectively)[140]. O-linked beta-N-acetylglucosamination (O-GlcNAcylation) of the uncleaved, precursor polypeptide of HCFC1 is essential for proteolysis to take place[141]. This PTM is added to HCFC1 by O-GlcNAc transferase (OGT), and interestingly, the activity of OGT is regulated by BAP1, which deubiquitinates OGT[133]. This series of events, i.e. the removal of a PTM activating an enzyme to add another PTM to another protein which then undergoes proteolysis in order to have active subunits is only a minimal slice of the vast regulatory possibilities PTMs and the proteins bearing/adding/removing the PTMs present. In addition to deubiquitinating OGT, BAP1 was also reported to catalyze the deubiquitination of K48-linked polyubiquitin chains on HCFC1[134,135]. However, abolishing the catalytic activity of BAP1 did not result in significant changes on HCFC1 levels or activity. In a ChIP-Sequencing study performed in mouse macrophages, 85% of BAP1 peaks showed co-localization with HCFC1 and 70% of OGT peaks were within 400 bp of HCFC1 peaks[133]. The authors propose that BAP1/HCFC1/OGT form a ternary complex and recruit the other known interactors of BAP1 to chromatin. BAP1 may have an effect on cell cycle by directly regulating HCFC1 activity or indirectly affecting the activity of HCFC1 by regulating the activity of OGT, or it does both, and/or it indeed forms a recruiter complex with these factors and controls gene expression. These possible scenarios remain to be investigated. Interestingly, one study reported that in S-phase, Ino80 is stabilized by BAP1 at active replication forks, which may represent another way for BAP1 to affect cell cycle progression[142]. In our interaction screens (chapter 3), not Ino80, but both HCFC1 and OGT are identified as BAP1 interactors.

In another study, BAP1 was reported to form a ternary complex with HCFC1 and the TF YY1[136]. The resulting complex was shown to activate the expression of a nuclear gene encoding a mitochondrial respiratory chain protein, cox7c.

In addition to HCFC1 and OGT, the mammalian PR-Dub complex has many more subunits compared to the Drosophila PR-Dub complex (see **table 2**):

TFs FoxK1 and FoxK2 interact with the mammalian PR-DUB. FoxK2-dependent recruitment of BAP1 to several target genes showed changes in gene expression (both

up and downregulation), in addition to a consistent decrease in H2AK119Ub[143]. More recently, it was shown that BAP1 mediates repression of FoxK2 targets; but the absence of BAP1 results in derepression only in the presence of BMI-RING1b (which ubiquitinate H2AK119)[144].

The Drosophila ASX has three homologues in mammals: ASXL1, 2 and 3. ASXL1 and 2 have been shown to interact with BAP1 (chapter 3,[133,135,136]). Interestingly, mutations in ASXL1 were reported in CMML patients (see above for Bap1 knock-out mice) and it is also classified as a driver gene in cancer along with BAP1 (also as a tumor suppressor)[132,145]. Both ASXL1 and ASXL2 were shown to regulate the activity of PPAR-gamma nuclear receptor in a reciprocal manner during adipocyte generation[146]. In addition, ASXL2 was shown to regulate the activity of the retinoic acid receptor positively and negatively by interacting with different chromatin modifiers[147]. Hence, ASXL1 and 2 seem to regulate transcription through interactions with members of the nuclear receptor family, but the exact mechanistic details are unknown. Moreover, both ASXL1 and ASXL2 were reported to interact with the PRC2 complex and mediate H3K27 methylation by PRC2[148,149]. ASXL2 is reported to be involved cardiac functioning, as mice that survived the knock-out of Asxl2 (which is partially embryonic lethal) displayed enlarged hearts and cardiac dysfunction in addition to reduced levels of H3K27me3[150,151].

KDM1B, a lysine demethylase, is another interactor that is also a chromatin modifier like BAP1. It demethylates di-methyl (me2) and mono-methyl (me) residues on Histone H3K4[152]. It is reported to regulate maternal imprinting as knocking out Kdm1B resulted in biallelic expression of various, normally imprinted genes in mouse oocytes[153].

The novel interactors, MBD5 and MBD6, are discussed in detail in the previous section and in chapter 3.

In summary, the human PR-Dub complex seems to be far more complex than the Drosophila counterpart. We do not know whether all the functional implications of the PR-Dub subunits summarized above take place within the PR-Dub complex context or whether these subunits have PR-Dub-independent interactions and functions.

BAP1 is also implicated in DNA damage response (see the next section for details). First, it was shown to be phosphorylated (at S592) in the S phase upon replicative stress and this resulted in the dissociation of BAP1 from chromatin[154]. Another study reported that BAP1 is phosphorylated on 6 sites upon ionizing radiation (IR) and the phosphorylation was required for BAP1 to promote the double strand break repair response (via BRCA1 and RAD51)[155]. Finally, another study showed BAP1 recruitment to IR-induced DNA damage sites along with ASXL1[156]. This study also reports that BAP1 facilitates BRCA1 and RAD51 foci formation at sites of damage.

Histone H2A (De)Ubiquitination

The first protein that was reported to be ubiquitinated was Histone H2A[161]. Interestingly, the ubiquitination of H2A was originally thought to be involved in transcriptional activation. Then, H2AUb was observed in heterochromatic regions in several studies. Finally, when PRC1 was shown to mediate repression via monoubiquitination of H2A, the functional association with this mark almost exclusively turned to repression[162].

In the context of Polycomb repression, as mentioned above, Histone H2AK119 monoubiquitination is well-known and well-studied. Put down by RING1B (and in some cases RING1A), it mediates the repression of gene expression during development and differentiation. Heterochromatin formation and X chromosome inactivation are processes

Box 3: Ubiquitin, (De)Ubiquitination and BAP1

Ubiquitin (Ub) is an extremely well conserved, ubiquitious protein. As a PTM, it has a remarkable size with 76 amino acids. Ubiquitination of target proteins is a three step process involving activation of Ub by an E1 activating enzyme, conjugation of Ub to an E2 conjugating enzyme and ligation of Ub to the target protein by an E3 ligase. Ub is added to lysine residues on target proteins where an isopeptide bond is formed between the last amino acid of Ub (glycine) and the lysine of the target protein[157]. There are 7 lysine residues on Ub itself on which polyubiquitin chains can be extended (K6, K11, K27, K29, K33, K48 and K63). Polyubiquitination is a trademark of a protein being marked for degradation by the Ubiquitin-Proteosome System (UPS). However, different homotypic chains of polyUb have been associated with different cellular processes besides protein degradation. There are also heterotypic chains where the polyUb formation does not extend on a single lysine on Ub.

The isopeptide bond between Ub and target proteins is hydrolyzed by specific proteins called deubiquitinases (Dubs). There are 5 classes of Dubs in eukaryotes: The Ub C-Terminal Hydrolase (UCH) Domain, The Ub-Specific Protease (USP) Domain, The Ovarian Tumor (OTU) Domain, the Josephin Domain, The Jab1/MPN/Mov34 metalloenzyme (JAMM) domain[158,159].

BAP1 is within the UCH domain of deubiquitinases. The Drosophila homologue of BAP1, CALYPSO, was preferentially shown to act on monoUb[121]. However, the mammalian BAP1 was reported to deubiquitinate the K48-linked polyUb chains of HCFC1[134,135]. Whether BAP1 has a preference for monoUb or polyUb remains to be clarified. So far the reported targets of deubiquitination by BAP1 are Histone H2A (monoUb)[121], OGT (mono or polyUb unknown)[133] and HCFC1 (polyUb)[134,135]. Identification of further targets, the nature of their ubiquitination and the regulatory consequences are of interest.

BAP1 itself was reported to be monoubiquitinated by the E2-E3 ligase UBE2O on its nuclear localization sequence on multiple lysines[160]. This results in sequestering BAP1 in the cytoplasm. BAP1, in return, deubiquitinates itself and then its nuclear localization is restored.

that involve gene silencing by PRC1. As a mode of repression, H2AK119Ub by PRC1 was reported to keep RNA Pol II from elongating at the promoters of a set of genes in ESCs[163]. Similarly, but in another context, H2AK119Ub mediated by another E3 ligase, 2A-HUB, was shown to interfere with transcriptional elongation and repress chemokine gene expression in macrophages[164].

There are several E3 ligases for H2AK119: RING1B, RING1A, 2A-HUB, CULLIN4B[165,166]. Recently, TRIM37 was identified as another E3 ligase for H2AK119 in breast cancer cell lines, it is postulated to act as an oncogene[167]. BRCA1/BARD1 heterodimer was shown to mediate monoubiquitination of H2A at other sites, namely at K127 and K129[168]. H2AUb mediated by BRCA1 was previously shown to be essential for maintaining the genome integrity via the silencing of microsatellite regions[169]. BRCA1 is also a DNA damage response protein[123].

In the context of DNA damage, Histone H2A and the variant histone H2A.X were shown to be polyubiquitinated on K13 and K15 (by RNF8/RNF168) and monoubiquitinated on K119 and K120 (by Ring1b)[170]. Both events seem to play a role in signaling for DNA damage and the recruitment of DNA damage response (DDR) machinery (i.e. foci formation). In fact, monoubiquitination on K119 was shown to be required for the phosphorylation of Serine

139 on histone H2A.X which is a major signal on chromatin for double strand breaks (DSBs)[170].

In the context of the regulation of cell cycle, Histone H2AK119Ub level is downregulated as the chromosomes are compacted for mitosis[171]. Deubiquitination of H2A by USP16 has been shown to be a prerequisite for the phosphorylation of Histone H3S10, which facilitates chromosome condensation[172]. USP3 and USP22 have also been shown to affect the cell cycle via their Dub activity on H2AK119Ub[171,173].

Other known Dubs for H2A are: BAP1, USP12, USP21, USP44, 2A-DUB and BRCC36. Some known functional associations with these proteins are that: 2A-DUB was reported as a positive co-regulator of the androgen receptor in addition to regulating the expression of a key hematopoietic TF GFI1[164,174]. Initially, BRCC36, USP3 and USP44 have all been reported to act as negative regulators of foci formation in the DDR mentioned above[171]; however, in a recent knock-down screen of Dubs, depletion of BRCC36 and 2A-DUB resulted in a delay in DDR[175].

So far, the general emerging picture is that H2AUb mediates gene repression: it needs to be removed to facilitate higher order chromatin formation and it forms a recruitment platform for DDR proteins at sites of DNA damage. The Dubs that have been characterized complement this picture in that: they seem to activate gene expression, they mediate the progression of the cell cycle when the levels of H2AUb need to be decreased and they negatively regulate foci formation at DNA damage sites via removing the poly/monoUb on H2A that signal damage. There are exceptional observations as mentioned above, especially regarding BAP1: It was reported to mediate gene repression, its role in cell cycle regulation needs clarification and it has been reported to facilitate DNA damage foci formation as discussed in the previous section.

Minibox 2: Double Strand Breaks

One of the most toxic types of DNA damage is Double Strand Breaks (DSBs). Ionizing radiation induces DSBs as utilized in chapter 3 of this thesis. DSBs are fixed by homologous recombination (HR) or nonhomologous end joining (NHEJ) pathways. HR can take place during S and G2 phases of the cell cycle. Although the 2 pathways differ in downstream players, the initial signaling of DNA damage is thought to be conserved and involves the ubiquitination of H2A and H2A.X.[170]

Another chromatin regulatory complex: NuRD

A major chromatin-associated complex that is conserved among metazoans is the Nucleosome Remodeling and Histone Deacetylation (NuRD) complex. The NuRD complex is a big, multisubunit protein complex that can manipulate chromatin in different ways: CHD3/4 subunit is the ATP-dependent nucleosome remodeler and HDAC1/2 deacetylases histones (and other proteins). In various purifications of the complex, LSD1, a lysine demethylase, appears as an interactor as well[176,177]. Other characterized subunits are the histone chaperones RBBP4/7, scaffold proteins GATAD2A/B, MTA1/2/3, MBD2/3 and DOC1 (see **table 3**). Some of these subunits are found within the complex in a mutually-exclusive manner, such as MBD2 and MBD3, and CHD3 and CHD4. Mutual exclusion status for other paralogous subunits is not known.

The fact that there are many possible distinct subunit compositions for the NuRD complex may be functionally-relevant: As mentioned in an earlier section, the expression profile of MBD2 and MBD3 seem to be different with MBD3 being expressed abundantly during the

pluripotent state and decreasing upon differentiation while the expression profile of MBD2 follows a reciprocal pattern[72]. MBD2 and MBD3 are a rather interesting paralogous couple of proteins as MBD2 has retained the affinity for methylated DNA but MBD3 has not[59]. This fact has led to speculations about MBD2 recruiting the NuRD complex to sites of methylated DNA, whereas the MBD3/NuRD complex cannot be targeted in that fashion. A study with a ChIP-Sequencing profile for MBD2 revealed that MBD2/NuRD is indeed enriched at methylated CpGs[178]. However, there are also nonmethylated target sites that MBD2/NuRD occupies, suggesting that MBD2 does not always undertake the recruiting function within the complex[179].

Subunit	Characteristic domain
CHD3 and 4	PHD type zinc finger 1,2; helicase; chromodomain 1,2
HDAC1 and 2	Deacetylase
RBBP4 and 7	WD repeat
GATAD2A and 2B	GATA type zinc finger
MTA1, 2 and 3	Zinc finger, BAH, ELM, SANT
MBD2	MBD
MBD3	MBD*
DOC1	-

Table 3: NuRD subunits and characteristic domains. (*) indicates loss of methyl-CpG binding capability.

Indeed, the question of recruitment for the NuRD complex is an area of active investigation as it is for the other chromatin associated complexes (discussed previously). In addition to all the subunits mentioned above, there are a number of NuRD interactors (Zmynd8, Znf512, Znf592, Znf687, CUTL1, Ikaros, etc.)(chapter 5, [180]). Among these, many have zinc fingers and other known DNA binding domains (such as the homeobox of CUTL1)[181]; therefore some of these interactors are expected to be tissue-specific recruiters of the complex.

Functionally, The NuRD complex was originally associated with gene repression, especially in light of the deacetylation acitivity it possesses. However, genome-wide profiling of deacetylases has revealed that histone deacetylation also takes place at actively expressed genes[182]. In line with this observation, some NuRD subunits were found at an active promoter that exhibited cyclical expression in response to activation with estrogen[183]. In addition, the NuRD complex has been reported to be required for activation of gene expression in other contexts as well (chapter 6,[184]). Whether these observations are due to the necessity of HDAC activity to 'reset' chromatin in-between cycles of gene expression or there is a compositional and/or functional switch that takes place within the complex to change its effect on transcription is not known yet (see chapter 6 for more discussion on the subject).

The NuRD subunits have been studied in the pluripotency context as well: MBD3 has been reported to interact with the core pluripotency network (e.g. OCT4)[185–187]; however we, in our group, have not been able to reproduce this interaction. Nonetheless, whether it is through a direct interaction with the pluripotency network or not, MBD3 seems to be enhancing the process of iPSC formation from differentiated cells[188]. It is also required for differentiation as MBD3 knock-out is embryonic lethal[189,190]. In addition, the repressive function of the NuRD complex has been suggested to create a threshold for gene activation that keeps aberrant gene expression to a minimum in the more open chromatin structure of the pluripotent cells[191]. The open chromatin structure is thought to contribute to the

plasticity within pluripotent cells that is required to maintain a high potential to differentiate into various lineages. This property comes with the price that spurious and nonspecific transcriptional activity would be easier to initiate. The NuRD complex is hypothesized to make up for this downfall by increasing the threshold for transcriptional noise.

As the complex is involved in regulating differentiation, it is no surprise that multiple NuRD subunits, MTAs, DOC1, Rbbp4/7, have been linked to various types of cancer[191,192]. Of note, whether the pathological role of these proteins is within the NuRD complex context is not known. However, considering the high regulatory capacity of the complex, it would be reasonable to expect that mutations found in NuRD subunits deregulating the complex activity has severe consequences. In addition to potential defects in transcriptional regulation promoting/maintaining tumorigenesis; NuRD complex has also been linked to DNA damage response[193–197]. Deregulation of NuRD activity resulting in defective DNA damage repair could be another functional aspect linking NuRD subunit mutations to various types of malignancies.

Mass Spectrometry-Based Proteomics

Classical molecular biology and biochemistry techniques investigate the functional and structural properties of biological entities of interest (e.g. a certain transcript or a protein) one at a time, singularly focusing on one type of molecule (a population of the molecule to be more exact). Various effectors and/or interactors could be taken into account, but usually the numbers of entities investigated remain in single digits. With the generation of high throughput techniques in the last decade, information on hundreds/thousands of biological molecules could be collected in single experiments. Particularly the hybridization or sequencing based technological advances in the genomics area have made it possible to analyze entire transcriptomes, first in a population of cells, and now in single cell settings[198]. Whole transcriptome datasets, while extremely informative, do not reflect the final population of functional agents: The end product of mRNA molecules are proteins and the dynamism in regulatory networks controlled by proteins could only be addressed by collecting information on the proteome of cells.

Mass spectrometers are instruments used to measure mass and charge of chemicals for identification purposes (**figure 6**). A revolutionary breakthrough in mass spectrometry was the generation of ionization techniques that allowed the measurement of biological molecules in mass spectrometers: Matrix assisted Laser Desorption/Ionization (MALDI) and ElectroSpray Ionization (ESI)[199,200]. MALDI involves embedding the biological molecules to be measured into a solid organic matrix. In the mass spectrometer, the analyte is excited along with matrix molecules (hence desorption) and ionized, followed by the measurement of its mass and charge. ESI involves the spraying of biological analytes with a liquid, which evaporates in the mass spectrometer, leaving the biological entities as gaseous molecules, and the mass and charge of these gaseous ions are measured. Prior to the generation of these techniques, biological entities were destroyed in the harsh vacuum environment of mass spectrometers before they could be measured[201].

Initially, mostly purified proteins were measured in the mass spectrometer for identification purposes, and information on the mass and charge of a few peptides was usually enough to identify a protein (mass fingerprinting)[203]. However, enormous leaps have been made regarding sample fractionation, preparation, measurement speed and sensitivity of modern mass spectrometers in addition to the generation of elegant algorithms for fast and accurate identification of peptides and proteins[201]. Nowadays, complex cellular mixtures from

mammalian cells are easily measured by fractionation (into 6-8 fractions) and within two days, 60-70% of the whole proteome is identified[204]. All of these advances have allowed three main areas of investigation in proteomics to be addressed by mass spectrometry-based techniques in a high throughput manner: Deep expression profiling (i.e. whole proteome analysis), interaction proteomics and PTM analysis[201].

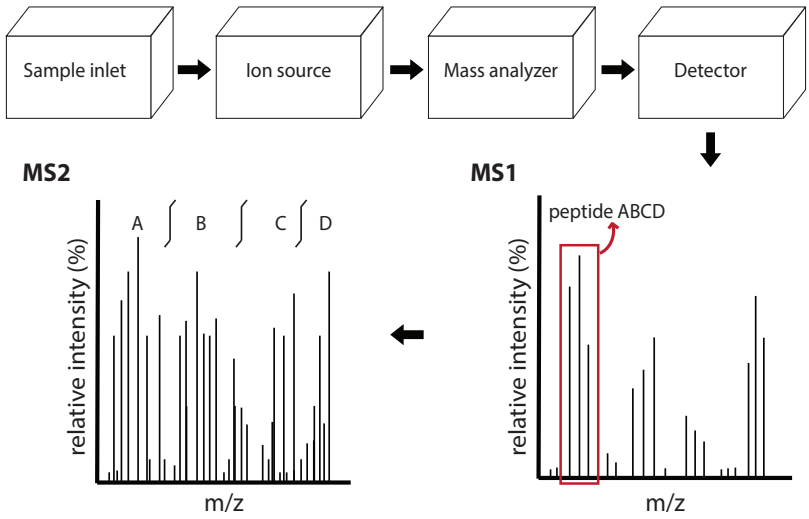


Figure 6: Generic components of a mass spectrometer. Precursor ions (e.g. peptide ABCD) are detected and isolated in MS1 and fragmented in MS2. Figure adapted from Yuan et al., 2014[202].

Typically, bottom-up proteomics is used to investigate all three questions. Bottom-up proteomics refers to digesting the proteome with a protease to generate peptides which are usually separated by on-line (directly connected to the mass spectrometer) nano-liquid chromatography, have their mass-to-charge ratio measured (MS1), fractionated into further ions via collision with an inert gas (MS/MS or MS2); and the information from both MS1 and MS2 scans is used to match the peptide ID to a database of peptides (where the proteins are theoretically digested with the protease that is used in the experiment to provide the mass and charge information)[205]. For large scale PTM analysis, in addition to this generic workflow, an enrichment of the PTM of interest may be necessary[206,207]. Alternatively, top-down proteomics could be used to investigate the PTMs on a particular protein or a protein complex. Top-down proteomics refers to using whole proteins for the MS1 stage, however, the necessity to purify proteins (or protein complexes) and the difficulty to ionize whole proteins as opposed to peptides are major drawbacks of this method[208]. The advantages of top-down proteomics is that PTMs on the same protein molecule could be inquired.

In recent years, another method that attempts to exploit the advantages of both bottom-up and top-down proteomics has been developed: Middle-down proteomics[209]. In this method, longer peptides are generated by 'selective' digestion with proteases. The longer peptides provide an elegant compromise between investigating intact proteins and looking at short peptic fragments: PTMs co-occuring proximally on the same peptide species can be identified. This method has been instrumental in investigating various histone PTMs that are impossible to detect with bottom-up methodologies and difficult to study in top-down

approaches (such as Histone H3K4 and K9).

So far, the identification aspect of mass spectrometry-based proteomics has been introduced. While it is essential to know the components of the proteome in different cells, studying a perturbation in the same cellular system requires quantification of proteins along with identification. Elegant quantification methods have enhanced interaction proteomics as well. However, as the famous saying goes, *"mass spectrometry is not inherently quantitative"*. Mass spectrometry is not inherently quantitative, because the ion intensities measured in MS1 and MS2 scans depend on the peptide composition which affects the ionization efficiency of the peptides. But the intensity, however loosely, *is* correlated with abundance, which is why some quantification methods use the ion intensity information and try to normalize that according to the size of the proteins[204,210,211]. However, there is no normalization trick that can fully make up for the differences of peptide composition that result in uneven 'mass spectrometric measurability'. Therefore, quantification methods that rely on the comparison of the same peptide species in different conditions give the most accurate results.

Quantification methods can be divided into 2 main branches[211,212]: 1. Methods based on labeling proteins or peptides: By integrating stable isotopes into proteins (Stable Isotope Labeling By Aminoacids in Cell Culture (see chapter 2 for details)) or peptides (dimethyl labeling, iTRAQ, etc.). 2. Label-free methods: Such as spectral counting, based on the number of MS1 scans of a peptide; or label free quantification (LFQ) algorithms based on measuring and comparing peptide intensities of proteins in 2 different states.

Importantly, all these quantification methods provide relative quantifications, comparing control vs perturbed states, or specific IPs to control IPs in the context of interaction proteomics. It should be noted that some label free methods could be utilized to gather absolute abundance information, but high accuracy absolute quantification methods rely on the spiking in of known amounts of proteins/peptides and comparing the measurement of related species to those of the spiked in entities[213,214].

In this thesis, a method to perform quantitative interaction proteomics on nuclear extracts from a population of cells is presented in chapter 2. In the rest of the chapters, quantitative interaction proteomics is performed to screen for interactors of proteins of interest (e.g. MBD5, MBD6, BAP1, ASXL2, CDK2AP1, MBD3, etc.) and of a histone PTM of interest (Histone H2AK119Ub) utilizing both SILAC (label-based) and label-free quantification methods.

Scope of this thesis

Chromatin regulatory complexes are part of epigenetic mechanisms that are required for the appropriate propagation of transcriptional control during self-renewal and differentiation of cells of an organism. The next chapters in this thesis focus on various chromatin regulatory complexes. A mass spectrometry-based quantitative proteomics approach for identifying nuclear protein-protein interactions is described in **chapter 2**. This method is utilized in **chapter 3** for the characterization of the MBD5 and MBD6 interactomes. Interestingly, the Methyl-CpG-binding domain (that has lost the affinity for meCpG) mediates the interaction with the PR-Dub complex for both proteins. In addition, we show that MBD6 is recruited to DNA damage sites upon microirradiation independently of the reported interaction with PR-Dub. In **chapter 4**, we characterize the H2AK118/119Ub interactome in a nucleosomal setting *in vitro*. Based on the interactors that include PRC2 members, a novel regulatory mechanism is proposed in which JARID2/AEBP2/PRC2 complex interacts with H2AK118/119Ub and catalyzes H3K27 methylation. In **chapter 5,** subunit dynamics and stability within the NuRD complex is investigated. Protein-protein interactions are quantified using label-free and SILAC-based methods. **Chapter 6** is a perspective on chromatin regulatory complexes that generally act to repress gene transcription. **Chapter 7** is a discussion of various aspects of these findings and several general questions in the epigenetics field.

References

1. Alberts, B. *et al. Molecular Biology of the Cell, 5th Edition*. *Shock* **30,** (2007).
2. Moreira, D. & López-García, P. Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* **7,** 306–11 (2009).
3. Becerra, A., Delaye, L., Islas, S. & Lazcano, A. The Very Early Stages of Biological Evolution and the Nature of the Last Common Ancestor of the Three Major Cell Domains. *Annu. Rev. Ecol. Evol. Syst.* **38,** 361–379 (2007).
4. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4576–4579 (1990).
5. CRICK, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **12,** 138–163 (1958).
6. Baltimore, D. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* **40,** 481–482 (1985).
7. Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13363–13383 (1998).
8. Gerstein, M. B. *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17,** 669–681 (2007).
9. Karthigeyan Dhanasekaran, Sujata Kumari, and C. K. in *Epigenetics: Development and Disease* **61,** 343–372 (2013).
10. Lee, J. Epigenetic Regulation by Long Noncoding RNAs. *Science (80-.).* **338,** 1435–1439 (2012).
11. Marchese, F. P. & Huarte, M. Long non-coding RNAs and chromatin modifiers: Their place in the epigenetic code. *Epigenetics* **9,** 21–26 (2014).
12. Wolpert, L. *et al. Principles of Development*. *Journal of Anatomy* **74,** (Current Biology, 1940).
13. Woodcock, C. L. & Ghosh, R. P. Chromatin Higher-order Structure and Dynamics. 1–25 (2010).
14. Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Curr. Opin. Genet. Dev.* **21,** 175–186 (2011).
15. Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13,** 436–447 (2012).
16. Crow, J. F., Dove, W. F. & Zacharias, H. Anecdotal , Historical And Critical Commentaries on Genetics. (1995).
17. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15,** 703–708 (2014).
18. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16,** 178–189 (2015).
19. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78,** 273–304 (2009).
20. Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154,** 490–503 (2013).
21. Kouzarides, T. SnapShot: Histone-Modifying Enzymes. *Cell* **131,** 10–11 (2007).
22. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
23. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23,** 781–783 (2009).
24. Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annu. Rev. Genet.* **43,** 559–599 (2009).
25. Thomas, M. C. & Chiang, C.-M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41,** 105–178 (2006).
26. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10,** 252–263 (2009).
27. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **16,** 155–166 (2015).
28. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in

metazoans. *Nat. Rev. Genet.* **13,** 720–31 (2012).
29. Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annu. Rev. Genet.* **47,** 483–508 (2013).
30. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 11–13 (2015).
31. Müller, F. & Tora, L. Chromatin and DNA sequences in defining promoters for transcription initiation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839,** 118–128 (2014).
32. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13,** 613–626 (2012).
33. Stadhouders, R. *et al.* Transcription regulation by distal enhancers: who's in the loop? *Transcription* **3,** 181–186 (2012).
34. De Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502,** 499–506 (2013).
35. Sexton, T. & Cavalli, G. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell* **160,** 1049–1059 (2015).
36. Koster, M. J. E., Snel, B. & Timmers, H. T. M. Genesis of Chromatin and Transcription Dynamics in the Origin of Species. *Cell* **161,** 724–736 (2015).
37. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16,** 190–202 (2015).
38. Schmitz, R. J. & Ecker, J. R. Epigenetic and epigenomic variation in Arabidopsis thaliana. *Trends Plant Sci.* **17,** 149–154 (2012).
39. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–20 (2013).
40. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–476 (2008).
41. Schübeler, D. Function and information content of DNA methylation. (2015). doi:10.1038/nature14192
42. Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74,** 481–514 (2005).
43. Bird, A. The dinucleotide CG as a genomic signalling module. *J. Mol. Biol.* **409,** 47–53 (2011).
44. Jones, P. a. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).
45. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat. Rev. Genet.* **13,** 153–162 (2012).
46. Burggren, W. W. Epigenetics as a source of variation in comparative animal physiology - or - Lamarck is lookin' pretty good these days. *J. Exp. Biol.* **217,** 682–9 (2014).
47. Maher, E. R. Genomic Imprinting. *J. Med. Genet.* **28,** 647 (1991).
48. Lee, H. J., Hore, T. a. & Reik, W. Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell* **14,** 710–719 (2014).
49. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511,** 611–615 (2014).
50. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–81 (2013).
51. Feldmann, A. *et al.* Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. *PLoS Genet.* **9,** (2013).
52. Krebs, A. R., Dessus-Babus, S., Burger, L. & Schübeler, D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *Elife* **3,** 1–18 (2014).
53. Baubec, T. & Schübeler, D. Genomic patterns and context specific interpretation of DNA methylation. *Curr. Opin. Genet. Dev.* **25,** 85–92 (2014).
54. Tiedemann, R. L. *et al.* Acute Depletion Redefines the Division of Labor among DNA Methyltransferases in Methylating the Human Genome. *Cell Rep.* **9,** 1554–1566 (2014).

55. Wu, H. & Zhang, Y. Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell* **156,** 45–68 (2014).
56. Guibert, S. & Weber, M. *Functions of DNA Methylation and Hydroxymethylation in Mammalian Development*. *Current Topics in Developmental Biology* **104,** (Copyright © 2013 Elsevier Inc. All rights reserved., 2013).
57. Booth, M. J. *et al.* Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* **336,** 934–937 (2012).
58. Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* **28,** 1106–1114 (2010).
59. Defossez, P. A. & Stancheva, I. *Biological functions of methyl-CpG-binding proteins*. *Progress in Molecular Biology and Translational Science* **101,** (Elsevier Inc., 2011).
60. Spruijt, C. G. & Vermeulen, M. P ERS P ECTI V E DNA methylation : old dog , new tricks ? *Nat. Publ. Gr.* **21,** 949–954 (2014).
61. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (80-.).* **311,** 844–7 (2006).
62. Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2013,** 1–16 (2013).
63. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69,** 905–914 (1992).
64. Nagase, T., Hirosawa, M., Ishikawa, K., Kikuno, R. & Ohara, O. Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* **7,** 143–150 (2000).
65. Ausió, J., Paz, A. M. De & Esteller, M. MeCP2: the long trip from a chromatin protein to neurological disorders. *Trends Mol. Med.* **20,** 1–12 (2014).
66. Fujita, N. *et al.* Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J. Biol. Chem.* **278,** 24132–24138 (2003).
67. Liu, C. *et al.* Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. *Cell Stem Cell* **6,** 433–444 (2010).
68. Li, X. *et al.* Epigenetic regulation of the stem cell mitogen Fgf-2 by Mbd1 in adult neural stem/progenitor cells. *J. Biol. Chem.* **283,** 27644–27652 (2008).
69. Clouaire, T., de las Heras, J. I., Merusi, C. & Stancheva, I. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Res.* **38,** 4620–4634 (2010).
70. Sarraf, S. A. & Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15,** 595–605 (2004).
71. Guezennec, X. Le *et al.* MBD2 / NuRD and MBD3 / NuRD , Two Distinct Complexes with Different Biochemical and Functional Properties. *Mol. Cell. Biol.* **26,** 843–851 (2006).
72. Spruijt, C. G. *et al.* Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
73. Sjolund, A. B., Senejani, A. G. & Sweasy, J. B. MBD4 and TDG: Multifaceted DNA glycosylases with ever expanding biological roles. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **743-744,** 12–25 (2013).
74. Rosenfeld, J. a *et al.* Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics* **10,** 143 (2009).
75. Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mescs. *Cell Stem Cell* **8,** 676–687 (2011).
76. Leung, D. *et al.* Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6690–5 (2014).

77. Xu, P.-F. *et al.* Setdb2 restricts dorsal organizer territory and regulates left-right asymmetry through suppressing fgf8 activity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2521–2526 (2010).
78. Guetg, C. *et al.* The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J.* **29,** 2135–2146 (2010).
79. Laget, S. *et al.* The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One* **5,** (2010).
80. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells* **11,** 805–814 (2006).
81. Wu, H. *et al.* Structural and histone binding ability characterizations of human PWWP domains. *PLoS One* **6,** (2011).
82. Motobayashi, M. *et al.* Neurodevelopmental features in 2q23.1 microdeletion syndrome: Report of a new patient with intractable seizures and review of literature. *Am. J. Med. Genet. Part A* **158 A,** 861–868 (2012).
83. Bonnet, C. *et al.* Extended spectrum of MBD5 mutations in neurodevelopmental disorders. *Eur. J. Hum. Genet.* **21,** 1457–61 (2013).
84. Mullegama, S. V *et al.* Reciprocal deletion and duplication at 2q23.1 indicates a role for MBD5 in autism spectrum disorder. *Eur. J. Hum. Genet.* **22,** 57–63 (2014).
85. Shichiji, M. *et al.* A cryptic microdeletion including MBD5 occurring within the breakpoint of a reciprocal translocation between chromosomes 2 and 5 in a patient with developmental delay and obesity. *Am. J. Med. Genet. Part A* **161,** 850–855 (2013).
86. Williams, S. R. *et al. Haploinsufficiency of MBD5 associated with a syndrome involving microcephaly, intellectual disabilities, severe speech impairment, and seizures. European journal of human genetics : EJHG* **18,** 436–441 (2010).
87. Jaillard, S. *et al.* 2q23.1 microdeletion identified by array comparative genomic hybridisation: an emerging phenotype with Angelman-like features? *Journal of medical genetics* **46,** 847–855 (2009).
88. Hodge, J. C. *et al.* Disruption of MBD5 contributes to a spectrum of psychopathology and neurodevelopmental abnormalities. *Mol. Psychiatry* **19,** 368–79 (2014).
89. Kleefstra, T. *et al.* Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am. J. Hum. Genet.* **91,** 73–82 (2012).
90. Du, Y. *et al.* The Essential Role of Mbd5 in the Regulation of Somatic Growth and Glucose Homeostasis in Mice. *PLoS One* **7,** 1–10 (2012).
91. Müller, J. & Verrijzer, P. Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr. Opin. Genet. Dev.* **19,** 150–158 (2009).
92. Lee, T. I. *et al.* Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* **125,** 301–313 (2006).
93. Whitcomb, S. J., Basu, A., Allis, C. D. & Bernstein, E. Polycomb Group proteins: an evolutionary perspective. *Trends Genet.* **23,** 494–502 (2007).
94. Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* **14,** 853–64 (2013).
95. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431,** 873–878 (2004).
96. De Napoles, M. *et al.* Polycomb group proteins ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7,** 663–676 (2004).
97. Müller, J. *et al.* Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111,** 197–208 (2002).
98. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–1043 (2002).
99. Eskeland, R. *et al.* Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol. Cell* **38,** 452–464 (2010).
100. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group

protein complex. *Science* **306,** 1574–1577 (2004).
101. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* **45,** 344–356 (2012).
102. Simon, J. a. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49,** 808–824 (2013).
103. Van Kruijsbergen, I., Hontelez, S. & Veenstra, G. J. C. Recruiting polycomb to chromatin. *Int. J. Biochem. Cell Biol.* (2015). doi:10.1016/j.biocel.2015.05.006
104. Kassis, J. a. & Brown, J. L. *Polycomb Group Response Elements in Drosophila and Vertebrates*. *Advances in Genetics* **81,** (Elsevier Inc., 2013).
105. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes Dev.* **17,** 1870–1881 (2003).
106. Min, J., Zhang, Y. & Xu, R. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27 Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. 1823–1828 (2003). doi:10.1101/gad.269603
107. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10,** 1291–1300 (2008).
108. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461,** 762–767 (2009).
109. Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell* **14,** 637–646 (2004).
110. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25,** 3110–3122 (2006).
111. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24,** 265–276 (2010).
112. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148,** 664–678 (2012).
113. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell. Biol.* **27,** 3769–3779 (2007).
114. Cooper, S. *et al.* Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Rep.* **7,** 1456–1470 (2014).
115. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157,** 1445–1459 (2014).
116. Gao, Z. *et al.* An AUTS2–Polycomb complex activates gene expression in the CNS. *Nature* **516,** 349–354 (2014).
117. Kondo, T. *et al.* Polycomb Potentiates Meis2 Activation in Midbrain by Mediating Interaction of the Promoter with a Tissue-Specific Enhancer. *Dev. Cell* **28,** 94–101 (2014).
118. Mousavi, K., Zare, H., Wang, A. H. & Sartorelli, V. Polycomb Protein Ezh1 Promotes RNA Polymerase II Elongation. *Mol. Cell* **45,** 255–262 (2012).
119. Frangini, A. *et al.* The Aurora B Kinase and the Polycomb Protein Ring1B Combine to Regulate Active Promoters in Quiescent Lymphocytes. *Mol. Cell* **51,** 647–661 (2013).
120. Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81,** 65–95 (2012).
121. Scheuermann, J. C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465,** 243–247 (2010).
122. Jensen, D. E. *et al.* BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16,** 1097–1112 (1998).
123. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12,** 68–78 (2011).
124. Nishikawa, H. *et al.* BRCA1-associated protein 1 interferes with BRCA1/BARD1 RING heterodimer activity. *Cancer Res.* **69,** 111–119 (2009).

125. Greenberg, R. a *et al.* Multifactorial contributions to an acute DNA damage response by BRCA1 / BARD1-containing complexes. *Genes Dev.* 34–46 (2006). doi:10.1101/gad.1381306.)
126. Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44,** 751–759 (2012).
127. Harbour, J. W. *et al.* Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330,** 1410–1413 (2010).
128. Abdel-Rahman, M. H. *et al.* Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. *J. Med. Genet.* **48,** 856–859 (2011).
129. Bott, M. *et al.* The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. *Nat. Genet.* **43,** 668–672 (2011).
130. Carbone, M. *et al.* BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs. *J. Transl. Med.* **10,** 179 (2012).
131. Carbone, M. *et al.* BAP1 and cancer. *Nat. Rev. Cancer* **13,** 153–9 (2013).
132. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339,** 1546–58 (2013).
133. Dey, a. *et al.* Loss of the Tumor Suppressor BAP1 Causes Myeloid Transformation. *Science (80-.).* **337,** 1541–1546 (2012).
134. Misaghi, S. *et al.* Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol. Cell. Biol.* **29,** 2181–2192 (2009).
135. Machida, Y. J., Machida, Y., Vashisht, A. a., Wohlschlegel, J. a. & Dutta, A. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J. Biol. Chem.* **284,** 34179–34188 (2009).
136. Yu, H. *et al.* The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell. Biol.* **30,** 5071–5085 (2010).
137. Pan, H. *et al.* BAP1 regulates cell cycle progression through E2F1 target genes and mediates transcriptional silencing via H2A monoubiquitination in uveal melanoma cells. *Int. J. Biochem. Cell Biol.* **60,** 176–184 (2015).
138. Tyagi, S., Chabes, A. L., Wysocka, J. & Herr, W. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Mol. Cell* **27,** 107–119 (2007).
139. Tyagi, S. & Herr, W. E2F1 mediates DNA damage and apoptosis through HCF-1 and the MLL family of histone methyltransferases. *EMBO J.* **28,** 3185–3195 (2009).
140. Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144,** 376–388 (2011).
141. Lazarus, M. B. *et al.* HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* **342,** 1235–9 (2013).
142. Lee, H.-S., Lee, S.-A., Hur, S.-K., Seo, J.-W. & Kwon, J. Stabilization and targeting of INO80 to replication forks by BAP1 during normal DNA synthesis. *Nat. Commun.* **5,** 5128 (2014).
143. Ji, Z. *et al.* The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Res.* **42,** 6232–6242 (2014).
144. Okino, Y., Machida, Y., Frankland-Searby, S. & Machida, Y. J. BRCA1-associated Protein 1 (BAP1) Deubiquitinase Antagonizes the Ubiquitin-mediated Activation of FoxK2 Target Genes. *J. Biol. Chem.* **290,** 1580–1591 (2015).
145. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* **12,** 599–612 (2012).
146. Park, U. H., Yoon, S. K., Park, T., Kim, E. J. & Um, S. J. Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor ?? *J. Biol. Chem.* **286,** 1354–1363 (2011).
147. Khan, F. F., Li, Y., Balyan, A. & Wang, Q. T. WTIP interacts with ASXL2 and blocks ASXL2-mediated activation of retinoic acid signaling. *Biochem. Biophys. Res. Commun.* **451,** 101–106 (2014).
148. Abdel-Wahab, O. *et al.* ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. *Cancer Cell* **22,** 180–193 (2012).

149. Lai, H. L. & Wang, Q. T. Additional Sex Combs-Like 2 Is Required for Polycomb Repressive Complex 2 Binding at Select Targets. *PLoS One* **8,** 1–13 (2013).
150. Lai, H. L. *et al.* Maintenance of adult cardiac function requires the chromatin factor Asxl2. *J. Mol. Cell. Cardiol.* **53,** 734–741 (2012).
151. Baskind, H. a. *et al.* Functional conservation of Asxl2, a murine homolog for the Drosophila Enhancer of trithorax and polycombgroup gene Asx. *PLoS One* **4,** (2009).
152. Yang, Z. *et al.* AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function. *Cell Res.* **20,** 276–287 (2010).
153. Lei, H. *et al.* KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. **461,** (2009).
154. Eletr, Z. M., Yin, L. & Wilkinson, K. D. BAP1 is phosphorylated at serine 592 in S-phase following DNA damage. *FEBS Lett.* **587,** 3906–3911 (2013).
155. Yu, H. *et al.* Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 285–90 (2014).
156. Ismail, I. H. *et al.* Germline mutations in BAP1 impair its function in DNA double-strand break repair. *Cancer Res.* **74,** 4282–4294 (2014).
157. Pickart, C. M. & Eddins, M. J. Ubiquitin: Structures, functions, mechanisms. *Biochimica et Biophysica Acta - Molecular Cell Research* **1695,** 55–72 (2004).
158. Reyes-Turcu, F. E., Ventii, K. H. & Wilkinson, K. D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu. Rev. Biochem.* **78,** 363–397 (2009).
159. Nijman, S. M. B. *et al.* A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123,** 773–786 (2005).
160. Mashtalir, N. *et al.* Autodeubiquitination protects the tumor suppressor BAP1 from cytoplasmic sequestration mediated by the atypical ubiquitin ligase UBE2O. *Mol. Cell* **54,** 392–406 (2014).
161. Fleming, A. & Osley, M. A. Silence of the rings. *Cell* **119,** 449–451 (2004).
162. Napoles, M. De *et al.* Polycomb Group Proteins Ring1A / B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Proteins* **7,** 663–676 (2004).
163. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* **9,** 1428–1435 (2007).
164. Zhu, P. *et al.* A Histone H2A Deubiquitinase Complex Coordinating Histone Acetylation and H1 Dissociation in Transcriptional Regulation. *Mol. Cell* **27,** 609–621 (2007).
165. Higashi, M., Inoue, S. & Ito, T. Core histone H2A ubiquitylation and transcriptional regulation. *Exp. Cell Res.* **316,** 2707–2712 (2010).
166. Zhou, W., Wang, X. & Rosenfeld, M. G. Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. *Int. J. Biochem. Cell Biol.* **41,** 12–15 (2009).
167. Bhatnagar, S. *et al.* TRIM37 is a new histone H2A ubiquitin liagse and breast cancer oncoprotein. *Nature* **516,** 116–120 (2014).
168. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. J. & Hiom, K. BRCA1 Is a Histone-H2A-Specific Ubiquitin Ligase. *Cell Rep.* **8,** 999–1005 (2014).
169. Zhu, Q. *et al.* BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477,** 179–184 (2011).
170. Smeenk, G. & van Attikum, H. The chromatin response to DNA breaks: leaving a mark on genome integrity. *Annu. Rev. Biochem.* **82,** 55–80 (2013).
171. Belle, J. I. & Nijnik, A. H2A-DUBbing the mammalian epigenome: Expanding frontiers for histone H2A deubiquitinating enzymes in cell biology and physiology. *Int. J. Biochem. Cell Biol.* **50,** 161–174 (2014).
172. Joo, H.-Y. *et al.* Regulation of cell cycle progression and gene expression by H2A deubiquitination. *Nature* **449,** 1068–1072 (2007).
173. Nicassio, F. *et al.* Human USP3 Is a Chromatin Modifier Required for S Phase Progression and Genome Stability. *Curr. Biol.* **17,** 1972–1977 (2007).
174. Wang, T. *et al.* The control of hematopoietic stem cell maintenance, self-renewal, and

differentiation by Mysm1-mediated epigenetic regulation. *Blood* **122,** 2812–22 (2013).
175. Nishi, R. *et al.* Systematic characterization of deubiquitylating enzymes for roles in maintaining genome integrity. *Nat. Cell Biol.* (2014). doi:10.1038/ncb3028
176. Wang, Y. *et al.* LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer. *Cell* **138,** 660–672 (2009).
177. Whyte, W. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482,** 221–225 (2012).
178. Menafra, R. *et al.* Genome-wide binding of MBD2 reveals strong preference for highly methylated loci. *PLoS One* **9,** (2014).
179. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153,** 480–92 (2013).
180. Zhang, J. *et al.* Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat. Immunol.* **13,** 86–94 (2012).
181. Ramdzan, Z. M. & Nepveu, A. CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. *Nat. Rev. Cancer* **14,** 673–682 (2014).
182. Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138,** 1019–31 (2009).
183. Métivier, R. *et al.* Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115,** 751–763 (2003).
184. Reynolds, N., O'Shaughnessy, A. & Hendrich, B. Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140,** 505–512 (2013).
185. Ding, J., Xu, H., Faiola, F., Ma'ayan, A. & Wang, J. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Research* **22,** 155–167 (2012).
186. Pardo, M. *et al.* An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. *Cell Stem Cell* **6,** 382–395 (2010).
187. Van den Berg, D. L. C. *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6,** 369–381 (2010).
188. Dos Santos, R. L. *et al.* MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. *Cell Stem Cell* **15,** 102–110 (2014).
189. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* **15,** 710–723 (2001).
190. Kaji, K., Nichols, J. & Hendrich, B. Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* **134,** 1123–1132 (2007).
191. Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14,** 735–751 (2014).
192. Lai, A. Y. & Wade, P. a. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat. Rev. Cancer* **11,** 588–596 (2011).
193. Chou, D. M. *et al.* A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 18475–18480 (2010).
194. Polo, S. E., Kaidi, A., Baskcomb, L., Galanty, Y. & Jackson, S. P. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J.* **29,** 3130–3139 (2010).
195. Smeenk, G. *et al.* The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. *J. Cell Biol.* **190,** 741–749 (2010).
196. O'Shaughnessy, A. & Hendrich, B. CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. *Biochem. Soc. Trans.* **41,** 777–82 (2013).
197. Gong, F. *et al.* Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. *Genes Dev.* **29,** 197–211, (2015).
198. Junker, J. P. & Van Oudenaarden, A. Every cell is special: genome-wide studies add a new

dimension to single-cell biology. *Cell* **157,** 8–11 (2014).
199. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60,** 2299–2301 (1988).
200. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-.).* **246,** 64–71 (1989).
201. Marco Y. Hein, Kirti Sharma, J. C. and M. M. in *Handbook of Systems Biology: Concepts and Insights* 3–25 (2012).
202. Yuan, Z.-F., Arnaudo, A. M. & Garcia, B. a. Mass spectrometric analysis of histone proteoforms. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* **7,** 113–28 (2014).
203. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. **422,** (2003).
204. Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80,** 273–299 (2011).
205. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5,** 699–711 (2004).
206. Udeshi, N. D. *et al.* Refined Preparation and Use of Anti-diglycine Remnant (K- -GG) Antibody Enables Routine Quantification of 10,000s of Ubiquitination Sites in Single Proteomics Experiments. *Mol. Cell. Proteomics* **12,** 825–831 (2012).
207. Olsen, J. V & Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics* **12,** 3444–52 (2013).
208. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications* **445,** 683–693 (2014).
209. Garcia, B. a. What Does the Future Hold for Top Down Mass Spectrometry? *J. Am. Soc. Mass Spectrom.* **21,** 193–202 (2010).
210. Vermeulen, M. *et al.* Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142,** 967–980 (2010).
211. Wasinger, V. C., Zeng, M. & Yau, Y. Current status and advances in quantitative proteomic mass spectrometry. *Int. J. Proteomics* **2013,** 180605 (2013).
212. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389,** 1017–31 (2007).
213. Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J. Proteome Res.* **7,** 1118–1130 (2008).
214. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).

1

Chapter 2

Identifying nuclear protein-protein interactions using GFP affinity purification and SILAC-based quantitative mass spectrometry

H. Irem Baymaz[1], Cornelia G. Spruijt[1] and Michiel Vermeulen[1#]

1. Molecular Cancer Research and Cancer GenomiCs Netherlands, University Medical Center Utrecht, The Netherlands
Present address: Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands

Corresponding author

Methods in Molecular Biology, 2014, vol. 1188, pages 207-226

Abstract
Many cellular proteins assemble into macromolecular protein complexes. Therefore, identifying protein-protein interactions (PPIs) is essential to gain insight into the function of proteins. Recently established quantitative mass spectrometry-based techniques have significantly improved the unbiased search for PPIs. In this chapter, we describe a single-step GFP affinity purification method combined with SILAC-based quantitative mass spectrometry that can be used to identify nuclear PPIs in mammalian cells.

1. Introduction

Proteins drive all processes in a cell through a complex and dynamic network of protein-protein interactions (PPIs). Identifying these interactions is therefore essential to gain insight into the function of proteins. There are several approaches available to identify cellular protein-protein interactions. When putative interactors are known, co-immunoprecipitation experiments followed by Western blot analysis can be used to validate true interactors among the list of candidates. However, it is quite common that potential interactors of a protein are not known. In this case, unbiased interaction screening approaches are needed. One such approach is the yeast-2-hybrid (Y2H) system which identifies interactors of a protein of interest using a library of 'prey' proteins.[1] The major disadvantage of this method, however, is that Y2H tends to result in many false positive interactions. Another limitation is that mammalian proteins expressed in yeast may lack some post-translational modifications that mediate PPIs. Furthermore, Y2H only detects direct protein-protein interactions and thus cannot be used to determine all components of larger protein complexes. An alternative to Y2H is the (tandem) affinity purification of the protein of interest followed by mass spectrometry (AP-MS).[2] This method requires extensive purification of the protein of interest (the bait) under high stringency conditions to minimize non-specifically interacting proteins. However, even when using very stringent conditions, high-abundant background binders are not completely removed. This problem is particularly relevant when making use of modern mass spectrometers, which are very sensitive and which can sequence proteins when present in a sample in femtomole amounts.[3] As a consequence, scientists still need to resort to other methods such as those mentioned above to distinguish true interactors from background binders. Therefore, although AP-MS is robust in identifying the proteins that are co-purified with the bait, the fact that interactions are not quantified compromises the ability to discriminate true interactors from background binders. In addition, the high stringency conditions can result in the loss of relatively weak but biologically relevant interactions. To overcome these problems, quantitative mass spectrometry methods have been established in recent years.[4] In most of these methods, differential stable isotopic labeling, either on protein or peptide level, is used in the specific and the control affinity purification. Prior to mass spectrometry analysis, the specific and control pull-downs are combined. Each peptide then has a 'light' and a 'heavy' intensity, and the ratio between these two states indicates the relative abundance of a peptide and the corresponding protein in the specific and control affinity purification. As a result, the bait and its interactors have a high ratio, whereas background proteins have a ratio close to one. Here we describe an application of this principle, a SILAC-based GFP affinity purification method from mammalian nuclear extracts. The method starts with SILAC labeling of cells that (either stably or transiently) express the GFP-tagged protein of interest.[5] As a control, wild-type cells lacking the GFP-tagged bait are labeled in parallel. Nuclear extracts generated from these cells are then used for GFP affinity purifications followed by quantitative mass spectrometry to identify and quantify protein-protein interactions.

2. Materials

For preparing buffers, ultrapure water (18.5 MΩ cm resistance, total organic carbon < 12 parts per billion) which will be referred to as Milli-Q is used. In addition, to avoid polymer accumulation in samples, do not use autoclaved pipette tips and keep all buffers in high quality glass bottles. Tabletop centrifuges with cooling capacity for 50-ml tubes and microcentrifuge tubes are required throughout the protocol.

2.1 SILAC Labeling

1. SILAC Dulbecco's Modified Eagle Medium (DMEM) lacking arginine and lysine
2. L-arginine monohydrochloride light and heavy ($^{13}C_6{}^{15}N_4$) each dissolved to a final concentration of 84 mg/ml in Milli-Q
3. L-lysine monohydrochloride light and heavy ($^{13}C_6{}^{15}N_2$) each dissolved to a final concentration of 146 mg/ml in Milli-Q
4. Dialyzed fetal bovine serum (D-FBS)
5. 200 mM L-glutamine
6. 100 units/ml Penicillin/Streptomycin
7. 50 ml syringes and 0.22 µm filters
8. Dulbecco's Phosphate Buffered Saline (PBS): 0.2 g/L KCl, 0.2 g/L KH_2PO, 8 g/L NaCl, 2.16 g/L $HNa_2O_4P.7H_2O$
9. Trypsin-EDTA: 200 mg/L EDTA, 170.000 units/L Trypsin
10. 94x16 mm and 145x20 mm cell culture dishes (referred to as 10-cm and 15-cm dishes respectively)

Optional (see Section 3.1):

11. SILAC RPMI medium lacking arginine and lysine
12. 100x Non-essential amino acids
13. 2i inhibitors (Axon Medchem, CHIR99021 and PD0325901)
14. Leukemia inhibitory factor (LIF, 1000000 units/ml)
15. β-mercaptoethanol
16. 100 mM Sodium pyruvate
17. Accutase

2.2 Transient Transfection of SILAC-labeled Cells

Polyethylenimine, linear (PEI, Polysciences), dissolved to a final concentration of 1 mg/ml in Milli-Q and neutralized with HCl

2.3 Nuclear Extraction

1. PBS
2. SILAC DMEM
3. Trypsin-EDTA
4. Glass dounce homogenizer with type B pestle (tight); depending on the amount of cells, different sizes can be used: 500 µl, 2 ml, 7 ml, or 40 ml
5. Buffer A: 10 mM KCl, 1.5 mM $MgCl_2$, 10 mM Hepes-KOH pH 7.9
6. Buffer C: 420 mM NaCl, 20 mM Hepes-KOH pH 7.9, 20% (v/v) glycerol, 2 mM $MgCl_2$, 0.2 mM EDTA
7. Nonidet P40 (NP-40), 10% stock solution
8. Dithiotreitol (DTT), 500 mM stock solution
9. Complete EDTA-free protease inhibitor cocktail tablets (CPI; Roche, 1 tablet/ml = 50x

stock solution)
Optional (see Section 3.3):
10. 5 M NaCl
11. 100% Glycerol

2.4 Bradford Assay For Protein Quantification

1. Albumin from bovine serum (BSA) dissolved in Milli-Q to a concentation of 1 mg/ml
2. BioRad Protein Assay Dye Reagent Concentrate (BioRad, 5x)

2.5 GFP Affinity Purification

1. GFP-binder beads (e.g. GFP-Trap_A from Chromotek)
2. Buffer C*: 300 mM NaCl, 20 mM Hepes-KOH pH 7.9, 20% (v/v) glycerol, 2 mM $MgCl_2$, 0.2 mM EDTA
3. PBS
4. NP-40, 10% stock solution
5. DTT, 500 mM stock solution
6. Complete EDTA-free protease inhibitor cocktail tablets
7. Ethidium bromide, 10 mg/ml stock solution
8. Gel-loader tips

Optional (see Section 3.5):
Blocked agarose beads (e.g. from Chromotek)

2.6 On-Bead Digestion of Proteins

1. Digestion buffer: 2 M urea dissolved in 100 mM Tris-HCl pH 7.5
2. DTT, 500 mM stock solution
3. Iodoacetamide (IAA) or chloroacetamide (CAA) dissolved to a final concentration of 550 mM in 50 mM ammonium bicarbonate
4. Trypsin dissolved to a final concentration of 0.1 mg/ml in 50 mM acetic acid
5. Thermoshaker

2.7 Desalting and Purification of Peptides for Mass Spectrometry

1. Blunt-ended syringe needle (1.2 mm diameter) with a nanotubing end inserted as a plunger
2. C18 material (Empore)
3. Methanol (ultra pure)
4. Buffer A: 0.5% (v/v) acetic acid in Milli-Q
5. Buffer B: 0.5% (v/v) acetic acid, 80% (v/v) acetonitrile (HPLC grade) in Milli-Q
6. 10% (v/v) trifluoroacetic acid (TFA) in Milli-Q

2.8 Elution of Peptides

1. Buffer A: 0.5% (v/v) acetic acid in Milli-Q
2. Buffer B: 0.5% (v/v) acetic acid, 80% (v/v) acetonitrile in Milli-Q
3. Eppendorf Combitip plus 2.5 ml
4. Speed vacuum concentrator (with 96-well plate compatible rotor)
5. HPLC autosampler 96 well plate

2.9 Mass Spectrometry

1. EASY nLC (Thermo Scientific)
2. High performance mass spectrometer (e.g. LTQ-Orbitrap Velos; Thermo Scientific)

2.10 Data Analysis

1. Windows-operated PC (at least 8 GB of RAM and multiple cores are recommended)
2. MaxQuant software package

3. Methods

3.1 SILAC Labeling

The most commonly used metabolic labeling strategy is SILAC (Stable Isotope Labeling by Amino acids in Cell culture).[5] In this method, cells are grown in the presence of normal (light) or heavy stable isotopic versions of certain amino acids (usually arginine and lysine). During cell culture, these light and heavy amino acids are incorporated into the proteins, enabling relative quantification of proteins between two functional states. The workflow described in this chapter consists of a so-called 'forward' and 'reverse' pull-down. In the forward pull-down, the cell line stably expressing a GFP-tagged transgene (referred to as GFP cells) is labeled 'heavy' and the corresponding wild-type (WT) cell line is labeled 'light'. In the reverse pull-down, the light and heavy labels are swapped. Therefore, four different cultures have to be labeled and expanded (see **Figure 1a**, **Notes 1 and 2**).

1. Prepare a bottle of 'heavy' and a bottle of 'light' medium. Once made, SILAC media can be kept for up to six weeks at 4°C. For each condition (light and heavy):
   a. Take a bottle of DMEM lacking arginine, lysine, and glutamine (see **Note 3**).
   b. Transfer 20--30 ml of this medium into a 50-ml tube and add 15 mg of light or heavy arginine and 36.5 mg of light or heavy lysine (see **Note 4**).
   c. Filter-sterilize these aliquots using a syringe and a 0.22 µm filter back into the DMEM bottle.
   d. Add 50 ml of D-FBS, 2 mM of L-glutamine, and 550 units of penicillin/streptomycin.
2. Trypsinize the WT and GFP cells, neutralize trypsin by adding DMEM and divide each cell suspension into two 15-ml tubes.
3. Spin down the cells at 400 x *g* for 5 minutes (min). This step is necessary to remove trypsin from the cells. Trypsin provides a source of non-labeled amino acids and residual trypsin can therefore compromise SILAC labeling.
4. Aspirate the supernatant and resuspend the cells in light or heavy medium (WT cells and GFP cells are both labeled heavy and light). Transfer 25% of the cells of each tube into a 10-cm culture dish and add the appropriate amount of medium (heavy or light) to each dish.
5. Culture cells for at least eight doublings in SILAC medium. Keep in mind that some cell lines tend to grow more slowly in SILAC medium compared to normal DMEM due to the use of D-FBS. While labeling, cells can be expanded in order to end up with the appropriate amount of labeled cells after eight cell doublings. Although it depends on the cell line, a confluent 15-cm dish usually yields around 400 µg of nuclear extract.

3.2 Transient Transfection of SILAC-labeled Cells

When a stable cell line expressing a GFP-tagged protein of interest is not available, transient transfections can be considered (see **Note 5**). In this case, WT cells are light and heavy labeled and expanded. Half of the light and half of the heavy cells are then transfected with the plasmid expressing the GFP-tagged bait, while the other two halves are transfected with a control plasmid (empty GFP plasmid). In the end, there are four cell populations (light control, light GFP, heavy control, heavy GFP; see **Figure 1b**). The method described below requires a total of 20 x 15-cm dishes, but this scale may be adjusted according to the amount of nuclear extract that is needed.

1. Expand light and heavy labeled cells to the required amount, in this case 10 x 15-cm dishes of light labeled cells and 10 x 15-cm dishes of heavy labeled cells. When cells reach about 60% confluency (see **Note 6**), half of the heavy and half of the light labeled cells (5 x 15-cm dishes of light and 5 x 15-cm dishes of heavy cells) are transfected with the GFP-tagged bait while the other half is transfected with the control GFP plasmid as follows:
2. Transfer 15 ml of DMEM without lysine, arginine, and glutamine (i.e. SILAC DMEM without anything added to it) into a 50-ml tube for each plasmid (bait and control).
3. Add 150 µg of the plasmid (bait or control) and 450 µl of PEI to each tube (see **Note 7**). Vortex for 10 s and incubate at room temperature (RT) for 30 min.
4. Pipette 1.5 ml of the transfection mix dropwise into each of five 15-cm dishes containing light cells and five 15-cm dishes containing heavy cells. Make sure to label the dishes.
5. Culture the cells for an additional 24-48 h at 37°C. In the end, there are four batches of cells to harvest: control light and heavy, GFP light and heavy. Prior to harvesting, expression of the GFP transgene can be checked by fluorescence microscopy.

3.3 Nuclear Extraction

The next step in the workflow is the generation of nuclear extracts (see **Note 8**). The following method is based on Dignam *et al.* and is suitable for cells harvested from multiple 15-cm dishes.[6] However, the method can be adjusted to smaller or larger scale cultures. The generation of nuclear extracts is a critical step in the procedure. Differential nuclear extraction of different batches of cells results in a variation of individual protein abundance and this introduces more noise into the experiment (i.e. background protein ratios strongly deviating from 1). Therefore, the different batches of cells should be extracted equally.

1. Wash the cells once with 15 ml of PBS.
2. Add 2 ml of trypsin-EDTA to each dish and incubate at 37°C for about 5 min. Long incubation with trypsin results in cell lysis. Do not trypsinize more than 10 dishes simultaneously.
3. Neutralize trypsin with 10 ml of medium and collect the cells of each batch in a 50-ml tube. Keep the cells on ice from now on.
4. Rinse the dishes with PBS to collect the remaining cells.
5. Centrifuge the cells at 400 x *g* for 5 min at 4°C and wash them with 20 ml of PBS. Centrifuge at 400 x *g* for 5 min at 4°C again.
6. Resuspend the pellet with 15 ml of ice-cold PBS and transfer the cell suspension to a 15-ml tube. It is possible to leave the cells in this state for about 30 min.
7. Centrifuge at 400 x *g* for 5 min at 4°C.
8. Aspirate the supernatant and estimate the volume of the pellet. Resuspend the pellet in 5 volumes of ice-cold buffer A.
9. Incubate the cell suspension on ice for 10 min. Cells swell during this incubation due to the osmotic uptake of water. The extent of swelling, however, varies between different cell lines. HeLa cells, for example, almost double their volume whereas HEK293T cells hardly swell.
10. Centrifuge for 5 min at 400 x *g* at 4°C.
11. Aspirate the supernatant, resuspend cells in 2 volumes of buffer A containing 1x CPI and 0.15% NP-40 and transfer the suspension to a dounce homogenizer. Keep the dounce homogenizer on ice at all times.

12. Dounce for 30-40 times. Wait for 30-60 seconds after every ten strokes of douncing to minimize the temperature increase due to friction.
13. Transfer the suspension into a 15-ml tube and centrifuge for 15 min at 3,200 x *g* at 4°C.
14. The supernatant is the cytoplasmic extract, which can be aliquoted (See **Note 9**), snap-frozen and kept for other purposes. The pellet consists of crude nuclei.
15. Gently add 5 volumes of ice-cold PBS to the pellet and detach the pellet from the tube by flicking it a few times. Do not resuspend the pellet since this will result in partial lysis of the nuclei.
16. Centrifuge for 5 min at 3,200 x *g* at 4°C. Discard the supernatant and remove any residual liquid by placing the tubes upside down on a tissue paper for 1 min.
17. Estimate the volume of the nuclei and add 2 volumes of buffer C containing 1x CPI, 0.1% NP-40 and 0.5 mM DTT. Resuspend the nuclei and transfer the suspension to a microcentrifuge tube.
18. Homogenize the nuclei by pipetting up and down 10-15 times.
19. Incubate the suspension at 4°C on a rotating wheel for 60 min.
20. Centrifuge the microcentrifuge tubes for 40 min in a pre-cooled tabletop centrifuge at maximum speed (about 17,900 x *g*) at 4°C.

Aliquot the supernatant (100-200 µl fractions), which is the nuclear extract, snap-freeze in liquid nitrogen and store at -80°C. Multiple freeze-thaw cycles should be avoided to preserve the quality of the nuclear extract. The pellet represents the insoluble chromatin fraction which could also be snap-frozen and kept for other purposes.

3.4 Bradford Assay for Protein Quantification

1. Dilute the extracts 1:10 in Milli-Q to a final volume of 20 µl.
2. Take 4 and 10 µl aliquots from the 1:10 dilution and add 1 ml of 1x Bradford protein assay solution to each sample.
3. Take 0, 1, 2, 5, 7, and 10 µl aliquots of 1 mg/ml BSA and add 1 ml of 1x Bradford protein assay solution to each sample.
4. Measure the samples at 595 nm using a spectrophotometer.
5. Generate a standard curve using the BSA samples and calculate the protein concentration of the extracts using linear regression. Note that the extracts were diluted 1:10 in step 1.

3.5 GFP Affinity Purification

The next step in the workflow is the GFP affinity purification (see **Notes 10 and 11**). Four extracts have been generated in section 3.3: light and heavy WT extracts and light and heavy GFP extracts (see **Note 12**). These four extracts are incubated with GFP-binder beads and then combined after the incubation and wash steps to generate the forward (WT light + GFP heavy) and reverse (WT heavy + GFP light) experiment.

1. Add 15 µl of GFP-binder beads (from a 50% slurry) into 4 microcentrifuge tubes. Cut the tips of 200 µl pipette tips when handling the agarose beads. Centrifuge the beads at 1500 x *g* for 2 min after each wash and never vortex the beads.
2. Wash the beads three times with 1 ml of buffer C* containing 1x CPI, 0.1% NP-40 and 0.5 mM DTT (see **Note 13**). Beads are washed by adding buffer, inverting the tubes 5 times and centrifugation.
3. Aspirate the supernatant. For each of the four extracts, calculate the volume for 1 mg of protein. The affinity purification is performed in a final volume of 400 µl that

contains the nuclear extract, buffer C* and ethidium bromide at a final concentration of 50 µg/ml (see **Notes 14 and 15, Figures 2 and 3**). First add the necessary amount of buffer C* to the beads, followed by the ethidium bromide and finally the nuclear extract. If the amount of nuclear extract to be added exceeds 400 µl, the final volume can be increased accordingly but it should be kept constant for all the samples.

4. Incubate the samples on a rotating wheel for 90 min at 4°C.

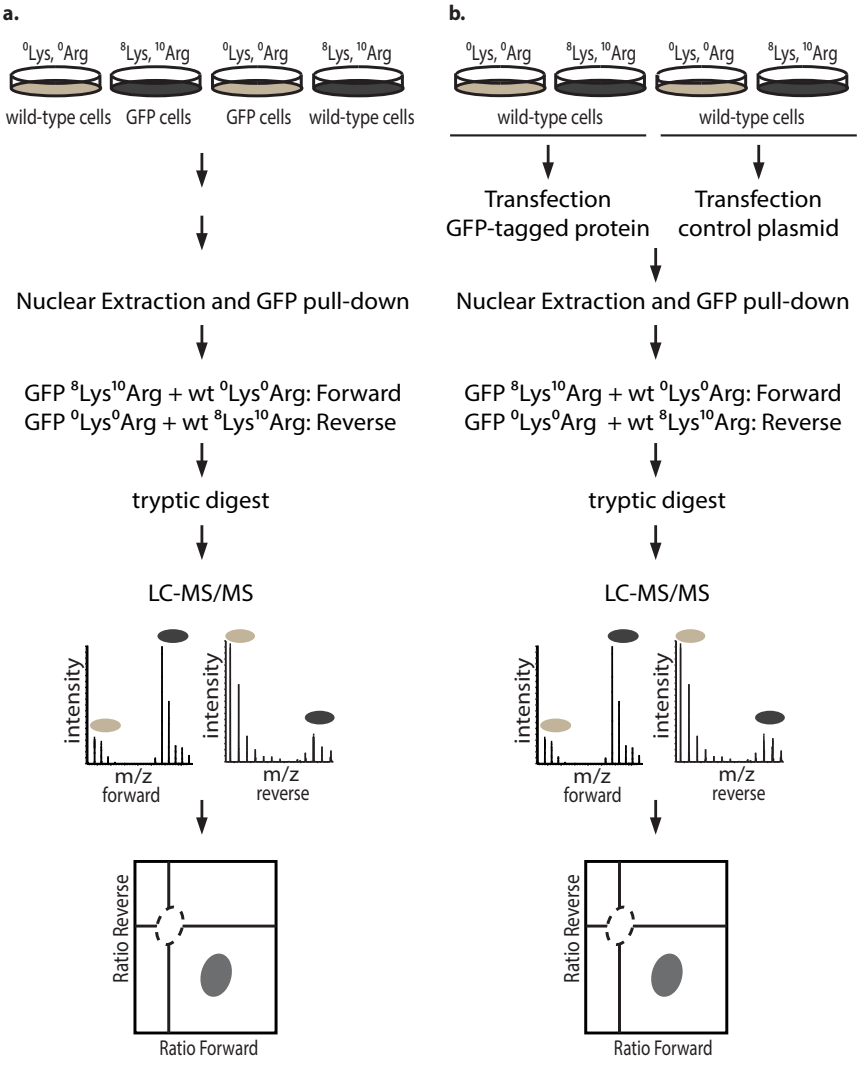


Figure 1: Schematic representation of the method including the SILAC labeling, nuclear extraction, and GFP-affinity purification followed by quantitative mass spectrometry. **a.** The workflow for cells stably expressing the GFP-tagged protein of interest and the corresponding wild-type cells. **b.** The workflow for transient expression of a GFP-tagged bait in SILAC-labeled cells.

5. Centrifuge. Wash the beads twice with 1 ml of buffer C* containing 1x CPI, 0.5% NP-40 and 0.5 mM DTT (see **Note 16**).
6. Wash the beads twice with 1 ml of PBS containing 0.5% NP-40.
7. Wash the beads once with 1 ml of PBS.
8. Add 1 ml of PBS to the light WT beads and transfer the beads to the microcentrifuge tube containing the GFP heavy beads. This is the forward experiment. Combine the heavy WT and light GFP beads in a similar way. This is the reverse experiment.
9. Centrifuge the samples and aspirate the supernatant completely using a gel-loader tip.

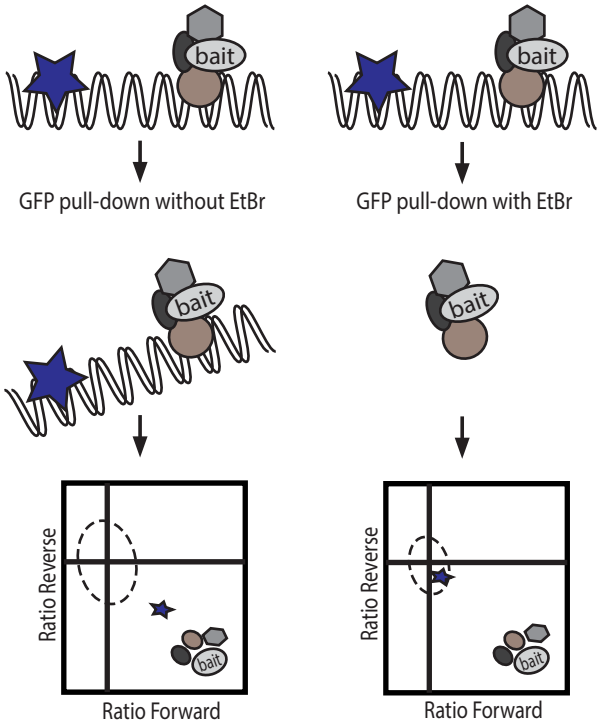


Figure 2: Preventing DNA-mediated protein interactions. The presence of ethidium bromide during the GFP affinity purification eliminates co-purification of proteins bound to DNA in close proximity to the bait and its interactors. In addition, the cloud of background proteins is less tight in the absence of ethidium bromide.

3.6 On-bead Digestion of Proteins

At the end of section 3.5, two samples remain (forward pull-down and reverse pull-down). The next step is digesting the proteins off the beads with trypsin. This method is adapted from Hubner *et al*.[7,8]

Always prepare urea solutions fresh before use and never cool the solutions to avoid urea precipitation. Furthermore, do not heat urea-containing samples to avoid adduct formation. Thaw and keep trypsin on ice at all times to minimize self-digestion.

1. Add 50 µl of digestion buffer to the samples.
2. Add DTT to a final concentration of 10 mM and shake the samples at 1,400 rpm on a thermoshaker for 20 min at RT.
3. Add IAA or CAA to a final concentration of 50 mM and incubate the samples for 20 min shaking at 1,400 rpm on a thermoshaker. Note that the stock solutions and IAA-containing samples need to be kept in the dark. CAA should be used if a post-translational modification analysis (of ubiquitin in particular) will be performed afterwards. IAA can generate adducts on lysine residues that mimic ubiquitination.[9]
4. Add 2.5 µl of trypsin to the beads. Incubate for 2 h on the shaker at 1,400 rpm at RT.
5. Centrifuge the samples at 2,000 x *g* for 2 min and transfer the supernatants into new

microcentrifuge tubes (avoid taking up any beads). To the remaining beads, add 50 µl of digestion buffer and incubate at RT for 5 min while shaking.

6. Centrifuge the beads, collect the supernatant and add it to the one collected in the previous step for each sample.
7. Add 1 µl of trypsin to the combined supernatant and incubate overnight at RT.

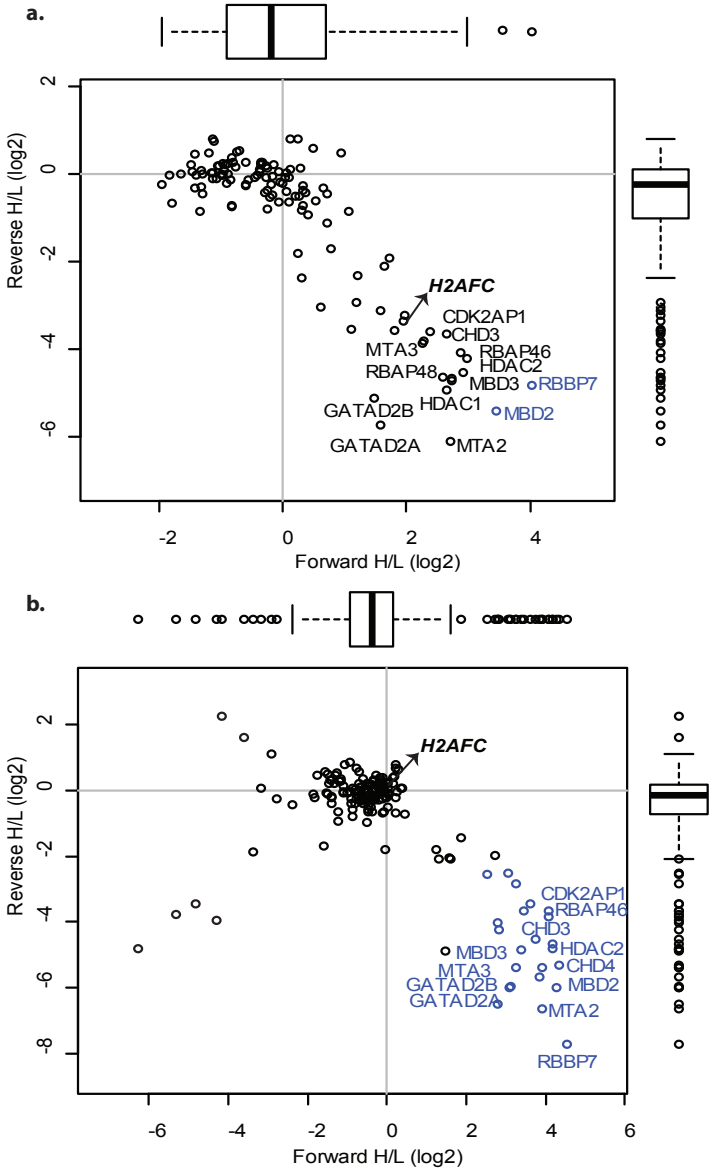


Figure 3: GFP-CDK2AP1 affinity purifications in the absence or presence of ethidium bromide using wild-type and GFP-tagged CDK2AP1-expressing HeLa cells. In both panels, known interactors of CDK2AP1 (NuRD complex subunits) are indicated and statistically significant interactors are shown in blue. **a.** The affinity purification was performed without ethidium bromide. A wide spreading of the background cloud and interactors is observed in the scatterplot. This spreading affects the boxplot statistics used for the significance calculation of interactors. In this pull-down, H2AFC, a histone variant, shows a high forward and low reverse ratio clustering with known CDK2AP1 interactors. **b.** The affinity purification was performed in the presence of ethidium bromide. In this case, the background cloud is more compact and tightly clustered around the origin of the scatterplot. As a consequence, NuRD complex subunits, which are known interactors of CDK2AP1, are significant outliers. Note that in this pull-down, H2AFC is identified as a background protein.

3.7 Desalting and Purification of Peptides for Mass Spectrometry

Following digestion of the proteins, the tryptic peptides have to be desalted prior to mass spectrometry analysis. This method is adapted from Rappsilber *et al.*[10,11] In brief, C18 material inserted into a p200 pipette tip (referred to as StageTips) is used to capture and purify peptides. After each centrifugation step, check the StageTips to ensure that all the liquid has flowed through before proceeding with the next step.

1. For each sample, puncture out a small disc of double-layered C18 material using a blunt-ended syringe needle and transfer the C18 material into a p200 pipette tip. Push the disc into the tip and fix it at the end but do not apply too much pressure.
2. Insert the tip into a microcentrifuge tube with a hole in its cap (see **Figure 4**). The pipette tip on the tube should be stable enough but it should not touch the bottom of the tube where the flow-through solvents will be collected.
3. Activate the C18 material by adding 50 µl of methanol on top of it and centrifuge at 1,500 x g for 3-5 min in a tabletop centrifuge. Check the tip to make sure that all the methanol has flowed through.
4. Add 50 µl of buffer B to the tip and centrifuge at 1,500 x g for 3-5 min. Discard the flow-through.
5. Load 50 µl of buffer A to the tip and centrifuge at 1,500 x g for 3-5 min.
6. Repeat step 5.
7. To each of the samples that were digested overnight with trypsin (section 3.6), add 10 µl of 10% TFA and resuspend. Centrifuge the samples at 2,000 x g for 2 min. Make sure not to take up any residual beads that may have been carried over when supernatants were collected in the previous section.
8. Load the samples onto the (labeled) StageTips and centrifuge at 400 x g for 10-15 min. It is important to centrifuge more slowly during this step to ensure efficient binding of peptides to the C18 plug.
9. When all the sample has flowed through the StageTip, load 50 µl of buffer A and centrifuge at 1,500 x g for 3-5 min. The StageTips can be stored at 4°C for months at this point.

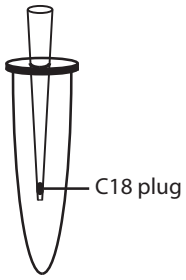


Figure 4: Schematic representation of a p200 pipette tip with a C18 plug inserted (StageTip) mounted on a 2-ml microcentrifuge tube.

3.8 Elution of Peptides

1. When proceeding directly after desalting and concentration of peptides using StageTips, start from step 2 of this section. If the samples have been stored at 4°C on StageTips, rehydrate StageTips by loading 30 µl of buffer A and centrifuge at 1,500 x *g* for 3-5 min.
2. Elute the peptides into a 96-well autosampler plate that is compatible with the nanoHPLC connected to the mass spectrometer. To elute, add 30 µl of buffer B and fit an air-filled Eppendorf combitip to the back of the StageTip. Apply pressure on the combitip to slowly force the solvent through the StageTip directly into the autosampler plate.
3. Concentrate the eluted samples until the volume is about 5 µl using the speed vacuum concentrator. Add 7 µl of buffer A to the samples and transfer them to the autosampler

plate of the nanoHPLC.

3.9 Mass Spectrometry

The following section provides guidelines for measuring GFP affinity purification samples by nanoHPLC-MS/MS, e.g. an EASY nLC coupled to an LTQ-Orbitrap Velos mass spectrometer. Expertise is required to operate these machines, therefore the following steps should be performed under the supervision of an experienced mass spectrometrist.

1. Program for an injection volume of 5 µl of the sample into the nanoHPLC.
2. Peptides are eluted from the nano-column packed with C18 using a 5--30% acetonitrile (v/v) gradient followed by a sharp increase to 60% acetonitrile in 10 min with a flow rate of 250 nl/min. Total elution time is around 120 min, but this can be increased if the sample is very complex.
3. The recommended settings for data acquisition are: for precursor MS spectra, *m/z* range 300--1750 with a resolution of 60,000 and a target value of 1 million ions per full scan. For MS/MS spectra, CID is selected as the fragmentation method and the 15 most intense precursor ions are selected for fragmentation from each full MS scan with a minimal ion count target value of 500. Dynamic exclusion is set to 30 seconds (both repeat duration and exclusion duration), list size 500, early expiration enabled (count 2, S/N threshold 2). MS/MS scans are acquired in the centroid mode in the dual pressure linear ion trap with a normalized collision energy of 35%.

3.10 Data Analysis

After data acquisition is complete, raw data are transferred to a Windows-operated PC. We use the MaxQuant software package to analyze the raw data which can be freely downloaded at www.maxquant.org *(11)*. The software also contains a module for downstream data analysis (statistical tests, filtering, clustering, GO term enrichments etc.) called Perseus. New versions of the software are continously being generated and made available at the website. There is also an active Google group (MaxQuant) for posting questions.

1. Open the MaxQuant.exe program.
2. Load the raw data files (using load files option) that were copied to a local disc in a separate data folder containing only the files that are to be analyzed together (in this case, only the raw data files of the forward and the reverse experiment). Do not use spaces when creating data folders and raw files since this generates an error while running MaxQuant.
3. Click on 'Exp. Design'. Then use Excel to open the 'experimentalDesignTemplate.txt' that is written into the 'combined' folder within the data folder. Specify the forward and reverse experiments in the 'experiment' column and save the changes.
4. Go to the 'Group-specific parameters' tab and check the protease that is used (trypsin(P)), the multiplicity (2) and the labeled amino acids (Arg10 and Lys8). If necessary, adjust these options according to the experiment.
5. Go to the 'MS/MS & Sequences' tab and upload the FASTA file database of the appropriate organism. These FASTA files can be downloaded at the MaxQuant website.
6. Go to the 'Identification and quantification' tab and upload the 'experimentalDesignTemplate.txt'.
7. Go to the 'Misc.' tab and click the 'Match between runs' and 'Re-quantify' options.
8. Start the analysis by clicking on 'Start'.

At the end of the analysis, the forward and reverse ratios of proteins are reported in 'ProteinGroups.txt' (within the txt folder that is written into the data folder). Using Perseus (or other programs such as R, MatLab etc.), the contaminants, reverse hits and proteins with less than 3 reported peptide ratios (Ratio H/L count) can be filtered out. After this step, log-transform (base 2) the protein ratios and calculate significance B (in Perseus). Significance B indicates the probability of a protein being a significant outlier from the background cloud based on intensity and ratio. After these steps, plot the log2-transformed normalized forward and reverse ratios against each other. The non-specific background binders (with 1:1 ratios in both experiments) cluster around the origin of the graph while the specific interactors have high forward and low reverse ratios (see Figure 1). Contaminant proteins (such as keratins and serum proteins) are easily distinguished since they have a low forward and a low reverse ratio (as they are not SILAC-labeled).

4. Notes

1. When a cell line is being labeled for the first time, it is recommended to perform a SILAC label incorporation check before proceeding with large scale experiments. This can be done by digesting a small amount of heavy-labeled nuclear or whole cell extract (10 µg) with trypsin using standard in solution digestion protocols or the filter-aided sample preparation (FASP) method followed by LC-MS/MS.[12] If incorporation efficiency is lower than 95%, the maximum observable ratios decrease significantly (i.e. 90% incorporation results in a maximum ratio of 9, at 80% the maximum ratio is 4, etc.). The minimum recommended incorporation efficiency is 95%. Note that the labeling efficiency will never reach 100% due to trace amounts of non-labeled amino acids in the culture medium. Another potential problem is arginine-to-proline conversion, which can be investigated during the incorporation check. When a heavy arginine (Arg10) is converted to proline, this proline is 6 Daltons heavier compared to normal proline. For heavy SILAC-labeled peptides containing one or more internal prolines, a third isotope cluster appears in the mass spectrum. The intensity of this third isotope cluster is not taken into account during quantification and this results in an underestimation of the peptide ratio. A small percentage of arginine-to-proline conversion can easily be normalized for, but as a rule of thumb, heavy proline peaks should not have an intensity of more than 5% of the normal heavy peak. Some cell lines are more prone to this problem than others and in some cases, titrating the amount of arginine and proline in the SILAC medium can reduce the amount of conversion. However, proline to arginine conversion also occurs and this potentially compromises arginine labeling efficency. Therefore, the titration should be performed carefully. As a last resort, lysine-only labeling can be used to circumvent this problem. In this case, LysC instead of trypsin is used to digest the proteins.
2. When working with a stable cell line expressing a GFP-tagged protein of interest, it is recommended to SILAC-label this cell line and the WT control cells simultaneously. Both lines will be labeled light and heavy. In the end, four batches of cells are harvested: light and heavy WT cells and light and heavy GFP cells. While culturing and harvesting, it is important to handle all these batches of cells as reproducibly as possible in order to minimize variations between the extracts.
3. For suspension cells, SILAC RPMI medium lacking lysine, arginine and glutamine is available. For labeling of mouse embryonic stem cells (mESCs), SILAC DMEM (with 15% D-FBS, L-glutamine, and penicillin/streptomycin) can be used with the necessary additions: 1x non-essential amino acids, 1 mM sodium pyruvate, lysine and arginine (light or heavy in amounts described), 1000 units/ml LIF, 4.2 µl of 99% (v/v) β-mercaptoethanol and 2i inhibitors (3 µM and 1 µM of CHIR99021 and PD0325901 respectively). 2i inhibitors consist of small molecule inhibitors of GSK-3 and ERK1/2 signaling and thus prevent differentiation.[13] Note that the mESCs should not be grown on feeder cells since these are not SILAC-labeled. Several mESC lines such as IB10 and R1 can grow in the absence of feeder cells when culture dishes are coated with 0.15% gelatin. Instead of trypsin-EDTA, accutase is used for detaching the cells which is a mixture of proteolytic and collagenolytic enzymes effective in detaching primary fibroblasts, neurons, endothetial cells and ESCs.[14]
4. The amount of labeled amino acids to be added to the medium is optimized for SILAC labeling of commonly used cell lines such as HeLa, HEK293T and MCF7. These amounts may need to be altered for labeling other cell lines.

5. We prefer to use cell lines stably transfected with a bacterial artificial chromosome (BAC) construct containing the gene of interest.[15] These BACs have been recombined to express the protein of interest with a GFP tag. Since the BAC is expected to contain the proximal regulatory regions for expression, the expression level of the GFP-tagged bait is at near endogenous levels. In this workflow, the parental WT cells are used as control cell line. An alternative approach is to generate a stable cell line in which the expression of the GFP-tagged bait is inducible. In this case, uninduced cells serve as the control. If no stable cell line is available, the GFP-tagged bait can be transiently over-expressed. However, the disadvantage of transient transfection is that the expression level of the GFP-tagged bait cannot be controlled. Although the expression level of the bait depends on the strength of the promoter in the plasmid, it usually exceeds endogenous levels. While this may not necessarily be problematic, a gross over-expression may induce false positive interactions. In addition, when the majority of the bait is not associated with interactors due to over-expression, this may compromise the depth of sample sequencing, thereby reducing the identification of substoichiometric interactors.
6. The confluency required at the time of transient transfection depends on the proliferation rate of the cells. It is recommended to keep the cells in culture after transfection for at least 24 hours and at most 48 hours. The confluency at the time of transfection should be adjusted accordingly to prevent overgrowth of the cells.
7. Usually 10-20 µg of DNA per 15-cm dish is used for transfection. The amount may be reduced in order to lower the amount of GFP-tagged bait in the nuclear extract in case of very strong over-expression.
8. When working with nuclear proteins, making nuclear extracts is highly recommended since this is an efficient way to get rid of a lot of highly abundant cytoplasmic proteins. This step decreases the background in the pull-down and in the mass spectrometer, facilitating the sequencing and identification of the GFP-tagged bait and interactors. The affinity purification method described here, however, is not restricted to nuclear extracts and can also be applied in combination with cytoplasmic or whole cell extracts. In these cases, the stringency of the buffer used during extraction and of Buffer C* used during the affinity purification should be increased and more input material should be used.
9. The cytoplasmic extract should be snap-frozen after adding glycerol to a final volume of 10% and NaCl to a final concentration of 150 mM.
10. In this chapter, the GFP tag is used for affinity purification. Although GFP is a large protein tag compared to other commonly used tags such as Flag, HA, etc., tagging proteins with GFP rarely results in a non-functional fusion protein.[15] In our hands, the GFP tag works well and does not often interfere with PPIs. This is true even when tagging very small proteins.[16] Depending on the domain structure of the bait, N- or C-terminal tagging may be preferred.
11. The GFP-binder beads used for affinity purification only bear the epitope recognition domain of a high-affinity monoclonal GFP antibody. This small, 13 kDa fragment is expressed in *E.coli* and then covalently cross-linked to agarose beads.[17] The result is a high-affinity enrichment resin lacking the heavy and light immunoglobin chains present in conventional antibodies. When using conventional antibodies for affinity purification, the heavy and light chains are digested together with the affinity-purified proteins and dominate the MS spectra due to their abundance. In that case, a specific

elution protocol (peptide elution or acidic elution) or in-gel digestion is recommended instead of on-bead digestion.[18]

12. When a 'wild-type' cell line (the parental cell line that was used to generate the transgenic cell line) is not available, it is possible to use the nuclear extracts from cells expressing the GFP-tagged transgene in combination with blocked agarose beads (BAB) as a negative control. These BAB are commercially available (see section 2.5). This may also be necessary when expression of the GFP-tagged transgene significantly changes the growth rate and/or morphology of the parental cell line. In this case, the experimental set-up changes. Instead of four different cultures, only the GFP-tagged cell line is labeled light and heavy but expanded to twice the amount. Four affinity purifications are then performed as follows: light and heavy nuclear extracts incubated with GFP-binder beads and light and heavy nuclear extracts incubated with BAB. The forward experiment becomes heavy GFP-binder pull-down + light BAB pull-down and the reverse experiment becomes light GFP-binder pull-down + heavy BAB pull-down.
13. Note that the buffer C* used for affinity purification contains 300 mM NaCl as opposed to the buffer C (420 mM NaCl) used for nuclear extraction. At the last step of nuclear extraction, addition of two volumes of buffer C with 420 mM NaCl to the crude nuclei results in a final salt concentration of approximately 300 mM. Since the protein concentrations of different batches of nuclear extracts may vary, the volume of nuclear extract to be used for affinity purification may also vary (in order to use the same amount of protein). By using a buffer with 300 mM salt for the GFP pull-down, the final salt concentration is kept equal in all affinity purifications.
14. Using 1 mg of nuclear extract is usually sufficient to identify the bait and interactors. For low-expressed bait proteins or when using whole cell lysates, this amount can be increased to a maximum of 5 mg. The amount of beads used during the pull-down need not be changed.
15. Using ethidium bromide is very critical to eliminate DNA-mediated indirect interactions. Although the bulk of the DNA ends up in the insoluble chromatin fraction during nuclear extraction, the nuclear extract still contains a fair amount of double-stranded DNA. In the absence of ethidium bromide, the pull-downs are more "noisy" and more proteins are identified as interactors of the bait, many of which are false positive, DNA-mediated secondary interactions (see **Figures 2 and 3**). As an alternative, DNA and RNA in the extract can be digested with an endonuclease such as MNase prior to the pull-down.
16. The amount of detergent and salt in the wash buffer may be altered. To identify low affinity or sub-stoichiometric interactors, the final NP-40 concentration can be decreased to 0.2%; to identify very strong interactors, it can be raised to 1%. The key point is to find a balance between specificity and sensitivity and this may vary depending on the bait.

Acknowledgments
We thank members of the Vermeulen group and Nina Hubner for critical reading of the manuscript and fruitful discussions. Work in the Vermeulen group is supported by grants from the Netherlands Organisation for Scientific Research (VIDI and Cancer Genomics Centre (CGC)) and the Dutch Cancer Society (KWF).

5. References

1. Fields S. and Song O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245
2. Gingras A. C., Gstaiger M., Raught B. *et al.* (2007) Analysis of protein complexes using mass spectrometry. *Nature Reviews Molecular Cell Biology* **8**, 645
3. Altelaar A. F. and Heck A. J. (2012) Trends in ultrasensitive proteomics. *Current opinion in chemical biology* **16**, 206
4. Vermeulen M., Hubner N. C. and Mann M. (2008) High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol* **19**, 331
5. Ong S. E., Blagoev B., Kratchmarova I. *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **1**, 376
6. 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 research* **11**, 1475
7. Hubner N. C., Bird A. W., Cox J. *et al.* (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *The Journal of cell biology* **189**, 739
8. Hubner N. C. and Mann M. (2011) Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods* **53**, 453
9. Nielsen M. L., Vermeulen M., Bonaldi T. *et al.* (2008) Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nature methods* **5**, 459
10. Rappsilber J., Mann M. and Ishihama Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896
11. Cox J., Matic I., Hilger M. *et al.* (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat Protoc* **4**, 698
12. Wisniewski J. R., Zougman A., Nagaraj N. *et al.* (2009) Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359
13. Lee K. H., Chuang C. K., Guo S. F. *et al.* (2012) Simple and efficient derivation of mouse embryonic stem cell lines using differentiation inhibitors or proliferation stimulators. *Stem cells and development* **21**, 373
14. Zhang P., Wu X., Hu C. *et al.* (2012) Rho kinase inhibitor Y-27632 and Accutase dramatically increase mouse embryonic stem cell derivation. *In vitro cellular & developmental biology. Animal* **48**, 30
15. Poser I., Sarov M., Hutchins J. R. *et al.* (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature methods* **5**, 409
16. Spruijt C. G., Bartels S. J., Brinkman A. B. *et al.* (2010) CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. *Mol Biosyst* **6**, 1700
17. Rothbauer U., Zolghadr K., Muyldermans S. *et al.* (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Molecular & cellular proteomics : MCP* **7**, 282
18. Paul F. E., Hosp F. and Selbach M. (2011) Analyzing protein-protein interactions by quantitative mass spectrometry. *Methods* **54**, 387

Chapter 3

MBD5 and MBD6 interact with the human PR-Dub complex through their methyl-CpG binding domain

H. Irem Baymaz[1#], Alexandra Fournier[2#], Sophie Laget[2], Zongling Ji[3], Pascal W.T.C. Jansen[1], Arne H. Smits[1], Laure Ferry[2], Anneloes Mensinga[1], Ina Poser[4], Andrew Sharrocks[3], Pierre-Antoine Defossez[2]* and Michiel Vermeulen[1]*

1. Molecular Cancer Research and Cancer GenomiCs Netherlands, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands
Present address: Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands
2. University Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, UMR 7216 CNRS, 75013 Paris, France
3. Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, United Kingdom
4. Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauer Str. 108, 01307 Dresden, Germany

# Equal contribution
* Corresponding authors

Proteomics, October 2014,
Volume 14, Issue 19, pages 2179-89

Abstract

MBD5 and MBD6 are two members of the methyl-CpG binding domain (MBD) family of proteins that are poorly characterized. Studies performed thus far have failed to show binding of the MBD5 and MBD6 MBD domain to methylated DNA. Here, we show that both MBD5 and MBD6 interact with the mammalian PR-DUB Polycomb protein complex in a mutually exclusive manner. Strikingly, the MBD domain of MBD5 and MBD6 is both necessary and sufficient to mediate this interaction. ChIP analyses reveal that MBD6 and FOXK2/PR-DUB share a subset of genomic target genes, suggesting a functional interaction in vivo. Finally, we show that MBD6, but not MBD5, gets recruited to sites of DNA damage in a PR-DUB independent manner. Our study thus implies a shared function for MBD5 and MBD6 through an interaction with PR-DUB, as well as an MBD6-specific recruitment to sites of DNA damage.

Introduction

DNA methylation is a stable epigenetic mark found in vertebrate and plant genomes. In mammals, the fifth carbon position of cytosine is methylated throughout the genome and mostly in a CpG dinucleotide context (5-meCpGs) [1, 2]. 5-meCpGs in the promoter of genes are generally associated with their repression and are involved in processes such as regulating the expression of imprinted genes, X chromosome inactivation and higher order chromatin compaction [3]. The molecular pathways via which DNA methylation mediates transcriptional silencing are not completely understood. One important mechanism is thought to involve the specific recruitment of transcriptionally-repressive protein complexes to methylated CpG dinucleotides [2, 3]. Initially, three classes of proteins were found to interact with 5-meCpG: the **M**ethyl-CpG **B**inding **D**omain (MBD) family of proteins, Kaiso and Kaiso-like proteins (ZBTB38 and ZBTB4), and **S**ET and **R**ing finger **A**ssociated (SRA) domain proteins (UHRF1 and UHRF2)[3]. More recently, it was shown that other proteins not belonging to these three protein classes can interact with methylated CpG containing DNA. Examples include certain homeobox proteins and proteins carrying a winged helix domain [4, 5].

The first discovered member of the MBD family of proteins is MeCP2 [6]. The domain within MeCP2 responsible for binding to methylated CpGs was narrowed down and named the **m**ethyl-CpG-**b**inding **d**omain [7]. Later on, based on sequence homology, more members of the MBD family of proteins were identified, which were designated MBD1-6. In addition, four other proteins carry an MBD domain (SETDB1, SETDB2, BAZ2A and BAZ2B) (**Figure 1A**). However, biochemical studies have revealed that not all members of the MBD family are capable of binding to 5-meCpG. MBD5 and MBD6 represent two such examples of MBD-containing proteins that cannot bind 5-meCpG [1, 3, 8].

MBD5 has two isoforms: a long isoform carrying a PWWP domain near its carboxy-terminal, and a shorter isoform lacking this domain. MBD6 has no other annotated domains in addition to the MBD. MBD5 and MBD6 localize to heterochromatin foci in mouse nuclei in a DNA methylation independent manner [9]. A microdeletion syndrome including part of the Mbd5 locus has been associated with neurological disorders [10]. Furthermore, Mbd5 knock-out mice show deregulation in glucose homeostasis, growth retardation and ultimately, preweaning lethality [11]. A detailed functional and biochemical characterization of these proteins is still lacking.

Here, we show that MBD5 and MBD6 interact with the human Polycomb repressive complex PR-DUB. PR-DUB was first identified in *Drosophila melanogaster* and the complex catalyzes deubiquitination of H2A at lysine 118 (119 in mammals) [12]. The mammalian counterpart of the complex is known to consist of BAP1, ASXL1/2, KDM1B, FOXK1/2 and HCFC1[13-15]. We find these proteins to co-purify with MBD5 and MBD6. Interestingly, the MBD domain of MBD5 and MBD6 is necessary and sufficient for the interaction with the PR-DUB complex, showing the first example of an MBD acting as a protein-protein interaction domain. MBD5 and MBD6 interact with the PR-DUB complex in a mutually-exclusive manner. Using ChIP, we find that MBD6 co-localizes to a subset of genomic FOXK2 and BAP1 target genes, suggesting a functional link between these proteins in vivo. Finally, we find that MBD6, but not MBD5, gets recruited to sites of DNA damage in a PR-DUB independent manner, thus revealing a functional distinction between MBD5 and MBD6.

Materials and methods

Cell culture and SILAC labeling

293T cells and HeLa cells were cultured in DMEM (high glucose, 4.5 g/l) with L-glutamine (Lonza) supplemented with 10% FBS (Thermo Scientific) and 550 units of penicillin-streptomycin (Lonza). For SILAC labeling, cells were cultured for at least 8 doublings in heavy and light SILAC medium (SILAC DMEM (PAA), 10% v/v dialyzed FBS (Gibco), 1 unit/ml of penicillin-streptomycin (Gibco) and 2 mM L-glutamine and either light or heavy forms of arginine and lysine at a final concentration of 27 mM and 66 mM, respectively), before expanding them to the desired amounts.

Transient transfection and nuclear extraction

15 mL of SILAC DMEM (without supplements) was mixed with 450 µl of PEI (1 mg/ml) and 150 µg of plasmid encoding for the protein of interest tagged with eGFP. This mixture was vortexed for 15 seconds and then incubated at room temperature for 25 minutes. Next, 5x15 cm dishes of light labeled and 5x15 cm dishes of heavy-labeled 293T or HeLa cells were transfected with 1.5 ml of the transfection mixture and incubated at 37°C for 36-40 hours. As a control, 5 light labeled 15 cm dishes and 5 heavy labeled 15 cm dishes of 293T or HeLa cells were mock transfected. Nuclear extracts were generated as described [16].

MBD5 isoform 1, isoform 2 and MBD6 were amplified from plasmids with gateway-compatible recombination arms [9]. Destination vectors resulting in N-terminal eGFP tagging (pcDNA5/FRT/TO), were prepared according to the manufacturer's instructions (Invitrogen). The destination vectors were then transfected into HeLa cells with a single FRT site (referred to as HeLa FRT cell line) along with the pOG44 recombinase (Invitrogen). Cells were selected for stable incorporation of the construct of interest into the FRT site using hygromycin (Roche). Stable cell lines were tested for expression of the transgene by western blotting for GFP. 14-16 hours prior to harvesting cells to generate nuclear extracts, 1 µg/ml of doxycycline was added to culture medium to induce protein expression.

Stable knock-down of BAP1 in MBD6-GFP HeLa FRT line

BAP1 knock-down using shRNA constructs against BAP1 and a scrambled shRNA construct as a control in the MBD6-GFP HeLa FRT cell line was performed essentially as described [17].

GFP Immunoprecipitation, on-bead digestion of proteins and mass spectrometry

Depending on the nuclear extracts used, i.e. from SILAC labeled cells or non-labeled cells, the GFP pull-down set-up can be seen in Figure 2a or Figure 4a, respectively.
In the case of SILAC extracts, 4 different IPs were performed using GFP nanotrap beads (Chromotek) and 4 different extracts (light non-transfected; light transfected; heavy non-transfected; heavy transfected). In the case of label-free extracts, 3 specific pull-downs were performed with GFP beads and 3 control pull-downs using non-GFP beads (Chromotek) using the same GFP extract. 1-2 µg of extract was used for each IP. IPs and subsequent on-bead digestion, LC-MS/MS analysis and data analysis were performed essentially as described [18]. Data was analyzed using the MaxQuant software [19].

Chromatin Immunoprecipitation

Approximately 20 million cells (one confluent 150 mm culture dish) were crosslinked by adding 10% vol/vol Buffer A (0.1 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 50 mM Hepes, freshly added before use: 11% final concentration formaldehyde) and rocking at room temperature

for 10 minutes. Crosslinking was quenched by the addition of 0.125 M glycine. After this step the dishes were kept on ice. Cells were scraped and collected in PBS following 1 PBS wash and then pelleted by centrifugation (1600xg, 7 minutes, 4°C). 5 ml of lysis buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 0.2% NP-40, 0.2% Triton-X, 1 mM EDTA, 0.5 mM EGTA) was added to the cell pellet, followed by centrifugation (1600xg, 5 minutes, 4°C). 5 ml of buffer C (0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 50 mM Hepes) was added to the cell pellet and the cell suspension was incubated at 4°C for 10 minutes on a rotating wheel, followed by centrifugation (1600xg, 5 minutes, 4°C). The pellet was resuspended in 900 µl of ChIP incubation buffer (0.2% SDS, 1% Triton-X, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl (pH 8)) and sonicated for 28 cycles (Diagenode Bioruptor, setting high; 30 seconds on, 30 seconds off). The sonicated cell suspension was cleared by centrifugation (3500xg, 20 minutes, 4°C) and the supernatant was snap-frozen as the chromatin fraction. For each ChIP, 300-400 µl from the chromatin fraction was incubated with protein A/G agarose beads (Santa Cruz), bovine serum albumin dissolved in Milli Q, ChIP incubation buffer, protease inhibitor cocktail and 2-3 µg of antibody overnight at 4°C on a rotating wheel. The next day beads were washed twice with wash buffer 1 (0.1% SDS, 0.1% NaDOC, 1% Triton-X, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl (pH 8)), once with wash buffer 2 (0.1% SDS, 0.1% NaDOC, 1% Triton-X, 0.5 M NaCl, 1 mM EDTA, 10 mM Tris-HCl (pH 8)), once with wash buffer 3 (0.25 M LiCl, 0.5% NaDOC, 0.5% NP-40, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl (pH 8)) and twice with wash buffer 4 (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl (pH 8)) in a volume of 200 µl. Each wash consisted of 5 minute incubation on a rotating wheel at 4°C followed by centrifugation (2000xg, 1 minute, 4°C). After the final wash, immunocomplexes were eluted off the beads by the addition of elution buffer (1% SDS, 0.1 M $NaHCO_3$) and incubation at room temperature on a rotating wheel for 15 minutes followed by centrifugation. The supernatant was collected and one more round of elution was performed on the beads, after which the collected supernatant was added to the first (final volume 250 µl). 10 µl of 5M NaCl was added to the samples after which samples were decrosslinked at 65°C for 4 hours / overnight, shaking at 1400 rpm. After decrosslinking, 10 mM EDTA, 40 mM of Tris-HCl (pH 6.5) and 20 µg of proteinase K (Roche) were added to the samples followed by incubation for 1 hour at 45°C, shaking at 1400 rpm. After this step, DNA was purified using PCR purification kit (Qiagen) and eluted in 200 µl Milli Q. Input DNA and immunoprecipitated samples were analyzed using qPCR. Primer sequences are available upon request.

Local laser irradiation and live microscopy

Cells were seeded on 12 mm round coverslips and sensitized for microirradiation by incubation with 10 µM BrdU for 20–24h at 37°C. Microirradiation was carried out with a 405 nm laser diode (30 mW) coupled into a LSM710 confocal microscope (Zeiss) equipped with a thermostated chamber. Laser was focused through a Plan-Apochromat 63X/1.4 oil objective to locally irradiate preselected lines of 1 µm in width within the nucleus, using the following settings: 40% power, 50 iterations, scan speed 12.6 µsec/pixel. For live-cell imaging, confocal image series of one mid z-section were taken before and after microirradiation at several time intervals using 488 nm and 514 nm Argon laser lines (25 mW), with a pixel size of 90 nm. Images were adjusted for brightness and contrast using Fiji (Fiji Is Just ImageJ) software.

Results

MBD5 and MBD6 interact with the human PR-DUB complex

To investigate which proteins interact with MBD5 and MBD6, we transfected GFP-tagged full length cDNA clones of these proteins into SILAC-labeled 293T cells. Forty hours after transfection, nuclear extracts were prepared and these were subjected to GFP pull-downs followed by on-bead digestion and LC-MS/MS (**Figure 2A**). Both GFP-MBD5 and GFP-MBD6 were found to co-purify with subunits of the human PR-DUB complex (BAP1, ASXL2 and KDM1B) (**Figure 2B and 2C**). Interestingly, MBD5 was not detected in the MBD6 pull-down and vice versa. Importantly, this association with the PR-DUB complex was not observed for other MBD proteins we tested (MeCP2, MBD1, MBD2, MBD3, MBD4, data not shown). Next, we generated HeLa Kyoto cells stably expressing PR-DUB subunits (BAP1 and ASXL2) with a C-terminal GFP tag at near endogenous levels and including naturally occurring splice variants using BAC TransGenOmics technology [20]. These cell lines were subsequently SILAC-labeled and subjected to GFP affinity enrichments coupled to quantitative mass spectrometry as described previously [18]. Endogenous MBD5 and MBD6 were found to co-purify with BAP1-GFP and ASXL2-GFP (**Figure 2D and 2E**). Most of the other interactors observed in the pull-downs such as FOXK1/FOXK2 and HCFC1 are known PR-DUB interactors [13, 14, 21]. In summary, these results clearly establish MBD5 and MBD6 as novel interactors of the mammalian Polycomb group protein complex PR-DUB.

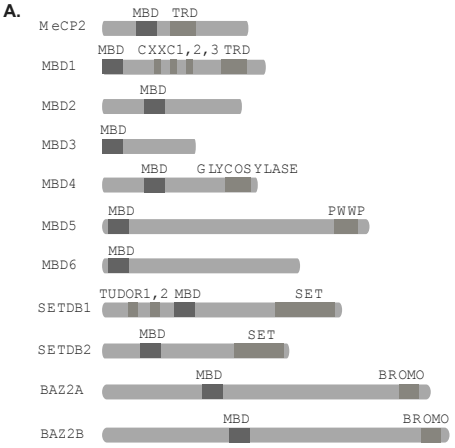


Figure 1. The mammalian family of MBD proteins. **A**. Schematic overview of the 11 human proteins that are known to contain an MBD or MBD-like domain. Other known domains in each of these proteins are also indicated. **B**. Sequence alignment of the 11 mammalian MBD domains.

B.

```
                10        20        30        40        50        60        70        80
                |         |         |         |         |         |         |         |
MECP2   --DRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGR-------
MBD4    --ATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSK-------
MBD1    --MAEDWLDCPALGPGWKRREVFRKSGATCGRSDTYYQSPTGDRIRSKVELTRYLGPACDLTLFDFKQGIL-----------
MBD2    --ESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVYYFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKM-----------
MBD3    --MERKRWECPALPQGWEREEVPRRSGLSAGHRDVFYYSPSGKKFRSKPQLARYLGGSMDLSTFDFRTGKMLMS--------
SETDB1  --YRGKNPLLVPLLYDFRRMTARRVNRKMG-FHVIYKTPCGLCLRTMQEIERYLFETGCDFLFLEMFCLDPYVL-------
SETDB2  LNLKGENPLQLPIKCHFQRRHAKTN-SHSSA-LHVSYKTPCGRSLRNVEEVFRYLLETECNFLFTDNFSFNTYVQ-------
BAZ2A   --IATPEEVRLPLQHGWRREVRIKKGSHRWQ-GETWYYGPCGKRMKQFPEVIKYLSRNVVHSVRREHFSFSPRMP-------
BAZ2B   --VTDERELRIPLEYGWQRETRIRNFGGRLQ-GEVAYYAPCGKKLRQYPEVIKYLSRNGIMDISRDNFSFSAKIR-------
MBD5    --DKEGGLPAIQVPVGWQRRVDQNG---------VLYVSPSGSLLSCLEQVKTYLLTDGTCKCGLECPLILPKVFNFDPGAA
MBD6    --DRAGGPVATSVPIGWQRCVREGA---------VLYISPSGTELSSLEQTRSYLLSDGTCKCGLECPLNVPKVFNFDPLAP
```

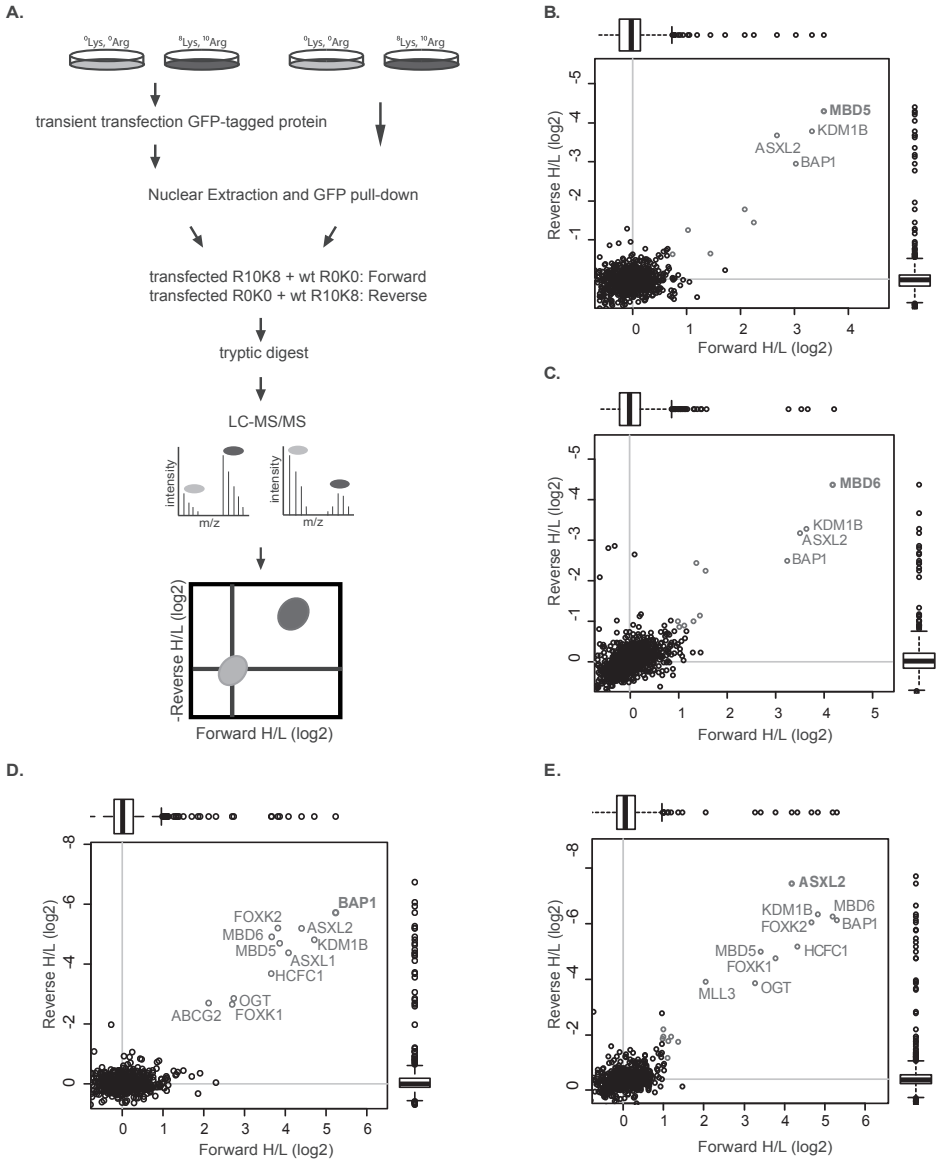


Figure 2. MBD5 and MBD6 both interact with the human PR-DUB complex. **A**. Schematic overview of the workflow. Light and heavy-labeled 293T cells were transiently transfected with GFP-MBD5 (**B**) or GFP-MBD6 (**C**) and subjected to single affinity purifications in a 'forward' and 'reverse' pull-down using GFP nanotrap beads. As a control, light and heavy-labeled 293T cells were mock transfected. In the figures, the ratio of all the proteins in the forward pull-down (X axis) is plotted against their ratio in the reverse pull-down (Y axis). MBD5 and MBD6 associated proteins appear in the upper right quadrant of Figure 2B and 2C, respectively, showing a high ratio in the forward and a low ratio in the reverse pull-down. Boxplot statistics was applied to determine statistically-significant interactors. Non-specific proteins appear around the origin of the figures (ratio ~ 1 in both pull-downs). HeLa cells stably expressing BAP1-GFP (**D**) or ASXL2-GFP (**E**) were also subjected to SILAC-labeling and single affinity purification as described above.

MBD5 and MBD6 interact with the human PR-DUB complex via their MBD

Having established that MBD5 and MBD6 both interact with the human PR-DUB complex, we were interested to determine the region within MBD5 and MBD6 that mediates this interaction. Two isoforms for MBD5 are known to exist, the second of which contains a substantial C-terminal truncation [9]. This shorter MBD5 isoform (amino acids 1-851) still interacts with the PR-DUB complex (data not shown). This suggests that the C-terminus of the full-length MBD5 protein is not mediating the interaction with PR-DUB. In support of this observation, a deletion mutant of MBD5 containing only the first 94 amino acids of the protein still maintained its interaction with the PR-DUB complex (**Figure 3A**). This N-terminal fragment of the MBD5 protein mainly contains the MBD domain. Similar results were obtained for MBD6 (data not shown). Finally, we were not able to observe interactions between PR-DUB and MBD5 or MBD6 when using constructs lacking the MBD domain (**Figure 3B** and data not shown). Based on these observations, we conclude that both MBD5 and MBD6 interact with the PR-DUB complex through their highly homologous MBD domains. This is the first example of an MBD domain mediating a protein-protein interaction as opposed to protein-DNA and protein-RNA interactions, which have been observed for other members of the MBD family of proteins [22, 23].

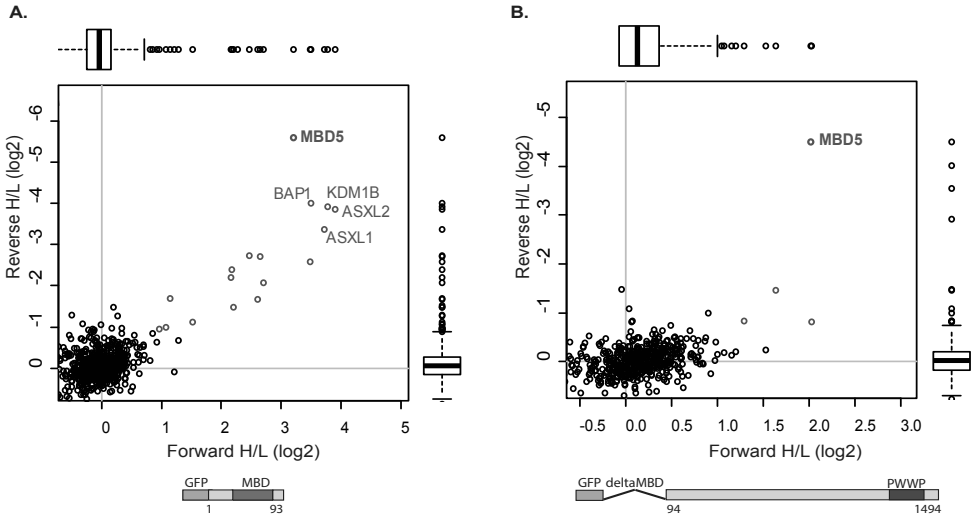


Figure 3. The MBD domain of MBD5 is both necessary and sufficient to mediate an interaction with the human PR-DUB complex**.** Light and heavy labeled 293T cells were transiently transfected with MBDonly_MBD5-GFP (**A**) or deltaMBD_MBD5-GFP (**B**) and subjected to single GFP affinity purifications as described in Figure 2.

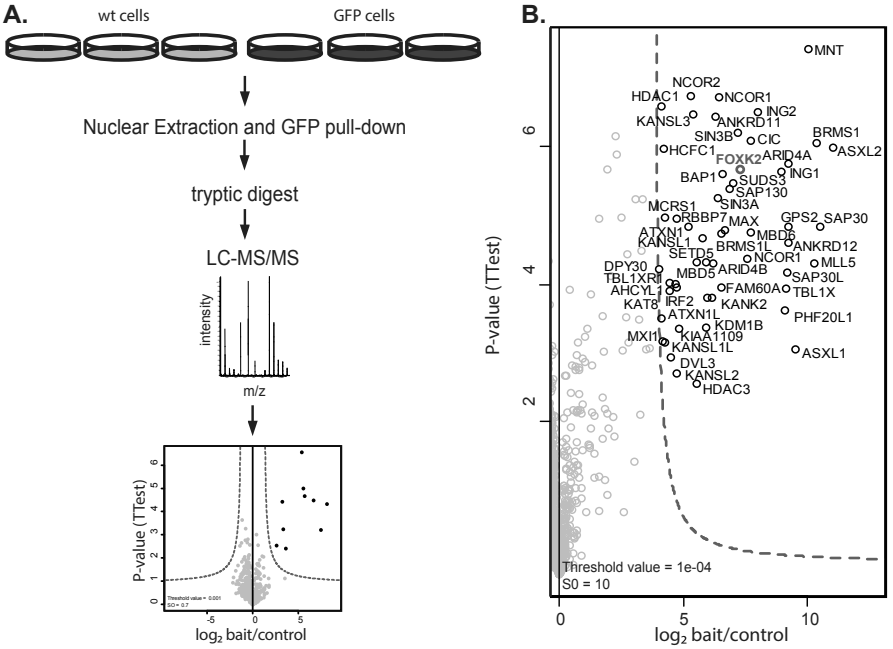
MBD6 co-localizes with BAP1 to a subset of FOXK2 target genes

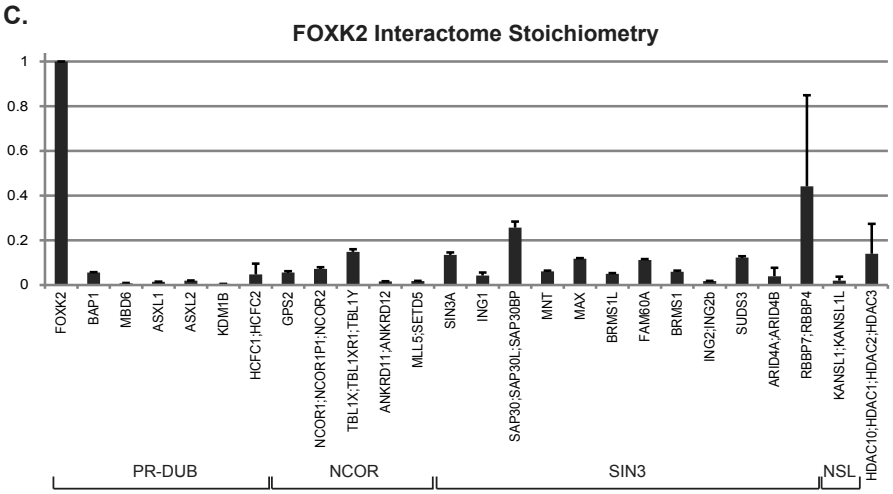
GFP-based purifications of the mammalian PR-DUB complex consistently resulted in the co-purification of FOXK1 and FOXK2. These proteins are part of the Forkhead box transcription factor family, all of which contain a Forkhead box DNA binding domain, also known as a winged helix. None of the other PR-DUB subunits contain a clear DNA binding domain. These two proteins are therefore candidates for recruiting the PR-DUB complex to target genes in the genome. Interestingly, FOXK1/2 have recently been shown to interact with

methyl cytosine and its oxidized derivatives [5, 24]. In addition to being part of the human PR-DUB complex, FOXK proteins have previously also been linked to the Sin3/HDAC complex [25-27]. To investigate the interaction between FOXK2 and PR-DUB in more detail, we performed label-free quantification of FOXK2 interactors using a stable and inducible HeLa FRT cell line expressing GFP-tagged FOXK2 [28]. Nuclear extracts from control and GFP-tagged cells were subjected to GFP affinity enrichment in triplicate, followed by on-bead trypsin digestion and LC-MS/MS. The GFP-tagged bait, specific interaction partners and background proteins were plotted in a volcano plot (**Figure 4A and B, Supplementary Figure 1A**). For stoichiometry estimations, the iBAQ algorithm was used [18, 29]. A large number of proteins were found to co-purify with FOXK2, including subunits of the Sin3/HDAC, NCOR, PR-DUB and NSL complexes. Stoichiometry analysis revealed that roughly equal amounts of Sin3/HDAC and N-CoR/SMRT complex co-purify with FOXK2 whereas somewhat lower levels of PR-DUB/MBD6 and NSL complex are detected (**Figure 4C**). Using a stable HeLa FRT cell line expressing GFP-tagged MBD6, we performed GFP chromatin immunoprecipitation followed by targeted qPCR for genes that are known to be FOXK2 targets and regulated by both BAP1 and FOXK2 [15, 30]. Out of several target loci tested, CGGBP1, PHB2 and JRK show clear MBD6 enrichment in the MBD6-GFP HeLa FRT cells compared to the negative control, empty HeLa FRT cells (**Figure 4D**). These results therefore establish a physical and functional link between PR-DUB, FOXK2, and MBD6 in vivo.

MBD6 is recruited to sites of DNA damage in a PR-DUB independent manner

The experiments described thus far have shown that MBD5 and MBD6 both interact with the mammalian PR-DUB complex through their MBD domain. The C-terminus of MBD5 and MBD6, however, is not highly conserved between the two proteins, suggesting that MBD5 and MBD6 may have specific functions that are not overlapping. As ubiquitination / deubiquitination signaling, and thus DUB proteins –including BAP1– are crucial for the assembly and disassembly of repair proteins at sites of DNA damage, we tested a potential involvement of MBD5 and MBD6 in this context [15, 31-33]. We observed that upon laser microirradiation, MBD6 was recruited to sites of laser-induced DNA damage; the recruitment was rapid (less than 30 seconds), and was visible for 5 to 10 minutes (**Figure 5A and Supplementary Figure 2**). We never observed recruitment of MBD5 isoforms to laser-induced DNA damage sites (**Figure 5A**), even though the cells did contain DNA lesions, as evidenced by positive 53BP1 labeling after irradiation (**Figure 5B**). As BAP1, a key component of PR-DUB, was recently shown to be involved in the DNA damage response, we investigated whether the recruitment of MBD6 to sites of DNA damage is dependent on BAP1[34]. We repeated the microirradiation experiment in a BAP1 knock-down background (**Supplementary figure 1B**), and observed that MBD6 localization to sites of DNA damage persisted in the absence of BAP1, albeit with possibly slower kinetics (**Figure 5C**). Next, we tested the deltaMBD deletion mutant of MBD6 that does not interact with PR-DUB. As shown in **Figure 5D**, this protein lacking its MBD was still efficiently recruited to sites of DNA damage. Therefore, it is the C-terminus of MBD6 (aminoacids 85-1003), and not its MBD, that distinguishes it from MBD5 and recruits it to DNA lesions. It remains to be seen whether the C-terminal-dependent recruitment of MBD6 to DNA damage sites coordinates the recruitment of PR-DUB to damage sites in an MBD domain-dependent manner.


A.
wt cells
GFP cells
Nuclear Extraction and GFP pull-down
tryptic digest
LC-MS/MS
intensity
m/z
P-value (TTest)
log2 bait/control
B.
P-value (TTest)
log2 bait/control
Threshold value = 1e-04
S0 = 10
MNT
NCOR2
HDAC1
NCOR1
ING2
KANSL3
ANKRD11
SIN3B
CIC
BRMS1
HCFC1
FOXK2
ARID4A
ASXL2
BAP1
SUDS3
ING1
SAP130
MCRS1
SIN3A
RBBP7
MAX
GPS2
SAP30
ATXN1
KANSL1
MBD6
BRMS1L
ANKRD12
SETD5
NCOR1
MLL5
DPY30
MBD5
ARID4B
SAP30L
TBL1XR1
AHCYL1
FAM60A
TBL1X
KAT8
IRF2
KANK2
PHF20L1
ATXN1L
KDM1B
MXI1
KIAA1109
KANSL1L
ASXL1
DVL3
KANSL2
HDAC3


C.
FOXK2 Interactome Stoichiometry
FOXK2
BAP1
MBD6
ASXL1
ASXL2
KDM1B
HCFC1;HCFC2
GPS2
NCOR1;NCOR1P1;NCOR2
TBL1X;TBL1XR1;TBL1Y
ANKRD11;ANKRD12
MLL5;SETD5
SIN3A
ING1
SAP30;SAP30L;SAP30BP
MNT
MAX
BRMS1L
FAM60A
BRMS1
ING2;ING2b
SUDS3
ARID4A;ARID4B
RBBP7;RBBP4
KANSL1;KANSL1L
HDAC10;HDAC1;HDAC2;HDAC3
PR-DUB
NCOR
SIN3
NSL

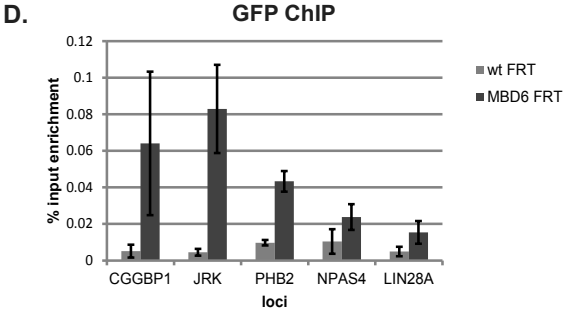
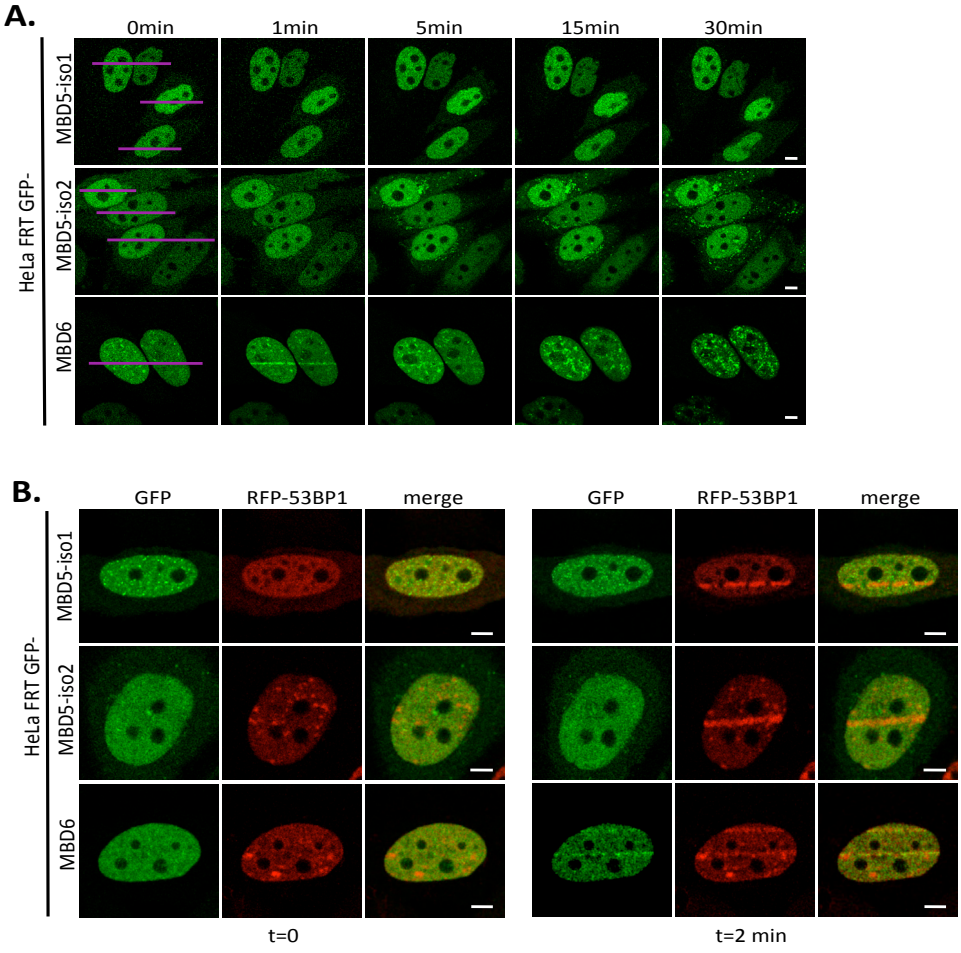

D.
GFP ChIP
% input enrichment
wt FRT
MBD6 FRT
CGGBP1
JRK
PHB2
NPAS4
LIN28A
loci

Figure 4. FOXK2 interacts with PR-DUB and MBD6 co-occupies a subset of FOXK2/PR-DUB target genes. **A.** Schematic overview of the workflow. **B.** Volcano plot showing GFP-FOXK2 interactors as revealed by label free quantitative mass spectrometry-based proteomics. Significant interactors of FOXK2-GFP are identified by permutation-based FDR corrected T-test (threshold: p=0.0001, S0=10). The LFQ intensity of the GFP pull-down relative to the control is plotted against the –Log10 of the p-value. The red line indicates the permutation-based FDR threshold. **C.** Visualization of the stoichiometry values in the FOXK2-GFP pull-down. Proteins that were identified as significant FOXK2 interactors but with a stoichiometry value relative to FOXK2 of < 1% were omitted from this figure, with the exception of KDM1B and MBD6. **D.** ChIP-qPCR analysis of GFP-MBD6 binding to a subset of FOXK2 target genes. Error bars show the standard deviation between technical triplicates.

Figure 5. MBD6, but not MBD5, is recruited to sites of DNA damage in a PR-DUB independent manner. **A.** MBD6, but not MBD5, is recruited to sites of laser-induced DNA damage. Live-cell imaging of laser-microirradiated HeLa FRT cells stably expressing GFP-MBD5 isoform 1, GFP-MBD5 isoform 2 or GFP-MBD6. **B.** MBD5 is not recruited to the regions of damaged DNA labeled by 53BP1. Cells were transfected with GFP-tagged MBD5 and MBD6 constructs as in A. Cells were also transfected with RFP-53BP1. Scale bars, 5µm.



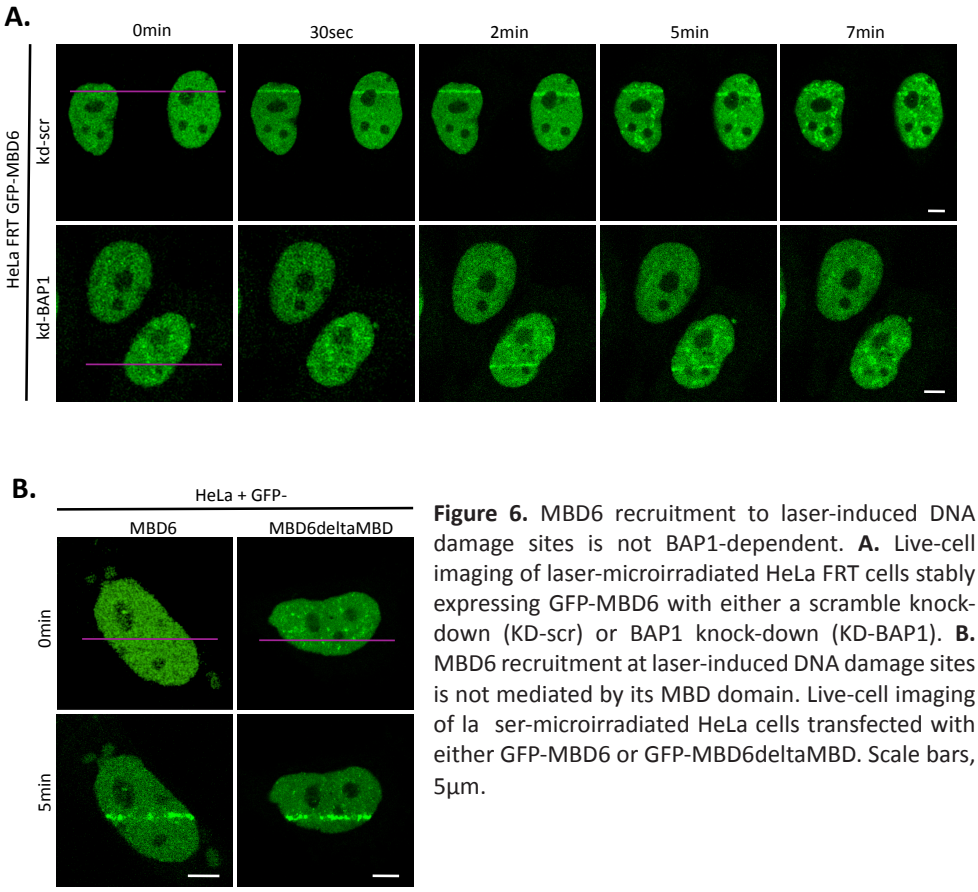


Figure 6. MBD6 recruitment to laser-induced DNA damage sites is not BAP1-dependent. **A.** Live-cell imaging of laser-microirradiated HeLa FRT cells stably expressing GFP-MBD6 with either a scramble knock-down (KD-scr) or BAP1 knock-down (KD-BAP1). **B.** MBD6 recruitment at laser-induced DNA damage sites is not mediated by its MBD domain. Live-cell imaging of la ser-microirradiated HeLa cells transfected with either GFP-MBD6 or GFP-MBD6deltaMBD. Scale bars, 5µm.

Discussion

Here, we have shown that two relatively uncharacterized members of the MBD family of proteins, MBD5 and MBD6, interact with the human Polycomb group protein complex PR-DUB through their MBD domain. Furthermore, we show that FOXK2/PR-DUB and MBD6 share a subset of genomic target genes. Finally, we show that MBD6, but not MBD5, gets recruited to sites of DNA damage upon microirradiation in a PR-DUB-independent manner. Recently, PR-DUB subunits have received a lot of attention given their seemingly broad involvement in a number of human malignancies. For example, BAP1 is proposed to cause a cancer syndrome characterized by predisposition to uveal melanoma and mesothelioma [21, 35]. BAP1 has also been linked to renal cell carcinoma and cutaneous melanoma [21, 35, 36]. Furthermore, conditional BAP1 knock-out mice display splenomegaly and develop myeloid malignancies [14]. Likewise, ASXL1 is mutated in a subset of myeloid malignancies [37]. In agreement with this emerging link between PR-DUB and cancer, both BAP1 and ASXL1 were recently reported as one of 138 human 'driver' genes that are frequently mutated and/or deregulated in cancer [38]. In addition to this tumorigenesis link, BAP1 interacts with HCFC1 and this interaction is known to regulate cell cycle progression [15, 39-41]. MBD5 and MBD6 have thus far not been linked to cancer. Further studies are needed to elucidate the molecular mechanisms underlying the association between deregulated

PR-DUB and cancer/cell cycle progression.

The Methyl-CpG binding domain family of proteins consists of 7 members (MeCP2 and MBD1-6). In addition, four other proteins are known to harbor an MBD-like domain (SETDB1A/B and BAZ2A/B). This more inclusive superfamily is also known as the TAM family (*T*TF-IIP5/*A*RBP/*M*eCP2) [1, 42, 43]. Within the human MBD family, only MeCP2, MBD1, MBD2 and MBD4 have been shown to bind to methylated DNA. In other vertebrates such as zebrafish, MBD3 also binds methylated DNA (M.V., manuscript in preparation). In mammals, two mutations in the MBD3 MBD have resulted in a loss of high affinity methyl- CpG binding [44]. Interestingly, several eukaryotes such as *Drosophila melanogaster* and *Caenorhabditis elegans* do not have high levels of CpG methylation yet they do contain MBD proteins. The question therefore is whether the MBD domain evolved as a *bona fide* methyl-CpG binding domain or whether the domain originated as having a different function and later acquired methyl-CpG binding activity. Detailed structural analysis suggests that the MBD domain originally evolved as a methyl-CpG binding domain and that in most cases, such as *C. elegans*, loss of methyl-CpG binding co-evolved with a loss of CpG methylation while general DNA binding affinity was not affected [43]. The conserved similarity/homology between MBD sequences indicates that while meCpG binding function was not conserved for some members of the family, the domain was nevertheless retained, possibly evolving to function in other ways. MBD5 and MBD6 represent two good examples of this as their MBD domains seem to have evolved, following a loss of meCpG binding activity, as a protein-protein interaction module. A better understanding of this evolutionary process could come from analyzing MBD5 and MBD6 across different species. The few known members of the family that are present in invertebrates, such as flt-1 in *C.elegans* and Toutatis in *Drosophila melanogaster*, are considered homologues of MBD2/MBD3, which is considered the ancient MBD protein [1, 43]. It is possible that more homologues of the mammalian MBDs are yet to be discovered, such as the recently reported orthologue of MBD5 in *Drosophila melanogaster*, sba [1, 45]. Thus far, however, an interaction between *sba* and the Drosophila PR-DUB complex has not been reported.

In summary, the results described in this study have revealed both shared and specific functions for MBD5 and MBD6. Both proteins interact with PR-DUB through their MBD domain. Furthermore, MBD6 gets recruited to sites of DNA damage in a PR-DUB independent manner, something which was not observed for MBD5. Several key questions still remain. What is the exact function of MBD5 and MBD6 in relation to PR-DUB? What are the molecular mechanisms underlying the link between MBD5/PR-DUB and brain function? Is there functional redundancy between MBD5 and MBD6, and if there is, has this redundancy resulted in a tissue-specific expression pattern for MBD5 and MBD6? How is MBD6 recruited to sites of DNA damage and what is its exact role in DNA repair? Further studies are required to answer these questions.

Acknowledgements

We thank the Vermeulen and Marc Timmers's groups and Berend Snel for fruitful discussions. We thank Anthony A. Hyman for support and Sophie Polo for useful advice on DNA damage experiments. Steve Jackson and Yaron Galanty are acknowledged for providing the RFP-53BP1 plasmid. The microscopy experiments were performed at the imaging facility (ImagoSeine) at the Jacques Monod Institute (Paris, France). Work in the Vermeulen lab is supported by the FP7 Framework program '4DCellFate' and by the NWO Gravitation Program Cancer GenomiCs Netherlands. Work in the Defossez lab is supported by Association pour

la Recherche contre le Cancer (ARC), Institut National du Cancer (INCa), Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), and and by ANR-11-LABX-0071 under program ANR-11-IDEX-0005-01. Work in the Sharrocks lab is supported by the Wellcome Trust. The work of IP is supported by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreements n° 241548 (MitoSys).

References

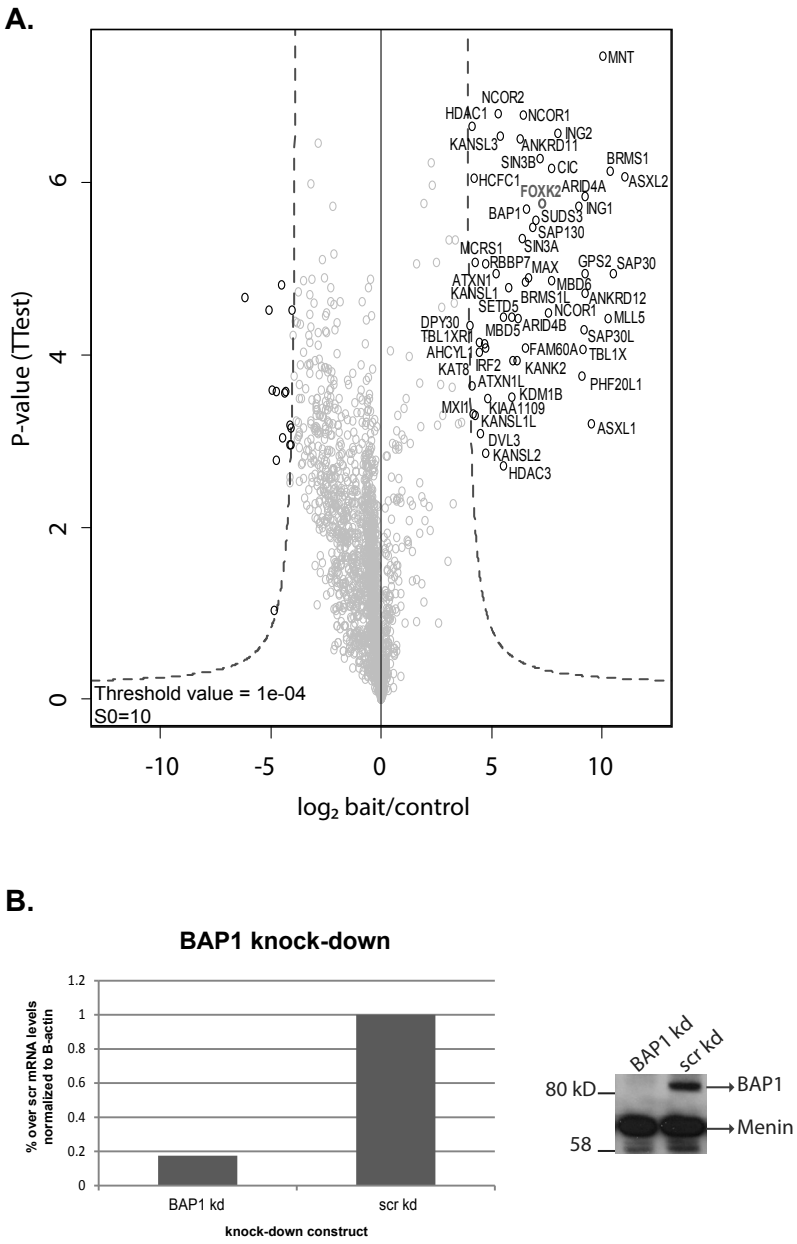
1. Hendrich, B. and S. Tweedie, (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet,* **19** (5):269-77.
2. Bird, A., (2011) The dinucleotide CG as a genomic signalling module. *J Mol Biol,* **409** (1):47-53.
3. Defossez, P.A. and I. Stancheva, Biological functions of methyl-CpG-binding proteins. *Prog Mol Biol Transl Sci,* **101** 377-98.
4. Hu, S., J. Wan, Y. Su, et al., (2013) DNA methylation presents distinct binding sites for human transcription factors. *Elife,* **2** e00726.
5. Spruijt, C.G., F. Gnerlich, A.H. Smits, et al., (2013) Dynamic readers for 5-(hydroxy) methylcytosine and its oxidized derivatives. *Cell,* **152** (5):1146-59.
6. Lewis, J.D., R.R. Meehan, W.J. Henzel, et al., (1992) Purification, Sequence, and Cellular-Localization of a Novel Chromosomal Protein That Binds to Methylated DNA. *Cell,* **69** (6):905-914.
7. Nan, X.S., R.R. Meehan, and A. Bird, (1993) Dissection of the Methyl-Cpg Binding Domain from the Chromosomal Protein Mecp2. *Nucleic Acids Res,* **21** (21):4886-4892.
8. Fournier, A., N. Sasai, M. Nakao, et al., The role of methyl-binding proteins in chromatin organization and epigenome maintenance. *Brief Funct Genomics,* **11** (3):251-64.
9. Laget, S., M. Joulie, F. Le Masson, et al., The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One,* **5** (8):e11982.
10. Motobayashi, M., A. Nishimura-Tadaki, Y. Inaba, et al., (2012) Neurodevelopmental features in 2q23.1 microdeletion syndrome: report of a new patient with intractable seizures and review of literature. *Am J Med Genet A,* **158A** (4):861-8.
11. Du, Y., B. Liu, F. Guo, et al., The essential role of Mbd5 in the regulation of somatic growth and glucose homeostasis in mice. *PLoS One,* **7** (10):e47358.
12. Scheuermann, J.C., A.G.D. Alonso, K. Oktaba, et al., (2010) Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature,* **465** (7295):243-U138.
13. Sowa, M.E., E.J. Bennett, S.P. Gygi, et al., (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell,* **138** (2):389-403.
14. Dey, A., D. Seshasayee, R. Noubade, et al., (2012) Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science,* **337** (6101):1541-6.
15. Yu, H., N. Mashtalir, S. Daou, et al., (2010) The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol,* **30** (21):5071-85.
16. Spruijt, C.G., H.I. Baymaz, and M. Vermeulen, (2013) Identifying specific protein-DNA interactions using SILAC-based quantitative proteomics. *Methods Mol Biol,* **977** 137-57.
17. Schram, A.W., R. Baas, P.W. Jansen, et al., (2013) A dual role for SAGA-associated factor 29 (SGF29) in ER stress survival by coordination of both histone H3 acetylation and histone H3 lysine-4 trimethylation. *PLoS One,* **8** (7):e70035.
18. Smits, A.H., P.W. Jansen, I. Poser, et al., (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res,* **41** (1):e28.
19. Cox, J. and M. Mann, (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol,* **26** (12):1367-72.
20. Poser, I., M. Sarov, J.R. Hutchins, et al., (2008) BAC TransgeneOmics: a high-throughput

method for exploration of protein function in mammals. *Nat Methods,* **5** (5):409-15.
21. Carbone, M., H.N. Yang, H.I. Pass, et al., (2013) BAP1 and cancer. *Nature Reviews Cancer,* **13** (3):153-159.
22. Zou, X., W. Ma, I.A. Solov'yov, et al., (2012) Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res,* **40** (6):2747-58.
23. Mayer, C., K.M. Schmitz, J. Li, et al., (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell,* **22** (3):351-61.
24. Iurlaro, M., G. Ficz, D. Oxley, et al., (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol,* **14** (10):R119.
25. Yang, Q., Y. Kong, B. Rothermel, et al., (2000) The winged-helix/forkhead protein myocyte nuclear factor beta (MNF-beta) forms a co-repressor complex with mammalian sin3B. *Biochem J,* **345 Pt 2** 335-43.
26. Shi, X. and D.J. Garry, (2012) Sin3 interacts with Foxk1 and regulates myogenic progenitors. *Mol Cell Biochem,* **366** (1-2):251-8.
27. Shi, X., D.C. Seldin, and D.J. Garry, (2012) Foxk1 recruits the Sds3 complex and represses gene expression in myogenic progenitors. *Biochem J,* **446** (3):349-57.
28. Marais, A., Z. Ji, E.S. Child, et al., (2010) Cell cycle-dependent regulation of the forkhead transcription factor FOXK2 by CDK.cyclin complexes. *J Biol Chem,* **285** (46):35728-39.
29. Schwanhausser, B., D. Busse, N. Li, et al., (2011) Global quantification of mammalian gene expression control. *Nature,* **473** (7347):337-42.
30. Ji, Z., I.J. Donaldson, J. Liu, et al., (2012) The forkhead transcription factor FOXK2 promotes AP-1-mediated transcriptional regulation. *Mol Cell Biol,* **32** (2):385-98.
31. Harper, J.W. and S.J. Elledge, (2007) The DNA damage response: ten years after. *Mol Cell,* **28** (5):739-45.
32. Zhou, W., X. Wang, and M.G. Rosenfeld, (2009) Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. *Int J Biochem Cell Biol,* **41** (1):12-5.
33. Jacq, X., M. Kemp, N.M. Martin, et al., (2013) Deubiquitylating enzymes and DNA damage response pathways. *Cell Biochem Biophys,* **67** (1):25-43.
34. Yu, H., H. Pak, I. Hammond-Martel, et al., (2014) Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. *Proc Natl Acad Sci U S A,* **111** (1):285-90.
35. Carbone, M., L.K. Ferris, F. Baumann, et al., (2012) BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs. *Journal of Translational Medicine,* **10**
36. Pena-Llopis, S., S. Vega-Rubin-de-Celis, A. Liao, et al., (2012) BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet,* **44** (7):751-9.
37. Abdel-Wahab, O. and A. Dey, (2013) The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics. *Leukemia,* **27** (1):10-15.
38. Vogelstein, B., N. Papadopoulos, V.E. Velculescu, et al., (2013) Cancer genome landscapes. *Science,* **339** (6127):1546-58.
39. Misaghi, S., S. Ottosen, A. Izrael-Tomasevic, et al., (2009) Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol Cell Biol,* **29** (8):2181-92.
40. Machida, Y.J., Y. Machida, A.A. Vashisht, et al., (2009) The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J Biol Chem,* **284** (49):34179-88.
41. Eletr, Z.M. and K.D. Wilkinson, (2011) An emerging model for BAP1's role in regulating cell cycle progression. *Cell Biochem Biophys,* **60** (1-2):3-11.
42. Aravind, L. and D. Landsman, (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res,* **26** (19):4413-21.
43. Iyer, L.M., S. Abhiman, and L. Aravind, (2011) Natural history of eukaryotic DNA methylation systems. *Prog Mol Biol Transl Sci,* **101** 25-104.
44. Saito, M. and F. Ishikawa, (2002) The mCpG-binding domain of human MBD3 does not bind

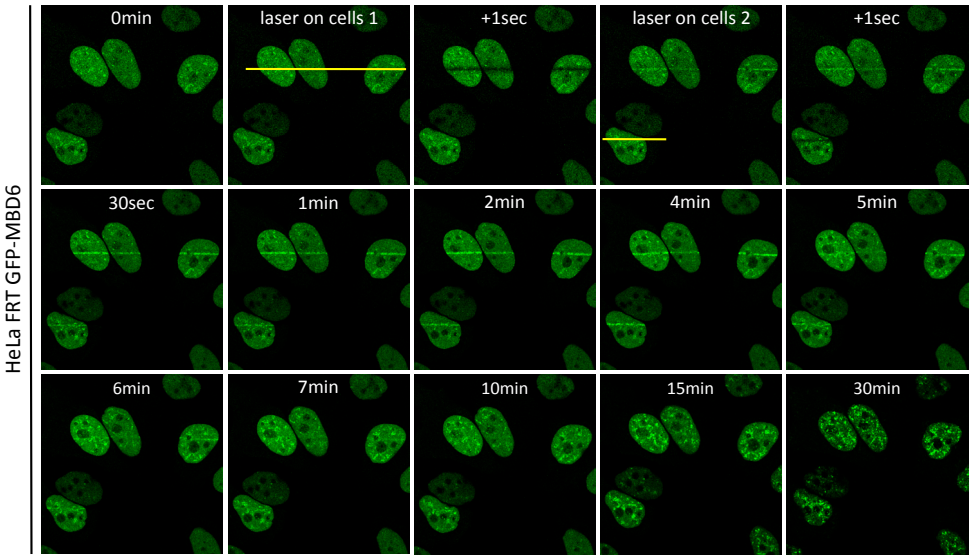
to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem,* **277** (38):35434-9.

45. Kleefstra, T., J.M. Kramer, K. Neveling, et al., (2012) Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am J Hum Genet,* **91** (1):73-82.

Supplementary Material



Supplementary figure 1A. GFP-FOXK2 interactors visualized in a volcano plot. GFP-FOXK2 interactors are localized in the right part of the figure. **1B.** Validation of shRNA mediated BAP1 knock-down in GFP-MBD6 HeLa FRT cells by qPCR (left) and Western blotting (right) .



Supplementary figure 2. Detailed kinetics of MBD6 recruitment to sites of laser-induced DNA damage. Live-cell imaging of laser-microirradiated HeLa FRT cells stably expressing GFP-MBD6 at the indicated time points.

Supplementary Table 1: Proteingroups output tables (Maxquant) of the SILAC and label-free based GFP affinity purifications and LC-MS/MS analyses. For the SILAC-based pull-downs (MBD5-GFP, MBD6-GFP, BAP1-GFP, ASXL2-GFP, MBDonly_MBD5-GFP , deltaMBD_MBD5-GFP), proteins are sorted according to the normalized forward SILAC ratio. For the label-free pull-down (FOXK2-GFP) according to their LFQ intensity (GFP1).
Available online: http://onlinelibrary.wiley.com/doi/10.1002/pmic.201400013/suppinfo

Addendum to Chapter 3

Part I: Characterization of the PR-Dub complex in mouse embryonic stem cells and neural progenitor cells
Part II: A mass spectrometry-based approach for investigating DNA damage-specific interactors of MBD6

H. Irem Baymaz[1], Pascal W.T.C. Jansen[1], Ina Poser[2], Michiel Vermeulen[1]

1. Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands
2. Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauer Str. 108, 01307 Dresden, Germany

Abstract

Part I

Polycomb complexes are highly expressed and functional in mouse embryonic stem cells, i.e. they repress genes that should be activated later in development[1]. The chromatin marks that are catalyzed by the two major Polycomb complexes, PRC1 and PRC2, are Histone H3K27me3 and Histone H2AK119Ub1. These marks are present abundantly in the mouse ESC genome[2,3]. Here we set out to investigate whether the complex that deubiquitinates Histone H2AK119Ub1, the PR-Dub complex, is also present in mESCs and also in a more differentiated lineage that is the neural progenitor cells (mNPCs). We characterize the PR-Dub complex in both settings and show that the interactions observed in purifications from HeLa cells are retained in pluripotent and multipotent cells.

Part II

In the previous chapter, we show that MBD6 is recruited to sites of damaged DNA. Here we investigate whether there are novel interactions induced by DNA damage using a mass spectrometry-based approach.

Part I: Characterization of the PR-Dub complex in mouse embryonic stem cells and neural progenitor cells

Introduction

The components of the human, mouse and Drosophila PR-Dub complexes have recently been identified and reported[4–6]. In our group, we have purified the human PR-Dub complex in HeLa cells stably transfected with a GFP-tagged BAP1 BAC and a GFP-tagged ASXL2 BAC as shown in chapter 3. We further wished to characterize the mouse PR-Dub complex in embryonic stem cells (mESCs) and neural progenitor cells (mNPCs). We stably transfected the BAP1 BAC that is recombined to express a C-terminally GFP-tagged fusion BAP1 protein[7] into mES cells. The same method was used for generating mESCs that stably express ASXL2-GFP fusion protein. Nuclear extracts generated from these lines were subjected to GFP affinity purification followed by mass spectrometry analysis (GFP AP-ms). NonGFP agarose beads were used in the control purification. Label-free quantitation algorithm was used to determine interactors (see figure 4A in the previous chapter for an overview of the method).

Results

Both human BAP1 protein and human ASXL2 proteins that are GFP-tagged are incorporated into the mouse ESC PR-Dub complex as a clear enrichment for interactors is observed in **Figure 1**. For BAP1, there are 2 potential novel interactors: Poldip2 and Fhl3. Poldip2 is a polymerase delta-interacting protein and Fhl3 may be a protein involved in muscle development[8]. Whether these proteins are true interactors of the mESC PR-Dub complex is unknown. To address this question, these proteins should be enriched reproducibly and specifically in subsequent PR-Dub subunit purifications. Interestingly, Fhl3 protein is enriched as a PR-Dub interactor in mouse NPCs as well (**Figure 2**) and its stoichiometry slightly increases in mNPCs (**Figure 3**).

The ASXL2-GFP AP-ms shows additional interactors (**Figure 1B**): Nr0b1, Nrip1 and Nanog. Nr0b1 and Nrip1 are proteins involved in nuclear receptor signaling[8]. ASXL2 is reported to be involved in peroxisome proliferator-activated receptor and retinoic acid receptor signaling[9,10]. Therefore, it is plausible that ASXL2 interacts with other proteins and complexes involved in these signaling pathways. It should be noted that the stoichiometry values for these proteins are rather low (<10%, data not shown) suggesting that ASXL2 is found most predominantly in the PR-Dub complex. An interaction between Nanog and ASXL2 has not been reported, however, Nr0b1 interacts with Nanog[11]. It is possible that ASXL2 indirectly interacts with Nanog via Nr0b1. Finally, as mentioned in chapter 3, ASXL1 and ASXL2 seem to be mutually-exclusive within the PR-Dub complex as ASXL1 is only identified in BAP1 purifications but not in ASXL2 purifications.

In summary, in this section we show the PR-Dub complex in mESCs using BAP1-GFP and ASXL2-GFP fusion proteins as baits. In addition, we show the PR-Dub complex in mNPCs using BAP1-GFP as a bait. Although there are candidate novel interactors in the three AP-ms experiments shown in Figure 1 and 2, the interactors with the highest stoichiometry do not change: Bap1, Asxl1, Asxl2, Kdm1b, Foxk1/2, Hcfc1, Ogt and Mbd6 (**Figure 3**). Mbd5 is consistently identified as an interactor albeit with low stoichiometry. The enrichment for Hcfc1 and Ogt show variation between experiments (not detected in ASXL2-GFP AP-ms and close to the background cloud in BAP1 AP-ms experiments (**Figure 1 and 2**)), however the stoichiometry for Hcfc1 is consistently high. For Ogt, the stoichiometry seems to decrease in mNPCs (**Figure 3**). Finally, based on the PR-Dub stoichiometry data in mESCs and mNPCs,

PR-Dub complex dynamics during differentiation seem to be more stable in comparison to PRC1 dynamics (S.L. Kloet, unpublished observations; [12]).

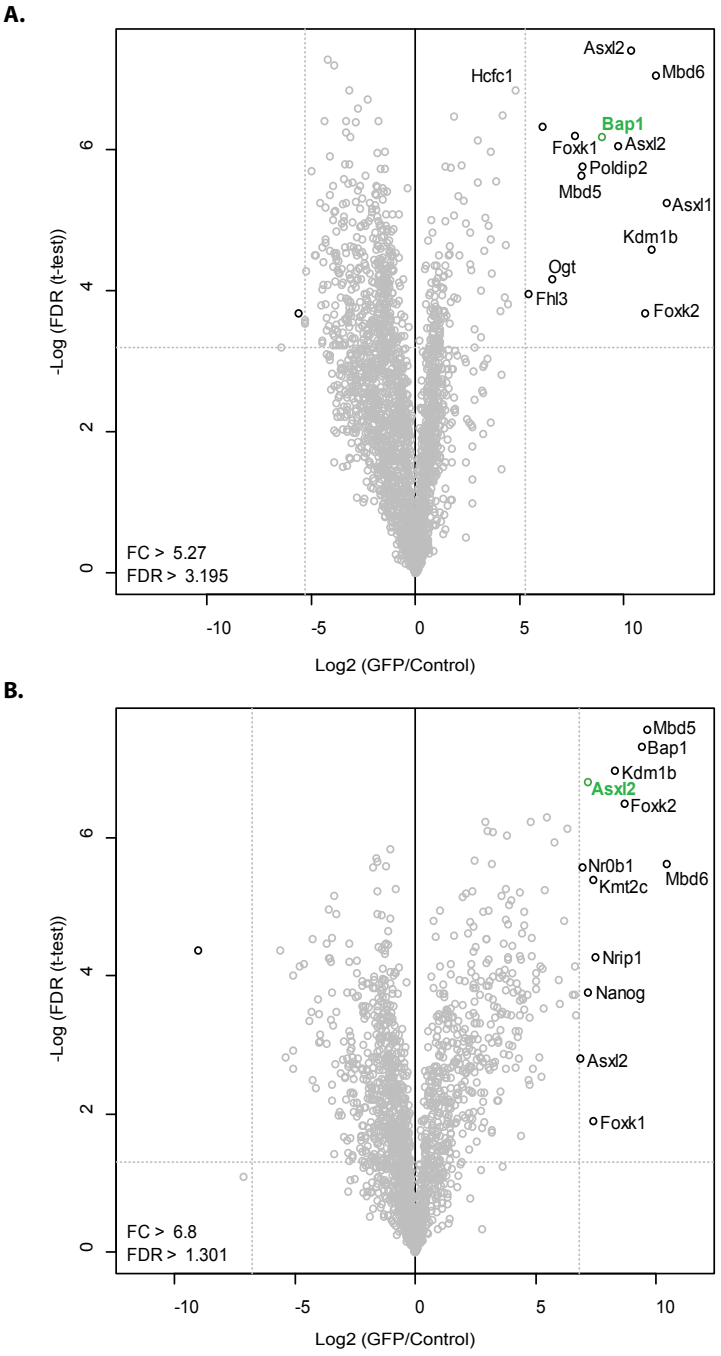


Figure 1: GFP affinity purification of PR-Dub subunits in mouse embryonic stem cells. **Panel A** shows the interactors of BAP1-GFP and **Panel B** shows the interactors of ASXL2-GFP in mESCs. The GFP tagged baits are shown in green. X-axis shows the ratio of enrichment between the specific and control affinity purifications. Y-axis shows the reproducibility of quantification between three replicates.

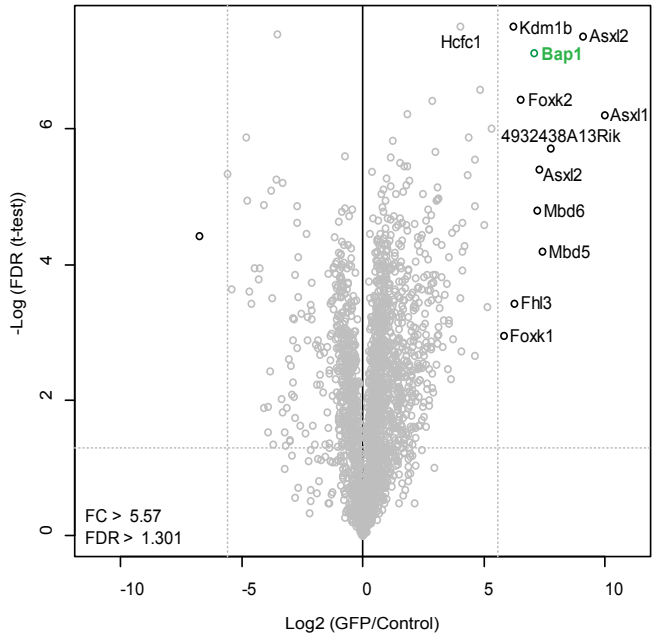
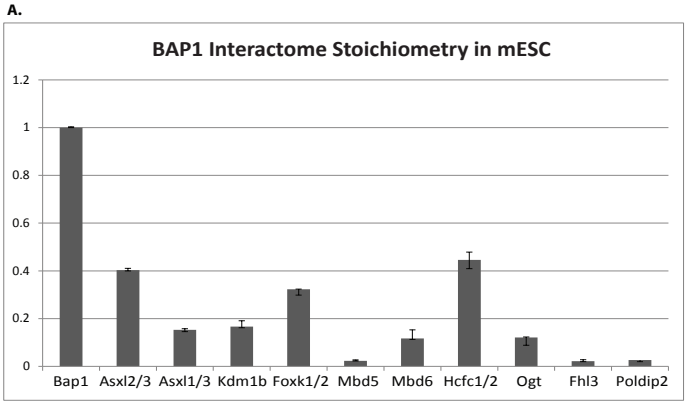


Figure 2: GFP affinity purification of BAP1-GFP in mouse neural progenitor cells.



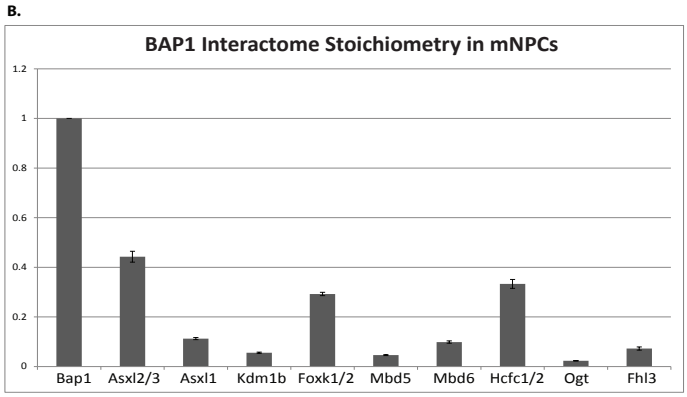


Figure 3: The stoichiometry of BAP1 interactors in mouse ESCs **(A)** and mouse NPCs **(B)**. The stoichiometry plots are based on the affinity purification experiments shown in Figure 1A and Figure 2. The stoichiometry values of interactors are normalized to the bait and the stoichiometry of the bait is set to 1. "4932438A13Rik" protein observed in Figure 2 is excluded from stoichiometry calculations shown in **panel B** due to low number of peptide counts (n=2). Some interactors (e.g. Foxk1 and 2) are collapsed due to the contribution of shared peptides to stoichiometry values.

Materials and methods

R1 mESCs were cultured in DMEM high glucose, FBS, LIF, pen/strep, L-glutamine and 2i inhibitors. Human BAP1 and ASXL2 BACs recombined to express a GFP-tagged fusion protein were transfected into R1 mESCs using Lipofectamine 2000, according to manufacturer's instructions. 200 ug/mL Geneticin was used to select for cells stably expressing the fusion protein. mESCs expressing BAP1-GFP (referred to as BAP1-GFP mESCs) were differentiated into mNPCS (referred to as BAP1-GFP mNPCs) in N2B27 medium. After differentiation, mNPCS were cultured in NSA medium with 1% L-glutamine, 1x N2 supplement, 10 ng/mL bFGF and 10 ng/mL EGF[13]. Nuclear extraction, GFP affinity purification, mass spectrometry measurement and analysis were performed as decribed in chapter 3. Stoichiometry analysis was performed as described in Smits et. al[14].

Part II: A mass spectrometry-based approach for investigating DNA damage-specific interactors of MBD6

Introduction

MBD6 localizes to microirradiated sites on DNA as shown in the previous chapter. While we have shown that this localization does not depend on BAP1, we do not have further information on the interactors of MBD6 at damage sites and whether any novel interactions are induced upon DNA damage. Is there a novel interactor that recruits MBD6 to sites of DNA damage? Are there novel interactors that MBD6 recruits to sites of DNA damage? To try to address these questions, we performed quantitative interaction proteomics in gamma irradiated MBD6 HeLa FRT cells.

Results

MBD6-GFP expression was induced in all four groups of SILAC-labeled MBD6 HeLa FRT cells. One heavy labelled and one light labelled group of cells were subjected to 10 Gys of gamma irradiation for 10 minutes (see **Figure 4A** for an overview of the workflow). Subsequent steps of nuclear extraction and GFP AP-ms were performed as described in chapter 2. In this experimental set-up, the bait (MBD6-GFP) and interactors (PR-Dub) observed under normal conditions (no induction of DNA damage) would end up in the background cloud in the origin of the scatter plot as the GFP-tagged protein is present in all 4 groups. Only DNA damage-specific interactors would be enriched in the right upper quadrant with a high forward and low reverse ratio. As seen in **Figure 4B**, there are no novel, irradiation-specific interactors detected in this screen. The dots that stand out in the left upper quadrant correspond to contaminant proteins.

In summary, we could not detect any damage-specific interactors when we performed AP-ms using MBD6-GFP as a bait after exposing the cells to gamma irradiation. It is possible that the subsequent steps of preparing nuclear extracts and performing GFP AP following gamma irradiation hinder the detection of interactions formed transiently after DNA damage induction. As seen in chapter 3, the dynamics of MBD6 recruitment to and dissociation from sites of irradiated DNA may be too fast for capturing damage-induced interactions with the methodology that we use.

Materials and methods

Inducible MBD6-GFP HeLa FRT cells were created as described in van Nuland et al.[15]. MBD6-GFP HeLa FRT cells were cultured in DMEM (high glucose), 1% pen-strep, 10% FBS. Approximately ~15 hours before gamma irradiation, the expression of MBD6-GFP was induced by the addition of doxycycline (1 ug/ml final) to culture medium. Half of the cells were exposed to 10 Grays of gamma irradiation for 10 minutes. Nuclear extraction and GFP AP-ms were performed as described in chapter 2.

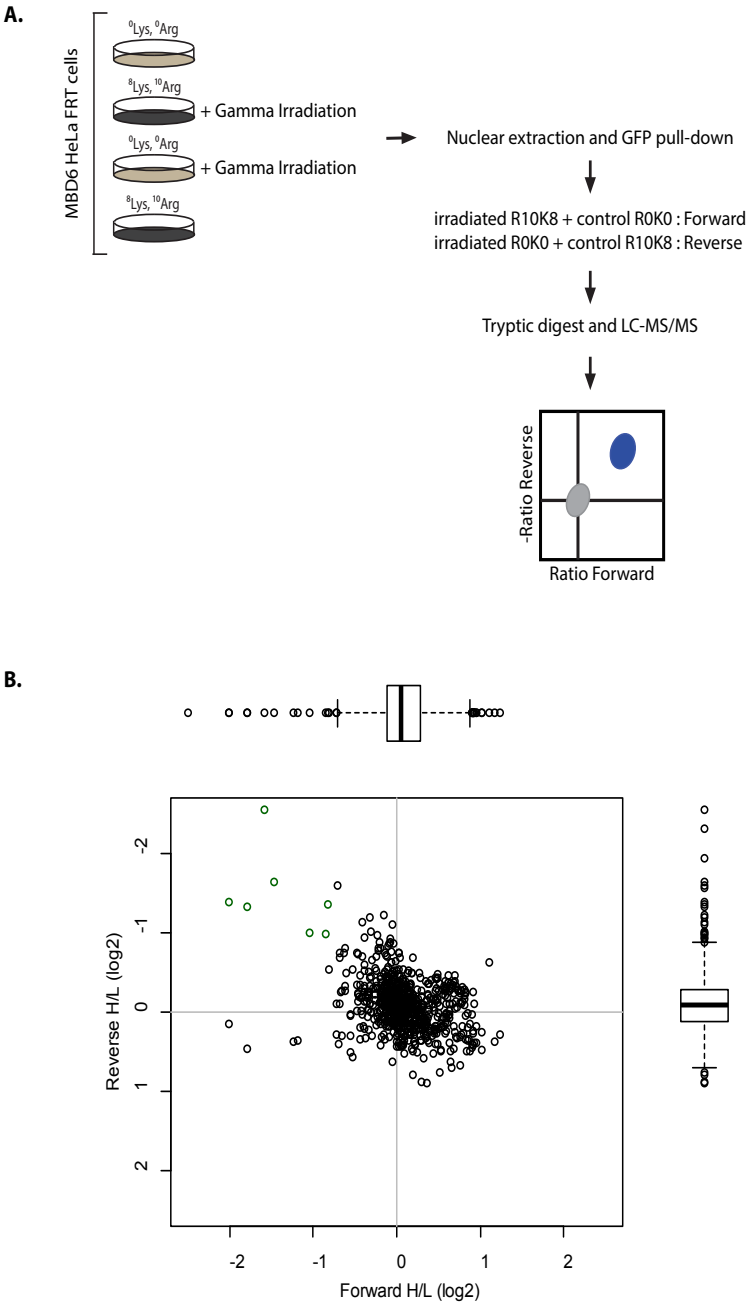


Figure 4: A quantitative interaction screen for MBD6-GFP upon DNA damage
Panel A shows the overview of the protocol. Half of SILAC labeled MBD6-GFP HeLa FRT cells were exposed to gamma irradiation. In this set-up, only the damage-induced interactors would stand out from the background cloud. **Panel B** shows the SILAC scatterplot. MBD6-GFP and other known interactors are in the background cloud. There are no specific, DNA-damage induced interactions.

Acknowledgments
We thank Radhika A. Warrier for setting up the gamma irradiation assay and Arne H. Smits for the R scripts that were used to create the interaction and stoichiometry plots.

References

1. Boyer, L. a *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
2. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).
3. Farcas, A. M. *et al.* KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *Elife* **2012,** 1–26 (2012).
4. Machida, Y. J., Machida, Y., Vashisht, A. a., Wohlschlegel, J. a. & Dutta, A. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J. Biol. Chem.* **284,** 34179–34188 (2009).
5. Yu, H. *et al.* The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell. Biol.* **30,** 5071–5085 (2010).
6. Dey, a. *et al.* Loss of the Tumor Suppressor BAP1 Causes Myeloid Transformation. *Science (80-.).* **337,** 1541–1546 (2012).
7. Poser, I. *et al.* BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5,** 409–415 (2008).
8. http://www.uniprot.org.
9. Park, U. H., Yoon, S. K., Park, T., Kim, E. J. & Um, S. J. Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor. *J. Biol. Chem.* **286,** 1354–1363 (2011).
10. Khan, F. F., Li, Y., Balyan, A. & Wang, Q. T. WTIP interacts with ASXL2 and blocks ASXL2-mediated activation of retinoic acid signaling. *Biochem. Biophys. Res. Commun.* **451,** 101–106 (2014).
11. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444,** 364–368 (2006).
12. Morey, L. *et al.* Nonoverlapping functions of the polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10,** 47–62 (2012).
13. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3,** 1594–1606 (2005).
14. Smits, A. H., Jansen, P. W. T. C., Poser, I., Hyman, A. a. & Vermeulen, M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.* **41,** 1–8 (2013).
15. Van Nuland, R. *et al.* Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol. Cell. Biol.* **33,** 2067–77 (2013).

Chapter 4

Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression

Reinhard Kalb[1], Sebastian Latwiel[2,4], H. Irem Baymaz[3,4], Pascal W.T.C. Jansen[3], Christoph W. Müller[2], Michiel Vermeulen[3] and Jürg Müller[1,5]

1. Max-Planck Institute of Biochemistry, Laboratory of Chromatin and Chromosome Biology, Martinsried, Germany
2. European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, Germany
3. Molecular Cancer Research and Cancer Genomics Netherlands, University Medical Center Utrecht, Utrecht, The Netherlands
present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands

4. These two authors contributed equally to this work
5. Corresponding author

Nature Structural & Molecular Biology, June 2014,
Volume 21, Number 6, pages 569-572

Abstract

A key step in gene repression by Polycomb is trimethylation of lysine 27 in histone H3 (H3-K27me3) by PRC2. H3-K27me3 provides a binding surface for PRC1. We show that monoubiquitination of histone H2A (H2Aub) by PRC1-type complexes in turn creates a binding site for Jarid2-Aebp2-containing PRC2 and promotes H3-K27me3 deposition on H2Aub nucleosomes. Jarid2, Aebp2 and H2Aub thus constitute components of a positive feedback loop that establishes H3-K27me3 chromatin domains.

Nucleosomes constitute the building block of eukaryotic chromosomes. They consist of a core of histone proteins around which DNA is wrapped in two helical turns. The post-translational modification of histones is a key step for the regulation of diverse processes that occur on nucleosomal DNA. Specific histone modifications often decorate arrays of nucleosomes that comprise many kilobases of DNA, but how such extended stretches of chromatin become modified is not well understood. A paradigm for a long-range chromatin modification mechanism is transcriptional repression by Polycomb protein complexes[1, 2]. The Polycomb system generates two distinct histone modifications: methylation of lysine 27 in histone H3 (H3-K27me) and monoubiquitination of lysine 119 in histone H2A (H2Aub) in vertebrates and at the corresponding lysine 118 in *Drosophila* H2A. Polycomb Repressive Complex 2 (PRC2) catalyzes mono-, di- and tri-methylation at H3-K27 (H3-K27me1/2/3)[1, 2]. At inactive Polycomb target genes, H3-K27me3 typically decorates nucleosomes across the entire upstream, promoter and coding region (www.modencode.org) and is essential for their repression[3]. The H3-K27me3 modification is recognized by Polycomb (Pc), a subunit of the canonical Polycomb repressive complex 1 (PRC1) and is thought to promote PRC1 interaction with chromatin across the entire length of repressed genes. PRC1 has been proposed to repress transcription through chromatin compaction and also through its ubiquitin ligase activity for H2A monoubiquitination[1,2]. To gain insight into the function of H2Aub, we set out to identify interactors of this modification.

We reconstituted arrays of four nucleosomes (referred to as oligonucleosomes) with recombinant *Drosophila* or *Xenopus* histones and monoubiquitinated H2A in these templates using appropriate recombinant enzymes (**Supplementary Figure 1**). We then used *Drosophila* H2A-K118ub oligonucleosomes and the corresponding unmodified oligonucleosome control template for affinity purification of H2A-K118ub binding proteins from *Drosophila* embryo nuclear extracts (**Figure 1a**). In parallel, we used *Xenopus* H2A-K119ub and unmodified control oligonucleosomes to identify vertebrate H2A-K119ub interactors in nuclear extracts from mouse embryonic stem cells (**Figure 1b**). In both experiments, quantitative mass spectrometry analyses identified PRC2 subunits as or among the most highly enriched H2Aub interactors (**Figure 1a, b**). Jarid2 and Aebp2 were the PRC2 subunits showing highest enrichment in both cases (**Figure 1a, b**).

The identification of PRC2 as an H2Aub interactor in both flies and vertebrates prompted us to analyze PRC2 histone methyltransferase (HMTase) activity on H2Aub nucleosomes. We reconstituted recombinant human PRC2 containing EED, EZH2, SUZ12, RBBP4 (here referred to as PRC2) and assemblies of the same complex that in addition contained AEBP2 (AEBP2–PRC2), or JARID2 (JARID2–PRC2) or both JARID2 and AEBP2 (JARID2–AEBP2–PRC2 (**Supplementary Figure 2a**). As substrates we used *Xenopus* mononucleosomes that were either unmodified or monoubiquitinated at H2A-K119 (**Supplementary Figure 2b**), and in all cases we used western blot analyses with antibodies against H3-K27me1 and H3-K27me3 to monitor PRC2 activity. We first performed a time course experiment to compare the activity of PRC2 and JARID2–AEBP2–PRC2 on H2A and H2Aub nucleosomes. Consistent with earlier reports[4,5], we found that the catalytic activity of PRC2 alone is largely unchanged on H2Aub nucleosome templates (**Figure 2a** lanes 1-8). As expected[6-9], inclusion of JARID2 and AEBP2 in PRC2 resulted in stronger activity for H3-K27 methylation already on unmodified nucleosome templates (**Figure 2a**, compare lanes 9-12 with lanes 1-4; see also refs. 6-9). However, we observed a much stronger increase in H3-K27me3 formation when we used JARID2–AEBP2–PRC2 for HMTase reactions on H2Aub nucleosomes (**Figure 2a**, compare lanes 13-16 with lanes 9-12). We estimate that JARID2–AEBP2–PRC2 tri-methylates H3-K27

in H2Aub nucleosomes with 25-fold higher efficiency compared to PRC2 (**Figure 2a**, compare lanes 4 and 16; see also **Supplementary Figure S2c,d**). To assess the contributions of JARID2 and AEBP2 to this stimulation of HMTase activity, we next compared the catalytic activity of all four forms of PRC2 on H2A and H2Aub nucleosome substrates (**Figure 2b**). JARID2–PRC2 showed higher H3-K27 methyltransferase activity than PRC2 on unmodified nucleosomes, as previously reported[9], but this was not any further increased on H2Aub nucleosomes (**Figure 2b**, compare lanes 5-8 and lanes 21-24). In contrast, AEBP2–PRC2 methylated H3-K27 in H2Aub nucleosomes with considerably higher efficiency than in unmodified nucleosomes (**Figure 2b**, compare lanes 9-12 with lanes 25-28). This suggests that AEBP2 is critical for the specific activation of PRC2 by H2Aub, whereas JARID2 has a more general function in boosting PRC2 HMTase activity, independently of the H2A modification state.

The work reported here uncovers that Jarid2-Aebp2-containing PRC2 binds to H2Aub nucleosomes and demonstrates that H3-K27 tri-methylation by this complex is strongly enhanced on H2Aub nucleosomes. This establishes H2Aub, Aebp2 and Jarid2 as components of a positive feedback loop where H2Aub promotes PRC2 binding and H3-K27me3 deposition, and in turn H3-K27me3 promotes binding of the canonical PRC1 via the chromodomain of Polycomb (**Figure 2c**). It is currently not clear whether canonical PRC1 indeed has E3 ligase activity for H2A monoubiquitination, or whether this modification is only generated by forms of PRC1 lacking Pc (e.g. refs. 10, 11). Intriguingly, in ES cells, we also identified the PRC1-type complexes, PRC1.1 and PRC1.6 (refs. 12, 13), as H2Aub interactors (**Figure 1b**), suggesting an additional feedback loop for H2Aub deposition in vertebrates. The positive feedback loop for H3-K27me3 formation by H2Aub uncovered here provides a rationale for how extended domains of Polycomb-repressed chromatin could be generated in both *Drosophila* and vertebrates (**Figure 2c**). Our findings could explain why H3-K27me3 levels at Polycomb target genes are reduced in murine embryonic stem cells where H2A-K119ub has been depleted[14]. However, we previously found that bulk H3-K27me3 levels were undiminished in late stage *Drosophila* larvae where bulk H2Aub levels had been depleted[15], suggesting that maintenance of H3-K27me3-modified chromatin domains does not strictly depend on H2Aub. The H2Aub-mediated feedback loop may thus primarily be required for the initial formation of H3-K27me3 chromatin domains when Polycomb repression is first established during the early stages of embryogenesis.

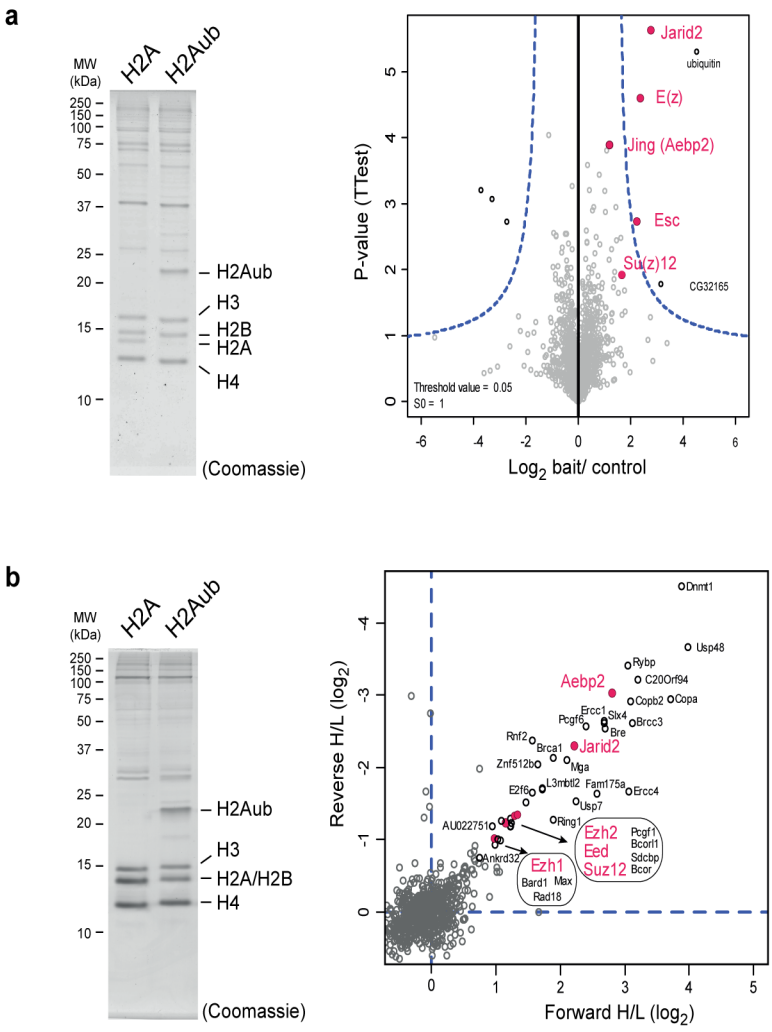


Figure 1: H2Aub nucleosomes bind Jarid2–Aebp2–PRC2 (**a**) Jarid2–Aebp2–PRC2 is the major H2Aub interactor in Drosophila. Left: Proteins from *Drosophila* embryo nuclear extracts, affinity-purified with unmodified (left lane) or H2Aub (right lane) nucleosomes were separated on a 16% polyacrylamide gel and visualized by coomassie staining; Supplementary Figure S3 shows all lanes of the triplicate pull-down reactions. Right: H2Aub interactors were identified by mass spectrometry using permutation-based FDR corrected T-test (threshold: p=0.05 & S0=1). The LFQ intensity of proteins in the H2Aub pull-down over the H2A control is plotted against the –Log10 (p-value) of the T-test in a volcano plot. The blue line indicates the permutation-based FDR threshold. Esc, E(z) and Jarid2 were identified as significant H2Aub interactors.

(**b**) Jarid2–Aebp2–PRC2 is an H2Aub interactor in mammalian cells. Left: as in (a) but with proteins from mouse embryonic stem cell nuclear extracts that were *s*table *i*sotope *l*abeled by *a*mino acids in *c*ell culture (SILAC)[16]; Supplementary Figure S3 shows all lanes of the SILAC pull-down. Right: Proteins identified by mass spectrometry in the H2A/H2Aub pull-downs were plotted by their log2 transformed normalized SILAC-ratios in the forward (x-axis) and reverse (y-axis) experiment (see Online Methods). Note that apart from PRC2, also PRC1.1 and PRC1.6 subunits and Dnmt1 were identified as specific H2Aub interactors (see also Supplementary table 1).

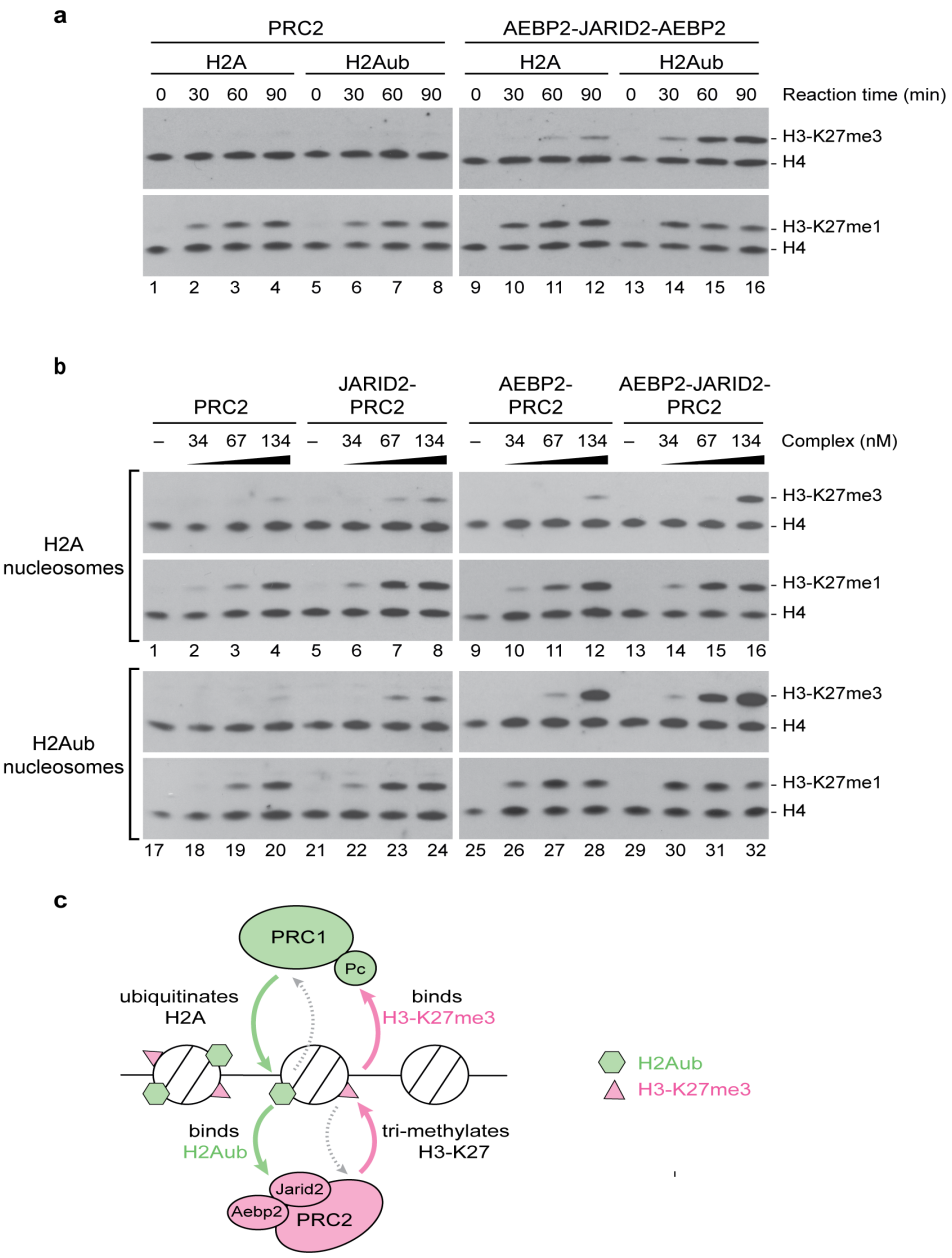


Figure 2: Aebp2 and Jarid2 stimulate H3-K27 methylation by PRC2 on H2Aub nucleosomes
(a) Western blot analysis of H3-K27me1 and -me3 formation in HMTase reactions with PRC2 (134 nM, lanes 1-8) or JARID2–AEBP2–PRC2 (134 nM, lanes 9-16) on unmodified (lanes 1-4 and 9-11) or H2Aub (lanes 5-8 and 13-16) mononucleosomes (460 nM) after indicated reaction times; histone H4 signal served as loading control. JARID2–AEBP2–PRC2 shows higher activity for H3-K27me3 formation on H2Aub than on unmodified nucleosomes (compare lanes 13-16 with 9-12) whereas PRC2 activity is comparable on both substrates (compare lanes 5-8 with 1-4); see also (b).

(b) Western blot analysis as in (a) of HMTase reactions with increasing amounts of the indicated PRC2 complexes on unmodified (lanes 1-16) or H2Aub (lanes 17-32) mononucleosomes (460 nM); reaction time was 90min in all cases. Comparison of H3-K27 methylation by each complex on H2Aub versus unmodified nucleosomes shows that only AEBP2–PRC2 and JARID2–AEBP2–PRC2 show higher activity on H2Aub nucleosomes. Supplementary Figures S2c,d shows results on oligonucleosome substrates and Supplementary Figure S4 shows original blots.
(c) Positive feedback loop model for generating Polycomb-repressed chromatin. Monoubiquitination of H2A by PRC1-type complexes promotes Jarid2–Aebp2–PRC2 binding and H3-K27me3 deposition; H3-K27me3 in turn promotes binding of canonical PRC1 via its Polycomb (Pc) subunit (solid green and magenta arrows). PRC1.1 and PRC1.6 are additional H2Aub interactors (Figure 1b) and PRC2 binds to and is stimulated by H3-K27me3 (refs 1,2), suggesting additional positive feedback loops for H2Aub and H3-K27me3 formation (dashed arrows).

Acknowledgments

This project was supported by the European Commission 7th Framework Program 4DCellFate grant number 277899 to C.W.M., M.V. and J.M. R.K. and J.M. acknowledge support by the Microchemistry core facility at the Max-Planck Institute of Biochemistry and financial support from the Max-Planck-Society. S.L. and C.W.M. acknowledge support by the Proteomics Core Facility at the European Molecular Biology Laboratory (EMBL). C.W.M acknowledges support by EMBL. M.V. acknowledges support from the Netherlands Organization for Scientific Research (NWO-VIDI and Cancer Genomics Netherlands). M.V. thanks C.G. Spruijt and A.H. Smits in his lab for providing SILAC labeled mouse ES cell extracts and for help with data analysis, respectively.

Author contributions

R.K. generated the unmodified and H2Aub nucleosome substrates, prepared Drosophila nuclear extracts, performed the pull-down experiments and the HMTase assays. H.I.B., P.W.T.C.J. performed the mass spectrometry analyses. S.L. prepared PRC2 and AEBP2-PRC2, and R.K. prepared JARID2. J.M., M.V. and C.W.M. supervised the project. J.M. and R.K. wrote the paper.

References

1. Margueron, R., Reinberg, D. *Nature* **469**, 343-49 (2011).
2. Simon, J.A., Kingston, R.E., *Mol Cell* **49**, 808-24 (2013).
3. Pengelly, R.A. *et al*, *Science* **339**, 698-99 (2013).
4. Whitcomb, S.J. *et al*, *J Biol Chem* **287**, 23718-25 (2012).
5. Yuan, G. *et al*, *J Biol Chem* **288**, 30832-42 (2013).
6. Cao, R. and Zhang, Y., *Mol Cell* **15**, 57-67 (2004).
7. Li, G. *et al*, *Genes Dev* **24**, 368-80 (2010).
8. Zhang Z. *et al*, *Stem Cells* **29**, 229-40 (2011).
9. Son, J. *et al*, *Genes Dev* **27**, 2663-2677 (2013).
10. Wang, H. *et al*, *Nature* **431**, 873-878 (2004).
11. Lagarou, A. *et al*, *Genes Dev* **22**, 2799-2810 (2008).
12. Gearhart, M.D. *et al*, *Mol Cell Biol* **26**, 6880-6889 (2006).
13. Gao, Z. *et al*, *Mol Cell* **45**, 344-356 (2012).
14. Endoh, M. *et al*, *PLoS Genet* **8**, e1002774 (2012).
15. Gutierrez, L. *et al*, *Development* **139**, 117-127 (2012).
16. Ong, S.E. *et al*, *Mol Cell Proteomics* **1**, 376-86 (2002).

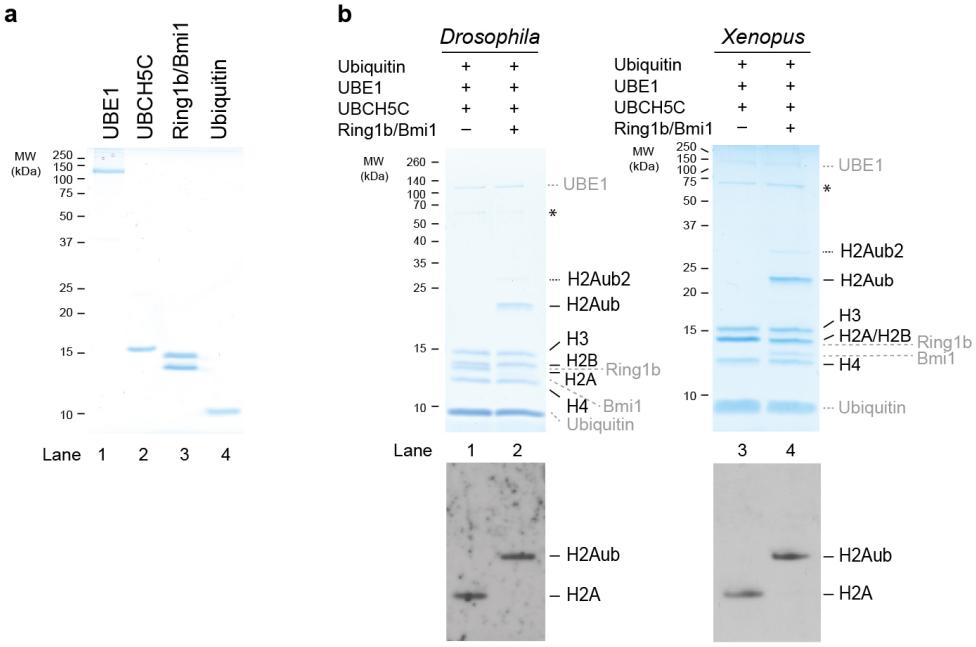
Supplementary Material

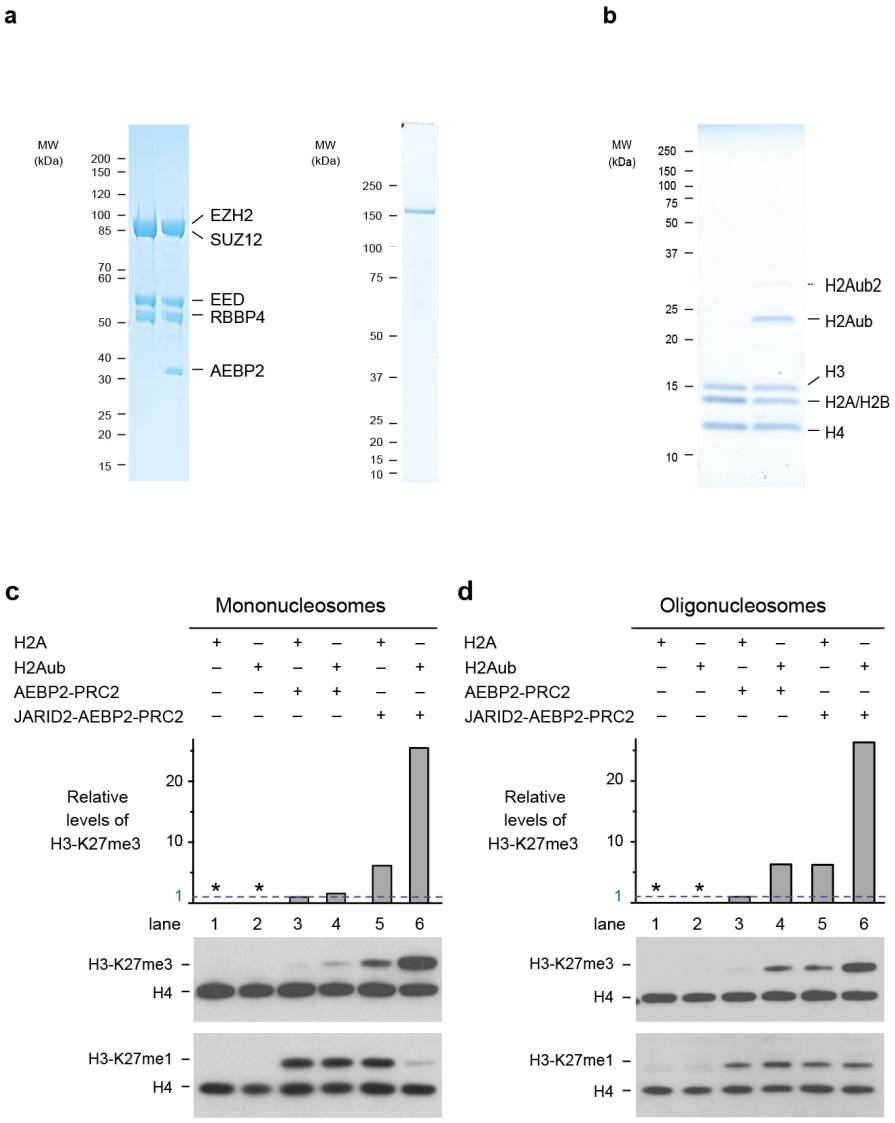
Supplementary Figure 1:

Recombinant proteins used for in vitro ubiquitination and methylation of nucleosomes.

(a) Coomassie-stained 4-12% polyacrylamide gel showing purified recombinant human ubiquitin activating enzyme UBE1 (lane1), human ubiquitin-conjugating enzyme UBCH5C (lane 2), mouse ubiquitin ligase Ring1b1-130/Bmi11-109 (lane 3) and Ubiquitin (lane 4).

(b) Reconstituted recombinant Drosophila or Xenopus oligonucleosomes were subjected to ubiquitination (+ Ring1b1-130/Bmi11-109, lane 2 in both cases) or mock (– Ring1b1-130/Bmi11-109, lane 1 in both cases) reactions and products were visualized by Coomassie staining after separation on denaturing 16% Tris-Glycine polyacrylamide gels (top) Note the size shift of the unmodified H2A band in lanes 1 and 3 to the monoubiquitinated H2A band in lanes 2 and 4, respectively; in both cases, a very small fraction appears to become di-ubiquitinated (H2Aub2). The band marked by an asterisk is a contaminant that sometimes is observed in UBE1 preparations. Bottom: Western blot analysis of the same reactions shown above with antibody against unmodified histone H2A. Reaction products were separated on 4-12% BT MES polyacrylamide gels. Note the size shift of the unmodified H2A band in lanes 1 and 3 to the monoubiquitinated H2A band in lanes 2 and 4, respectively.





Supplementary Figure 2:
Recombinant proteins and nucleosomes used for HMTase assays on mono- and oligonucleosomes
(a) Coomassie stained polyacrylamide gels with purified recombinant PRC2 (left), AEBP2-PRC2 (middle) and JARID2 (right). Recombinant purified PRC2 consisting of EZH2, SUZ12, EED, RBBP4 (left). AEBP2-PRC2 additional contains AEBP2 (middle). These PRC2 complexes and JARID2 were used for the experiments shown in Fig. 2 and Figure S2.
(b) Non-ubiquitinated and ubiquitinated mononucleosomes after coupling to streptavidin coated Dynabeads (M-280).
(c) and **(d)** JARID2 and H2Aub stimulate H3-K27 methyltransferase activity of AEBP2-PRC2. (c) Western blot analysis of H3-K27me1 and -me3 formation in HMTase reactions performed with AEBP2-PRC2 (134 nM, lanes 3-6) in the absence (lanes 3, 4) or presence of JARID2 (120 nM, lanes 5,6) on recombinant Xenopus mononucleosomes (460 nM) that were unmodified (lanes 1, 3, 5)

or contained H2Aub (lanes 2, 4, 6) (see Figure S1); histone H4 signal served as loading control. Histogram shows quantification of H3-K27me3 chemiluminescence signal by ImageJ; no signal is detected in lanes 1, 2 (asterisk). Note the increase of H3-K27me3 tri-methylation when H2A is monoubiquitinated and JARID2 is added to the reaction. The reduction of H3-K27me1 signal in lane 6 suggests that most H3-K27 residues became di- or tri-methylated. (d) as in (c) but with recombinant Xenopus 4-mer oligonucleosomes (460 nM) as substrate. The reaction contained 67 nM AEBP2-PRC2 and 60 nM JARID2. Supplementary Figure S5 shows original blots.

Figure 1a

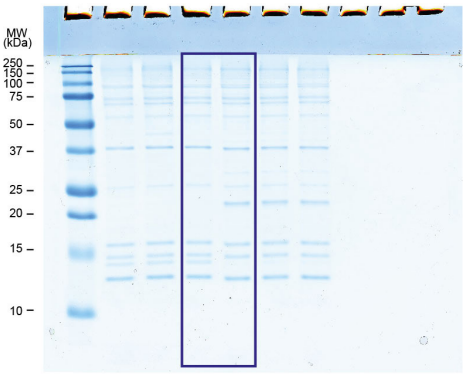
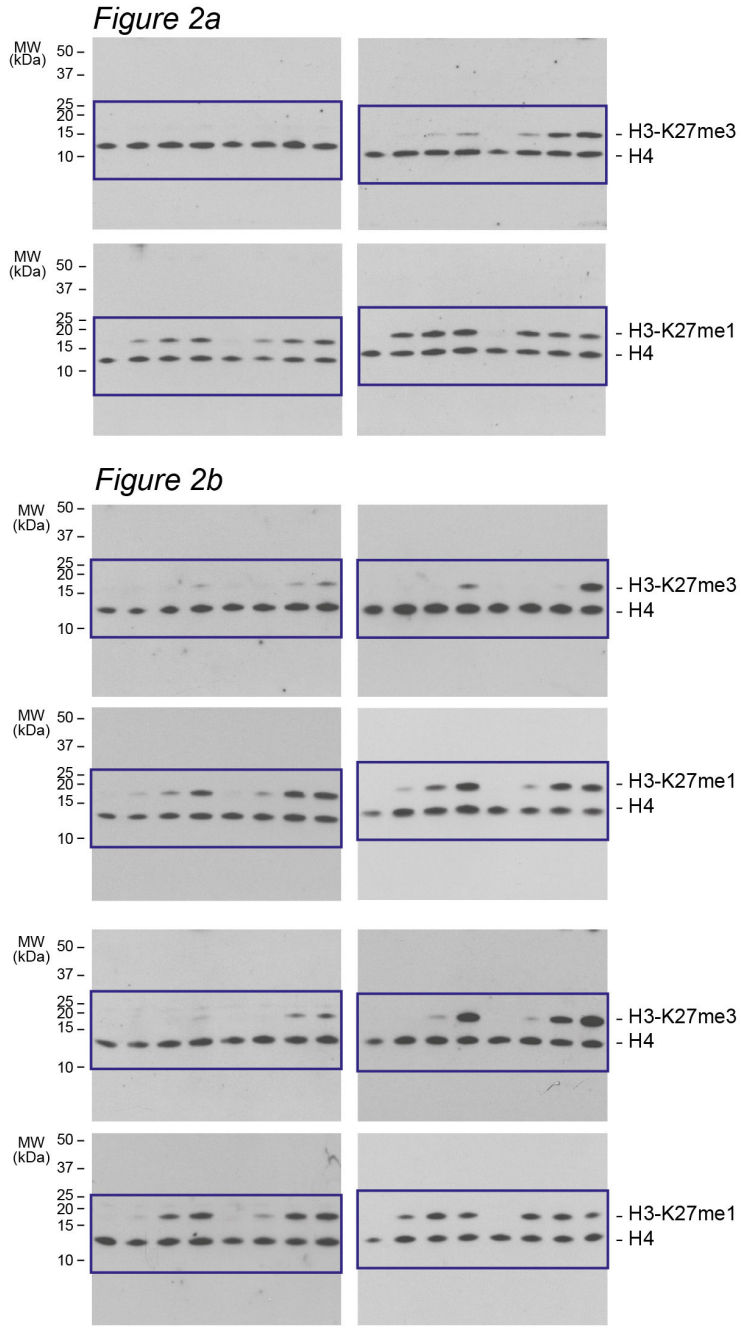
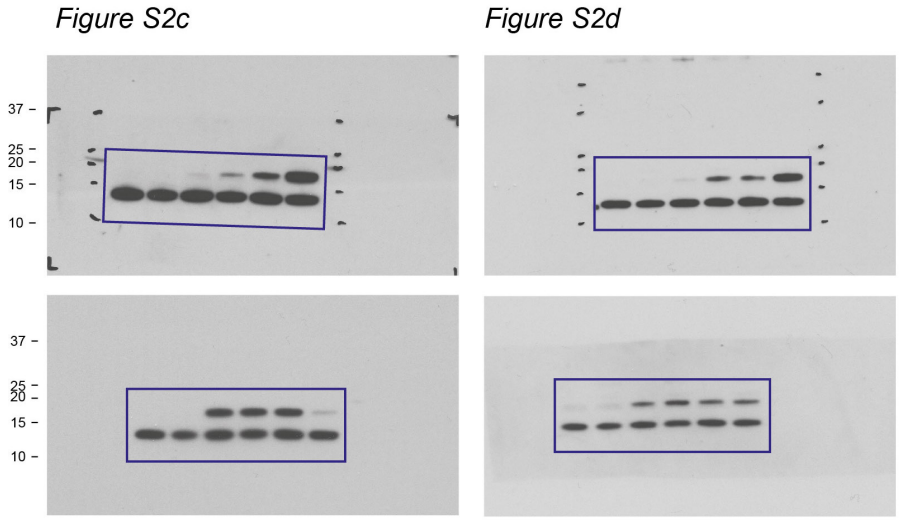


Figure 1b

Supplementary Figure 3:
Original Coomassie gels for pull-down reactions.
Left: Full gel image of the triplicate pull-down reactions from Drosophila nuclear extracts. Area of the gel boxed in blue is shown in Figure 1a.
Right: Full gel image of the SILAC pull-down experiment from mouse embryonic stem cell nuclear extracts. Area of the gel boxed in blue is shown in Figure 1b.



Supplementary Figure 4:
Full-size scans of western blot membranes shown in Figure 2. Scans are arranged in the same order as shown in Figure 2a,b and cropped portions are boxed in blue.



Supplementary Figure 5:
Full-size scans of western blot membranes shown in Figure S2. Scans are arranged in the same order as shown in Figure S2c,d and cropped portions are boxed in blue.

Supplementary Table 1:
Identification and quantification of proteins detected in two SILAC-based H2A vs H2Aub nucleosome pulldowns. **Available online:**
http://www.nature.com/nsmb/journal/v21/n6/full/nsmb.2833.html#supplementary-information

Chapter 5

Towards elucidating the stability, dynamics and architecture of the NuRD complex using quantitative interaction proteomics

Susan L. Kloet[1]*, H. Irem Baymaz[1]*, Matthew Makowski[1]*, Vincent Groenewold[2], Pascal Jansen[1], Madeleine Berendsen[1], Hassin Niazi[1], Geert J. Kops[2,3] and Michiel Vermeulen[1-3]#

1. Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen
2. Department of Molecular Cancer Research, UMC Utrecht
3. Cancer GenomiCs Netherlands

* Equal contribution
# Corresponding author

FEBS Journal, May 2015,
Volume 282, Issue 9, pages 1774–1785

Abstract

The Nucleosome Remodeling and Deacetylase (NuRD) complex is an evolutionary conserved chromatin-associated protein complex. Although the subunit composition of the mammalian complex is fairly well characterized, less is known about the stability and dynamics of these interactions. Furthermore, detailed information regarding protein-protein interaction surfaces within the complex is largely still lacking. Here, we show that the NuRD complex interacts with a number of substoichiometric zinc finger-containing proteins. Some of these interactions are salt sensitive (ZNF512B, SALL4), whereas others (ZMYND8) are not. The stoichiometry of the core subunits is not affected by high salt, indicating that the core complex is stabilized by hydrophobic interactions. Interestingly, the RBBP4 and -7 proteins are sensitive to high non-ionic detergent concentrations during affinity purification. In a subunit exchange assay using SILAC labeled nuclear extracts, the RBBP4 and -7 proteins are identified as dynamic core subunits of the NuRD complex, consistent with their proposed role as histone chaperones. Finally, using cross-linking mass spectrometry, we uncover novel features of NuRD molecular architecture that complement our AP-MS/MS data. Altogether, these findings extend our understanding of MBD3/NuRD structure and stability.

Introduction

The Nucleosome Remodeling and Deacetylase (NuRD) complex is a conserved chromatin-associated protein complex that was first purified and characterized from Xenopus laevis and human cells in the late 1990s [1-3]. As the name implies, the NuRD complex contains both histone deacetylase (HDAC) and ATP-dependent chromatin remodeling activity. In the human complex, HDAC activity is catalyzed by HDAC1 and HDAC2. The large CHD3 and CHD4 subunits contain ATPase activity, which is used to move the position of nucleosomes on DNA. In addition to these enzymatic activities, the NuRD complex contains a number of other core subunits: MTA1-3, GATAD2A and –B, MBD2-3, RBBP4 and -7 and CDK2AP-1 [4, 5]. Some of these proteins, such as MBD2 and MBD3, are mutually exclusive within the NuRD complex. For GATAD2A/B, RBBP4 and -7 and MTA1-3, it is currently not known whether these proteins exclusively form heterodimers/trimers or mutually exclusive homodimers/trimers. In addition to the well-described core subunits, a large number of proteins have been reported to interact with the NuRD complex. Examples include SALL4, FOG1 and ZMYND8 [6-8]. These proteins may recruit the NuRD complex to its target sites in the genome.

Due to the presence of HDACs, which are generally associated with gene silencing, the NuRD complex has long been thought of as a transcriptional co-repressor complex. Indeed, in luciferase assays, NuRD subunits repress transcription of reporter constructs [4]. Furthermore, the methyl CpG binding protein MBD2 links MBD2-containing NuRD to methylated CpG islands and induces gene silencing [9]. However, recent genome wide profiling studies have revealed that NuRD complexes, such as MBD3/NuRD, are also found at active enhancers and promoters, indicating that NuRD-mediated regulation of transcription is more diverse than previously thought [10, 11]. Further functional experiments have revealed an important role for the NuRD complex in regulating cell fate decisions [12, 13]. These different biological functions of the NuRD complex may be driven by subtle changes in subunit composition.

Recently, a number of mass spectrometry-based methods were developed which can be used not only to confidently identify protein-protein interactions, but also to obtain information about the stoichiometry, dynamics and the interaction surfaces between subunits [14-16]. Here, we applied these different methods to gain further insight into the structure and dynamics of the MBD3/NuRD complex in HeLa cells. We applied label free quantitative mass spectrometry to investigate the stoichiometry of the MBD3/NuRD complex using different buffer stringencies during affinity purification. A SILAC-based subunit exchange assay was used to show that the RBBP4 and -7 proteins are dynamic core subunits within NuRD. Finally, cross-linking mass spectrometry was used to identify inter-protein contacts within MBD3/NuRD. Altogether, these experiments increase our understanding of the stability, dynamics and architecture of the NuRD complex.

Results

RBBP4 and -7 interaction with NuRD is salt stable but detergent sensitive

To facilitate the purification of MBD3/NuRD, we created a stable cell line with doxycycline-inducible GFP-tagged MBD3. Cells were incubated for 16 hours in the presence or absence of doxycycline. Expression of GFP-MBD3 occurs only after addition of doxycycline (dox) to the medium (**Figure 1A**). Nuclear extracts from this cell line were used for GFP affinity purifications combined with label-free LC-MS/MS analysis. Briefly, triplicate pulldowns were performed with GFP-Trap_A beads and GFP-MBD3-containing nuclear extracts (+Dox). Simultaneously, triplicate pulldowns were performed with GFP beads on control nuclear

extracts (-Dox). Proteins that bind specifically to GFP-MBD3 will be enriched in the GFP pulldowns and appear on the right side of the plot. Proteins that bind nonspecifically to the GFP beads will appear in the background cloud.

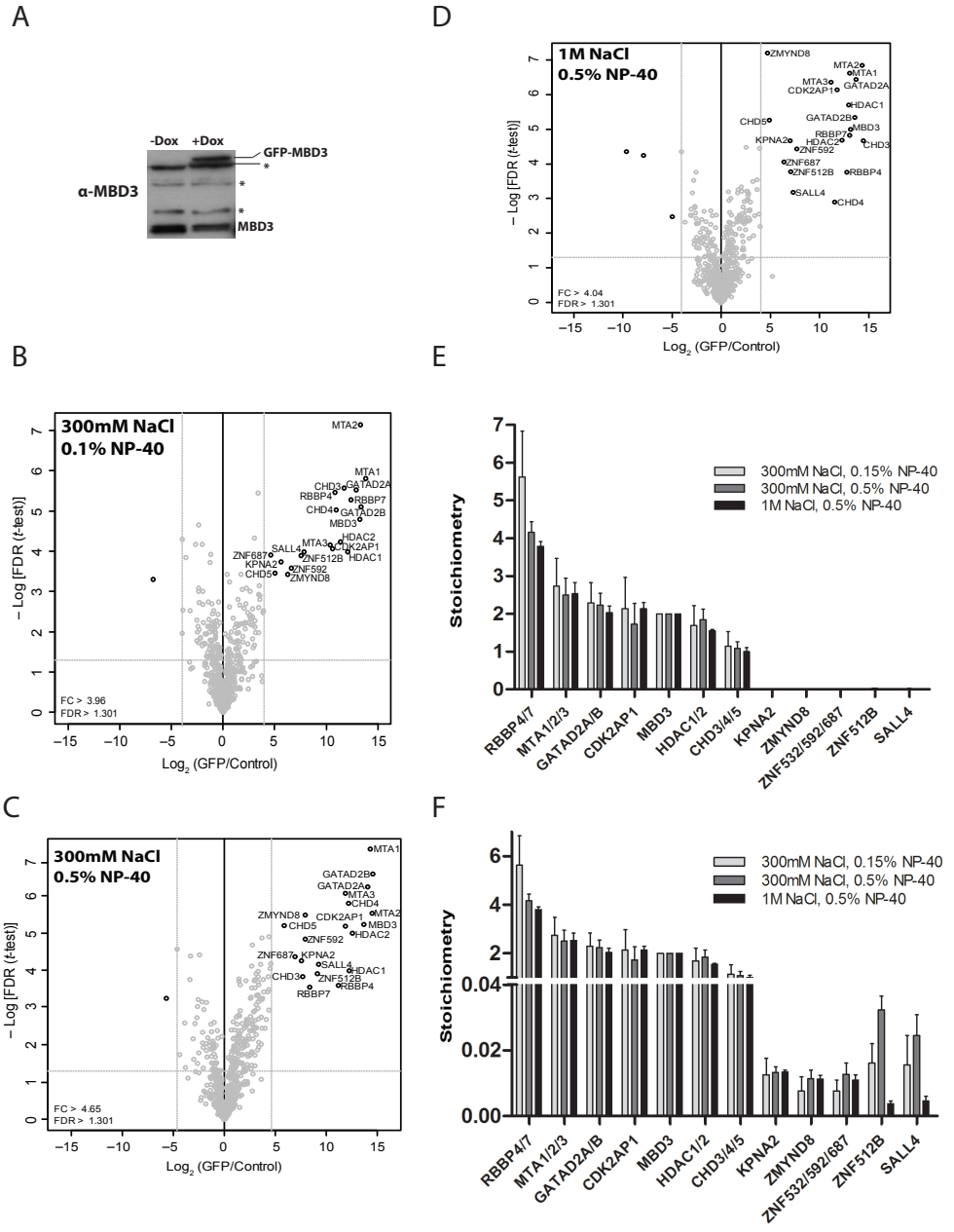


Figure 1: Label-free proteomics reveal RBBP4 and -7 interaction with NuRD is NP-40 sensitive **(A)** An anti-MBD3 Western blot of GFP-MBD3 HeLa cells after 16 hours incubation with Doxycycline (+Dox) or without (-Dox). Nonspecific bands are marked with an *. **(B-D)** Volcano plots from label-

free GFP pulldowns of GFP-MBD3 HeLa nuclear extracts with varying salt and NP-40 concentration. Statistically enriched proteins in the GFP-MBD3 pull-down are identified by a permutation-based FDR-corrected t-test. The LFQ intensity of the GFP pull-down over the control (fold change (FC), X axis) is plotted against the –log 10 transformed p-value of the t-test (Y axis). The proteins in the upper right corner represent the bait and its interactors. **(E)** Stoichiometry of NuRD core subunits. The iBAQ value of each protein group is divided by the iBAQ value of the GFP-MBD3 bait, which was set to 2. **(F)** Stoichiometry of NuRD complex interactors relative to the GFP-MBD3 bait using the same data as in (E) but with a differently scaled axis.

To determine the stability of the NuRD complex, several NP-40 and NaCl concentrations were used during the affinity purifications (**Figure 1B-D**). In each of these pull-downs, known NuRD core subunits were identified. In addition, several previously described NuRD interactors were also detected, including ZMYND8, ZNF592, and SALL4 [8, 14, 17]. Interestingly, we did not pulldown LSD1 in any of our MBD3-GFP purifications. LSD1 has previously been reported as a putative subunit of the NuRD complex [18] but our data on MBD3/NuRD does not agree with this observation. Next, the intensity-based absolute quantification (iBAQ) values were compared between all NuRD subunits and interactors. The iBAQ algorithm normalizes the total mass spec intensity for each protein according to the number of theoretically observable peptides [19]. This allows estimation of the relative abundance of large and small proteins detected in affinity purifications. Since the NuRD complex contains many paralogs that overlap at the peptide level, iBAQ values were summed for all paralogs (**Table S1**). Comparison of the iBAQ values relative to the GFP-MBD3 bait revealed that most core subunits of NuRD remain tightly bound to each other and to MBD3 despite the presence of high salt (1M NaCl) and NP-40 concentration (0.5%) in the wash steps (**Figure 1E**). Surprisingly, the core RBBP4 and -7 subunits were very sensitive to NP-40. Increasing the NP-40 concentration from 0.15% to 0.5% led to a decrease in the amount of RBBP4 and -7 associated with the complex (stoichiometry of 5.5 reduced to 4).

In addition, the association of several zinc finger domain-containing proteins with NuRD was very sensitive to high salt concentrations. The association of ZNF512B decreased nearly 8-fold when the salt concentration in the wash was increased from 300mM to 1M NaCl (**Figure 1F**). A reduction in SALL4 binding was also observed with increased salt. In contrast, other substoichiometric interactors such as ZMYND8 and KPNA2 remained tightly associated with the core complex under increased salt and NP-40 concentration. Together, these results suggest that the core NuRD complex is very stable when challenged by high salt concentrations, indicating that interactions within the NuRD complex are at least partially driven by hydrophobic interactions. The partial dissociation of RBBP4 and -7 observed at high NP-40 concentration may suggest that these proteins are less tightly associated to the rest of the core complex. Finally, although other detected interactors have a low iBAQ value relative to the core complex (1-2% relative to the core subunits), some of these interactions are very stable under high salt and NP-40 concentrations, suggesting a functional role for these proteins in relation to the NuRD complex.

RBBP4 and RBBP7 partially dissociate from the NuRD complex in a subunit exchange assay

To further study the potential dynamics of NuRD subunit interactions, a SILAC-based subunit exchange assay was used (**Figure 2A**) [15]. HeLa cells expressing GFP-CDK2AP-1 were labeled in culture with heavy amino acids (Forward) or light amino acids (Reverse). Similarly, WT HeLa cells were also labeled with light or heavy amino acids. GFP pulldowns were

performed immediately after mixing the nuclear extracts (T_0) or after overnight incubation (T_{ON}). Proteins that are more dynamically associated with NuRD will dissociate from the complex during the overnight incubation step and may be replaced by proteins from the other, differentially labeled extract. This eventually results in a decrease of detected SILAC ratios, where dynamic core subunits and/or interactors will move towards the background cloud in the scatter plot.

At T_0, all NuRD core subunits are significantly enriched according to boxplot statistics (**Figure 2B**). However, after overnight incubation, RBBP4 and -7 clearly separate from the other NuRD subunits and migrate towards the background cloud (**Figure 2C**). To more directly compare the 2 scatter plots, the difference in forward and reverse ratios between the plots was visualized in a graph (**Figure 2D**). A protein with no change in ratios between experiments would have a value of 0. This graph clearly shows that RBBP4 and -7 are the most dynamic NuRD core subunits. These observations are in agreement with recent structural studies, which suggest that MTA and histone H4 compete for RBBP binding [20]. Furthermore, RBBP4 and -7 are part of many different protein complexes other than NuRD (Sin3 complex, PRC2), which also could explain their observed dynamic behavior [21, 22].

Cross-linking mass spectrometry reveals novel architectural features of the human NuRD complex

To further characterize the structural interactions between NuRD complex core subunits, we conducted cross-linking mass spectrometry to identify inter-protein interaction surfaces (**Figure 3A**). NuRD complex was isolated from HeLa nuclear extracts using tandem affinity purification with a GFP-His-MBD3 bait and subsequently cross-linked with BS3 for LC-MS/MS analysis. Crosslinking was monitored by silver staining to ensure complete cross-linking of all NuRD subunits (**Figure 3B**). High confidence cross-linked peptides (FDR <0.05) were then identified by pLink using default settings and searching against a database of NuRD core components (**Figure 3C and 3D**) [23]. In total, 336 spectra matching with high confidence to 16 non-ambiguous inter-protein cross-links, 87 non-ambiguous intra-protein cross-links, and 46 ambiguous cross-links encompassing a variety of homo- or heteromers or intra-protein cross-links were identified (**Table S2**).

Structural validation was confounded by lack of overlap between the few existing NuRD crystal structures and our identified cross-links. For example, although 5 HDAC1 and 11 HDAC2 intra-protein cross-links as well as 4 ambiguous HDAC1/2 cross-links were identified, all of them were located in the undetermined region of the known HDAC1 crystal structure (PDB 4BKX). However, seven cross-links mapping to the CHD4 chromodomain-ATPase region (residues 500-1298) were identified, which has a yeast CHD1 ortholog with known structure (PDB 3MWY). To our knowledge, this represents the highest number of mappable cross-links from our data set to any single structure. Structural validation of our CHD4 chromodomain-ATPase cross-links was performed using a homology modelling approach. The human CHD4 sequence was aligned to the yeast CHD1 template using HHPred and homology models were constructed using Modeller (Human CHD4 Uniprot: Q14839) [24-27]. In ten overlapping homology models, well-ordered regions from the crystal structure yield consistently low divergence models, while short alignment gaps or structurally undetermined loops are clearly disordered (**Figure 3E**). All homology models were highly similar except for very small, flexible regions, so one model was chosen as representative and cross-links were validated by calculating solvent accessible surface (SAS) distance using Xwalk [28]. All seven cross-links showed an SAS under 30 Å, as expected from the maximum length of the BS3

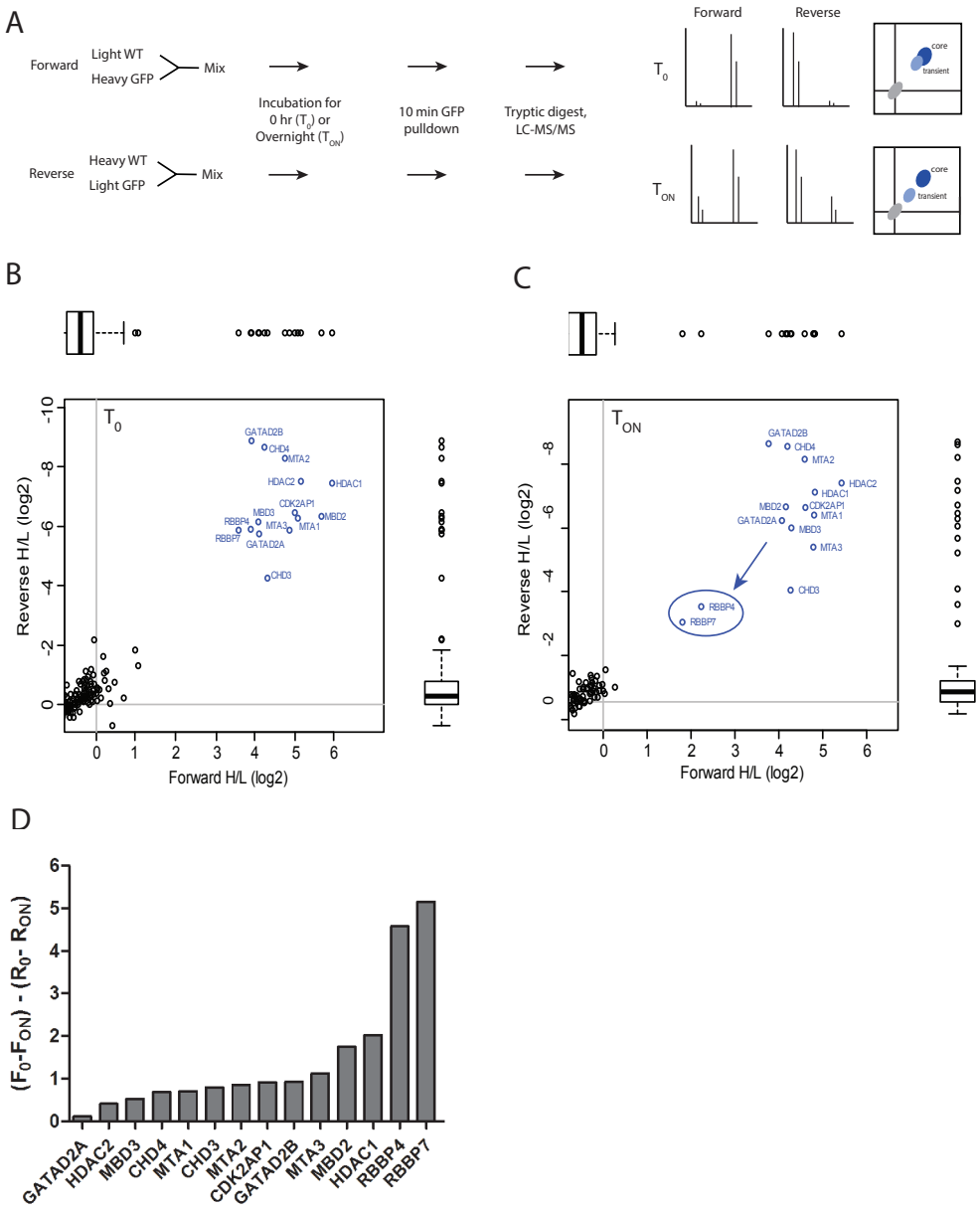


Figure 2: RBBP4 and -7 are the most dynamic NuRD core subunits
(A) Schematic representation of the SILAC subunit exchange assay. **(B,C)** Scatter plots of GFP-CDK2AP-1 forward and reverse pulldowns immediately after extract mixing (**B,** T_0) and after overnight incubation (**C,** T_{ON}). The SILAC ratio for each protein in the forward pulldown is plotted against the ratio in the reverse pulldown. Boxplot statistics were used to determine statistically significant GFP-CDK2AP-1 interactors. The arrow and circle in (C) illustrate the movement of RBBP4 and -7 towards the background cloud, indicating that these proteins exchange between the heavy and light extract **(D)** Bar graph showing the difference in forward and reverse SILAC ratios between T_0 and T_{ON}.

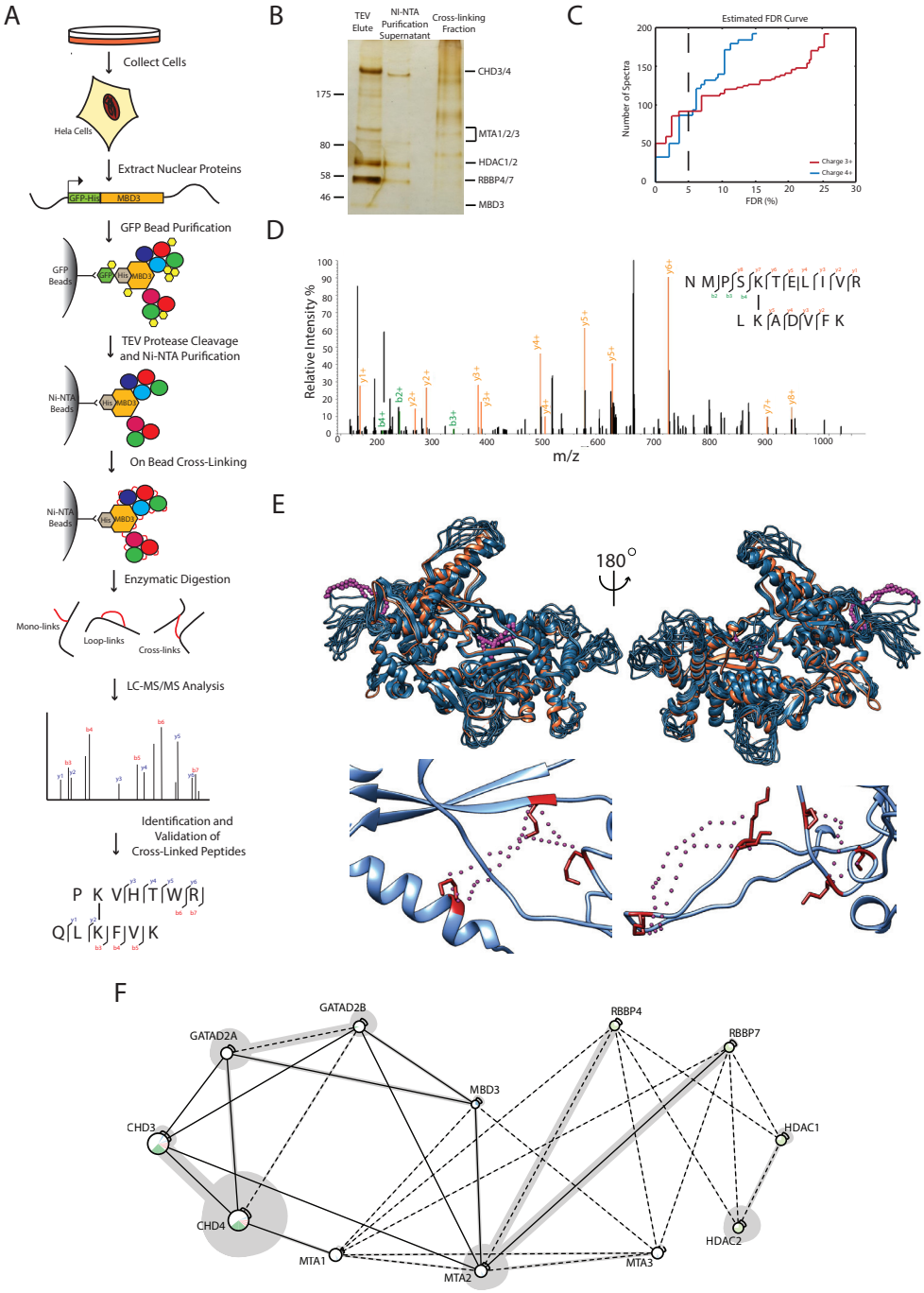


Figure 3: Cross-linking mass spectrometry offers insight into NuRD molecular architecture
(A) MBD3-NuRD was obtained by tandem affinity purification (GFP affinity enrichment, TEV cleavage and Ni-NTA enrichment) and cross-linked on-bead with BS3. Proteolytic digestion and analysis by AP-MS/MS was followed by identification and validation of high-confidence cross-linked

peptides (FDR < 0.05) using pLink. **(B)** NuRD complex purification and cross-linking was monitored by silver staining to ensure complete cross-linking of all complex components. Note that after the cross-linking reaction, NuRD subunits migrate at an apparent higher molecular weight in the SDS PAGE gel. **(C)** High confidence spectra were matched to putative cross-links using a target-decoy approach in pLink. Representative FDR curves are shown for cross-links of charge state +3 and +4. **(D)** An annotated high-scoring spectrum from a CHD4 intra-protein cross-link, K959-K969. **(E)** Structural validation was performed using a human CHD4 chromodomain-ATPase homology model aligned to a yeast CHD1 ortholog template structure. Yeast CHD1 is colored in red with ten overlapping CHD4 homology models in blue. Validated cross-links are indicated in purple. Below the homology models are close-up views of seven CHD4 cross-links in the Chromodomain-ATPase region using a representative CHD4 homology model; the lower left image shows crosslinks between K959-K969, K959-K1189, and K969-K1189 in a well ordered region, while the lower right image shows crosslinks between K690-K693 and K693-K696 in a gapped region and cross-links between K833-K838 and K833-839 in a structurally undetermined region. **(F)** Cross-link map for human NuRD subunits. Subunits connected by at least one unambiguous cross-link are indicated by a solid line, while subunits connected only by ambiguous cross-links are indicated by a dashed line. In our data, all ambiguous cross-links are ambiguous for peptides from a paralogous subunit. The relative number of inter- or intra-protein cross-links is designated by the area of gray shading surrounding that link.

cross-linker. Three cross-links mapped to a well-ordered region, and four mapped to a disordered region that likely possesses a highly flexible structural conformation. Therefore, all seven CHD4 chromodomain-ATPase cross-links seem to be structurally plausible, and we conclude that our cross-linking data is of good quality.

Few NuRD subcomplexes have resolved structures, most notably the MBD2-GATAD2A coiled-coil subcomplex, the RBBP4-MTA1 subcomplex, and the HDAC1-MTA1 subcomplex [20, 29, 30]. None of these studies describe comprehensive protein structures for both binding partners, so the complete interaction interfaces remain unknown. As such, our inter-protein cross-linking data reveal novel features of the NuRD complex architecture and extend our understanding of NuRD structure-function relationships (**Figure 3F**). For example, robust contacts were observed between MTA1-3 and RBBP4 and -7 subunits. Our data cannot discern whether these RBBP4 and -7 cross-link sites indicate binding of a single MTA paralog or a multimer, nor can we distinguish between potentially dynamic or paralog-specific interactions at each site. However, mapping cross-linked residues onto the known RBBP4 structure shows K25 and K307 fall on either side of the known MTA1 interaction interface (PDB 4PC0 [20], **Figure 4A**). As RBBP4 K25 maps relatively close to the known MTA1 binding pocket while K307 sits across a large, adjacent hydrophobic region, these cross-links could outline the opposite ends of the MTA1-3 binding platform (**Figure 4B**). Intriguingly, this would explain our observation that MTA1-3 binding to RBBP4 and -7 is stabilized by hydrophobic interactions under high salt conditions. Similarly, although the MBD3 structure is not known, we identified a short MBD3 region (residues 92-124) that contained 5 cross-links with GATAD2A/B, two of which represented unique cross-links for GATAD2A, one which was unique for GATAD2B and two of which were ambiguous for the GATAD2A/B peptide (**Figure 4C**). The MBD2 and -3 sequences were aligned in EBI's Clustal Omega in order to map any MBD3-GATAD2A crosslinks to the solved MBD2-GATAD2A coiled-coil structure (PDB 2L2L, [30, 31]). Only two cross-linked lysine residues could be mapped to a representative conformer GATAD2A structure (K163 and K178), but no crosslinks could be mapped to the MBD2 structure (**Figure 4D**). Furthermore, one GATAD2A cross-link site, K178, was located

at the tail end of the conformer, which exhibited a highly flexible conformation and thus offers little structural information. It is possible we observe low cross-link coverage within the MDB2-GATAD2A coiled-coil domain because, as the authors of that study note, this interaction hides at least 1,337 Å² of solvent accessible area, potentially hindering BS3 accessibility. Thus, our cross-links may outline an interaction interface rather than reveal a binding footprint. However, in agreement with previous work, our data depict a single binding interface between MBD3 and GATAD2A/B paralogs. Overall, this cross-linking mass spectrometry data greatly improves our understanding of the NuRD complex subunit architecture and will help direct future structural studies.

Discussion

Here we have used a variety of mass spectrometry-based methods to increase our understanding of the dynamics, stability and architecture of the MBD3/NuRD complex. GFP-based affinity purifications combined with a label-free method to estimate protein complex stoichiometry revealed that the RBBP4 and -7 core subunits of the NuRD complex are sensitive to high detergent concentration. Increasing the NP-40 concentration from 0.1 to 0.5% in the wash buffer resulted in a reduction of RBBP4 and -7 co-purified with MBD3. This result illustrates an important limitation of AP-MS experiments, namely that part of the interactions that occur in vivo may be lost during affinity purification. Conversely, interactions can be formed in the extract that may not occur in vivo. To overcome these limitations, cross-linking prior to cell lysis can be used as an alternative approach. However, this approach is technically very challenging due to experimental and computational issues.

Except for MBD3 and HDAC1/2, the stoichiometry values we obtained in our GFP-MBD3 purifications are generally in good agreement with our previously published NuRD stoichiometry results [14] [32]. Previously, we made use of a BAC GFP-MBD3 HeLa cell line, which expresses MBD3 with a GFP tag at near endogenous levels. In the current study we used a doxycycline-inducible GFP-MBD3 cell line. This cell line may be expressing MBD3 at slightly higher levels, which may result in slightly different MBD3 and HDAC1/2 stoichiometry values. It is important to note that the final stoichiometry values we obtained represent an average value based on a large number of potentially heterogeneous complexes. Based on our data, we cannot be sure whether paralogous proteins such as MTA1-3 or GATAD2A-B are present within the same complex or whether these proteins are in fact mutually exclusive as we have previously shown for MBD2 and-3 [33]. It is highly likely that a large number of distinct NuRD complexes exist. However, our cross-linking data cannot distinguish unambiguously between homomeric and heteromeric interactions for HDAC1/2, MTA1-3 and GATAD2A-B. Interestingly, we were able to detect unambiguous interprotein crosslinks for CHD3 and CHD4, indicating that these proteins may form a heterodimer. Nevertheless, NuRD complex heterogeneity most likely creates a significant challenge towards high-resolution determination of the NuRD molecular architecture.

In addition to the core NuRD subunits, a number of substoichiometric interactors were identified in the MBD3-GFP purifications. Interestingly, these interactors showed differential sensitivity to salt and detergent. ZMYND8 and ZNF532/592/687 co-purified with GFP-MBD3 even when challenged with 1 M NaCl and 0.5% NP-40. The stability of this interaction indicates that it may be functionally relevant. Interestingly, these proteins have previously been described to interact with each other as a central hub in a large transcriptional network [34]. Given the presence of a large number of putative DNA binding zinc fingers in the ZMYND8/ZNF module, we hypothesize that this module may recruit the NuRD complex to a

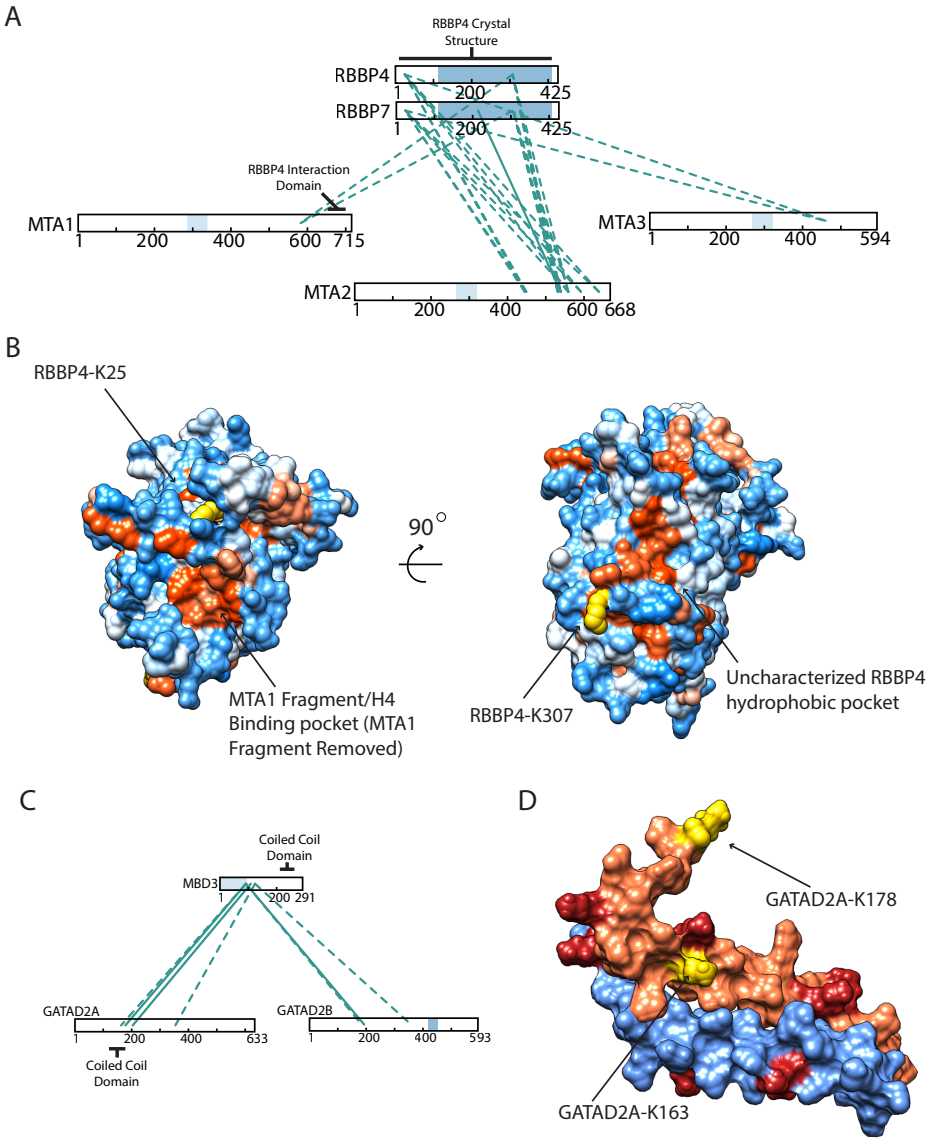


Figure 4: Structural analysis of cross-linked residues reveals novel features of NuRD subcomplex interactions **(A)** Twelve identified cross-links between RBBP4 and -7 and MTA1-3 map to two residues on RBBP4, K25 and K307 (K24 and K306 on RBBP7). **(B)** Residues from (A) are mapped onto the known crystal structure for RBBP4. These residues surround the binding site of the MTA1 peptide. RBBP4-K25 sits near the MTA1 binding site, while RBBP4-K307 lies across a large hydrophobic region representing a potential MTA1 binding platform. Hydrophobic regions are indicated in red, while hydrophilic regions are indicated in blue. **(C)** Five cross-links identified between MBD3 and GATAD2A/B map to MBD3 residues K92, K109, and K124. **(D)** Coiled-coil conformer model of GATAD2A and MBD2. The GATAD2A coil is colored in light red and the MBD2 coil is colored in blue. All lysine resides are colored dark red, while the cross-linked residues are indicated and colored in yellow.

subset of its target genes. The same holds true for other detected substoichiometric interactors. Further work is needed to determine the functional significance of the detected ZMYND8/ZNF – NuRD interaction.

Materials and Methods

Cloning

A tandem affinity purification (TAP) tag was cloned into pcDNA5 FRT/TO. This tag consists of eGFP followed by 2 TEV protease cleavage sites and a 6xHis tag. The TEV sites and His tag were added to eGFP via PCR. MBD3 was amplified using BamHI and XhoI and was cloned into the MCS of pcDNA5 FRT/TO.

Cell Culture and SILAC Labeling

HeLa FRT cells were transfected with pcDNA5 FRT/TO TAP-MBD3 using Lipofectamine LTX Plus (Invitrogen). Cells underwent selection for 10 days with hygromycin, and single colonies were picked, expanded and screened for expression of TAP-MBD3. TAP-MBD3 expression was induced with 1ug/mL doxycycline treatment for 16 hours. SILAC labelling of HeLa cells was done as described [35].

Nuclear extracts

Nuclear extracts were prepared essentially according to Dignam et al. [36]. Cells were harvested with trypsin, washed with PBS 2x and centrifuged at 400xg for 5 min at 4˚C. Cells were swelled for 10 min at 4˚C in 5 volumes of Buffer A [10mM Hepes KOH pH 7.9, 1.5mM $MgCl_2$, 10mM KCl] and then pelleted at 400xg for 5 min at 4˚C. Cells were resuspended in 2 volumes Buffer A plus protease inhibitors and 0.15% NP-40 and transferred to a dounce homogenizer. After 30-40 strokes with a Type B pestle, the lysates were spun at 3900 rpm for 15 min at 4˚C. The nuclear pellet was washed 1x with PBS and spun at 3900 rpm for 5 min at 4˚C. The nuclear pellet was resuspended in Buffer C [420mM NaCl, 20mM Hepes KOH pH 7.9, 20% v/v glycerol, 2 mM MgCl2, 0.2 mM EDTA] with 0.1% NP-40, protease inhibitors and 0.5mM DTT. The suspension was incubated with rotation for one hour at 4˚C, then spun at 14000rpm for 15 min at 4˚C. The supernatant was aliquoted and stored at -80°C until further usage.

Label free pulldowns and cross-linking mass spectrometry

Label-free GFP pulldowns, LC-MS/MS and data analysis were done essentially as described [14]. For cross-linking mass spectrometry, 8 mL of HeLa nuclear extract (5 mg/mL plus 50 microgram/mL ethidium bromide) containing doxycycline induced TAP-MBD3 was incubated with 400 microliters of GFP-trap beads (Chromotek) for 2 hours at 4°C in a rotation wheel. Beads were then washed three times with 10 mL wash buffer (300 mM NaCl, 0.5% NP-40, 20 mM Hepes KOH pH 7.9, 0.5 mM DTT, 2 mM MgCl2, 0.2 mM EDTA and complete protease inhibitors) and twice with 5 mL of TEV cleavage buffer (150 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.1% NP-40 and 1 mM DTT). Beads were then resuspended in 0.5 mL TEV cleavage buffer containing 100 units GST-TEV protease (US Biological) and incubated overnight at 4°C in a rotation wheel. The supernatant was collected and GST-TEV was removed from the eluate by adding 10ml GST beads and incubating for 15 minutes 4 °C in a rotation wheel. The eluate was then incubated with 20 ml Ni-NTA beads (Qiagen) for 45 minutes at 4 °C in a rotation wheel. Beads were washed once with 1 mL TEV cleavage buffer and three times with 1 mL cross-linking buffer (25 mM Hepes KOH pH 7.9, 300 mM KCl, 5% glycerol, 0.05% Tween-20).

Beads were then resuspended in 50 ml cross-linking buffer containing 250 mM BS3 cross-linker (Sigma) and incubated for 30 min at 37°C in a thermoshaker (1100 rpm). The cross-linking reaction was quenched by adding 7 ml 0.5 M ammonium bicarbonate and incubating for 5 minutes at RT. Finally, beads were washed twice with 0.5 mL PBS after which bound cross-linked proteins were digested overnight using on-bead LysC/trypsin digestion. Tryptic peptides were desalted and purified using stagetips prior to LC-MS/MS analysis.

Analysis of cross-linking mass spectrometry data

Thermo RAW MS files were converted to .mgf format using the MSConvert tool from ProteoWizard and analyzed using pLink with default settings for HCD using BS3 as a cross-linking agent [23, 37]. Cross-linked peptides matched to spectra with an FDR <0.05 were considered high confidence and were used for further analysis. Spectra were visually analyzed and annotated using the pLabel software [38]. Template structures for homology modelling were searched using HHPred, and homology models were built using the Modeller interface in UCSF Chimera [24, 26, 39]. Yeast CHD1 chromodomain and ATPase residues 38-800 (PDB 3MWY) were used as a template for residues 500-1298 from human CHD4 (Uniprot Q14839) for homology modelling [27]. RMSD calculations were performed in UCSF Chimera. Solvent accessible surface distances were calculated for a representative CHD4 model using Xwalk. xiNET was used to visualize high-confidence intra- and inter-protein cross-links in network form (Rappsilber Lab--crosslinkviewer.org). All protein structure visualizations were produced using UCSF Chimera.

Acknowledgements

This work is supported by the EU FP7 framework program 4DCellFate. We thank Ina Poser and Anthony Hyman for providing the BAC GFP-CDK2AP-1 cell line. We thank Nina Vermeulen-Hubner and Henk Stunnenberg for providing access to the Q Exactive mass spectrometer.

References

1. Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. (1998) The dermatomyositis- specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities, *Cell.* **95**, 279-89.
2. Wade, P. A., Jones, P. L., Vermaak, D. & Wolffe, A. P. (1998) A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase, *Curr Biol.* **8**, 843-6.
3. Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E. & Schreiber, S. L. (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex, *Nature.* **395**, 917-21.
4. Spruijt, C. G., Bartels, S. J., Brinkman, A. B., Tjeertes, J. V., Poser, I., Stunnenberg, H. G. & Vermeulen, M. CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex, *Molecular bioSystems.* **6**, 1700-6.
5. Allen, H. F., Wade, P. A. & Kutateladze, T. G. The NuRD architecture, *Cell Mol Life Sci.* **70**, 3513-24.
6. Miccio, A., Wang, Y., Hong, W., Gregory, G. D., Wang, H., Yu, X., Choi, J. K., Shelat, S., Tong, W., Poncz, M. & Blobel, G. A. NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development, *The EMBO journal.* **29**, 442-56.
7. Lauberth, S. M. & Rauchman, M. (2006) A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex, *The Journal of biological chemistry.* **281**, 23922-31.
8. Eberl, H. C., Spruijt, C. G., Kelstrup, C. D., Vermeulen, M. & Mann, M. A map of general and specialized

chromatin readers in mouse tissues generated by label-free interaction proteomics, *Molecular cell.* **49**, 368-78.
9. Baubec, T., Ivanek, R., Lienert, F. & Schubeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family, *Cell.* **153**, 480-92.
10. Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., Cowley, S. M. & Young, R. A. (2012) Enhancer decommissioning by LSD1 during embryonic stem cell differentiation, *Nature.* **482**, 221-5.
11. Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., Costello, I., Humphreys, P., Mansfield, W., Nakagawa, K., Strouboulis, J., Behrens, A., Bertone, P. & Hendrich, B. NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment, *Cell stem cell.* **10**, 583-94.
12. Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A. A., Caspi, I., Krupalnik, V., Zerbib, M., Maza, I., Mor, N., Baran, D., Weinberger, L., Jaitin, D. A., Lara-Astiaso, D., Blecher-Gonen, R., Shipony, Z., Mukamel, Z., Hagai, T., Gilad, S., Amann-Zalcenstein, D., Tanay, A., Amit, I., Novershtern, N. & Hanna, J. H. Deterministic direct reprogramming of somatic cells to pluripotency, *Nature.* **502**, 65-70.
13. Dos Santos, R. L., Tosti, L., Radzisheuskaya, A., Caballero, I. M., Kaji, K., Hendrich, B. & Silva, J. C. MBD3/NuRD Facilitates Induction of Pluripotency in a Context-Dependent Manner, *Cell stem cell.* **15**, 102-10.
14. Smits, A. H., Jansen, P. W., Poser, I., Hyman, A. A. & Vermeulen, M. (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics, *Nucleic acids research.* **41**, e28.
15. Mousson, F., Kolkman, A., Pijnappel, W. W., Timmers, H. T. & Heck, A. J. (2008) Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes, *Mol Cell Proteomics.* **7**, 845-52.
16. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure, *Current opinion in structural biology.* **23**, 252-60.
17. van den Berg, D. L., Snoek, T., Mullin, N. P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I. & Poot, R. A. (2010) An Oct4-centered protein interaction network in embryonic stem cells, *Cell stem cell.* **6**, 369-81.
18. Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., Li, R., Li, Y., Zhang, Y., Li, Q., Yi, X. & Shang, Y. (2009) LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer, *Cell.* **138**, 660-72.
19. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. & Selbach, M. (2011) Global quantification of mammalian gene expression control, *Nature.* **473**, 337-42.
20. 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., Wang, H., Shepherd, N. E., Stokes, P. H., Blobel, G. A., Vermeulen, M., Glover, D. M., Mackay, J. P. & Laue, E. D. (2014) Insight into the architecture of the NuRD complex: Structure of the RbAp48-MTA1 sub-complex, *The Journal of biological chemistry*.
21. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein, *Genes & development.* **16**, 2893-905.
22. Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. (1997) Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex, *Cell.* **89**, 357-64.
23. Yang, B., Wu, Y. J., Zhu, M., Fan, S. B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y. X., Chen, H. F., Luo, S. K., Ding, Y. H., Wang, L. H., Hao, Z., Xiu, L. Y., Chen, S., Ye, K., He, S. M. & Dong, M. Q. (2012) Identification of cross-linked peptides from complex samples, *Nature methods.* **9**, 904-6.
24. Soding, J., Biegert, A. & Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction, *Nucleic acids research.* **33**, W244-8.
25. Sali, A. & Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *Journal of molecular biology.* **234**, 779-815.
26. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper,

U. & Sali, A. (2007) Comparative protein structure modeling using MODELLER, *Current protocols in protein science / editorial board, John E Coligan [et al].* **Chapter 2**, Unit 2 9.
27. Hauk, G., McKnight, J. N., Nodelman, I. M. & Bowman, G. D. (2010) The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor, *Molecular cell.* **39**, 711-23.
28. Kahraman, A., Malmstrom, L. & Aebersold, R. (2011) Xwalk: computing and visualizing distances in cross-linking experiments, *Bioinformatics.* **27**, 2163-4.
29. Millard, C. J., Watson, P. J., Celardo, I., Gordiyenko, Y., Cowley, S. M., Robinson, C. V., Fairall, L. & Schwabe, J. W. (2013) Class I HDACs share a common mechanism of regulation by inositol phosphates, *Molecular cell.* **51**, 57-67.
30. 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. & 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, *Proceedings of the National Academy of Sciences of the United States of America.* **108**, 7487-92.
31. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI, *Nucleic acids research.* **38**, W695-9.
32. Smits, A. H. & Vermeulen, M. (2014) Exploring Chromatin Readers Using High-Accuracy Quantitative Mass Spectrometry-Based Proteomics in *Systems Analysis of Chromatin-Related Protein Complexes in Cancer* pp. 133-148, Springer.
33. Le Guezennec, X., Vermeulen, M., Brinkman, A. B., Hoeijmakers, W. A., Cohen, A., Lasonder, E. & Stunnenberg, H. G. (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties, *Molecular and cellular biology.* **26**, 843-51.
34. Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y., Le, N. T., Chan, D. W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., Kim, B. J., Li, C., Chen, R., Li, W., Wang, Y., O'Malley, B. W. & Qin, J. (2011) Analysis of the human endogenous coregulator complexome, *Cell.* **145**, 787-99.
35. Baymaz, H. I., Fournier, A., Laget, S., Ji, Z., Jansen, P. W., Smits, A. H., Ferry, L., Mensinga, A., Poser, I., Sharrocks, A., Defossez, P. A. & Vermeulen, M. (2014) MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain, *Proteomics*.
36. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei, *Nucleic acids research.* **11**, 1475-89.
37. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. (2008) ProteoWizard: open source software for rapid proteomics tools development, *Bioinformatics.* **24**, 2534-6.
38. Wang, L. H., Li, D. Q., Fu, Y., Wang, H. P., Zhang, J. F., Yuan, Z. F., Sun, R. X., Zeng, R., He, S. M. & Gao, W. (2007) pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry, *Rapid communications in mass spectrometry : RCM.* **21**, 2985-91.
39. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry.* **25**, 1605-12.

Supporting Information

Supplementary Table 1: iBAQ values used to calculate NuRD subunit stoichiometry.
For a more comprehensive version including the iBAQ values of proteins from each affinity purification, see: http://onlinelibrary.wiley.com/doi/10.1111/febs.12972/suppinfo

Table S1 iBAQ values used to calculate NuRD subunit stoichiometry									
	300mM NaCl, 0.1% NP-40								
Names	**GFP1-WT**	**GFP2-WT**	**GFP3-WT**	**GFP1-WT scale**	**GFP2-WT scale**	**GFP3-WT scale**	**GFP-WT scale mean**	**GFP-WT scale SD**
GATAD2A/B	2.1E+09	3.52E+09	1.55E+09	1.940007242	2.908406137	2.019299419	2.289237599	0.537679338
CHD3/4/5	9.6E+08	1.92E+09	7.26E+08	0.899050093	1.587777372	0.947921924	1.144916463	0.384306457
CDK2AP1	2E+09	3.7E+09	1.1E+09	1.89445614	3.059694458	1.443072702	2.132407767	0.834165576
MBD2/3	2.1E+09	2.42E+09	1.53E+09	2	2	2	2	0
RBBP4/7	5E+09	8.47E+09	3.96E+09	4.708025513	6.995594096	5.174447536	5.626022382	1.208793773
MTA1/2/3	2.5E+09	4.33E+09	1.75E+09	2.348935663	3.580974464	2.283807499	2.737905875	0.730844651
HDAC1/2	1.3E+09	2.74E+09	1.2E+09	1.236212721	2.263030435	1.572999341	1.690747499	0.523437796
ZMYND8	8118274	14479265	2684736	0.007596046	0.011965786	0.003507396	0.007689743	0.004229974
ZNF532/592/687	8294565	13275880	3249308	0.007760996	0.010971299	0.004244964	0.007659087	0.003364325
ZNF512B	2.1E+07	23016670	7155470	0.020030569	0.01902117	0.009348057	0.016133265	0.005897797
SALL4	1.3E+07	31092000	6440500	0.012538947	0.02569469	0.008414005	0.015549214	0.009025065
KPNA2	1.2E+07	21928563	6222763	0.011417603	0.018121949	0.008129549	0.012556367	0.005092603
	300mM NaCl, 0.5% NP-40								
Names	**GFP1-WT**	**GFP2-WT**	**GFP3-WT**	**GFP1-WT scale**	**GFP2-WT scale**	**GFP3-WT scale**	**GFP-WT scale mean**	**GFP-WT scale SD**
GATAD2A/B	4E+09	6.69E+09	7.74E+09	1.990957591	2.121174199	2.58349256	2.231874783	0.311392657
CHD3/4/5	1.9E+09	3.27E+09	3.84E+09	0.927235269	1.035990831	1.281486859	1.081570986	0.181470962
CDK2AP1	2.8E+09	4.51E+09	7.06E+09	1.383627518	1.428693327	2.358112037	1.723477628	0.550071229
MBD2/3	4E+09	6.31E+09	5.99E+09	2	2	2	2	0
RBBP4/7	8.2E+09	1.23E+10	1.34E+10	4.087539177	3.913107838	4.463020766	4.154555927	0.281015107
MTA1/2/3	4.3E+09	7.4E+09	9E+09	2.155061415	2.34346786	3.004599247	2.501042841	0.446151362
HDAC1/2	3.1E+09	5.77E+09	6.38E+09	1.564817738	1.829117274	2.13108174	1.84167225	0.283340697
ZMYND8	1.7E+07	37610750	40857220	0.008512676	0.011918427	0.013641125	0.011357409	0.002609847
ZNF532/592/687	1.9E+07	39719470	48317400	0.009402177	0.012586657	0.016131878	0.012706904	0.003366462
ZNF512B	5.5E+07	1.1E+08	1.03E+08	0.027574897	0.035000372	0.034472393	0.032349221	0.004143105
SALL4	3.5E+07	83476000	88536000	0.017435439	0.026452613	0.029559785	0.024482612	0.006297667
KPNA2	2.3E+07	43671000	44065000	0.011399941	0.013838852	0.014712116	0.01331697	0.001716653

1M NaCl, 0.5% NP-40								
Names	**GFP1-WT**	**GFP2-WT**	**GFP3-WT**	**GFP1-WT scale**	**GFP2-WT scale**	**GFP3-WT scale**	**GFP-WT scale mean**	**GFP-WT scale SD**
GATAD2A/B	5.4E+09	5.8E+09	4.93E+09	2.149275717	2.110534944	1.829781911	2.029864191	0.174355674
CHD3/4/5	2.7E+09	2.77E+09	2.42E+09	1.10124518	1.008331544	0.900067805	1.003214843	0.100686243
CDK2AP1	5.4E+09	5.39E+09	6.13E+09	2.167015925	1.963172618	2.276771099	2.135653214	0.159134272
MBD2/3	5E+09	5.49E+09	5.38E+09	2	2	2	2	0
RBBP4/7	9.2E+09	1.08E+10	1.02E+10	3.679963366	3.921090779	3.772575501	3.791209882	0.121638964
MTA1/2/3	7.1E+09	6.79E+09	6.08E+09	2.853844559	2.471821326	2.257678416	2.527781434	0.301996967
HDAC1/2	4E+09	4.24E+09	4.04E+09	1.593791809	1.542261744	1.502207455	1.546087003	0.04591185
ZMYND8	3E+07	32573180	26573690	0.012016847	0.011858465	0.009871143	0.011248818	0.001195727
ZNF532/592/687	3E+07	31916430	24684520	0.012047593	0.011619372	0.009169386	0.01094545	0.001552947
ZNF512B	9026083	12742183	8492883	0.00362102	0.00463887	0.003154792	0.003804894	0.000758933
SALL4	1.4E+07	15151000	8340600	0.005420649	0.005515814	0.003098224	0.004678229	0.001369151
KPNA2	3.4E+07	37039000	34128000	0.013839263	0.013484275	0.01267729	0.013333609	0.000595458
				Scale to MBD3 iBAQ	BAIT-1M	2492690495	2746829265	2692058065
					BAIT-300mM	1996852450	3155680730	2995150310
					BAIT-0.15% NP	1068750000	1210055465	765450000

Supplementary Table 2: Overview of all assigned cross-links with FDR <0.05 as calculated in pLink.

Non-ambiguous Intra-protein Links			
sp|O94776|MTA2_HUMAN	460	sp|O94776|MTA2_HUMAN	504
sp|O94776|MTA2_HUMAN	460	sp|O94776|MTA2_HUMAN	508
sp|O94776|MTA2_HUMAN	492	sp|O94776|MTA2_HUMAN	504
sp|O94776|MTA2_HUMAN	501	sp|O94776|MTA2_HUMAN	508
sp|O94776|MTA2_HUMAN	504	sp|O94776|MTA2_HUMAN	559
sp|O94776|MTA2_HUMAN	508	sp|O94776|MTA2_HUMAN	539
sp|O94776|MTA2_HUMAN	508	sp|O94776|MTA2_HUMAN	559
sp|O94776|MTA2_HUMAN	522	sp|O94776|MTA2_HUMAN	539
sp|O94776|MTA2_HUMAN	531	sp|O94776|MTA2_HUMAN	539
sp|O94776|MTA2_HUMAN	531	sp|O94776|MTA2_HUMAN	559
sp|O94776|MTA2_HUMAN	533	sp|O94776|MTA2_HUMAN	539
sp|O94776|MTA2_HUMAN	533	sp|O94776|MTA2_HUMAN	559
sp|O94776|MTA2_HUMAN	539	sp|O94776|MTA2_HUMAN	559
sp|O94776|MTA2_HUMAN	539	sp|O94776|MTA2_HUMAN	592
sp|O94776|MTA2_HUMAN	559	sp|O94776|MTA2_HUMAN	592
sp|O94776|MTA2_HUMAN	611	sp|O94776|MTA2_HUMAN	618
sp|O95983|MBD3_HUMAN	46	sp|O95983|MBD3_HUMAN	68
sp|O95983|MBD3_HUMAN	92	sp|O95983|MBD3_HUMAN	109
sp|O95983|MBD3_HUMAN	129	sp|O95983|MBD3_HUMAN	142
sp|Q09028|RBBP4_HUMAN	156	sp|Q09028|RBBP4_HUMAN	215
sp|Q12873|CHD3_HUMAN	288	sp|Q12873|CHD3_HUMAN	295
sp|Q12873|CHD3_HUMAN	839	sp|Q12873|CHD3_HUMAN	843
sp|Q13547|HDAC1_HUMAN	438	sp|Q13547|HDAC1_HUMAN	444
sp|Q13547|HDAC1_HUMAN	444	sp|Q13547|HDAC1_HUMAN	451
sp|Q13547|HDAC1_HUMAN	451	sp|Q13547|HDAC1_HUMAN	457
sp|Q13547|HDAC1_HUMAN	466	sp|Q13547|HDAC1_HUMAN	476
sp|Q13547|HDAC1_HUMAN	469	sp|Q13547|HDAC1_HUMAN	476
sp|Q14839|CHD4_HUMAN	123	sp|Q14839|CHD4_HUMAN	126
sp|Q14839|CHD4_HUMAN	263	sp|Q14839|CHD4_HUMAN	273
sp|Q14839|CHD4_HUMAN	266	sp|Q14839|CHD4_HUMAN	273
sp|Q14839|CHD4_HUMAN	266	sp|Q14839|CHD4_HUMAN	285
sp|Q14839|CHD4_HUMAN	266	sp|Q14839|CHD4_HUMAN	287
sp|Q14839|CHD4_HUMAN	266	sp|Q14839|CHD4_HUMAN	290
sp|Q14839|CHD4_HUMAN	284	sp|Q14839|CHD4_HUMAN	287
sp|Q14839|CHD4_HUMAN	284	sp|Q14839|CHD4_HUMAN	289
sp|Q14839|CHD4_HUMAN	285	sp|Q14839|CHD4_HUMAN	290
sp|Q14839|CHD4_HUMAN	285	sp|Q14839|CHD4_HUMAN	295
sp|Q14839|CHD4_HUMAN	287	sp|Q14839|CHD4_HUMAN	290
sp|Q14839|CHD4_HUMAN	287	sp|Q14839|CHD4_HUMAN	295
sp|Q14839|CHD4_HUMAN	290	sp|Q14839|CHD4_HUMAN	297
sp|Q14839|CHD4_HUMAN	290	sp|Q14839|CHD4_HUMAN	304
sp|Q14839|CHD4_HUMAN	290	sp|Q14839|CHD4_HUMAN	501
sp|Q14839|CHD4_HUMAN	295	sp|Q14839|CHD4_HUMAN	304
sp|Q14839|CHD4_HUMAN	297	sp|Q14839|CHD4_HUMAN	504
sp|Q14839|CHD4_HUMAN	690	sp|Q14839|CHD4_HUMAN	693
sp|Q14839|CHD4_HUMAN	693	sp|Q14839|CHD4_HUMAN	696
sp|Q14839|CHD4_HUMAN	833	sp|Q14839|CHD4_HUMAN	838

sp|Q14839|CHD4_HUMAN	833	sp|Q14839|CHD4_HUMAN	839
sp|Q14839|CHD4_HUMAN	969	sp|Q14839|CHD4_HUMAN	1189
sp|Q14839|CHD4_HUMAN	1390	sp|Q14839|CHD4_HUMAN	1396
sp|Q14839|CHD4_HUMAN	1390	sp|Q14839|CHD4_HUMAN	1454
sp|Q14839|CHD4_HUMAN	1660	sp|Q14839|CHD4_HUMAN	1664
sp|Q14839|CHD4_HUMAN	1687	sp|Q14839|CHD4_HUMAN	1693
sp|Q14839|CHD4_HUMAN	1687	sp|Q14839|CHD4_HUMAN	1723
sp|Q14839|CHD4_HUMAN	1690	sp|Q14839|CHD4_HUMAN	1723
sp|Q14839|CHD4_HUMAN	1693	sp|Q14839|CHD4_HUMAN	1723
sp|Q16576|RBBP7_HUMAN	155	sp|Q16576|RBBP7_HUMAN	214
sp|Q16576|RBBP7_HUMAN	159	sp|Q16576|RBBP7_HUMAN	214
sp|Q86YP4|P66A_HUMAN	178	sp|Q86YP4|P66A_HUMAN	204
sp|Q86YP4|P66A_HUMAN	204	sp|Q86YP4|P66A_HUMAN	233
sp|Q86YP4|P66A_HUMAN	349	sp|Q86YP4|P66A_HUMAN	459
sp|Q86YP4|P66A_HUMAN	451	sp|Q86YP4|P66A_HUMAN	459
sp|Q86YP4|P66A_HUMAN	459	sp|Q86YP4|P66A_HUMAN	503
sp|Q86YP4|P66A_HUMAN	459	sp|Q86YP4|P66A_HUMAN	507
sp|Q86YP4|P66A_HUMAN	487	sp|Q86YP4|P66A_HUMAN	507
sp|Q86YP4|P66A_HUMAN	499	sp|Q86YP4|P66A_HUMAN	507
sp|Q86YP4|P66A_HUMAN	503	sp|Q86YP4|P66A_HUMAN	507
sp|Q86YP4|P66A_HUMAN	550	sp|Q86YP4|P66A_HUMAN	585
sp|Q86YP4|P66A_HUMAN	576	sp|Q86YP4|P66A_HUMAN	585
sp|Q8WXI9|P66B_HUMAN	15	sp|Q8WXI9|P66B_HUMAN	30
sp|Q8WXI9|P66B_HUMAN	15	sp|Q8WXI9|P66B_HUMAN	44
sp|Q8WXI9|P66B_HUMAN	44	sp|Q8WXI9|P66B_HUMAN	52
sp|Q8WXI9|P66B_HUMAN	147	sp|Q8WXI9|P66B_HUMAN	157
sp|Q8WXI9|P66B_HUMAN	281	sp|Q8WXI9|P66B_HUMAN	462
sp|Q8WXI9|P66B_HUMAN	353	sp|Q8WXI9|P66B_HUMAN	462
sp|Q8WXI9|P66B_HUMAN	454	sp|Q8WXI9|P66B_HUMAN	462
sp|Q92769|HDAC2_HUMAN	413	sp|Q92769|HDAC2_HUMAN	439
sp|Q92769|HDAC2_HUMAN	440	sp|Q92769|HDAC2_HUMAN	451
sp|Q92769|HDAC2_HUMAN	444	sp|Q92769|HDAC2_HUMAN	452
sp|Q92769|HDAC2_HUMAN	451	sp|Q92769|HDAC2_HUMAN	452
sp|Q92769|HDAC2_HUMAN	458	sp|Q92769|HDAC2_HUMAN	468
sp|Q92769|HDAC2_HUMAN	458	sp|Q92769|HDAC2_HUMAN	478
sp|Q92769|HDAC2_HUMAN	462	sp|Q92769|HDAC2_HUMAN	468
sp|Q92769|HDAC2_HUMAN	462	sp|Q92769|HDAC2_HUMAN	478
sp|Q92769|HDAC2_HUMAN	466	sp|Q92769|HDAC2_HUMAN	478
sp|Q92769|HDAC2_HUMAN	468	sp|Q92769|HDAC2_HUMAN	478
sp|Q92769|HDAC2_HUMAN	468	sp|Q92769|HDAC2_HUMAN	481
Non-Ambiguous Inter-protein Links			
sp|O94776|MTA2_HUMAN	41	sp|O95983|MBD3_HUMAN	46
sp|O94776|MTA2_HUMAN	639	sp|Q12873|CHD3_HUMAN	1067
sp|O94776|MTA2_HUMAN	539	sp|Q16576|RBBP7_HUMAN	214
sp|O94776|MTA2_HUMAN	639	sp|Q8WXI9|P66B_HUMAN	462
sp|O95983|MBD3_HUMAN	92	sp|Q86YP4|P66A_HUMAN	178
sp|O95983|MBD3_HUMAN	109	sp|Q86YP4|P66A_HUMAN	204

sp|O95983|MBD3_HUMAN	92	sp|Q8WXI9|P66B_HUMAN	193
sp|Q12873|CHD3_HUMAN	116	sp|Q14839|CHD4_HUMAN	123
sp|Q12873|CHD3_HUMAN	1199	sp|Q14839|CHD4_HUMAN	1865
sp|Q12873|CHD3_HUMAN	1385	sp|Q86YP4|P66A_HUMAN	27
sp|Q12873|CHD3_HUMAN	14	sp|Q8WXI9|P66B_HUMAN	15
sp|Q13330|MTA1_HUMAN	532	sp|Q14839|CHD4_HUMAN	266
sp|Q13330|MTA1_HUMAN	532	sp|Q14839|CHD4_HUMAN	290
sp|Q13330|MTA1_HUMAN	686	sp|Q14839|CHD4_HUMAN	1390
sp|Q14839|CHD4_HUMAN	1844	sp|Q86YP4|P66A_HUMAN	349
sp|Q14839|CHD4_HUMAN	1865	sp|Q86YP4|P66A_HUMAN	503

Ambiguous Inter/Intra-Protein Links			
sp|Q14839|CHD4_HUMAN	969	sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	969;959
sp|O95983|MBD3_HUMAN	92	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	163;184
sp|O94776|MTA2_HUMAN;sp|Q9BTC8|MTA3_HUMAN	285;288	sp|O94776|MTA2_HUMAN	323
sp|Q86YP4|P66A_HUMAN	349	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358
sp|Q86YP4|P66A_HUMAN	459	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	448;451
sp|Q92769|HDAC2_HUMAN	413	sp|Q13547|HDAC1_HUMAN;sp|Q92769|HDAC2_HUMAN	363;364
sp|Q92769|HDAC2_HUMAN	439	sp|Q13547|HDAC1_HUMAN;sp|Q92769|HDAC2_HUMAN	363;364
sp|Q8WXI9|P66B_HUMAN	462	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	448;451
sp|Q92769|HDAC2_HUMAN	468	sp|Q13547|HDAC1_HUMAN;sp|Q92769|HDAC2_HUMAN	363;364
sp|Q92769|HDAC2_HUMAN	478	sp|Q13547|HDAC1_HUMAN;sp|Q92769|HDAC2_HUMAN	363;364
sp|Q9BTC8|MTA3_HUMAN	463	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24
sp|Q86YP4|P66A_HUMAN	459	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358
sp|Q86YP4|P66A_HUMAN	487	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358
sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358	sp|Q86YP4|P66A_HUMAN	507
sp|Q8WXI9|P66B_HUMAN	435	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	448;451
sp|O94776|MTA2_HUMAN	533	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	307;306
sp|O94776|MTA2_HUMAN	539	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	307;306
sp|O94776|MTA2_HUMAN	559	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24
sp|O94776|MTA2_HUMAN	559	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	307;306
sp|O94776|MTA2_HUMAN	592	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24
sp|O94776|MTA2_HUMAN	639	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24
sp|O95983|MBD3_HUMAN	124	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358
sp|O94776|MTA2_HUMAN;sp|Q9BTC8|MTA3_HUMAN	285;288	sp|O94776|MTA2_HUMAN	330
sp|O94776|MTA2_HUMAN;sp|Q9BTC8|MTA3_HUMAN	285;288	sp|O95983|MBD3_HUMAN	46
sp|O94776|MTA2_HUMAN;sp|Q13330|MTA1_HUMAN	32;32	sp|O94776|MTA2_HUMAN	323
sp|O94776|MTA2_HUMAN;sp|Q13330|MTA1_HUMAN	32;32	sp|O95983|MBD3_HUMAN	46
sp|O94776|MTA2_HUMAN	445	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24
sp|O94776|MTA2_HUMAN	449	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24
sp|O94776|MTA2_HUMAN	531	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	307;306

sp|Q14839|CHD4_HUMAN	123	sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	113;126
sp|Q13330|MTA1_HUMAN	582	sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	307;306
sp|Q14839|CHD4_HUMAN	1844	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358
sp|Q14839|CHD4_HUMAN	1852	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358
sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	307;306	sp|O94776|MTA2_HUMAN	533
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1394;1398	sp|Q14839|CHD4_HUMAN	1390
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1394;1398	sp|Q14839|CHD4_HUMAN	1454
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	894;884	sp|Q14839|CHD4_HUMAN	290
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	969;959	sp|Q12873|CHD3_HUMAN	288
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	969;959	sp|Q14839|CHD4_HUMAN	1189
sp|Q13330|MTA1_HUMAN	305	sp|Q13330|MTA1_HUMAN;sp|Q9BTC8|MTA3_HUMAN	343;326
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1339;1335	sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	994;984
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1798;1787	sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1791;1780
sp|Q09028|RBBP4_HUMAN;sp|Q16576|RBBP7_HUMAN	25;24	sp|Q13547|HDAC1_HUMAN;sp|Q92769|HDAC2_HUMAN	363;364
sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1791;1780	sp|Q12873|CHD3_HUMAN;sp|Q14839|CHD4_HUMAN	1339;1335
sp|O94776|MTA2_HUMAN;sp|Q9BTC8|MTA3_HUMAN	285;288	sp|Q13330|MTA1_HUMAN;sp|Q9BTC8|MTA3_HUMAN	343;326
sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	354;358	sp|Q86YP4|P66A_HUMAN;sp|Q8WXI9|P66B_HUMAN	448;451

5

Chapter 6

Perspective on unraveling the versatility of 'co-repressor' complexes

H. Irem Baymaz[a*], Ino D. Karemaker[a*], and Michiel Vermeulen[a#]

a. Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, the Netherlands

* These authors contributed equally and are listed alphabetically
# Corresponding author

Biochimica et Biophysica Acta - Gene Regulatory Mechanisms
August 2015, Volume 1849, Issue 8, pages 1051–1056

Abstract

A multitude of post-translational modifications take place on histones, one of the best studied being acetylation on lysine residues, which is generally associated with gene activation. During the last decades, several so-called co-repressor protein complexes that carry out the reverse process, histone deacetylation, have been identified and characterized, such as the Sin3, N-CoR/SMRT and NuRD complexes. Although a repressive role for these complexes in regulating gene expression is well established, accumulating evidence also points to a role in gene activation. Here, we argue that integration of various state-of-the-art technologies, addressing different aspects of transcriptional regulation, is essential to unravel this apparent biological versatility of 'co-repressor' complexes.

Introduction

To facilitate the packaging of DNA into the cell nucleus, four types of histone proteins (H2A, H2B, H3 and H4) together form an octamer around which 147 base pairs of DNA is wrapped. This complex of histones and DNA is called the nucleosome. Both the core domains and the N-termini of histones, the so-called histone tails, contain a large number of amino acid residues that can be post-translationally modified. Because of their proximity to DNA, these modifications can influence gene expression. An important and frequent modification that occurs on histone tails is acetylation, in which an acetyl group ($COCH_3$) is covalently coupled to a lysine residue (K). Already in the 1960s, histone acetylation was proposed to be associated with gene activation[1], an assumption that has been supported by many findings since (e.g.[2]). The acetyl group can neutralise the positive charge of the histone, thus disrupting higher-order chromatin structure by weakening the interaction between the histone octamer and DNA[3]. In general, this leads to enhanced gene transcription *in cis*. In addition, acetylated histones are recognized by bromodomain-containing activator proteins[4]. Many proteins have been identified that have the ability to acetylate histones; these enzymes are called histone acetyltransferases (HATs, reviewed in[3]). Most protein complexes containing HATs are thus known transcriptional activators. In contrast, histone deacetylases (HDACs) remove acetyl groups from histones and these proteins are therefore thought to have a repressive effect on gene expression. Indeed, several complexes containing HDACs were described to induce transcriptional repression, such as the Sin3A/B, N-CoR/SMRT and NuRD complexes[5–11]. Although a repressive role for the these complexes in regulating gene expression is well established, accumulating evidence has revealed that co-repressor complexes also localize to actively transcribed genes and are sometimes required for their activation. In this perspective, we discuss some examples where the Sin3A/B, N-CoR/SMRT and NuRD complexes are associated with gene activation (for a more comprehensive review, refer to [12]). We also discuss recent technological advances that need to be further developed and integrated to decipher the molecular mechanisms behind the apparent biological versatility of 'co-repressor' complexes.

When co-repressors activate

Early on it was shown that upon inhibition or depletion of HDACs, an approximately equal number of genes are up- and downregulated[13–16]. While these observations may be attributed to indirect effects, subsequent studies proved otherwise[17]. Pioneering mechanistic studies in yeast clearly showed that the Sin3/Rpd3 complex is required for activation of some of its target genes upon osmostress and heat shock in a histone deacetylation-dependent manner[18,19]. The development of new genomics technology also brought about new perspectives on the function of histone acetylation and deacetylation, both in yeast (ChIP-on-chip) and mammals (ChIP-sequencing). The first global localization study for Rpd3, interestingly, showed enrichment both near repressed and active genes[20]. Similarly, in mammals, the genome-wide localization of several HATs and HDACs showed that binding of HATs as well as HDACs positively correlates with gene expression, RNA Polymerase II binding and acetylation levels[21]. The authors proposed that at active genes, the function of HDACs may be to remove acetyl groups during transcription, so that chromatin is 'reset' for the next round of transcription[22]. This suggests that the role of HDACs may not be solely repressive, but can in fact also support active gene expression.

Indeed, targeted studies focusing on individual HDAC-containing complexes provided further support for these initial observations. The mammalian orthologues of Sin3, Sin3A

and B, were shown to be enriched in the proximity of genes that are activated as myoblasts differentiate into myotubes[23]. A subset of these genes is downregulated upon Sin3A/B depletion, further implying the activating effect of Sin3 proteins on these genes. It is noteworthy that Sin3A/B target genes that are repressed during myoblast differentiation are also enriched for the transcription factor E2F4. Relevant to this observation, in a more recent report by the same group, Sin3A was reported to be enriched at promoters of active genes with high levels of H3K4me3, along with the H3K4me3 interactor ING1[24]. Another set of genes that are Sin3A targets were also enriched for E2F4. Upon perturbation of the H3K4me3/me1 ratio at the promoter proximal regions, decreased binding of both ING1 and Sin3A was observed in the former set of genes, whereas Sin3A/E2F4 targets were insensitive to the H3K4me3/me1 ratio. These observations classify Sin3 targets depending on different co-factors and/or the chromatin marks that are present *in cis*. Thus, recruitment of Sin3 to chromatin by different co-factors possibly results in different transcriptional outputs.

For the nuclear receptor co-repressors, it was first observed in 2000 that N-CoR was required for transcriptional activation mediated by the retinoic acid receptor-specific ligand (LG550). This ligand failed to activate target genes in N-CoR$^{-/-}$ MEFS and exogenous expression of N-CoR restored activation[25]. Similarly, the closely related SMRT co-repressor was observed to mediate full estrogen-dependent ER-alpha transcriptional activation, albeit in a cell type-specific manner[26]. In addition, a recent ChIP-sequencing study revealed that N-CoR and SMRT colocalize with known activators at Vitamin D3-activated gene enhancers. The genome-wide binding profile of N-CoR showed high correlation with the Vitamin D receptor binding upon Vitamin D3 stimulation, suggesting an activating role for N-CoR[27]. While most of the evidence for an activating role of N-CoR and SMRT is based on correlation, it is clear that gene regulation mediated by these co-repressors is more complex than originally described (reviewed in [28]).

For the NuRD complex, a number of studies have suggested a correlation with active transcription. For example, when the genome-wide localisation of the NuRD subunit MBD3 was studied in MCF-7, MDA-231 and mES cells, it was found that MBD3 preferentially localises to the active part of the genome, i.e. active enhancer regions, active promoters and gene bodies of actively transcribed genes[29,30]. Together with the finding that the NuRD subunit Mi-2β localises mainly to transcriptionally active genes[31], this suggests an activating role for the NuRD complex at these loci. Also in the haematopoietic system, an activating role for the NuRD complex was described. FOG-1, which itself is a co-factor for the haematopoietic factor GATA-1 and is required for regulation of most GATA-1 dependent genes, is known to bind to the NuRD complex and recruit it to FOG-1 target genes[32]. While the NuRD complex is present at target genes of GATA-1 and FOG-1 that are repressed, it is also recruited to some other GATA-1 and FOG-1 targets where it mediates activation. Indeed, it was shown that NuRD occupancy at these genes remained high or even increased upon FOG-1-dependent activation. Additionally, a construct containing a FOG-1-dependent reporter gene could no longer be activated by FOG-1 when the NuRD subunits MTA1-3 were knocked down, indicating that NuRD is directly required for transcriptional activation[32]. By studying mice homozygous for a mutated form of FOG-1 that is unable to interact with the NuRD complex, an activating role for NuRD was suggested for some FOG-1/GATA-1 target genes also *in vivo*[32]. Together, these findings show that the NuRD complex is directly involved in activation of gene expression in multiple instances, indicating that this may be a more general role than previously thought.

Data interpretation and future directions

When studying transcriptional regulation by multi-subunit protein complexes, there are many factors that need to be considered that are generally overlooked in genome-wide sequencing based approaches. Below we describe these factors, and discuss several methods and technologies that we think are required, in an integrative manner, to solve the molecular mechanisms underlying the apparent biological versatility of 'co-repressor' complexes (see **Figure 1**).

First of all, most genome-wide profiling studies such as ChIP-sequencing require large amounts of cells. Heterogeneity within a certain cell population that is used for the experiment may compromise correct interpretation of the data. In large heterogeneous cell populations, opposing individual signals may be averaged out and therefore escape notice. Also, signals present in a minority of cells may be lost due to overruling signals from the majority. Finally, proteins that appear to co-occupy certain genes in a ChIP-sequencing experiment, may actually bind to the same genes but in a mutually exclusive manner. It is thus important to realise that population effects can strongly affect data output and result in false, generalizing conclusions concerning individual cells.

Asynchronous cell populations cause another layer of heterogeneity. Cells that are in different phases of the cell cycle differ in the genes they express and the epigenetic marks that are present on their chromatin. Thus, pooling asynchronous cells possibly masks some cell cycle specific patterns. A 2003 study using synchronized cell populations showed alternating enrichment of factors driving waves of expression[33]. This led to the model of cyclical regulation of transcription, which was a remarkable leap from the analogy that visualizes transcriptional control as an on/off switch[28]. DNA methylation has also been shown to vary cyclically and thereby regulate gene expression[34,35]. These experiments in synchronized cells thus revealed that cyclical variation may be a more general phenomenon in the regulation of gene expression and that conclusions drawn from the analysis of asynchronous cell populations may not represent all regulatory patterns equally well.

An additional point that needs to be emphasised is that correlation does not necessarily mean causation. When two proteins or epigenetic marks are found to colocalise at a specific locus, this could indeed mean that the presence of one led to recruitment of the other, or that both interacted already before binding at the locus together. However, proteins or marks could also be attracted to the same site independently of each other, for example through binding other proteins. Other scenarios are also imaginable, such as one protein or mark actually causing repulsion of the other, but that this event had not yet taken place at the time of analysis. When examining correlation studies we therefore need to take these different possibilities into account, and appreciate the different conclusions they may propose. Information on the duration of the binding of factors may vastly add to the interpretation of correlation studies. It was first shown in 2000 that regulatory factors are not statically bound at the their target regulatory sites upon stimulation, but are rather in a state of constant exchange[36]. Additionally, in the 2003 study mentioned above, the dynamics and co-occurrence of factors on the estrogen-responsive pS2 gene promoter were identified over a series of time course experiments[33]. Knowing the dynamics of factor binding at target sites provides a more detailed picture of the *in vivo* situation and allows us to decipher the cause within the correlation.

Essential to keep in mind is the structural heterogeneity of the Sin3A/B, N-CoR/SMRT and NuRD complexes. Each complex consists of multiple core subunits, many of which have orthologues, resulting in a large number of possible complex compositions[28,37–39]. Evidence is

emerging that similar but distinct core subunits have different biological functions, implying that the exact subunit composition influences the activity of the complex[23,40,41]. In addition, various substoichiometric, often cell-type specific interactors have been identified[39]. These proteins also have the ability to alter the behaviour of the complex, for example by directing the complex to specific targets. Identifying the exact composition of the complexes in different contexts should also shed light on their functional heterogeneity. An additional variation on complex composition concerns different post-translational modifications (PTMs) of the subunits. For example, for the NuRD complex, progressive acetylation of HDAC1 was shown to block its catalytic activity during erythropoiesis[42–44]. The complex is hypothesized to switch function from repression to activation as the HDAC subunits lose their catalytic activity and the only chromatin modifying activity the complex can perform is remodelling. This example shows that PTMs contribute to the functional switch by modulating the catalytic activities of the complexes. In addition, PTMs could affect complex composition by influencing interactions with other subunits or interactors. Yet another layer of complication regarding the study of multisubunit complexes concerns potential complex-independent functions of the subunits. Therefore it may be inaccurate to extrapolate from data obtained by studying only one subunit to the whole complex.

Another area of transcriptional regulation that needs to be considered is at the stage of RNA Polymerase II (Pol II) elongation. For a long time, general transcription factors recruiting Pol II machinery was considered sufficient to generate transcriptional output, which is an assumption that mostly holds true for transcriptional control in *Saccharomyces cerevisiae*. Notably, chromatin modifying complexes do interact with the elongating Pol II, affecting its processivity in *Saccharomyces cerevisiae*; however this usually takes place on the gene bodies and increasingly toward the 3' end of genes[45]. In metazoans, on the other hand, there had been indications of extra levels of regulation between the recruitment of Pol II and the initiation of productive transcription at the 5' ends of genes. With the genome-wide maps of Pol II localization, the extent of paused versus elongating Pol II could be observed more globally[46,47]. Further work elucidated the presence of positive and negative elongation factors that strictly regulate transcriptional elongation by Pol II. So far, a link between Pol II stalling/elongation and various chromatin complexes has been shown for the NuRD and Sin3B complexes via specific interactors, IKAROS and PHF12 respectively[48,49]. NuRD positively regulates elongation via its remodeling activity whereas the presence of Sin3B/HDAC1 is suggested to block elongation due to histone deacetylation. More examples of this type of regulation undertaken by these chromatin complexes are expected to emerge and possibly clarify the context-dependency of their activating or repressing function.

Novel technology to investigate nuclear organization in three dimensions has revealed many new insights regarding regulation of transcription. Especially the generation of 'C technologies' consisting of chemical crosslinking of DNA followed by deep sequencing provided a genome-wide picture of chromosomal architecture[50,51]. There is a strong correlation between the position of a gene within the nucleus (peripheral or central) and its activity (silent or active respectively)[51], but the exact cause and effect relationship between the two aspects is not yet known. There are a few examples demonstrating enhanced transcription upon induced chromatin looping[52,53], but more data on different loci is required to fully address the causality question. However, classifications have emerged that qualify certain looping events such as permissive looping versus instructive looping[50]. As such, it is now conceivable that the cause and effect relationship between looping and transcription is not unilateral and can change context- or locus-dependently. A recent report showed that

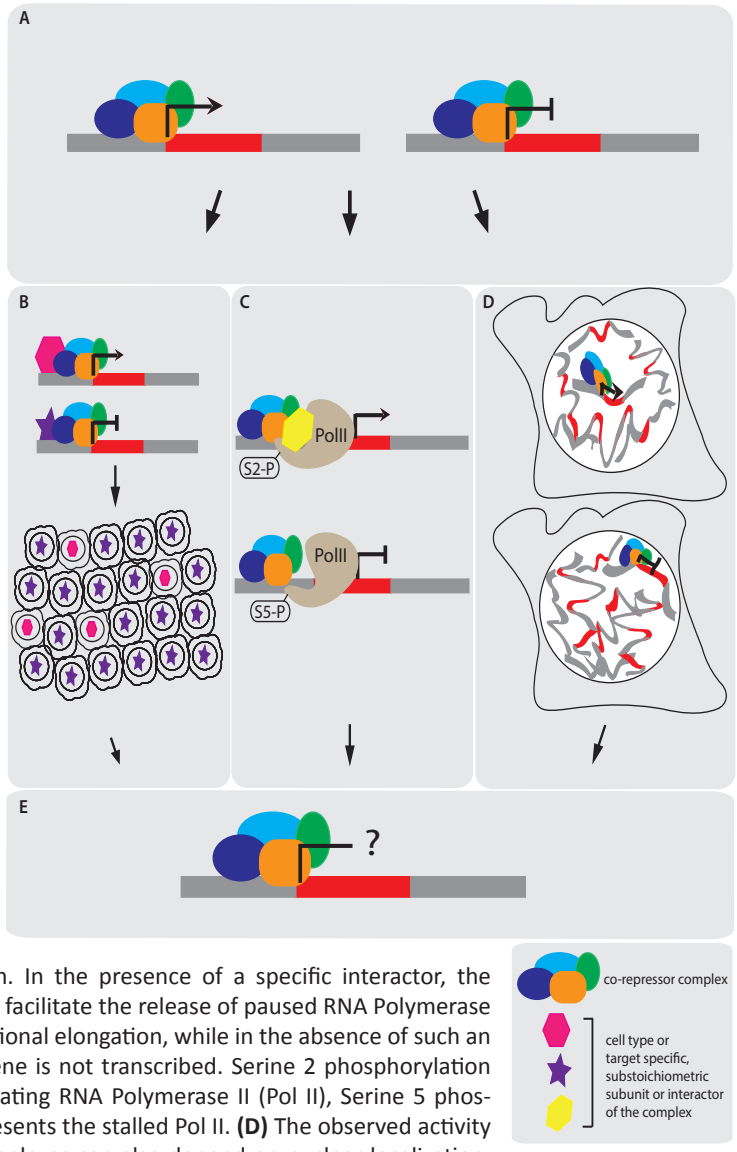
decondensation of chromatin was sufficient for a locus to move from the nuclear periphery to a more central nuclear position, suggesting that the change in chromatin compaction is driving the change in location within the nucleus, and hence, the 3D organization[54]. This study is exemplary in integrating multiple levels of regulation: chromatin compaction regulating the potential for transcription and, at the same time, affecting the position of the locus in question. For the purposes of understanding the function of the chromatin complexes (Sin3A/B, N-CoR/SMRT and NuRD), the next step would be to look at their global 3D distribution within the nucleus and see whether they affect the genome architecture in any way.

At the population level, an established approach that has given many insights as mentioned above is ChIP re-ChIP[33]. Whereas ChIP assays are designed to identify the DNA targets of a certain protein of interest, re-ChIP or sequential ChIP allows for the identification of proteins that are concomitantly binding to the same DNA sequence. Although this technically challenging method has been shown to yield qualitatively good data (e.g.[33,35]), it is restricted in the sense that it is a targeted approach, where proteins of interest are known in advance. Further advances may therefore mainly lie in unbiased single cell and even single locus approaches. Indeed, gene expression has been revealed as an essentially stochastic process, resulting in cell-to-cell variations in mRNA and protein levels[55]. Single cell biology, although a relatively young field, has already yielded many promising methods that have provided insights into the variability of cell populations, and many researchers are working on even more advanced techniques to look into single cell dynamics[56,57]. The stochasticity of transcription observed in single cell studies provides a very intriguing viewpoint, suggesting that there may be far more levels of regulation in play that lead to the deterministic pattern of expression observed at the population level[58,59]. In addition, effort has been put into developing methods facilitating unbiased identification of proteins and histone marks at a single genomic locus, although these approaches are still constrained to using populations of cells to obtain enough material[60–63]. Both single cell and single locus approaches have the potential to offer more details about the behaviour of the Sin3A/B, N-CoR/SMRT and NuRD complexes. This detailed insight seems indispensable for elucidating the apparent versatility of the complexes and for thoroughly explaining the different contexts of their function.

In an ideal world, we would be able to look at the exact complex composition on a single locus in a single cell, identify all other proteins present on the locus and their binding dynamics, know the 3D location of the locus within the nucleus, know the state of Pol II and the transcriptional output. By putting all this information together, we can study all aspects of transcriptional control at the single cell level as well as the population level, pinpoint the contribution of stochastic versus deterministic processes, and refine the concepts of activation and repression by ‘co-repressor’ complexes.

Figure 1: Brief overview of discussed methods

Panel **(A)** depicts the apparent versatility of the co-repressor complexes. In the panels below, different methods of addressing this question are shown. The read-out of these methods is influenced by the context characterized by differential interactors, the interaction with a paused or elongating Pol II and nuclear localization. **(B)** The action of the co-repressor complexes depends on association with different interactors which could be specific to cell type, to the phase of the cell cycle or to a subset of targets. Due to the heterogeneity of the cell population (e.g. different phases of the cell cycle and/or slight differences in developmental stage) the opposite outcomes mediated by differential interactors are not observed, i.e. the majority of the population of cells masks the opposite effect observed in the minority of cells. **(C)** The co-repressor complexes could be involved in regulating transcription elongation. In the presence of a specific interactor, the co-repressor complexes facilitate the release of paused RNA Polymerase II, resulting in transcriptional elongation, while in the absence of such an interactor the target gene is not transcribed. Serine 2 phosphorylation (S2-P) represents elongating RNA Polymerase II (Pol II), Serine 5 phosphorylation (S5-P) represents the stalled Pol II. **(D)** The observed activity of the co-repressor complexes can also depend on nuclear localization. When a target gene bound by a complex is located at the nuclear periphery it is silenced, while the same gene bound by the same complex could be activated when located more centrally in the nucleus. This instance of regulation could be explained by the repressive versus permissive chromatin context within the nucleus. **(E)** In summary, the caveats of the currently used methods prevent us from predicting the outcome of the action of the 'co-repressor' complexes as the activating and repressing functions are dependent on as yet ill-defined contexts. Technology integration could resolve the differential action of co-repressors and help in defining the contextual differences that cause the different outcomes in the first place.



Techniques

Chromatin immunoprecipitation is an immunoprecipitation technique that is used to study pro tein-DNA interactions in the context of the chromatin. The basic steps include crosslinking proteins and chromatin, after which the DNA is cut into small fragments. An antibody for a specific protein is then used to immunoprecipitate this protein together with the associated DNA. Finally, the crosslinks are reversed, the DNA is purified and can be subjected to a number of techniques. In **ChIP-seq**, the DNA is sequenced by massively parallel DNA sequencing. Alternatively, the DNA can be hybridised to a DNA microarray (**ChIP-on-chip**), another method to identify the fragments that were enriched by ChIP.

In **ChIP-reChIP**, immediately after the first immunoprecipitation the sample is subjected to a second round of immunoprecipitation with a different antibody. This approach is used to study the colocalisation of proteins.

C technologies are a collection of technologies that use chromatin conformation capture to study the 3D organisation of DNA inside the nucleus. Usually the protocol includes crosslinking followed by restriction digestion and intramolecular ligation. After the crosslinks are reversed, there are several methods to analyse the sample, typically containing an amplification and quantitation step.

RNA-sequencing is a method used to study the transcriptome of cells. Generally, RNA is isolated from cells and reverse-transcribed into cDNA, which is then subjected to sequencing. This provides information on the transcription status of genes at a particular time. Recently, the development of single-cell RNA-sequencing made it possible to study the transcriptome of individual cells.

Single-cell technologies cover a plethora of techniques, including methods to visualize transcription (e.g. RNA-FISH), histone modifications (e.g. ISH-PLA), individual protein localization (e.g. FRAP), or protein-DNA interactions (e.g. DamID) at the single-cell level.

Single locus proteomics aims to identify all the proteins associated with a single locus, however, this technique is not yet applicable to single cells.

Acknowledgments

The Vermeulen laboratory is supported by EU FP7 framework program 4DCellFate and a European Research Council Starting Grant (no. 309384).

References

1. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Actylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci.* **51,** 786–794 (1964).
2. Hebbes, T., Thorne, A. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7,** 1395–1402 (1988).
3. Sterner, D. E. & Berger, S. L. Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* **64,** 435–459 (2000).
4. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett.* **586,** 2692–2704 (2012).
5. Nasmyth, K., Stillman, D. & Kipling, D. Both positive and negative regulators of HO transcription are required for mother-cell-specific mating-type switching in yeast. *Cell* **48,** 579–587 (1987).
6. Sternberg, P. W., Stern, M. J., Clark, I. & Herskowitz, I. Activation of the yeast HO gene by release from multiple negative controls. *Cell* **48,** 567–577 (1987).
7. Laherty, C. D. *et al.* Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* **89,** 349–356 (1997).
8. Hörlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377,** 397–403 (1995).
9. Nagy, L. *et al.* Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell* **89,** 373–380 (1997).
10. Xue, Y. *et al.* NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* **2,** 851–861 (1998).
11. Kehle, J. *et al.* dMi-2, a Hunchback-Interacting Protein That Functions in Polycomb Repression. *Science (80-.).* **282,** 1897–1900 (1998).
12. Reynolds, N., O'Shaughnessy, A. & Hendrich, B. Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140,** 505–512 (2013).
13. Bernstein, B. E., Tong, J. K. & Schreiber, S. L. Genomewide studies of histone deacetylase function in yeast. *Proc. Natl. Acad. Sci.* **98,** 13708–13713 (2000).
14. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102,** 109–126 (2000).
15. Reid, G. *et al.* Multiple mechanisms induce transcriptional silencing of a subset of genes, including oestrogen receptor alpha, in response to deacetylase inhibition by valproic acid and trichostatin A. *Oncogene* **24,** 4894–4907 (2005).
16. Peart, M. J. *et al.* Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proc. Natl. Acad. Sci.* **102,** 3697–3702 (2005).
17. Deckert, J. & Struhl, K. Histone Acetylation at Promoters Is Differentially Affected by Specific Activators and Repressors. *Mol. Cell. Biol.* **21,** 2726–2735 (2001).
18. De Nadal, E. *et al.* The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. *Nature* **427,** 370–374 (2004).
19. Ruiz-Roig, C., Viéitez, C., Posas, F. & de Nadal, E. The Rpd3L HDAC complex is essential for the heat stress response in yeast. *Mol. Microbiol.* **76,** 1049–1062 (2010).
20. Kurdistani, S. K., Robyr, D., Tavazoie, S. & Grunstein, M. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat. Genet.* **31,** 248–54 (2002).
21. Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138,** 1019–31 (2009).
22. Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138,** 1019–31 (2009).
23. Van Oevelen, C. *et al.* The mammalian Sin3 proteins are required for muscle development and sarcomere specification. *Mol. Cell. Biol.* **30,** 5686–5697 (2010).
24. Cheng, J. *et al.* A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol. Cell* **53,** 979–992 (2014).
25. Jepsen, K. *et al.* Combinatorial roles of the nuclear receptor corepressor in transcription and

development. *Cell* **102,** 753–763 (2000).
26. Peterson, T. J., Karmakar, S., Pace, M. C., Gao, T. & Smith, C. L. The silencing mediator of retinoic acid and thyroid hormone receptor (SMRT) corepressor is required for full estrogen receptor alpha transcriptional activity. *Mol. Cell. Biol.* **27,** 5933–48 (2007).
27. Meyer, M. B. & Pike, J. W. Corepressors (NCoR and SMRT) as well as coactivators are recruited to positively regulated 1α,25-dihydroxyvitamin D3-responsive genes. *J. Steroid Biochem. Mol. Biol.* **136,** 120–4 (2013).
28. Perissi, V., Jepsen, K., Glass, C. K. & Rosenfeld, M. G. Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* **11,** 109–123 (2010).
29. Shimbo, T. *et al.* MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS Genet.* **9,** e1004028 (2013).
30. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153,** 480–492 (2013).
31. Zhang, J. *et al.* Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat. Immunol.* **13,** 86–94 (2012).
32. Miccio, A. *et al.* NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. *EMBO J.* **29,** 442–456 (2010).
33. Métivier, R. *et al.* Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115,** 751–763 (2003).
34. Kangaspeska, S. *et al.* Transient cyclical methylation of promoter DNA. *Nature* **452,** 112–115 (2008).
35. Métivier, R. *et al.* Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452,** 45–50 (2008).
36. McNally, J. G., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science (80-.).* **287,** 1262–1265 (2000).
37. Grzenda, A., Lomberk, G., Zhang, J. S. & Urrutia, R. Sin3: Master scaffold and transcriptional corepressor. *Biochim. Biophys. Acta* **1789,** 443–450 (2009).
38. Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. Mi-2/NuRD: Multiple complexes for many purposes. *Biochim. Biophys. Acta* **1677,** 52–57 (2004).
39. McDonel, P., Costello, I. & Hendrich, B. Keeping things quiet: Roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int. J. Biochem. Cell Biol.* **41,** 108–116 (2009).
40. Allen, H. F., Wade, P. A. & Kutateladze, T. G. The NuRD architecture. *Cell. Mol. Life Sci.* **70,** 3513–3524 (2013).
41. Laugesen, A. & Helin, K. Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell* **14,** 735–751 (2014).
42. Qiu, Y. *et al.* HDAC1 Acetylation Is Linked to Progressive Modulation of Steroid Receptor-Induced Gene Transcription. *Mol. Cell* **22,** 669–679 (2006).
43. Luo, Y. *et al.* Trans-regulation of histone deacetylase activities through acetylation. *J. Biol. Chem.* **284,** 34901–34910 (2009).
44. Yang, T. *et al.* Acetylation of histone deacetylase 1 regulates NuRD corepressor complex activity. *J. Biol. Chem.* **287,** 40279–40291 (2012).
45. Smith, E. & Shilatifard, A. The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes. *Mol. Cell* **40,** 689–701 (2010).
46. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130,** 77–88 (2007).
47. Muse, G. W. *et al.* RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39,** 1507–1511 (2007).
48. Bottardi, S. *et al.* The IKAROS Interaction with a Complex Including Chromatin Remodeling and Transcription Elongation Activities Is Required for Hematopoiesis. *PLoS Genet.* **10,** e1004827 (2014).

49. Jelinic, P., Pellegrino, J. & David, G. A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol. Cell. Biol.* **31,** 54–62 (2011).
50. De Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502,** 499–506 (2013).
51. Sexton, T. & Cavalli, G. Review The Role of Chromosome Domains in Shaping the Functional Genome. *Cell* **160,** 1049–1059 (2015).
52. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149,** 1233–1244 (2012).
53. Noordermeer, D. *et al.* Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat. Cell Biol.* **13,** 944–951 (2011).
54. Therizols, P. *et al.* Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science (80-.).* **346,** 1238–1242 (2014).
55. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135,** 216–226 (2008).
56. Junker, J. P. & Van Oudenaarden, A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* **157,** 8–11 (2014).
57. Bheda, P. & Schneider, R. Epigenetics reloaded: the single-cell revolution. *Trends Cell Biol.* **24,** 712–723 (2014).
58. Voss, T. C. *et al.* Combinatorial probabilistic chromatin interactions produce transcriptional heterogeneity. *J. Cell Sci.* **122,** 345–356 (2009).
59. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* **15,** 69–81 (2014).
60. Byrum, S. D., Raman, A., Taverna, S. D. & Tackett, A. J. ChAP-MS: A method for identification of proteins and histone posttranslational modifications at a single genomic locus. *Cell Rep.* **2,** 198–205 (2012).
61. Waldrip, Z. J. *et al.* A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* **9,** 37–41 (2014).
62. Déjardin, J. & Kingston, R. E. Purification of Proteins Associated with Specific Genomic Loci. *Cell* **136,** 175–186 (2009).
63. Pourfarzad, F. *et al.* Locus-specific proteomics by TChP: Targeted chromatin purification. *Cell Rep.* **4,** 589–600 (2013).

6

Chapter 7

Discussion

"When life itself seems lunatic, who knows where madness lies? Perhaps to be
too practical is madness. To surrender dreams - this may be madness.
Too much sanity may be madness - and maddest of all:
to see life as it is, and not as it should be!"

Cervantes, *Don Quixote* (1605, 1615)

Regulation of transcription is an essential process in driving life forward. It is involved in regulating cell metabolism via sensing intracellular metabolites and integrating extracellular signals provided by signaling pathways in both multicellular and single cell organisms. In addition, activation and repression of different transcriptional programs is essential for proliferation and differentiation in multicellular organisms. Hence, the appropriate response to environmental changes and the maintenance of tissue homeostasis are mediated by modulation of transcription.

As a complex and multistep process, the various steps of transcription, i.e. initiation, elongation and termination, are relatively well described with most of the components biochemically characterized. As DNA is wrapped around histone octamers in eukaryotic cells, transcriptional regulation requires alteration of the chromatin environment. A reciprocal correlation between 'chromatin density' and amenability to transcription dates back to 1920s[1]. However, the characterization of the transcription machinery preceded studies revealing the properties of chromatin around active transcription sites. This sequence of discoveries is mostly due to the technological challenges of studying chromatin and partly to the underestimation of the influence of chromatin components on DNA accessibility.

The chromatin environment, as introduced in chapter 1, affects all DNA-templated processes. In this thesis, well-known transcriptional regulatory complexes, NuRD, PRC1 and PRC2 as well as the relatively recently characterized PR-Dub complex, are the subject of study. The mammalian PR-Dub complex composition and the functional implication of MBD6 in DNA damage response are presented in chapter 3. Novel insight into the interplay between PRC1 and PRC2 complexes provided by the biochemical characterization of nucleosomal H2AK118/119Ub interactors is presented in chapter 4. Attempts at characterizing the stability of the NuRD complex, along with some novel interaction data, are presented in chapter 5. Chapter 2 outlines the method we utilize in our laboratory to characterize nuclear protein-protein interactions. Chapter 6 is a general perspective on chromatin regulatory complexes that were originally characterized as repressors.

In the following section, various questions raised by the work described in this thesis and several general points regarding the epigenetics field are discussed.

Polycomb targeting

As introduced in chapter 1, PcG and TrxG proteins are instrumental in maintaining the repressed or active state of chromatin respectively[2–4]. Both groups represent epigenetic regulators as they lay down the chromatin marks that have come to define active and repressed states. However, how these complexes are recruited to their tissue-specific targets remains an unanswered and critical question. In *Drosophila melanogaster*, the finding that there are Polycomb Response Elements (PREs), specific sequences with which some PcG proteins interact (via TFs), was critical as it presented a mechanism for sequence-specific targeting[5]. Interestingly, TrxG proteins are also found to be associated with some PREs, invoking the question of what mediates specificity at these shared PRE/TRE elements[6].

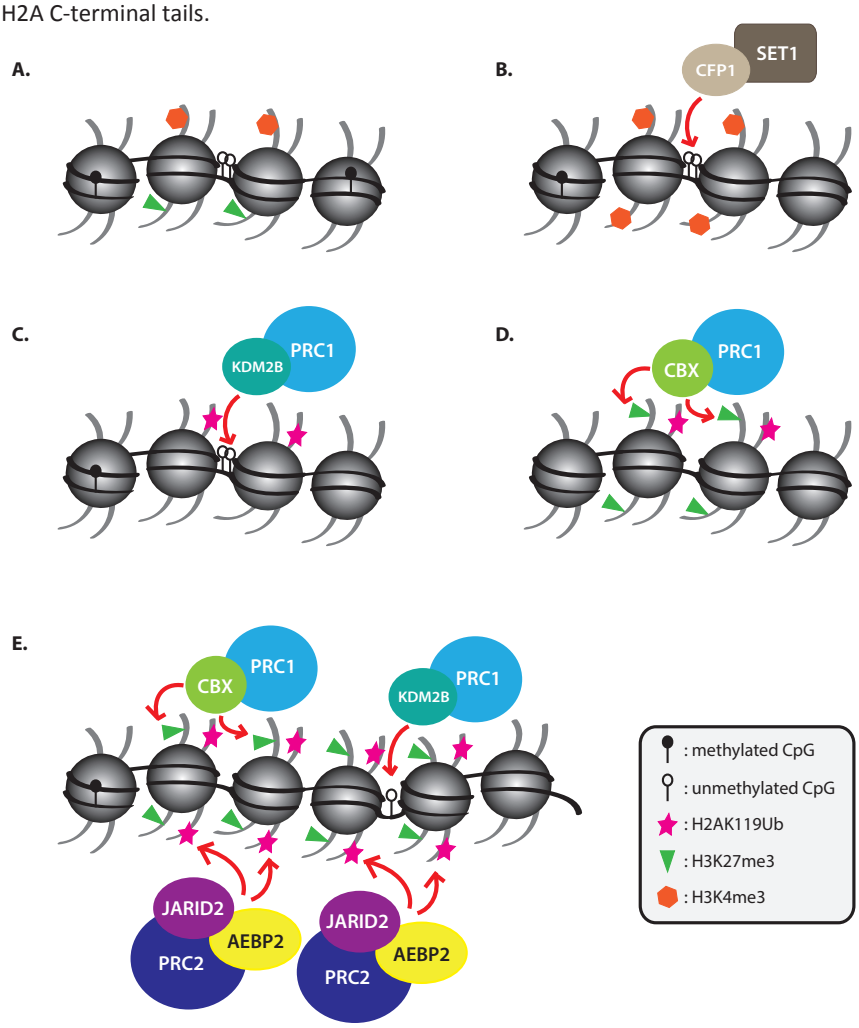
In the mammalian system, the number of PcG complex core subunits and interactors increase substantially compared to *Drosophila melanogaster* (see **table 1 and 2** in chapter 1). In addition, PREs appear to be absent in mammals (with a few exceptions[7]). Instead, specific interactions between PcGs and various long non-coding RNAs have been observed in addition to interactions with some TFs[3]. Therefore, in the mammalian system, targeting could be mediated by various interactors (especially if the expression of the interactor is tissue-specific) and long non-coding RNAs.

This hypothesis holds true for some targeting events in the genome, but it seems insufficient to explain all the loci that PcGs are found at that are revealed by genome-wide assays. However, taking another factor that is characteristic of mammalian genomes into account provides a lot of insight: ~60% of the mammalian promoters overlap with CpG islands (CpGIs). It was first proposed in 2008 that CpGIs could be the mammalian equivalent of PREs[8]. However, for PcG recruitment, the absence of any activating factor is a pre-requisite. In line with this observation, subsequent studies utilizing ectopic CpGI-like sequences inserted into the mESC genome have shown recruitment of both TrxG (CFP1) and PcG (PRC2) proteins to CpGIs[9,10]. Based on these studies, an alternative hypothesis is now emerging in the field which envisions the targeting of PcGs (as wells TrxGs) as 'responding' to chromatin environment as opposed to 'instructing' it[11,12]. The epigenetic regulators are hypothesized to 'sample' chromatin via specific interactors, and depending on the pre-existing chromatin state (activating or repressing marks and/or factors), their recruitment is stabilized or lost. This sampling is, as exemplified so far, mainly driven by CpG-rich promoters and the CXXC zinc finger-containing proteins that interact with unmethylated CpGs: KDM2B and CFP1 both have a CXXC zinc finger domain. KDM2B interacts with and recruits variant PRC1 complexes to unmethylated CpGs[13] and CFP1 is a subunit of the SET1 complex[4]. Since neither PRC1 nor SET1 is found at all unmethylated CpGIs, what seems to be determining whether their recruitment is maintained or not is the chromatin environment. For example, the presence of active transcriptional machinery could negatively affect PRC1 recruitment and result in its dissociation whereas a dense nucleosome environment may favor PRC1 activity and recruitment (as shown for PRC2[14]). On the other hand, for TrxG proteins, CFP1 could be sampling chromatin for active marks and activating factors to set the stage for activation. This is how 'chromatin sampling' by PcG and TrxG proteins may be mediated via CXXC-bearing interactors and complex subunits (**figure 1**).

It is important to note that the presence of factors on chromatin that may positively or negatively affect PcG recruitment are not necessarily mutually exclusive. The best examples are the so-called bivalent chromatin domains in which H3K4me3 co-localizes with H3K27me3[15]. Bivalent domains, found most predominantly in mESCs, are thought to prime genes for activation and most bivalent domains are resolved (i.e. either activated or repressed with the corresponding mark taking over) upon differentiation. While it was shown that the presence of H3K4me3 hinders PRC2 activity on the same histone H3 molecule[16], the existence of bivalent marks was proven by showing that nucleosomes can be asymmetrically modified: One Histone H3 molecule bearing the active mark and the other H3 within the same nucleosome carrying the repressive mark[17]. At bivalent domains, the predominance of the either histone mark and the proteins it recruits may be effective in tipping the balance towards activation versus repression. For example, H3K4me3 could recruit TFIID which would facilitate the recruitment of more TrxG proteins and amplify the H3K4me3 signal, which would recruit more TFIID and promote PIC formation[18,19]. Conversely, H3K27me3 may recruit PRC1 which monoubiquitinates Histone H2AK119[20–22]. Indeed, the PRC1 subunit RING1B is found at a subset of bivalent domains[8]. In chapter 4 of this thesis along with some recent reports[23,24], data that show H2AK119Ub recruiting PRC2 have been presented which form the basis of a mutual positive feedback loop between PRC1 and PRC2, facilitating the amplification of repressive signals from both sides by both complexes.

An integration of various genome-wide data sets has revealed that H3K4me3 is found mostly on CpG-promoters[8,25,26]. Co-occurence of H3K4me3 and H3K27me3 was exclusively observed on CpG-promoters in pluripotent cells (i.e. mESCs and iPSCs), an observation

Figure 1: Chromatin environment affects the recruitment of epigenetic regulators. **Panel A** shows the co-existence of H3K4me3 and H3K27me3 on the same nucleosome, hence a bivalent region. **Panel B** shows the recruitment of the SET1 complex via an interaction between CFP1 and unmethylated CpGs. **Panel C** shows the recruitment of PRC1 by KDM2B, also interacting with unmethylated CpGs. **Panel D** shows the recruitment of the PRC1 complex via an interaction between CBX proteins and H3K27me3. **Panel E** shows Polycomb domain formation by the mechanisms described in panel C and D, in addition to the recruitment of PRC2 via an interaction with H2AK119Ub. Note that the histone tails depicted in this figure are only H3 N-terminal tails and H2A C-terminal tails.



that links bivalent marks and chromatin sampling mediated by proteins with CXXC domains. Therefore, how chromatin may be 'sampled' by PRC1 and SET1 complexes at nonCpG promoters remains to be determined. In addition, while PRC2 recruitment to CpG promoters has been reported[8,10,26,27], a detailed mechanism with an intermediary interactor or subunit is not characterized yet for the PRC2 complex (like KDM2B and PRC1). At the moment, three components of chromatin sampling have been identified: unmethylated CpG promoters,

CXXC domain-containing proteins and the complexes that they recruit (PRC1 and SET1). Different contexts involving PRC2 and CpG promoters in addition to all the complexes (PcGs and TrxGs) at nonCpG promoters need to be investigated further. Interestingly, despite the conservation of PcG and TrxG proteins between *Drosophila melanogaster* and vertebrates, the *Drosophila melanogaster* genome lacks DNA methylation and hence CpG islands. In that case, sampling may be taking place at the PREs and other promoter elements. It is expected that different promoter elements and the factors that they recruit direct chromatin sampling at nonCpG promoters in the mammalian system as well.

In the DNA methylation section in chapter 1, a different interpretation of a subset of DNA methylation events was introduced as 'responding' to TF occupancy as opposed to instructing transcription[28]. Chromatin modifiers that put on marks associated with activation and transcription are also hypothesized to sample chromatin as discussed above[11,12]. Depending on the chromatin environment (i.e. marks and/or other proteins already present) their recruitment is stabilized or lost. This hypothesis presenting epigenetic factors as responding to chromatin environment and then proceeding to 'lock' the pre-existing state (and possibly amplify activating/repressing signals to fulfill some threshold requirements for both processes) seems to solve the specificity problem inherent in the hypothesis in which epigenetic factors instruct transcription. The emerging picture of transcriptional regulation is that the TFs lay down the initial regulatory signals, which are then amplified and maintained by the epigenetic machinery.

A closer look at the interplay between PRC1 and PRC2

As mentioned above, before the recruitment of PRC1 to CpGIs via KDM2B was characterized in detail in 2012, multiple studies reported PRC2 recruitment to CpGIs in the absence of active marks and/or activating factors[8,10,26]. The finding that the CBX proteins of PRC1 have an affinity for the H3K27me3 mark catalyzed by PRC2 precedes this observation[20–22]. In light of all this information, the hierarchical model for Polycomb recruitment is straightforward: PRC2 is recruited to target sites in the genome by 1. specific interactions with certain TFs or 2. long noncoding RNAs or 3. by the genetic component that are the CpGIs. Once PRC2 is recruited, it catalyzes the methylation of Histone H3K27. This mark serves as a recruiter of both PRC1 (via CBX proteins) and more PRC2 (via EED)[29,30], and more H3K27 methylation in addition to monoubiquitination of Histone H2AK119 is catalyzed. This model offers a stepwise explanation for how the two major PcG complexes are recruited to chromatin.

However, genome-wide binding profiles for PRC1 and PRC2 contradict this model in that the sites the complexes occupy do not fully overlap[31,32]. H2AK119Ub levels in the absence of PRC2 persist[32–34]. We and others have shown the hierarchical model could in fact be a mutual recruitment model in that the presence of H2AK119Ub engages PRC2 just as H3K27me3 engages PRC1 (chapter 4,[23,24]). The latest findings are interesting in this sense and reveal an extra layer of regulation of PcG recruitment. However, it should be kept in mind that, just as the H2AK119Ub mark is still present in the absence of PRC2 (and H3K27me3), H3K27me3 levels are not totally disrupted in Ring1 knock-out mESCs that lack H2AK119Ub[34,35]. Therefore, Histone H2AK119Ub serving as a recruiter for PRC2 seems to have regulatory symmetry with Histone H3K27me3 recruiting PRC1: Both mechanisms are at play at a subset of PRC1 and PRC2 targets. It may well be that the affinity of one complex for the mark laid by the other serves as another example of chromatin sampling by both complexes and may be effective mostly in the formation of Polycomb domains, but it is not the means of all targeting events of the either complex. That is to say, in addition to

the target genes they repress in a coordinated manner[31,34], PRC1 and PRC2 have targets on chromatin that they occupy independently of one other. For these 'individual' target sites, the means of recruitment could involve all the aforementioned targeting mechanisms other than recognizing the chromatin mark catalyzed by the other PcG complex.

Another important aspect to consider is the difference between establishing and maintaining the Polycomb domains. The positive feedback loop the two PRC complexes generate for each other with the histone marks they catalyze may be critical in establishing Polycomb domains in the genome at a specific developmental window. Maintaining these domains in successive cell divisions may not require the agonistic crosstalk as vigorously.

Another level of complexity is presented by different complex compositions. Clearly, the PRC1 complexes that can be recruited by H3K27me3 are those that contain CBX proteins. For PRC2, our *in vitro* interaction screen suggests AEBP2 and JARID2, known PRC2 interactors, as the key factors in H2AK119Ub recognition. Therefore, the PRC2 that could be recruited by H2AK119Ub could be a subset of all PRC2 that stably interacts with JARID2 and AEBP2. Very recently, the activity of PRC2 was shown to be stimulated by the methylation of JARID2 by PRC2, which is an intriguing instance of a complex post-translationally modifying an interactor to modulate its own activity[36].

An intriguing chromatin mark: Histone H2A(K119) monoubiquitination

As introduced in chapter 1, H2A monoubiquitination is associated with several processes: transcriptional regulation, cell cycle progression and the DNA damage response.

In the context of transcriptional regulation,

H2AK119Ub is catalyzed by the PRC1 complex and mediates gene silencing[37,38]. Monoubiquitination of H2A catalyzed by BRCA1 mediates the silencing of microsatellite repeats in the genome[39]; however the ubiquitination sites on Histone H2A by BRCA1 are on different lysines (K127 and K129)[40]. It would be of interest to know whether the different functional associations observed for some PTMs depending on the site of the modification on the same histone molecule (for example methylation (15 daltons x2/3 for me2/3) on Histone H3K4 vs H3K9) are also observed for a PTM as bulky as ubiquitination (~7 000 daltons).

There are a few reports showing that H2AK119Ub mediated by PRC1 or another E3 ligase, 2A-HUB, function in blocking the elongation of RNA Pol II. The catalytic subunit of PRC1, RING1B, was shown to be present in bivalent promoters in mESCs[8]. It was proposed that the RNA Pol II machinery present at these promoters is kept from elongating due to the activity of the PRC1 complex[41]. H2AK119Ub catalyzed by 2A-HUB was shown to be critical in repressing transcriptional elongation at a subset of chemokine genes in macrophages by inhibiting FACT recruitment[42]. The FACT complex facilitates transcriptional elongation via displacing nucleosomes[43]. As various genome-wide assays are generated, also for profiling paused versus elongating RNA Pol II, it would be interesting to see whether most instances of repression linked to H2AK119Ub are cases of RNA Pol II arrest at promoter proximal sites. At least a subset of PRC1-mediated repression events could indeed be the inhibition of elongation as there is a mild to moderate enrichment of RPII18 (a subunit of all RNA polymerases found in *Drosophila melanogaster*) in the chromatin fraction defined by the presence of PcGs in the model organism *Drosophila melanogaster*[44].

With the characterization of the PR-Dub complex in *Drosophila melanogaster*, a surprising finding was that removal of H2AK118Ub seemed necessary for transcriptional repression[45]. At first it was not exactly clear how PR-Dub contributed to repression: it was one of the

possibilities discussed in the original paper that the repression of the hox genes may be mediated by a temporal ubiquitination-deubiquitination cycle mediated by PRC1 and PR-Dub[45]. A follow-up paper from the same group further classified target genes of PRC1: The genes that required the H2A ubiquitin ligase activity of Sce (D.m. homologue of Ring1b) were called Class I genes and those that did not need H2A ubiquitination for repression were called Class II[46]. PR-Dub-mediated deubiquitination was necessary to ensure repression of Class I genes. A possible scenario is that the PR-Dub may be removing ubiquitin from non-target sites (Class II) to provide enough ubiquitin molecules to be utilized by PRC1 for the repression of Class I genes. Alternatively, a specific H2AK118Ub-binding repressor could be sequestered upon aberrant Histone H2A ubiquitination. PR-Dub, via deubiquitinating non-target sites, could be restoring this unknown repressor levels to be re-directed to correct target sites (Class I). In both of these scenarios, it seems that the PR-Dub complex is functioning to correct for targeting defects in play regarding PRC1 H2A ubiquitin ligase activity (which is only required at Class I targets). However, how such specific targeting of PR-Dub activity to Class II H2AK118Ub would be mediated is not known as genome-wide binding profiles for Sce and Calypso (D.m. Bap1) overlap at both Class I and II target genes. Other possible scenarios are that the H2AUb levels need to be tightly regulated locally for accurate repression and PR-Dub may be adjusting the concentration of H2AUb levels at specific loci. Or, as mentioned above, PR-Dub and PRC1 temporally regulate H2AK118Ubn levels. For a more detailed discussion of these scenarios, see Scheuermann et al[47]. Characterizing the exact nature of the repression mediated by H2AUb will possibly shed more light on PRC1/PR-Dub dynamics.

The mammalian PR-Dub activity on Histone H2A has been reported to result in an increase in local H2AK119Ub levels, but the exact effect of this increase on transcription was not clear[48]. In another study, H2A deubiquitination was reported to repress gene transcription, but in a PRC1-dependent manner[49]. That is, the derepression observed in BAP1 knock-down is lost in the absence of RING1b-BMI E3 ligases for H2AK119. This study claims that RING1-BMI is mediating gene activation and under normal conditions PR-Dub is antagonizing this activation effect. However, this observation of RING1 depletion compensating the upregulation observed upon BAP1 depletion was shown for 3 target genes (of both RING1b and BAP1). Therefore whether the proposed interplay between RING1 and BAP1 is valid at more sites or whether this is an isolated observation remains to be seen.

In the context of DNA damage,

H2A mono and polyubiquitination on various lysines have been reported to recruit DDR proteins and promote foci formation[50]. All deubiquitinases of H2A studied in this context, expectedly, were reported to negatively affect foci formation with the exception of BAP1[51]. However, a recent screen showed that depletion of 2A-DUB and BRCC36 resulted in DSB repair defects as well, suggesting that these two Dubs of H2AUb are also promoting DSB repair like BAP1[52]. BAP1 has been reported to localize to DNA damage sites by multiple studies and its presence positively affects foci formation[53,54]. Some possible explanations for these unexpected observations could be: 1. These three Dubs deubiquitinate H2AK119Ub on nucleosomes that are 'less proximal' to the damage site to supply ubiquitin for nucleosomes closer to the damage site. 2. Deubiquitination of H2AK119Ub facilitates nucleosome eviction to promote the recruitment of DDR machinery. 3. The Dubs target other DDR proteins and enhance their activity. 4. The presence of the Dubs at damage sites is independent of their deubiquitinase activity and serves another purpose to facilitate the damage response.

In chapter 3, our data showed that MBD6 recruitment to sites of irradiation is independent

of its interaction with BAP1. Whether BAP1 is recruited to irradiated loci via MBD6 is still unclear in our assays either due to BAP1 recruitment being more dynamic than we can capture or the eGFP signal being diffused throughout the nucleus obscuring the signal at sites of damage. An alternative explanation is that BAP1 is not recruited to sites of irradiated DNA in our assays using the HeLa cell line.

Finally, in the context of cell cycle progression,

Deubiquitinases for H2A monoubiquitin that are implicated in cell cycle regulation (USP16, USP3 and USP22) are required to make sure H2A is ubiquitin-free for chromosome compaction prior to metaphase[51,55]. Although the effect of BAP1 on the cell cycle has been studied by multiple groups due to the tumor suppressor function of BAP1[56–58], it has not been shown to effect the progression of the cell cycle in this context.

The mammalian PR-DUB: subunits and activities beyond deubiquitination of H2AK119Ub

The catalytic subunit of the mammalian PR-Dub complex, BAP1, has evolved to include an HCFC1-binding domain. HCFC1 is a multifunctional regulatory protein, interacting with various TFs and TrxG proteins, and reported to regulate the progression of the cell cycle[59,60]. It is reported to be deubiquitinated by BAP1, although whether this deubiquitination affects the stability of HCFC1 is not clear[56,57]. Interestingly, HCFC1 needs to undergo proteolysis to reach maturation. O-linked beta-N-acetylglucosamination (O-GlcNAcylation) by OGT is essential for proteolysis[61,62] and OGT activity was reported to be regulated by BAP1 deubiquitinating OGT[63]. This complex network of interactions and the addition/removal of PTMs between the 3 proteins need to be clarified in more detail (**figure 2**).

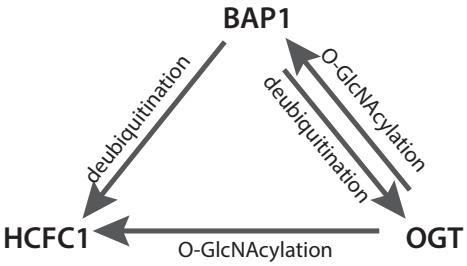


Figure 2: The regulatory circuit within the PR-Dub complex. Reported enzymatic activities of PR-Dub subunits on one another are shown. Crosstalk or a possible sequential relationship between these activities is not yet elucidated.

OGT is an enzyme of increasing popularity due to the emerging findings that O-GlcNAc seems to be a rather abundant modification on chromatin and other proteins (chromatin associated and otherwise)[64,65]. As mentioned above, at least in the case of HCFC1, the regulatory consequence of O-GlcNAcylation is remarkable. In addition, O-GlcNAcylation metabolism in cancer cells is deregulated[66]. Therefore, O-GlcNAc and the protein responsible for the modification (OGT) are of interest in cancer epigenetics[67].

More links to cancer for PR-Dub subunits are provided by BAP1 and ASXL1. As mentioned in chapter 1, BAP1 and ASXL1 are characterized as driver genes of cancer in a study that screened mutations in a number of different malignancies[68]. Both proteins are categorized as tumor suppressors. ASXL1 is mostly mutated in myeloid malignancies as opposed to various solid cancers in which BAP1 is mutated[69–71].

The mammalian PR-DUB complex has, in addition to BAP1 and OGT, one more chromatin modifier that is KDM1B (this thesis, [57,58,63]). The catalytic activity of this lysine demethylase has been shown to be directed to H3K4 mono- and dimethylation, chromatin marks found at enhancers and active transcription units[72,73]. It could be that KDM1B is demethylating this

mark at sites where BAP1 is mediating H2AK119Ub deubiquitination. Additionally, KDM1B has a CW-type zinc finger and a SWIRM domain, and both were shown to be required for the demethylase activity[74]. Whether these domains also mediate the recruitment of PR-Dub to chromatin will be interesting to find out. Another possibility is that the presence of KDM1B within the PR-Dub complex is required to act on other complex members and not on chromatin at target sites (or, possibly, it does both).

In total, there are two erasers (BAP1, KDM1B) and one writer (OGT) as chromatin modifiers in the PR-Dub complex. BAP1 is a tumor suppressor, along with another protein within the complex that does not have any enzymatic activity (that we know of), ASXL1. An important question to address is to what extent the reported tumor suppressor activities of BAP1 and ASXL1 and the putative oncogenic link of OGT are within the PR-Dub complex context. If a remarkable extent of subunit activities is conducted within the complex, targeting of only one protein could rescue the consequences of deregulated PR-Dub activity. If all/some of the proteins have complex-independent functions, individual targeting of enzymatic activities may be required. Either way, the cancer-related and enzymatic activity–bearing subunits of the PR-Dub complex present possibilities to extend the scope of epigenetic drugs.

Novel PR-Dub subunits/interactors: MBD5 and MBD6

It is of note that the neurological disorders associated with mutations in the *Mbd5* gene and microdeletion/duplication of the loci encompassing *Mbd5*, are all neurodevelopmental disorders, i.e. they are not the neurological illnesses that surface later in the life of an adult[75–77]. This observation is indicative of the fact that the neural differentiation program is compromised in the absence of (or in the dysfunctioning of) MBD5. Polycomb complexes are, as mentioned before, essential for the correct setting of differentiation. The fly homologue of MBD5, sba, was not found to interact with the fly PR-Dub complex([45], Jürg Müller personal communication). Therefore, any study looking into the potential involvement of PR-Dub in the Mbd5-related neurological defects should be designed using a mammalian system. Interestingly, knocking out the *Mbd5* gene in a mammalian system, i.e. mice, resulted not only in neurodevelopmental defects, but also in disturbed glucose homeostasis[78]. Therefore, tissue-specific interactors and activities of MBD5 are of particular interest (see next section).

As mentioned in chapter 1, there is one report showing the expression profile of Mbd5 and Mbd6 in various mouse tissues[79]. The second isoform of MBD5 is highly expressed in oocytes. This finding is intriguing in light of the study that reports KDM1B is functionally linked to maintaining the maternal imprints in oocytes[80]. It would be interesting to see if this function of KDM1B is within the context of the PR-Dub complex and whether and how MBD5 contributes to the maintenance of maternal imprints.

MBD6 was observed to be highly expressed in testes and brain. Although the heterogeneity of these tissues in the expression profiling should be taken into account, it is curious that MBD6 is expressed highly in the brain considering that we show it localizes to DSBs in chapter 3 in this thesis. The DDR machinery is most often studied in proliferating cells and it is intricately linked with cell cycle checkpoints (G1, G2, mitotic checkpoints) which are not active in postmitotic neurons (G0)[81]. All DNA damage response proteins and pathways are nonetheless present in postmitotic cells, albeit in lower levels[81]. However, the exact mechanisms of DDR are not as clear due to lack of data. Upon DNA damage, re-entry into G1 and S phases has been shown for postmitotic neurons, but no division follows the S phase entry. It seems the neurons are more likely to undergo apoptosis if the re-entry to G1 extends into the S phase as well[82,83]. It would be of interest to find out how MBD6 contributes to the

DDR and whether the PR-Dub interaction is involved: Although we have convincingly shown that the recruitment of MBD6 to DNA damage sites is PR-Dub-independent, we still do not know if MBD6 recruits PR-Dub to sites of damage. There are reports in literature claiming that PR-Dub localizes to IR sites, but our preliminary results are inconclusive in this aspect. Another possibility to consider is that MBD6 may have other functional roles that have not been characterized yet.

Challenges of studying multisubunit complexes

With high quality interaction proteomics, the subunits and interactors of multisubunit complexes are relatively easy to identify as presented in chapters 3 and 5. Of note, for some proteins in some of the complexes, which protein is a subunit of the complex and which is an interactor of the complex is not an easy distinction to make. To address this question, complex compositions in different cell types could be compared to identify tissue-specific interactors. In addition, if available, information on the relative stoichiometry of proteins within the complex could serve as a criterion for distinguishing subunits from interactors.

Possibly a more pressingly necessary distinction that is required concerns the function of proteins of a complex that are a) in the context of the complex and b) independent of the complex. This is a rather difficult question to tackle. Conventionally, the effect of deleting the domain on the protein of interest mediating the interaction with the complex of interest is observed. However, there is always the possibility that the deletion of a domain disrupts all functionality of the protein. In addition, the domain that mediates interaction with the complex could also be functional in a complex-independent context. Therefore, while this experimental set-up does provide some insight into the possible complex-independent functions of proteins, it has its caveats. For chromatin regulatory complexes, the overlap of binding profiles of various subunits on chromatin is informative in that if there's overlap between the subunits, it is highly likely that those sites are targeted by the complex of interest and not by particular subunits individually. The most informative method will be to identify all proteins on a locus of interest to see the co-localization of proteins (and hence complex composition as a whole) using methods such as ChIP-ms or derivatives.

Another interesting question is the tissue-specific interactors and functions of chromatin complexes. With the generation of CRISPR-Cas genome targeting, it has become significantly more feasible to manipulate the genomes of higher eukaryotes[84]. What would be an interesting experimental set-up is to knock-out various subunits or interactors of a complex one by one in an inducible manner. The induction of the knock-out could be driven by a tissue-specific promoter. When applied to multiple subunits one by one or in combination, these experiments can shed light on the function of tissue-specific interactors in addition to tissue-specific functions of the chromatin regulatory complexes like those studied in this thesis that are found abundantly in many tissue types and highly conserved among metazoans.

Mass Spectrometry-based Proteomics for Epigenetics

Mass spectrometry-based proteomics technologies have facilitated the investigation of many aspects of chromatin biology in a high-throughput and accurate manner. Particularly the generation of reliable quantification methods has enhanced interaction proteomics in such a way that the specificity of interactions are detected at an unprecedented accuracy. In affinity purification experiments, stretches of DNA or histone peptides bearing a modification of interest can be used as baits along with unmodified versions to serve as control[85,86].

Incubation of modified and unmodified baits with a cellular extract will provide an unbiased screen of all proteins that specifically interact with or are repelled by the modification of interest. While this methodology does not distinguish between direct and indirect interactors, many reader proteins and complexes for various chromatin marks have been discovered in these screens[87,88]. For protein-protein interactions (PPIs), affinity purifications for a tagged protein of interest can be performed along with a nonspecific purification as control and thus the specific interactors of the protein of interest are enriched. In this thesis, a detailed method for identifying PPIs using quantitative interaction proteomics has been described in chapter 2. This method has been applied to various proteins of interest. In addition, various readers and interactors of a chromatin mark, i.e. Histone H2AK119Ub in a nucleosomal setting have been characterized. Identification of novel interactors has provided functional clues for proteins of interest and extra layers of regulation concerning the chromatin mark in question.

Other epigenetic questions of interest are 1. the identification of histone PTMs, either of known PTMs at novel sites or of novel PTMs altogether. 2. Quantification of histone PTMs 3. The combination of PTMs on histones 4. Characterization and quantification of histone variants. Bottom up proteomics could be used to investigate the first and second questions, but the third and fourth questions require middle-down or top-down techniques. To investigate the combination of histone marks, it is obviously necessary to detect PTMs on the same peptide species, which is why middle-down proteomics would be the preferred method to look into this question[87,88]. Histone variants are rather difficult to quantify using bottom up proteomics due to the number of shared peptides: peptides that could be used to distinguish between variants and for quantification of variants decrease in number if the protein sequences are very similar, which is the case for histone variants. This is why top-down proteomics is more suitable to study histone variant dynamics[91–94]. However, compared to bottom-up proteomics, both middle-down and top-down techniques have remarkable difficulties regarding all steps of mass spectrometry-based proteomics, i.e. sample preparation, ionization and detection, and data analysis[95,96]. Therefore, further advances in middle-down and top-down techniques are very valuable for studying the main protein component of the chromatin environment that is histones.

A number of recent methods have been generated in the field to characterize and quantify the protein content on certain loci in the genome[97–100]. These methods are groundbreaking in the sense that they give a snapshot of the *in vivo* situation at a locus with all the associated proteins. At this stage, the technical challenges limit the use of these methods for loci in single cells. However, even if it may be only at the population level, it is rather exciting to think that soon we may be able visualize the whole chromatin landscape with all the protein components at a database like ENSEMBL (for visualizing genes and transcripts) in different cell types and under different conditions.

Another aspect of chromatin biology that should be considered is the PTMs found on chromatin regulatory complexes. Although some PTMs are reported on individual members of various chromatin complexes, a combinatorial PTM screening is yet to be reported for any complex. If and how different PTMs affect complex formation and function is an area that has not been explored yet due to technical challenges. It is up to the mass spectrometry-based proteomics field to overcome these challenges and shed light on this area of regulation where possibilities are too high in number to be tackled by any other method than one that is high throughput.

Emerging and evolving concept: Epigenetics, controversy and more controversy

Epigenetics refers to all heritable information stored on chromatin that is not contained within the DNA sequence. Depending on how conservatively interpreted, it could refer to information that is passed over cell divisions and/or across generations - or not passed on at all (which in fact goes against the definition of the term). Originally, the term was conceived by Conrad Waddington in 1942 to comprise the processes of genotype interpretation leading to a specific phenotype[101]. The implication of heredity (across generations as well as cell divisions) was linked to the term later on, especially in light of DNA methylation studies[102]. In 2000, Strahl and Allis suggested the term 'the histone code' to refer to the information stored in chromatin in the form of specific combinations of histone PTMs[103]. They claimed that the histone code is instrumental in phenotype formation. In a review in 2001, Jenuwein and Allis further postulated that the histone code is, just like DNA methylation, a very important player in epigenetic mechanisms and it is, like DNA methylation, (most likely) heritable across cell divisions[104]. In this review, the only mention of transgenerational inheritance of the histone code is in the context of imprinted genes: The authors propose that the imprinted regions may harbor histone PTMs in addition to methylated DNA, forming another combination (code) to make sure imprinted regions are faithfully transmitted across generations. A rather bold proposition of the review is *'that a "histone code" exists that may considerably extend the information potential of the genetic (DNA) code.'* Eventually, the strong implication that epigenetics presents other levels of information as opposed to presenting mechanisms of interpretation of the genetic information raised a lot of controversy in the field[105,106], especially since the hereditary aspects (across cell divisions and generations) of these marks were rather unknown. In 2007, Adrian Bird described epigenetics as 'the structural adaptation of chromosomal regions so as to register, signal, or perpetuate altered activity states'[107]. While this definition accommodates the functional implications and consequences of epigenetic marks, the nature of the 'perpetuation of altered activity states' remains unclear for the most part and most marks. In 2009, a group of scientists proposed what they called an operational definition of epigenetics as: "An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence'' and specified that the inheritance could be via mitosis or meiosis[108].

It is important to make distinctions between the three phenomena that are claimed, at one point or another, to be covered by epigenetics. 1. Does epigenetics refer to all the information on top of DNA that enable/disable a certain DNA-templated process? (e.g. acetylation of promoter nucleosomes before transcription.) 2. Does epigenetics refer to all the information mentioned in (1) only when it is stably maintained across cell divisions? 3. Does epigenetics refer to information mentioned in (1) only when it is inherited across generations?

From the most conservative perspective, if epigenetics, by definition, stands to include all 3 factors mentioned above, then it needs to perpetuate the altered activity state in the absence of the initial, causative signal across generations. Examples fulfilling this criterion, (i.e. transgenerational inheritance), have been well-documented in plants and *Caenorhabditis elegans* in mechanisms involving DNA methylation and long noncoding RNAs[109,110]. However, in mammals, these examples are refrained to the case of imprinted genes[111]. Less conservative interpretations of the term epigenetics (i.e. definition (1) above) encompass all changes on chromatin induced by intra or extracellular signaling pathways by cis or trans-acting mechanisms regardless of inheritance (intercellular or intergenerational). Even within this limited description, there has been some discussion calling into question

the cause versus consequence relationship between epigenetic marks and regulatory processes. This confusion mostly stems from equating the presence of an epigenetic mark to functional outcomes without accounting for other regulatory factors such as TFs and the regulatory loops that they initiate.

Transcription factors (TFs) are the undoubted master regulators of transcription. The literature is full of examples where the expression of transcription factors overrules the effect of chromatin marks[112,113,114]. One of the best and possibly the most revolutionary example of this phenomenon is the ectopic expression of OCT4, SOX2, KLF4 and MYC TFs being able to reprogram differentiated cells back into a pluripotent state[115]. In addition, TFs (by definition) have specific DNA sequences that they recognize and bind which confers them specificity – a property essential in regulation and lacking in chromatin regulatory proteins. Chromatin remodeler and modifiers depend on TFs for recruitment for their activities to be targeted.

However, numerous elegant studies from model organisms are accumulating in literature, emphasizing the importance of epigenetic mechanisms in establishing various phenotypes. For example, in honeybees, changing DNA methylation patterns affected the developmental fate of bees in becoming a worker versus queen bee[116]. In *Drosophila melanogaster*, a particular histone mark (H3K27me3) was shown to be responsible for all the repression mediated by Ez2 as the mutation of the histone perpetuated the Ez2 knock-out phenotype[117]. Conversely, another study in *D. melanogaster* resulted in viable flies when total Histone H3 in the organism was replaced with a mutant that cannot be methylated at the K4 position[118]. The mutant flies did display some developmental defects and in fact this observation could be used by both sides of the argument: One claiming that the mark is not that critical as the organism survives, the other side claiming that the mark is obviously functional due to the observed developmental difference. Such functional studies are yet to be performed in higher eukaryotes.

With increased emphasis on epigenetics, the Roadmap Epigenomics Project was launched in 2008 with the aim of mapping epigenetic marks in different human tissues. Some scientists expressed their concern about such ambitious projects being misdirected in their huge invesment in epigenetic information[119]. They felt that the mechanisms of gene regulation undertaken by TFs and the feedback loops that they generate (mechanisms already well-described decades ago) are being overlooked in an attempt of better characterizing the developmental expression profile regulation.

In my opinion, taking the argument to extremes is doing disservice to both sides of the discussion. Clearly, epigenetic mechanisms cannot be considered independently of the underlying DNA sequence and the regulatory circuits that are directed by sequence-specific elements. However, some chromatin marks are also emerging to be instrumental in regulatory pathways initiated by TFs. How they are put on chromatin, how they are interpreted and how specificity is mediated are questions that need to be answered in more detail. These answers will not devalue the already elucidated gene regulatory mechanisms, they will only reveal more mechanisms of how the genetic material is interpreted at different loci and during different developmental phases. As functional studies on chromatin marks increase in number and include more model organisms, we will appreciate better which marks are critical, to what extent exactly and which are just circumstantial to biological processes. I do expect more causal evidence to be discovered on the epigenetics side - when the term is considered within the sense that Waddington proposed. However, in the sense that includes transgenerational inheritance, I do expect less likeness to mechanisms observed in lower eukaryotic model organisms considering the waves of ‘resetting’ chromatin in the germline

of mammals[102,111,120].

As for the proposition of Jenuwein and Allis that a combinatorial histone code drives differentiation and maybe even exists alongside DNA methylation in imprinted regions across generations: they were right that numerous histone marks have been discovered and strongly linked to certain biological processes and outcomes. However, insofar as we know, the combinatorial complexity is not at all high considering the mathematical possibilities[121]. More accurately, the combinatorial complexity has been shown to be high to a certain extent, but the functional implications remain unresolved for the co-occurence of more than, approximately, 3 marks[89]. That is to say, the functional difference –if there is one- between H3K9me2K14acK23acK27me3 and H3K9me1K14acK23acK27me3 is not yet clear. Jenuwein and Allis were right that imprinted regions do contain some certain histone marks (H3K9me3) that seem to recruit certain reader proteins in mammals[122]. On the other hand, the mechanisms involved in the propagation of histone marks, across mitosis and meiosis, are not fully elucidated yet and remain an intense area of investigation[123–125].

On a more clinical note, many epigenetic regulators are dysregulated in cancer[68]. Before the exact (causative) functions of the epigenetic regulators are described, multiple epigenetic drugs have passed clinical trials and are in use today. Some examples that act on epigenetic writers are histone deacetylase inhibitors, DNA methyltransferase inhibitors, histone methyltransferase inhibitors and drugs that block epigenetic readers such as Bromo-domain inhibitors[126–130]. From the perspective of molecular biology, this development is bizarre in that only the global effect of these drugs is known but not exactly the individual targets and the specific outcomes at those sites. From a medical perspective, it is quite usual that the exact molecular mechanisms of drugs in use are unknown.

Interestingly, or rather expectedly, the controversy surrounding how much regulation could be ascribed to epigenetic information as well as how heritable it is has spread to the field of evolution: In the October 8 2014 issue of the journal Nature, a comment article containing opposing arguments from 2 different camps of scientists was published as the pinnacle of this disagreement[131]. One group proposed that in light of all the research of the past half-century, the standard evolutionary theory (SET), with its 'gene-centric focus', needs a make over which they called the extended evolutionary synthesis (EES). The additional factors within EES are listed as: 'how physical development influences the generation of variation (developmental bias); how the environment directly shapes organisms' traits (plasticity); how organisms modify environments (niche construction); and how organisms transmit more than genes across generations (extra-genetic inheritance).' Three out of these listed four factors (developmental bias, plasticity and extra-genetic inheritance) are -at least partially- epigenetic phenomena. The opposing group of scientists acknowledged the importance of all the listed phenomena in evolution and claimed that all of these are already studied intensively as part of the SET. They refused to treat these phenomena as newly emerging revolutionary mechanisms. While recognizing their value and contribution to evolution, they cautioned against putting these factors on par with the most powerful force in evolution that is genetic change.

Concluding Remarks

Many gene regulatory mechanisms have been characterized in detail in the past decades. The role chromatin plays in transcriptional regulation has been revealed to a large extent. However, many questions remain, especially regarding the combinatorial activity of transcriptional and epigenetic regulators and the dynamics of the *in vivo* situation in cells. More advances in genomics and proteomics technologies and an integration of the two areas will provide more answers.

References

1. Crow, J. F., Dove, W. F. & Zacharias, H. Anecdotal , Historical And Critical Commentaries on Genetics. (1995).
2. Simon, J. a. & Kingston, R. E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* **49,** 808–824 (2013).
3. Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* **14,** 853–64 (2013).
4. Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81,** 65–95 (2012).
5. Steffen, P. a & Ringrose, L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* **15,** 340–56 (2014).
6. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* **20,** 2041–2054 (2006).
7. Sing, A. *et al.* A Vertebrate Polycomb Response Element Governs Segmentation of the Posterior Hindbrain. *Cell* **138,** 885–897 (2009).
8. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4,** (2008).
9. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464,** 1082–1086 (2010).
10. Mendenhall, E. M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* **6,** 1–10 (2010).
11. Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev.* **27,** 1318–1338 (2013).
12. Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P. & Brockdorff, N. Chromatin Sampling-An Emerging Perspective on Targeting Polycomb Repressor Proteins. *PLoS Genet.* **9,** (2013).
13. Farcas, A. M. *et al.* KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *Elife* **2012,** 1–26 (2012).
14. Yuan, W. *et al.* Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. *Science (80-.).* **337,** 971–975 (2012).
15. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).
16. Schmitges, F. W. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* **42,** 330–341 (2011).
17. Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151,** 181–193 (2012).
18. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131,** 58–69 (2007).
19. Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152,** 1021–1036 (2013).
20. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes Dev.* **17,** 1870–1881 (2003).
21. Min, J., Zhang, Y. & Xu, R. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27 Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. 1823–1828 (2003). doi:10.1101/gad.269603
22. Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell* **14,** 637–646 (2004).
23. Cooper, S. *et al.* Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Rep.* **7,** 1456–1470 (2014).
24. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157,** 1445–1459 (2014).

25. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
26. Orlando, D. a., Guenther, M. G., Frampton, G. M. & Young, R. a. CpG island structure and trithorax/polycomb chromatin domains in human cells. *Genomics* **100,** 320–326 (2012).
27. Jermann, P., Hoerner, L., Burger, L. & Schubeler, D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proc. Natl. Acad. Sci.* 1400672111– (2014). doi:10.1073/pnas.1400672111
28. Schübeler, D. Function and information content of DNA methylation. (2015). doi:10.1038/nature14192
29. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10,** 1291–1300 (2008).
30. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461,** 762–767 (2009).
31. Boyer, L. a *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
32. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25,** 3110–3122 (2006).
33. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148,** 664–678 (2012).
34. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24,** 265–276 (2010).
35. Napoles, M. De *et al.* Polycomb Group Proteins Ring1A / B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Proteins* **7,** 663–676 (2004).
36. Sanulli, S. *et al.* Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation Article Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. 769–783 (2015). doi:10.1016/j.molcel.2014.12.020
37. De Napoles, M. *et al.* Polycomb group proteins ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7,** 663–676 (2004).
38. Endoh, M. *et al.* Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet.* **8,** (2012).
39. Zhu, Q. *et al.* BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477,** 179–184 (2011).
40. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. J. & Hiom, K. BRCA1 Is a Histone-H2A-Specific Ubiquitin Ligase. *Cell Rep.* **8,** 999–1005 (2014).
41. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* **9,** 1428–1435 (2007).
42. Zhou, W. *et al.* Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Mol. Cell* **29,** 69–80 (2008).
43. Formosa, T. The role of FACT in making and breaking nucleosomes. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819,** 247–255 (2012).
44. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143,** 212–24 (2010).
45. Scheuermann, J. C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465,** 243–247 (2010).
46. Gutierrez, L. *et al.* The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* **139,** 117–127 (2012).
47. Scheuermann, J. C., Gutiérrez, L. & Müller, J. Histone H2A monoubiquitination and Polycomb repression: the missing pieces of the puzzle. *Fly (Austin).* **6,** 162–168 (2012).
48. Ji, Z. *et al.* The forkhead transcription factor FOXK2 acts as a chromatin targeting factor for the BAP1-containing histone deubiquitinase complex. *Nucleic Acids Res.* **42,** 6232–6242 (2014).
49. Okino, Y., Machida, Y., Frankland-Searby, S. & Machida, Y. J. BRCA1-associated Protein 1

(BAP1) Deubiquitinase Antagonizes the Ubiquitin-mediated Activation of FoxK2 Target Genes. *J. Biol. Chem.* **290,** 1580–1591 (2015).
50. Smeenk, G. & van Attikum, H. The chromatin response to DNA breaks: leaving a mark on genome integrity. *Annu. Rev. Biochem.* **82,** 55–80 (2013).
51. Belle, J. I. & Nijnik, A. H2A-DUBbing the mammalian epigenome: Expanding frontiers for histone H2A deubiquitinating enzymes in cell biology and physiology. *Int. J. Biochem. Cell Biol.* **50,** 161–174 (2014).
52. Nishi, R. *et al.* Systematic characterization of deubiquitylating enzymes for roles in maintaining genome integrity. *Nat. Cell Biol.* (2014). doi:10.1038/ncb3028
53. Yu, H. *et al.* Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 285–90 (2014).
54. Ismail, I. H. *et al.* Germline mutations in BAP1 impair its function in DNA double-strand break repair. *Cancer Res.* **74,** 4282–4294 (2014).
55. Joo, H.-Y. *et al.* Regulation of cell cycle progression and gene expression by H2A deubiquitination. *Nature* **449,** 1068–1072 (2007).
56. Misaghi, S. *et al.* Association of C-terminal ubiquitin hydrolase BRCA1-associated protein 1 with cell cycle regulator host cell factor 1. *Mol. Cell. Biol.* **29,** 2181–2192 (2009).
57. Machida, Y. J., Machida, Y., Vashisht, A. a., Wohlschlegel, J. a. & Dutta, A. The deubiquitinating enzyme BAP1 regulates cell growth via interaction with HCF-1. *J. Biol. Chem.* **284,** 34179–34188 (2009).
58. Yu, H. *et al.* The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell. Biol.* **30,** 5071–5085 (2010).
59. Tyagi, S., Chabes, A. L., Wysocka, J. & Herr, W. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. *Mol. Cell* **27,** 107–119 (2007).
60. Tyagi, S. & Herr, W. E2F1 mediates DNA damage and apoptosis through HCF-1 and the MLL family of histone methyltransferases. *EMBO J.* **28,** 3185–3195 (2009).
61. Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144,** 376–388 (2011).
62. Lazarus, M. B. *et al.* HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* **342,** 1235–9 (2013).
63. Dey, a. *et al.* Loss of the Tumor Suppressor BAP1 Causes Myeloid Transformation. *Science (80-.).* **337,** 1541–1546 (2012).
64. Zhang, S., Roche, K., Nasheuer, H. P. & Lowndes, N. F. Modification of histones by sugar ??-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. *J. Biol. Chem.* **286,** 37483–37495 (2011).
65. Bond, M. R. & Hanover, J. a. A little sugar goes a long way: The cell biology of O-GlcNAc. *J. Cell Biol.* **208,** 869–880 (2015).
66. Slawson, C. & Hart, G. W. O-GlcNAc signalling: implications for cancer cell biology. *Nat. Rev. Cancer* **11,** 678–684 (2011).
67. Forma, E., Jóźwiak, P., Bryś, M. & Krześlak, A. The potential role of O-GlcNAc modification in cancer epigenetics. *Cell. Mol. Biol. Lett.* 1–23 (2014). doi:10.2478/s11658-014-0204-6
68. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339,** 1546–58 (2013).
69. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* **12,** 599–612 (2012).
70. Carbone, M. *et al.* BAP1 and cancer. *Nat. Rev. Cancer* **13,** 153–9 (2013).
71. Abdel-Wahab, O. & Dey, a. The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics. *Leukemia* **27,** 10–15 (2013).
72. Yang, Z. *et al.* AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function. *Cell Res.* **20,** 276–287 (2010).
73. Fang, R. *et al.* Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 Methylation. *Mol. Cell* **39,** 222–233 (2010).

74. Zhang, Q. *et al.* Structure-function analysis reveals a novel mechanism for regulation of histone demethylase LSD2/AOF1/KDM1b. *Cell Res.* **23,** 225–41 (2013).
75. Bonnet, C. *et al.* Extended spectrum of MBD5 mutations in neurodevelopmental disorders. *Eur. J. Hum. Genet.* **21,** 1457–61 (2013).
76. Mullegama, S. V *et al.* Reciprocal deletion and duplication at 2q23.1 indicates a role for MBD5 in autism spectrum disorder. *Eur. J. Hum. Genet.* **22,** 57–63 (2014).
77. Motobayashi, M. *et al.* Neurodevelopmental features in 2q23.1 microdeletion syndrome: Report of a new patient with intractable seizures and review of literature. *Am. J. Med. Genet. Part A* **158 A,** 861–868 (2012).
78. Du, Y. *et al.* The Essential Role of Mbd5 in the Regulation of Somatic Growth and Glucose Homeostasis in Mice. *PLoS One* **7,** 1–10 (2012).
79. Laget, S. *et al.* The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One* **5,** (2010).
80. Lei, H. *et al.* KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. **461,** (2009).
81. Kruman, I. I. *et al.* Cell Cycle Activation Linked to Neuronal Cell Death Initiated by DNA Damage. *Neuron* **41,** 549–561 (2004).
82. Schwartz, E. I. *et al.* Cell cycle activation in postmitotic neurons is essential for DNA repair. *Cell Cycle* **6,** 318–329 (2007).
83. Pan, L., Penney, J. & Tsai, L. Chromatin Regulation of DNA Damage Repair and Genome Integrity in the Central Nervous System. *J. Mol. Biol.* **426,** 3376–3388 (2014).
84. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346,** (2014).
85. Spruijt, C. G., Baymaz, H. I. & Vermeulen, M. Identifying specific protein-DNA interactions using SILAC-based quantitative proteomics. *Methods Mol. Biol.* **977,** 137–157 (2013).
86. Vermeulen, M. Identifying chromatin readers using a SILAC-based histone peptide pull-down approach. *Methods Enzymol.* **512,** 137–160 (2012).
87. Vermeulen, M. *et al.* Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142,** 967–980 (2010).
88. Spruijt, C. G. *et al.* Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
89. Young, N. L. *et al.* High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics* **8,** 2266–2284 (2009).
90. Sidoli, S. *et al.* Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *Proteomics* 2200–2211 (2014). doi:10.1002/pmic.201400084
91. Boyne, M. T., Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. Precise characterization of human histones in the H2A gene family by top down mass spectrometry. *J. Proteome Res.* **5,** 248–253 (2006).
92. Pesavento, J. J., Garcia, B. A., Streeky, J. A., Kelleher, N. L. & Mizzen, C. A. Mild performic acid oxidation enhances chromatographic and top down mass spectrometric analyses of histones. *Mol. Cell. Proteomics* **6,** 1510–1526 (2007).
93. Garcia, B. a, Pesavento, J. J., Mizzen, C. a & Kelleher, N. L. Pervasive combinatorial modification of histone H3 in human cells. *Nat. Methods* **4,** 487–489 (2007).
94. Pesavento, J. J., Bullock, C. R., LeDuc, R. D., Mizzen, C. A. & Kelleher, N. L. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J. Biol. Chem.* **283,** 14927–14937 (2008).
95. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications* **445,** 683–693 (2014).
96. Yuan, Z.-F., Arnaudo, A. M. & Garcia, B. a. Mass spectrometric analysis of histone proteoforms. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* **7,** 113–28 (2014).
97. Déjardin, J. & Kingston, R. E. Purification of Proteins Associated with Specific Genomic Loci.

Cell **136,** 175–186 (2009).
98. Pourfarzad, F. *et al.* Locus-specific proteomics by TChP: Targeted chromatin purification. *Cell Rep.* **4,** 589–600 (2013).
99. Byrum, S. D., Raman, A., Taverna, S. D. & Tackett, A. J. ChAP-MS: A method for identification of proteins and histone posttranslational modifications at a single genomic locus. *Cell Rep.* **2,** 198–205 (2012).
100. Waldrip, Z. J. *et al.* A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* **9,** 37–41 (2014).
101. WADDINGTON, C. H. Canalization of Development and the Inheritance of Acquired Characters. *Nature* **150,** 563–565 (1942).
102. Heard, E. & Martienssen, R. a. Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell* **157,** 95–109 (2014).
103. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).
104. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–1080 (2001).
105. Ptashne, M. Epigenetics: core misconcept. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 7101–3 (2013).
106. Ptashne, M. Faddish stuff: Epigenetics and the inheritance of acquired characteristics. *FASEB J.* **27,** 1–2 (2013).
107. Bird, A. Perceptions of epigenetics. *Nature* **447,** 396–398 (2007).
108. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23,** 781–783 (2009).
109. Jablonka, E. & Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* **84,** 131–176 (2009).
110. Weigel, D. & Colot, V. Epialleles in plant evolution. *Genome Biol.* **13,** 249 (2012).
111. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat. Rev. Genet.* **13,** 153–162 (2012).
112. Charlton, J., John, M. & Bird, A. P. Spl sites inthe mouse. *Genes Dev.* 2282–2292 (1994).
113. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462,** 587–594 (2009).
114. Arnold, P. *et al.* Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res.* **23,** 60–73 (2013).
115. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).
116. Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319,** 1827–1830 (2008).
117. Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A. & Müller, J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* **339,** 698–9 (2013).
118. Hödl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* **22,** 2253–2257 (2012).
119. MADHANI, H. D., NICOLE J. FRANCIS & ROBERT E. KINGSTON, ROGER D. KORNBERG, DANESH MOAZED, GEETA J. NARLIKAR, BARBARA PANNING, K. S. Epigenomics: A roadmap, but to where? *Science (80-.).* **322,** 43–44 (2008).
120. Lee, H. J., Hore, T. a. & Reik, W. Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell* **14,** 710–719 (2014).
121. Rando, O. J. Combinatorial complexity in chromatin structure and function: Revisiting the histone code. *Current Opinion in Genetics and Development* **22,** 148–155 (2012).
122. Smallwood, S. A. & Kelsey, G. De novo DNA methylation: A germ cell perspective. *Trends in Genetics* **28,** 33–42 (2012).
123. Probst, A. V, Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* **10,** 192–206 (2009).
124. Margueron, R. & Reinberg, D. Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* **11,** 285–296 (2010).
125. Campos, E. I., Stafford, J. M. & Reinberg, D. Epigenetic inheritance : histone bookmarks across

generations. *Trends Cell Biol.* **24,** 664–674 (2014).
126. Rodríguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nat. Med.* **17,** 330–9 (2011).
127. Dawson, M. a & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150,** 12–27 (2012).
128. Waldmann, T. & Schneider, R. Targeting histone modifications--epigenetics in cancer. *Curr. Opin. Cell Biol.* **25,** 184–9 (2013).
129. Tsai, H.-C. & Baylin, S. B. Cancer epigenetics: linking basic biology to clinical medicine. *Cell Res.* **21,** 502–17 (2011).
130. Easwaran, H., Tsai, H. C. & Baylin, S. B. Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. *Mol. Cell* **54,** 716–727 (2014).
131. Laland K. and Wray G.A., et al. Does evolutionary theory need a rethink ? *Nature* **514,** 161–164 (2014).

7

Appendix

Summary
Samenvatting
Curriculum Vitae
List of Publications
Acknowledgments

Her gün bir gün geçmeyor.*

Özdemir Asaf, *Mythe*

*It is not one day that passes every day.

Summary

Humans are in the animal kingdom within the eukaryotic domain. In eukaryotes, the genetic material (genome) is stored in a compartment within cells that is called the nucleus. Extremely sophisticated mechanisms have evolved to replicate and interpret the information stored in the genome. The relatively novel branch of science that is Molecular Biology has been addressing these mechanisms for the past century with increasingly more elegant tools and techniques.

Genetic information is utilized via gene expression: Genes are expressed to form functional agents called proteins that perform a plethora of structural and functional roles inside and outside of cells. Proteins can be categorized according to their function and/or structure as well as their location. This thesis focuses on proteins that function within the nucleus, playing roles in controlling gene expression via specific or non-specific interactions with the genome. Proteins very rarely act on their own: Specific and stable protein-protein interactions make up groups of proteins, referred to as protein complexes.

Chapter 1 of this thesis is an introduction to the eukaryotic genome, present in the form of chromatin. Components and several properties of chromatin along with various chromatin-interacting protein complexes are described in detail. A short introduction regarding the technique that is used in the subsequent chapters, mass spectrometry-based proteomics, is also enclosed.

Chapter 2 describes a detailed method for identifying protein-protein interactions. Chapter 3 presents interaction screens for proteins MBD5 and MBD6. The members of the PR-Dub complex, which is a chromatin regulatory complex, are identified as interactors. Further functional data on MBD5 and MBD6 are presented. In Chapter 4, specific interactors of a covalent modification on chromatin are investigated using quantitative proteomics. Preliminary experiments as to the functional consequence of the specificity of some of the interactors are shown. In Chapter 5, the stability and subunit composition of another chromatin regulatory complex, NuRD, are addressed using quantitative and cross-linking proteomics. Chapter 6 is a general discussion on the versatility of several chromatin regulatory complexes that have primarily been shown to repress gene expression. Various technical pitfalls and the necessity of data integration while studying these protein complexes are elaborated on.

Chapter 7 is a general discussion focusing on open questions in relation to the research described in the preceding chapters. A short section on mass spectrometry and the chromatin field, as well as a discussion about the controversy surrounding some of the terminology within the chromatin field, are included.

Samenvatting

De mens wordt geplaatst in het dierenrijk, binnen het domein Eukaryoten. In eukaryoten is het genetische materiaal (het genoom) opgeslagen in een compartiment van de cel dat de kern wordt genoemd. Om de informatie die is opgeslagen in het genoom te repliceren en interpreteren zijn uiterst geavanceerde mechanismen geëvolueerd. De relatief nieuwe tak van wetenschap die Moleculaire Biologie wordt genoemd heeft zich de afgelopen eeuw gericht op deze mechanismen, met alsmaar elegantere instrumenten en technieken.

Genetische informatie wordt gebruikt door middel van genexpressie: genen worden tot expressie gebracht om functionele eenheden (eiwitten genaamd) te vormen, die zowel binnen als buiten de cel een overvloed aan structurele en functionele taken uitvoeren. Eiwitten kunnen worden gecategoriseerd gebaseerd op hun functie, structuur en/of locatie. Dit proefschrift richt zich op eiwitten die binnen de kern functioneren, waar ze een rol spelen in het reguleren van genexpressie via specifieke of aspecifieke interacties met het genoom. Eiwitten werken slechts zelden alleen: specifieke en stabiele eiwit-eiwit interacties vormen groepen van eiwitten, die worden aangeduid als eiwitcomplexen.

Hoofdstuk 1 van dit proefschrift is een introductie in het eukaryotische genoom, in de vorm van chromatine. Componenten en diverse eigenschappen van chromatine worden in detail beschreven, evenals verschillende chromatine-interacterende eiwitcomplexen. Ook wordt een korte introductie gegeven in een techniek die in de daarop volgende hoofdstukken wordt gebruikt; proteomics gebaseerd op massa spectrometrie.

Hoofdstuk 2 beschrijft een gedetailleerde methode voor het identificeren van eiwit-eiwit interacties. Hoofdstuk 3 introduceert interactiescreenings voor de eiwitten MBD5 en MBD6. De onderdelen van het PR-Dub complex, wat een chromatine-regulerend complex is, worden geïdentificeerd als interactoren. Verdere functionele data wat betreft MBD5 en MBD6 worden gepresenteerd. In Hoofdstuk 4 worden specifieke interactoren van een covalente modificatie op chromatine onderzocht met behulp van kwantitatieve proteomics. Voorlopige resultaten van experimenten naar de functionele consequentie van de specificiteit van sommige interactoren worden gepresenteerd. In Hoofdstuk 5 wordt de stabiliteit en compositie van een ander chromatine-regulerend complex, NuRD, onderzocht met behulp van kwantitatieve en crosslinking proteomics. Hoofdstuk 6 is een algemene discussie over de veelzijdigheid van verschillende chromatine-regulerende complexen, waarvan hoofdzakelijk is getoond dat ze genexpressie onderdrukken. Er wordt uitgeweid over verscheidene technische valkuilen en de noodzaak tot de integratie van data bij het bestuderen van deze eiwitcomplexen.

Hoofdstuk 7 is een algemene discussie die zich richt op openstaande vragen met betrekking tot het onderzoek dat is beschreven in de voorafgaande hoofdstukken. Een korte sectie over massa spectrometrie en het chromatine veld is ook opgenomen, evenals een discussie over de controverse rond bepaalde terminologie in het chromatine veld.

Curriculum Vitae

H. İrem Baymaz was born in 1986 in Ankara, Turkey. She studied Molecular Biology and Genetics at Bilkent University between 2003 and 2007. She started her master's degree studies at Erasmus MC, Rotterdam in 2008. She completed her first internship at the Department of Virology, working in Professor Bart Haagman's group, with a focus on the human coronavirus, NL63. Her second internship was in the Department of Cell Biology, working with Dr. Eric Soler in Professor Frank Grosveld's group, focusing on key transcription factors and chromatin modifiers involved in erythropoiesis. She began her PhD study in February 2011 in Dr. Michiel Vermeulen's group at the UMC Utrecht Molecular Cancer Research Department. The PhD studies were completed at the Radboud Institute for Molecular Life Sciences in July 2015. Her research focused on several chromatin regulatory complexes and employed quantitative proteomics for identifying protein-protein interactions.

List of Publications

Perspective on unraveling the versatility of 'co-repressor' complexes.
Baymaz H.I.*, Karemaker I.D.*, Vermeulen M.
Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. 2015; Volume 1849, Issue 8, pages: 1051-1056

Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics.
Kloet S.L.*, **Baymaz H.I.***, Makowski M.*, Groenewold V., Jansen P.W., Berendsen M., Niazi H., Kops G.J., Vermeulen M.
FEBS Journal. 2015; Volume 282, Issue 9, pages: 1774-1785.

Identifying nuclear protein-protein interactions using GFP affinity purification and SILAC-based quantitative mass spectrometry.
Baymaz H.I., Spruijt C.G., Vermeulen M.
Methods in Molecular Biology. 2014; Volume 1188, pages: 207-226.

Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression.
Kalb R., Latwiel S.#, **Baymaz H.I.#**, Jansen P.W., Müller C.W., Vermeulen M., Müller J.
Nature Structural and Molecular Biology. 2014; Volume 21, Issue 6, pages: 569-571.

MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain.
Baymaz H.I.*, Fournier A.*, Laget S., Ji Z., Jansen P.W., Smits A.H., Ferry L., Mensinga A., Poser I., Sharrocks A., Defossez P.A., Vermeulen M.
Proteomics. 2014; Volume 14, Issue 19, pages: 2179-2189.

Identifying specific protein-DNA interactions using SILAC-based quantitative proteomics.
Spruijt C.G., **Baymaz H.I.**, Vermeulen M.
Methods in Molecular Biology. 2013; Volume 977, pages: 137-157.

* : shared first authourship

: shared second authorship

Acknowledgments

"It was the best of times, it was the worst of times…"

Nooo, not really -- this is a *far* more dramatic entrance than this period warrants, but I have always wanted to quote this opening! And what's more, my story spans two cities too. Even three in fact.

Let me try again, more seriously this time.

End of an era, yes.

Too many people have been generous with their time, expertise and/or friendship.

It all began with Michiel giving me a chance in his group of course: *Michiel*, my first impressions of you were: very enthusiastic about science, cheerful, positive, seems to think and talk at a fast pace. Over time, your incredible capacity for multitasking, strange uniprot-like memory for (obscure) protein names (KIAAxyz… etc), and your hand-in-hand confidence and modesty were added to the list; but my parting impressions of you are basically the same - which is quite remarkable considering the young and starting PI to the full-fledged professor transition you underwent in the past 4.5 years. I hope you know how much I appreciate all the supervision and support I could count on and the fact that you would always hear me out. I have learned so much science from you and always enjoyed the background stories you would tell with gusto about certain techniques and breakthroughs. Due to your numerous connections and the invitations you receive, I got to be part of many collaborations, meetings and papers, so I have been very lucky being one of your PhD students. One thing I still wish I could have picked up from you is your knack of getting along with everyone; but (alas) after 4.5 years of scrutiny and failure to replicate, I conclude that this is not an acquired trait. It has been a pleasure working with you, and I wish you and your lovely family all the best.

Marc, thank you for your valuable input during all the work discussions and also in my annual PhD committee meetings. Having you around in journal clubs was another highly educational experience that I have missed after the move to Nijmegen. I am very glad you agreed to be in my manuscript committee as you too have contributed to the end result with all the notes you have given me over the years. *Tobias,* thank you for being in my PhD committee and all your feedback, the questions you would ask always made me think of very different aspects of what I was doing that had never occurred to me. The other members of my manuscript committee, *Hans* and *Leonie*, thank you for the time you invested in reading my thesis and your positive feedback.

All the past and present members of the Vermeulen group, what-a-pleasure it has been! And that all starts with *Nelleke:* Ambitious, organized and so quick to take things in, understand and analyze. Also in the know of what's going on with whom all the time ;-). You have been a great source of information when it comes to where things were and how things were done. And most often you were the one to steer conversations back into English, seemingly simple and small but a very kind gesture. For another example, see raisin-free apple pies :-). Always a pleasure being roomies… *Radhika,* your calm and collected demeanour, with all your experimental experience helped me a great deal as I settled in the group. Being able to talk to you about anything has helped me through a few tough spots. Thank you for all your support and advice, talking down my worries, resolving my confusion about some stuff... Your opinion, scientific and otherwise, is important to me. I hope the difficult times you had to go through will come to an end soon... *Anneloes*, you were always very kind and helpful, and also quick with your lovely laughter :-). Thank you for all your help along the way. *Pascal,* very instrumental to the whole mass spec operation -possibly to

the whole Vermeulen operation!-, thank you for all the mass spec stuff you have taught me. *Arne,* what an easy-going and cheerfully helpful colleague you've been! And so (so, so, so) confident! I tried to present challenges along the way as you well know, but to no avail..:-) Kidding aside, I have always felt I was around a truly bright scientist giving me feedback on many different levels and thank you for all that. You have created such useful tools for the whole group that I have no doubt you'll be as much of an asset, if not more, wherever you go next. *Susan,* my dear, you are a force to be reckoned with! I admire how energetically you dove into all the new experiences life abroad presented, including the language. After Nelleke, with the move to Nijmegen, you became my logistics person to bug with 'Susan, where was … again?'. I very much appreciate all your help and scientific input -- almost as much as all the coffee breaks, delicious dinners and tons of trivia I would never have known if it weren't for you ;-). *Marijke,* always busy, fast, efficient and helpful. It was thanks to your and Pascal's organizational skills and hard work that moving the lab went as smoothly as it did. Thank you for all your help. *Matthew,* the 'structural edge' of the group, I believe I have learned something new and of interest almost every time we had a little conversation, which is very cool. As you had so much fun quoting, human life may be 'nasty, brutish and short'; but you seem to be making something interesting there ;-). Thanks again for all the deep and shallow conversations about (many) things deep and shallow. *Raghu*, every time an experiment did not work in my hands, your 'But why not? This is very easy!' exclamation is for sure in my 'top 10 most epic reactions a PhD student can get' list :-). All the best. *Ino,* cut-paste master extraordinaire, not a single sentence ever goes to waste! With your dry sense of humor, wry quips, efficiency and decisiveness, it was delightful writing with you. You are so low-key (seemingly) that it was a lot of fun to see (or hear about) the high notes you hit (how many pancakes..? - and this is the only hint of an example I'm willing to share in writing ;-)). *Xiaofei,* my U-neighbour, it was such a pleasure to discuss experimental set-ups with you and learn tons of ubiquitin trivia from you in the process. If there is anybody who could knock-down a most ubiquitious protein, it is you. It was motivating to witness your passion for that unbelievably complex system and I hope you're enjoying the stain-free bench you were looking forward to pipetting on :-). *Cristina,* my neighbour in the 'ghetto', your laid-back and getting-things-done-one-by-one attitude has been inspiring for me as I am not looking forward to starting from scratch again. You have done it calmly, beautifully and so independently! And in the meantime setting up an ambitious technique - I have so much respect for you. I hope the technique falls into place just as well as the other stuff have. To all the students who have passed through the group, it was a pleasure having you all around; fresh minds, so much talent and promise! Focusing on the talent I tried to help cultivate: *Deepani,* my first shared student, your enthusiasm for science, learning and trying new things was amazing to see. You are so adaptable that wherever you go it seems you've been there for a long time, and it felt like that in the lab as well. *Corina,* my cheerful, lively and lovely student, you can take on any problem and make it melt with your focus and obstinacy. It was a learning experience for me to try to manage you. Then I realized maybe you were managing me? :-) And a more particular thank you to *Anneloes, Maria* and *Arne* (especially *Arne* during the big move) for all the Dutch documents/voicemails you guys patiently translated for me over the years. *Ino* and *Nelleke,* thank you so much for translating the Dutch bits of the thesis. I would like all of you to remember that, if any Turkish document ever comes your way, my offer to translate is good for life. And of course another particular thank you to *Susan, Maria* and *Arne* for accepting to be my paranymphs. I hope we'll enjoy the day (assuming I get through the first part).

All the Timmers group members have contributed to my PhD experience substantially: Our shared work discussions and VerTim retreats were very fulfilling -scientifically and socially- thanks to you all. *Gianpiero* (wish you were around for a bit longer), *Pim* and *Nikolai,* it was a pleasure knowing you all. *Rick (van Nuland)* and *Andree,* you guys set a good example for making the best out of the PhD experience. *Petra,* always a source of useful advice and funny stories about the boys :-).

Roy, you have also been a very helpful and enthusiastic colleague, I hope wrapping things up goes smoothly. *Elfi,* it was a nice change to have your more clinical point of view. *Hetty* and *Richard,* thank you for all your help and very useful tips along the way. *Simona,* dearest co-paranymph, you are the most easy-going co-paranymph one could ask for and it was so much fun running around, trying to put things together with you. Always laid-back and cheerful, your attitude towards, well, everything, is inspiring :-).

I feel like I've forgotten someone…:-) *Maria,* my lovely, if it weren't for your help, I'd probably have gone without electricity and internet for my first year in Utrecht. But you see, as grateful as I am (and I am very grateful), these favors I do not take personally because I know you'd have helped anyone out with things like these (as a true altruist!). What I do take personally are all those times we hung out and talked about anything and everything with a solid conviction that we were on the same page – due to all the ways we are incredibly similar in and in spite of all the ways we are incredibly different in. Having you around was a stroke of luck for me. An unforgettable bit of us is of course 'I don't exactly know what you mean but I'm trying to understand?' :-) It's been a good and fun kind of trying on both parts, yes? I have another solid conviction that we'll keep in touch… (and you'll keep giving me sonication advice that I haven't asked for :-)))

A big thank you to people who make sure that things keep running: In Utrecht; *Marjoleine, Marcel, Cheuk, Huub, Cristina, Marianne* and *Mirjam,* thank you. *Wim, Marc, Eric and Dennis,* thank you for all the ICT support. *Betty,* thank you so much for all your kind e-mails, it is always good to see you again and catch up. In Nijmegen, a special thanks goes to *Maria, Lidwien, Josephine, Anita, Willem, Eva* and of course *Siebe.* And last but not least, *Marion,* thank you so much for all your help with everything.

Dear colleagues at the second and third floors of the Stratenum and the very same floors of the RIMLS, thank you for the friendly atmosphere you created. 3 pleasant years at the Stratenum and then 1,5 years in Nijmegen went by; and while the people in Utrecht were missed, new colleagues in Nijmegen could not have been more inclusive and welcoming. I wish all the best for all of you. To *Klaas* and all the Mulder group members, thank you for all your feedback during the Super Group meetings. *Luan* and *Nina,* thank you for the Q-Exactive support whenever Fusion failed to rise to the occasion. *Bilge,* bu balkonda o bira içilecek! ;-)

All the collaborators I was lucky to interact with: Dear *Sophie,* the MBD5/6 story all began with your very elegantly cloned expression plasmids. *Alexandra,* it was a pleasure to share results and brainstorm a bit over long and short e-mails. *Pierre,* thank you for all of your informative, enthusiastic and always very kind e-mails. *Reinhard* and *Jürg,* I was happy to be the interactor screen-er of the infamous and intriguing H2AUb. *Andrew* and *Zongling,* thank you for much needed input at a point where we were rather stuck.

Ve şimdi birazcık daha kişisel teşekkürler: *Nadire,* Hollanda zincirinin ilk halkası sendin. Başka pek çok muhtelif zincirin pek çok halkası olmakla birlikte… Anlaşılmanın lüks olduğu zamanlarımız olmadı pek bizim (en büyük lüks) ve bütün kartlarınla, e-maillarınla, bazen benim fark bile etmediğim duygularımı sezip, iyi ifade edemediğim düşüncelerimi (o muhteşem kaleminle) sözcüklere dökmenle her zaman çok destek oldun bana. Ve 'ama ben/ en çok şeyi/ en kısa zamanda/ sana söyledim/ yalnız sana.' Bendeki sen çok özel, biliyorsun... *Çiğdem,* şaka maka gümüş yıldönümümüzü kutlamamıza azıcık bir şey kaldı! Araya ne vakitler, ne kişiler, ne olaylar girer; bize ne gam (tamam tamam, arada gam da olmadı değil ama), bir muhabbete başlarız ve ne eğleniriz en ciddi dertleşmelerden sonra bile... Saatlerimiz ayarlı bond, ondan herhalde? *Seda,* politik entelim benim, aktivistim; senin o espri yeteneğin (ve sivri dilin ;-)) yok mu her defasında tam 12'den vuran! Vazgeçilmez. *Cansel*, en gırgır karikatür! Yanında olup eğlenmemek ne mümkün kuzum, enerjin muhteşem senin... *Gülin*, en pembe kalp diyeceğim elbette, sonsuz sevgi kaynağı! Ankara buluşmaları unutulmaz, ihmal edilemez, o buluşmalar Ankara dışına çıktığındaysa ef-sane! Demem o ki dostlar, sizinle paylaşılanlar, geçirilen zamanlar her zaman çok keyifli, iyileştirici

ve çok güzel işte... Muhabbetimiz daim olsun. Bu gruba da dahil ama Hollanda hikayesine daha bir (!) dahil *Nilhan*, hesapladım, senin 2589., benim 2588. günüm bu memlekette gözüm. Ne zamanlardı o Rotterdam günleri ve ne garip bir hikaye sonrası… Statik ben kepçe oldum bu minik kazanda, kök salansa dinamik sen... Bakalım bundan sonra nerelere savrulacağız, nasıl ironilere sebep (ve gark) olacağız... Fakat şeytan tüyün baki Nil! Bendeyse bazı başka şeyler.. *Deryacım,* sıra arkadaşlığından başlayıp Hollanda'da buluşmalara varan geçmişimizle biz çok eski dostuz diyebileceğimiz kadar sene geçti gitti arada, değil mi? Seninle ve *Fırat*'la (turistik fil :-)) buluşmalar, sohbetler hep çok keyifli... Bütün Erasmus MC güzellerine de, *Nesrin* (güzel Karadenizli), *Özge* (eski 7. kat komşum), *Merih* (orijinal Arnhemli :-)), *Sophia* (the loveliest of Sophias) ayrıca teşekkürler. Ve elbette artık çoğunluğu Eindhoven'da kalmamış olan Eindhoven grubu, *Beste, İso, Melike, Görkem, Sinan, Nimet, Tunç, Ezgi, Ulaş, Can, Ekin,* her zaman eğlenceli sosyal ortamlarınıza beni de kattığınız için çok teşekkürler… *Rick (Janssens)* - 'een kinderhand is gauw gevuld', yes, true; but then again a good, solid friendship is strangely underrated. Sometimes it only takes one other person with a similar sense of humor to take the edge off the 'people are strange when you're a stranger' syndrome - and that is exactly why a friend like you could never be overrated.

Güzel bacı, vesikalı konuşmalar falan.. Ne çok konuşuyoruz biz birbirimizle yahu, değil mi? Hem şikayet eder hem şımarırım; çünkü bilirim hep ne kadar şanslı olduğumu beni o kadar dinleyip, akıllar verip, yalnız bırakmayan bir ablam olduğu için. Hem de 'more universally right'! Daha ne olsun! Çok teşekkürler, gerçekten… *Greg,* the most supportive brother–in-law one could ask for, thank you so much for all the long distance calls and brainstorming along with me about the uncertainties of the future. Always interesting to hear your point of view (and a lot of fun to hear your twisted, twisted jokes ;-)). *Annem* ve *babam* - size ansiklopedi setinden aşağısı ithaf edilmezdi güzellerim, ama şimdilik bu kadar oldu, idare edin. *Baba,* sen olmasan kim bana Ankara/Amerika haberleri verecek, benimle dünyanın halini tartışacak ve en Malthusçu yorumlarıyla günüme gün katacak!.. Ve *güzel annem,* sen olmasan nasıl taşınacaktı o evler, bu kutular kendilerini toplamaz, toplamıyor gene.. Şaka bir yana, ikinize de bütün ilginiz, koşulsuz desteğiniz için; uzun uzun, ince ince dinleyip yaptığınız durum analizleri ve önerileriniz için; ve fakat son tahlilde 'kendi kararını kendin ver tabii' deyip büyümeme izin verdiğiniz için çok çok teşekkür ederim.

Herkese beni dinlediği için bu kadar teşekkür ettikten sonra ne kadar çok konuşan bir insan olduğum gerçeği ortaya çıktı tabii! :-) Anyhow, thank you once again dear colleagues, friends and family! Wishing you all the best and also that this won't be a lasting goodbye…

İrem.