Systematic analysis of the interactome of modified chromatin

PhD Thesis

in partial fulfillment of the requirements for the degree
"Doctor of Philosophy" (Ph.D.)
in the Molecular Biology Program
at the Georg August University Göttingen,
Faculty of Biology

submitted by

Miroslav Nikolov

Born in Pleven, Bulgaria

Göttingen 2012

Thesis committee members:

Prof. Dr. Henning Urlaub   Bioanalytical Mass Spectrometry Group,
(1st reviewer)   Max Planck Institute for Biophysical Chemistry, Göttingen
                                  Bioanalytics, Department of Clinical Chemistry,
                                  University Medical Center, Göttingen

Prof. Dr. Peter Rehling   Department of Biochemistry II,
(2nd reviewer)   Georg-August University, Göttingen

Prog. Dr. Dirk Görlich   Department of Cellular Logistics,
                                  Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 19th October 2012

Affidavit

I hereby declare that the presented thesis entitled "Systematic analysis of the interactome of modified chromatin" has been written independently and with no other sources and aids than quoted.

Göttingen, 19th September 2012

Miroslav Nikolov

List of publications (September 2012)

Jaspers MH, Nolde K, Behr M, Joo S, Plessmann U, **Nikolov M**, Urlaub H, Schuh R. "The Claudin Megatrachea Protein Complex." *J Biol Chem.* 2012 Aug 28. (Epub ahead of print)

Nikolov M, Schmidt C, Urlaub H; "Quantitative Mass Spectrometry-based proteomics: an overview" *Methods Mol Biol.* 2012; 893:85-100

Alkhaja AK, Jans DC, **Nikolov M**, Vukotic M, Lytovchenko O, Ludewig F, Schliebs W, Riedel D, Urlaub H, Jakobs S, Deckers M. "MINOS1 is a conserved component of mitofilin complexes and required for mitochondrial function and cristae organization." *Mol Biol Cell.* 2012 Jan; 23(2): 247-57. Epub 2011 Nov 23.

Nikolov M, Stützer A, Mosch K, Krasauskas A, Soeroes S, Stark H, Urlaub H, Fischle W. "Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns." *Mol Cell Proteomics.* 2011 Nov; 10(11): M110.005371. Epub 2011 Aug 11.

Ivanov I, Tsacheva I, Stoyanova V, **Nikolov M**, Tchorbadjieva MI, Petrova S, Christov L, Georgieva V, Georgiev G. "Chaperone-like effect of polyzwitterions on the interaction of C1q with IgG." *Z Naturforsch C.* 2009 Jan-Feb; 64(1-2):149-54.

Shema E*****, **Nikolov M***, Haj-Yahya M, Allemand E, Yamaguchi Y, Muchardt C, Urlaub H, Brik A, Oren M, Fischle W. "Systematic identification of proteins binding to chromatin-embedded ubiquitylated H2B reveals recruitment of the SWI/SNF complex to regulate transcription." under review in *Mol Cell*, *****equal contribution

Nikolov M, Fischle W. "Systematic analysis of histone modification readout." under review in *Mol Biosyst*, invited publication

Tirard M, Hsiao HH, **Nikolov M**, Urlaub H, Melchior F, Brose N. "In vivo identification of SUMOylated proteins in the brain of His_6-HA-SUMO1 knock-in mice." under review in *Proc Natl Acad Sci USA*

Acknowledgements

Table of contents

List of figures

List of tables

Abbreviations

%	percent	M	molar
1D	one dimensional	*m/z*	mass-to-charge ratio
2D	two dimensional	MALDI	matrix-assisted laser desorption ionization
aa	amino acid		
ACN	acetonitrile	min	minute
AFM	atomic force microscopy	MNase	Micrococcal nuclease
APS	ammonium peroxodisulfate	MS	mass spectrometry
ATP	adenosine triphosphate	MS/MS	tandem mass spectrometry
bp	base pair	Mw	molecular weight
BSA	bovine serum albumin	n	nano
C.elegans	*Caenorhabditis elegans*	NE	nuclear extract
C°	degree centigrade	OD	optical density
cm	centimeter	PAGE	polyacrylamide gel electrophoresis
Da	Dalton (g/mol)	PBS	phosphate buffered saline
DDA	data-dependent acquisition	PCI	phenol/chlorophorm/isoamyl alcohol
ddH$_2$O	double distilled water (sterile)		
DMEM	Dulbecco's Modified Eagle's medium	PEG	polyethylene glycol
		PMSF	phenylmethylsulfonyl fluoride
DNA	deoxyribonucleic acid	ppb	parts per billion
DNase	deoxyribonuclease	ppm	parts per million
DTT	dithiothreitol	PTM	posttranslational modification
E.coli	*Escherichia coli*	RNA	ribonucleic acid
e.g.	for example, *exempli gratia*	RP	reversed phase
EDTA	ethylendiaminetetraacetic acid	rpm	revolutions per minute
ESI	electrospray ionization	RT	retention time or room temperature
et al.	and others, *et alii*	s	second
FA	formic acid	SCX	strong cation-exchange
FBS	foetal bovine serum	SDS	sodium dodecyl sulfate
FW HM	full width at half maximum	SILAC	stable isotope labeling with amino acids in cell culture
g	gram or gravity force		
h	hour	snRNA	small-nuclear RNA
H	heavy (SILAC)	TB	tris/borate
His-tag	histidine tag	TBE	tris/borate/EDTA
HPLC	high pressure liquid chromatography	Tris	tris-(hydroxymethyl) aminoethane
		TSS	transcription start site
i.e.	that is, *id est*	U	unit
IPI	International Protein Index	V	volt
k	kilo	v/v	volume per volume
KDM	lysine demethylase	vol.	volume
KMT	lysine methyltransferase	w/v	weight per volume
l	liter	WT	wild type
L	light (SILAC)	α	anti-/antibody
LC	liquid chromatography	Δ	delta (deletion)
m	milli or meter	μ	micro

1 Abstract

Regulation of chromatin composition and structure is crucial for maintaining genome integrity and execution of the wide array of functions related to gene regulation and downstream cell signaling. Chemical modifications of chromatin are at the core of these processes and include methylation of the primary sequence of DNA, as well as various modifications of the proteinaceous chromatin packaging units - the histones. Recent genome-wide mapping approaches have been instrumental for characterization of the location and distribution of these marks relative to the different functional domains and gene regulatory elements in chromatin. The majority of the characterized chromatin modifications function as signaling platforms for recruitment of various protein complexes. Therefore, it is of equal importance to describe these sets of proteins for dissection of the functional consequence of their binding. More importantly, global analysis of the interactomes of functionally associated chromatin modifications might shed light on the proteins required for establishment and operation of specific chromatin domains.

In this study, a novel approach for identification of chromatin modification-dependent protein binding was established. *In vitro* reconstituted oligonucleosomal templates with homogeneous and defined modification status were used for affinity purification from SILAC-labeled nuclear extracts. The interactomes of ten individual chromatin species were investigated and the results provided valuable insight into chromatin biology on several levels. Investigation of the nature of the subproteomes recruited by single modifications provided evidence for their functional role. Additionally, the results offer a comprehensive catalogue of candidate proteins for further dissection of specific chromatin modification molecular readout. This was exemplified here with the demonstration of the recruitment of the SWI/SNF complex to monoubiquitylated H2B-containing chromatin for regulation of transcription of a specific set of genes. Investigation of the interactome of doubly modified chromatin identified a large number of factors whose recruitment presumably depends on the cooperative action of two modifications. Furthermore, the comparative analysis of individual datasets revealed novel relationships between the different modifications. On a global scale, this resulted in the identification of a limited set of proteins that likely play an important role for the function of heterochromatin domains. Lastly, the chromatin affinity purification approach was used for developing a SILAC internal standard method for direct quantitative comparison of recruitment to different chromatin modifications and combinations thereof.

2   Introduction

The genomic DNA of eukaryotes is confined within the membranes of the cell nucleus. Its physiological structural and functional form – chromatin, is an intricate deoxyribonucleoprotein complex that is at the core of regulation of virtually all DNA-templated processes. These are carried out by a large number of chromatin proteins and macromolecular complexes. An important feature of many of these complexes is their recruitment or regulation by chemical modifications of the chromatin structural proteins and the DNA itself. The chromatin modifications are crucial for reading and translating the information within the genome in a programmed and organized fashion, resulting in tightly regulated, specific and selective gene expression during different stages of development or in different cell types. Moreover, there is evidence for the involvement of these modifications in epigenetic phenomena – the propagation of heritable traits "coded" beyond the sequence of the DNA.

2.1   Chromatin structure and organization

2.1.1   The nucleosome core particle

The length of the DNA molecules in eukaryotic cells far exceeds the dimensions of the nucleus and therefore mechanisms exist that allow its compaction. This compaction can result in up to 10 000-fold decrease in size and is achieved at several structural levels. The basic structural unit of chromatin is the nucleosome core particle. It consists of 147 base pairs (bp) DNA wrapped around two copies of each histones H2A, H2B, H3 and H4 (Figure 2.1 A) [1]. The histones are relatively small (11-15 kDa) and basic proteins that are highly conserved among all eukaryotes. Despite having no apparent sequence homology, the four histones share high degree of structural homology – namely the histone fold motif. It consists of two short α-helices flanking a third long helix, which are connected by two loops. These structural motifs take part in the protein-protein, as well as in the protein-DNA interactions that form the nucleosome core particle. The histone proteins interact with each other as heterodimers. H2A interacts with H2B, and H3 – with H4. In addition, and under higher concentrations *in vitro*, two H3/H4 dimers form a tetramer (Figure 2.1 B) [3]. Two H2A/H2B dimers and one H3/H4 tetramer constitute the histone octamer, around which the

147 bp of in DNA is wrapped in a 1.65 tight superhelical turn (Figure 2.1 A). However, both in vivo and in vitro, these interactions are executed in a stepwise fashion. First, the H3/H4 tetramer interacts with the central stretch of DNA, followed by the H2A/H2B dimer. Importantly, the protein-DNA interactions within are not nucleotide-, and therefore sequence-specific [4]. Nevertheless, because of greater bending properties, A-T-rich sequences facing the histone octamer in the minor groove of DNA are preferred [3]. Within the structure of the nucleosome core particle, more than 75% of the DNA surface remains free and available for additional protein-DNA interactions [3]. The N-terminus of all four histones, as well as the C-terminus of H2A, are far less structured than the histone fold motif. They are arginine- and lysine-rich and are responsible for the majority of the sequence variability between different species. They protrude outwards from the nucleosome core and are not visible in any of the 3D crystal structures determined to date. Nevertheless, they are crucial for interactions with adjacent nucleosomes [5] and with non-histone chromatin proteins [6].
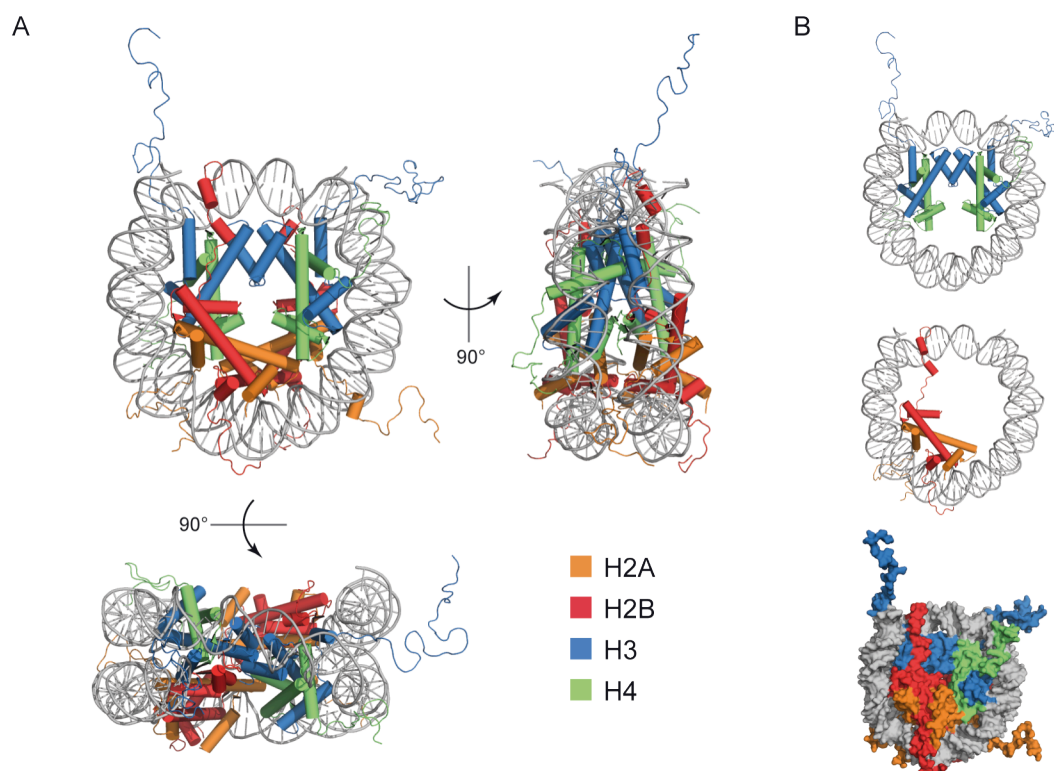
Figure 2.1 Nucleosome core particle structure.

A. Crystal structure of the nucleosome core particle with front and side views (PDB ID 1KX5) [2] visualized using MacPyMol (http://pymol.org). Histone proteins are coloured as indicated, DNA is coloured gray. Alpha-helical regions are represented as cylinders. **B.** As in A, but only the histone (H3/H4)$_2$ tetramer (top), H2A/H2B dimer (middle) or surface view of the whole structure (bottom) are shown.

Apart from the four canonical types of histones, there are a number of major non-allelic histone variants that bring about structural and functional heterogeneity in different nucleosome core particles. They are expressed in all tissue types and can be incorporated in chromatin during different stages of development and the cell cycle. All histones, but H4, have such variants and while some are differing from the canonical form by only a few amino acids (for example H3.1 and H3.3), others have additional domains (for example H2A and macroH2A) [3].

2.1.2 Higher order structures and chromatin domains

The nucleosome core particle is the fundamental unit of chromatin. The linear arrangement of these structures along the DNA constitutes the first level of DNA compaction – the 10-11 nm wide "beads on a string" form of chromatin (Figure 2.2 A). Individual nucleosome core particles are separated by 10-60 bp of linker DNA. The nucleosome core DNA, linker DNA and the histone octamer together form the nucleosome. A fifth class of histones – the linker histone H1, interacts extensively with the linker DNA and the nucleosome core particle, forming a structure called chromatosome, and is essential for higher levels of chromatin compaction [7]. Under physiological salt conditions the nucleosome beads on a string chain can fold into a higher order structure – the 30 nm fiber [7]. Several models of the architecture of this structure have been proposed [7, 8], however it's exact organization and physiological significance remain controversial [9]. The higher levels of chromatin organization are even more elusive and presumably involve long-range chromatin loop interactions.

Cytologically, chromatin can be divided in two main categories – euchromatin and heterochromatin. Historically, heterochromatin was described as the distinct part of chromatin that remained compacted throughout the cell cycle and was visible as more densely stained regions of the nucleus using dyes with affinity for DNA under light microscopy (and later observed as more electron-dense regions using electron microscopy, Figure 2.2 B), unlike euchromatin that undergoes general decondensation during interphase [10, 11]. Generally, heterochromatin is characterized with lower gene density, high content of repetitive sequences and late replication [7]. The centromeres and telomeres are part of constitutive heterochromatin, which is present in all cells at all times, while regions that are specifically compacted and silenced during development and differentiation are termed facultative heterochromatin. Euchromatin, on the other hand, is more relaxed, replicates early in S-phase and mostly includes gene-rich regions of the genome. Even though this near-century old classification of chromatin domains is backed up by molecular and

biochemical evidence, it is now clear that upon closer look, there are more chromatin subtypes with specific functional, structural and compositional properties [12].
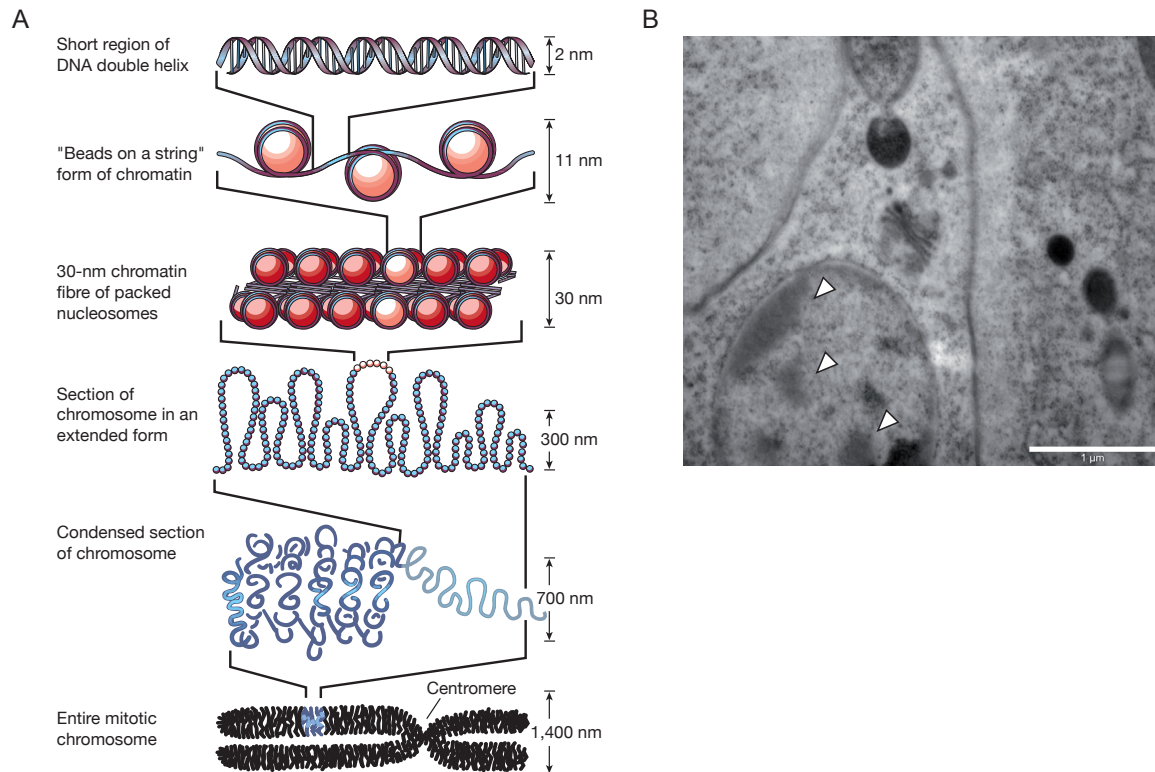
Figure 2.2 Higher order chromatin structures and domains

A. Several levels of chromatin condensation. Nucleosomes are arranged as "beads on a string", which interact to form the "30 nm fiber". These are then condensed by yet unknown mechanisms and interactions into mitotic chromosomes. Adapted from [13] with permission from the publisher. **B.** Electron micrograph of a *C. elegans* neuronal cell body. Arrows indicate darker regions of the nucleus, scale bar represents 1 μm. The micrograph was acquired on U/Pb-contrasted ultramicrotom sections of high-pressure frozen *C. elegans* in the laboratory of Dr. Stefan Eimer, ENI – Göttingen.

2.2 Chromatin modifications and their readers

The fundamental unit of chromatin – the nucleosome, is not only a static structural component with DNA-compaction functions. It is highly dynamic throughout the cell cycle and can exist in different forms. One degree of diversity is achieved by incorporation of histone variants. Additionally, differential and specific positioning of nucleosomes along the DNA molecule is possible from the action of multiple chromatin remodeling complexes [14]. The dynamics of these processes reflects the changing requirements for execution of different genetic programs at different stages of the cell cycle or variable environmental conditions. Post-translational modifications (PTM) of histones and modifications of DNA are an enormous source for such variability and play central role in regulating virtually all DNA-templated processes [15].

2.2.1 Types of chromatin modifications and associated functions

General properties and establishment of histone modifications

The fact that the histone proteins are chemically modified *in vivo* was recognized for the first time nearly 50 years ago [16]. However, especially with the advancements in mass spectrometry-based proteomics [17-19], it has only recently been possible to site-specifically identify (and quantify) the wide array of histone modifications. Virtually every known type of protein PTM has been found on histones. There are nearly 200 detected sites of modification and the number is still growing [20]. While the majority of these PTMs are found in the flexible N- (or C-) terminal tails, there are a number of modified sites in the histone core regions as well. There is functional evidence only about a fraction of those modifications, and while some types of PTM have been widely studied, there are only single reports for the existence of whole classes of others (e.g. tyrosine hydroxylation [20]).

Methylation and acetylation of lysine comprise the most abundant classes of histone modifications. Additionally, methylation of arginine, phosphorylation of serine and threonine, are well studied. Other classes of PTMs include crotonylation, glycosylation, hydroxylation, ADP-ribosylation, deimination, proline isomerization, citrulination and H3 tail clipping. Furthermore, the conjugation of whole proteins is also well documented, with the examples of monoubiquitylation and to a lesser extent – sumoylation [21]. The tremendous degree of diversity comes not only from the different positions of the modifications, but also from their degree. Methylation can be found in three varieties for both lysine (mono-, di- and tri-) and arginine (mono-, symmetric and asymmetric di-) residues.

The establishment of histone modifications is a highly dynamic process and the majority of them have a high turnover [15]. This is regulated by the opposing activities of the enzymes responsible for depositing or erasing these PTMs. The process of histone

acetylation and deacetylation is arguably most well studied [22, 23]. The enzymes depositing acetyl marks on histones are known as histone acetyl transferases (HATs), and those erasing the modification – histone deacetylases (HDACs). There are many different enzymes that catalyze these processes, and they are specific for certain histone residues or distinct molecular processes. High sequence and function specificity is even more evident for lysine methylation [15]. The enzymes catalyzing this modification are termed lysine methyl transferases (KMTs) and those erasing it – lysine demethylases (KDMs). Additional level of complexity stems from the finding that many lysine HMTs show specificity only for a certain degree of methylation, while others are capable of mono-, di- and trimethylation. The same holds true for the histone demethylases [15].

Functional consequences

Because of the wide variety and degrees of histone modifications they hold an enormous potential for functional responses. While histone hyperacetylation is generally associated with active transcription and hypoacetylation with repressed transcription [16, 24], the functional consequence of the majority of other modifications is specifically dependent on the site, degree and timing of modification. Generally, histone modifications have been linked to transcriptional regulation, DNA repair, cell cycle progression, alternative splicing, chromosome segregation, epigenetic silencing, nucleosome remodeling and other major cellular processes [25-27]. A brief description of the main functional consequences and modifying enzymes for more prominent histone modifications is given in Table 1. Histone modifications exert their function by two major mechanisms. The first mechanism relies on direct disruption of contacts between individual nucleosomes or between the histones and DNA. The best characterized such effect is by histone acetylation. Deposition of acetyl group has a strong effect as it neutralizes the positive charges of lysine residues. For example acetylation of lysine 16 in the N-terminal histone tail of H4 can significantly affect internucleosomal contacts and completely disrupt higher order chromatin condensation [28]. On the other hand, acetylation of H3K56 can affect $(H3/H4)_2$ tetramer contacts with DNA, but this does not contribute to destabilization of nucleosomes [29]. Lastly, ubiquitylation of H2B can interfere with chromatin compaction and results in more accessible chromatin arrays [30]. The second mode of action is by recruiting specific non-histone proteins to chromatin. In such manner, the histone modifications function as molecular beacons and depending on the site, type and degree of the modification, can recruit a wide variety of proteins and affect virtually all DNA-templated processes. The mechanisms of recruitment and functional consequences will be discussed in more detail in Chapter 2.2.2.

Histone	Residue/ modification	Function	Deposition enzyme
H2A	H2AR3me	Transcriptional repression	PRMT6
	H2AK5ac	Transcriptional activation	p300/CBP
	H2AK119ub	Spermatogenesis	HR6A
		Polycomb mediated gene silencing	RING1/2
	H2AXS139ph	DNA repair	ATR, ATM, DNA-PK
H2B	H2BK12ac	Transcriptional activation	p300/CBP
	H2BK120ub	Transcriptional activation and elongation	BRE1
H3	H3R2me2a	Transcriptional repression	PRMT6
	H3R2me2s	Euchromatin maintenance	
	H3K4me1/2	Transcriptional activation	MLL1
	H3K4me1/2/3	Transcriptional activation	MLL3
	H3K4me3	Transcriptional activation	MLL5
	H3K9ac	Transcriptional activation	Gcn5
	H3K9me1/2	Transcriptional silencing	G9a
	H3K9me2/3	Transcriptional silencing	Suv39H1/2
		Transcriptional silencing of euchromatin	SETDB1
	H3S10ph	Mitosis	Aurora-B
	H3K23ac	Transcriptional activation	Gcn5
	H3K27ac	Transcriptional activation	Gcn5
	H3K27me1/2/3	HOX gene silencing	EZH1/2
	H3K27me2	Transcriptional activation during development	NSD2
	H3K27me3	X-chromosome inactivation	EZH2
	H3K36me2	Transcriptional silencing	NSD1
	H3K36me3	Transcriptional elongation	SET2
	H3K79me1/2/3	Transcriptional activation	DOT1
H4	H4R3me	Transcriptional activation	PRMT1
	H4K5ac	Transcriptional activation	Tip60
	H4K20me1	Transcriptional silencing/activation	PR-Set7
	H4K20me2	Transcriptional activation	NSD1
	H4K20me2/3	Transcriptional silencing, DNA repair	SUV4-20H1/2
	H4sumo	Transcriptional repression	Ubv9

Table 1 Selected histone modifications, their functions and deposition enzymes

Representative histone modifications, their functions and deposition enzymes. ac – acetylation, me1/2/3 – mono-/di-/trimethylation, me2a – asymmetric dimethylation, me2s – symmetric dimethylation, ph – phosphorylation, ub – monoubiquitylation, sumo – sumoylation. Based on [21, 25, 26, 31, 32].

Distribution in chromatin and the genome

Taking together the functions of individual histone modifications (Table 1) and the concept of active euchromatin and inactive heterochromatin, it is not surprising that individual PTMs have specific distribution along the genome. There are modifications specific for one type of chromatin and excluded from the other type. The most prominent histone modifications for constitutive heterochromatin, are H3K9me3, H4K20me3 and to a lesser degree H3K27me3 [33]. Additionally, dimethylation of H3K9 and H3K27 are often found together with the respective trimethylation marks. High resolution mapping of histone modifications at these regions is, however, very difficult owing to the repetitive nature of the underlying DNA sequence. Nevertheless, advances in computational approaches for analysis of high-throughput sequencing data have shown that different regions of constitutive heterochromatin, including gene deserts, subtelomeres, centromeres and pericentromeric regions are not monolithic and can have varying levels of histone modifications, with prevalence for one or another mark (Figure 2.3 A) [34, 35]. A hallmark modification for facultative heterochromatin is H3K27me3, along with the associated Polycomb repressor complexes [36]. An important feature of chromatin are the boundary elements, or insulators, that prevent spreading of heterochromatic regions [37]. They contain high levels of the protein CTCF, as well as the H3K9me1 mark [38]. CTSF, in particular, has also been implicated in mediating higher order chromatin organization and regulating the dynamics and localization of heterochromatic domains [39].

Transcriptionally active euchromatin contains a large number of characteristic histone modifications. However, these are not distributed equally. There are regions, especially those employed at different stages of transcriptional regulation, that are particularly modification-rich. The general modification landscape of an active gene is represented in Figure 2.3 B. The transcriptional start site (TSS) contains high levels of methylated H3K4, particularly H3K4me3, as well as acetylation of H3K9. Methylated H3K79 is characteristic for the 5'-region of the active gene, while H3K36me3 – for the 3'-region. H3K9me2/3 and H3K27me2/3 are excluded from the active gene bodies [15]. Silent genes, on the other hand, contain high levels of H3K9me2/3 and H3K27me2/3 and are depleted from active marks, such as H3K4me and H3K9ac (although these could be found at low levels at the TSS) (Figure 2.3 C) [40]. Interestingly, a specific pattern consisting of H3K4me3 and H3K27me3 is present at genes that are expressed at low levels but poised for activation in ES cells. Upon differentiation, the modification status resolves in either H3K4me3 and activation of the gene, or in H3K27me3 leading to its silencing (Figure 2.3 D) [41].
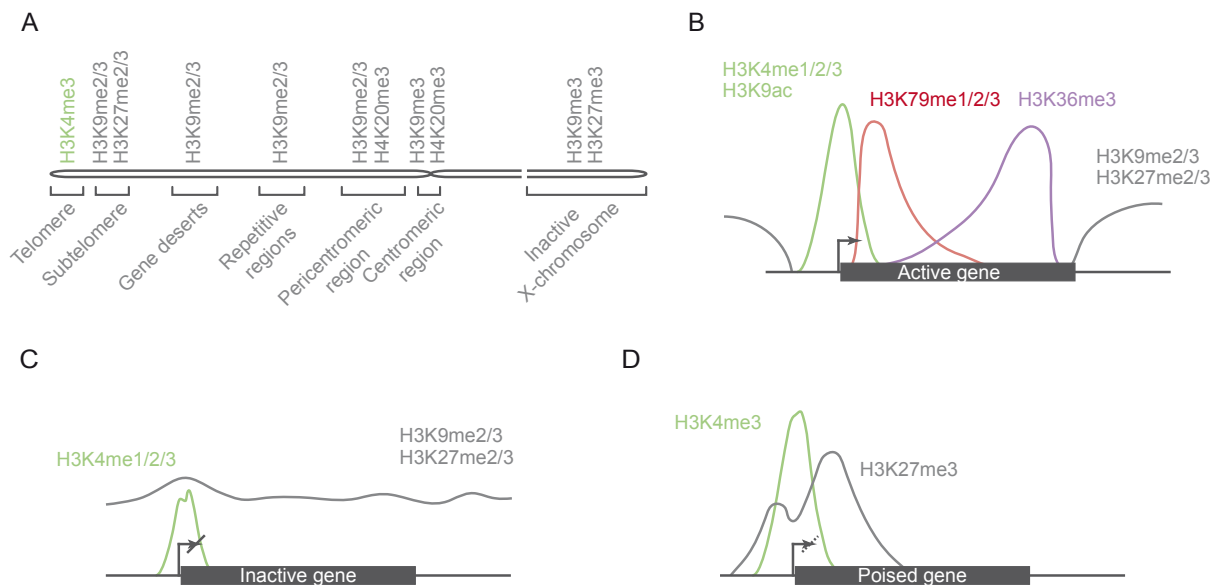
Figure 2.3 Distribution of histone modifications along the genome or at representative genes.

Levels of different histone modifications along different chromatin domains (**A**), active (**B**) inactive (**C**) and bivalent-poised (**D**) genes. A illustrates the distribution along the genome length and does not represent a mitotic chromosome. Transcriptional start site (TSS) is marked with an arrow and gene body with dark rectangle. This is a cartoon illustration and does not represent the true complexity and variability of the levels at different genes. A is based on information from [35]. B is adapted from [15] with permission from the publisher; C and D are based on data from [38, 40, 42];

Recent studies mapping the high-resolution profile of all known (at that time) sites of histone methylations and acetylations [38, 43] have significantly expanded our understanding of the chromatin landscape. It was shown that histone acetylation in general is associated only with active genes. It appears that the histone modification landscape of active genes is more complex than previously appreciated (Figure 2.4). Multiple methylation and acetylation marks have specific positions with respect to gene controlling elements. While H3K4me3 and H3K9ac are associated mainly with the TSS, H3K4me2, H3K9me1, H3K27me1, H4K20me1 are characteristic also for the gene body. Strikingly, the distribution of the monomethyl modification level of H3K9 and H3K27 is in stark contrast to the distribution of their di- and tri-methyl counterparts, mainly associated with silent loci. H3K4me1, on the other hand, has highest levels at enhancer elements. Insulators show higher CTCF binding, as well as H3K4me1/2/3, H3K9me1 and several acetylation marks.
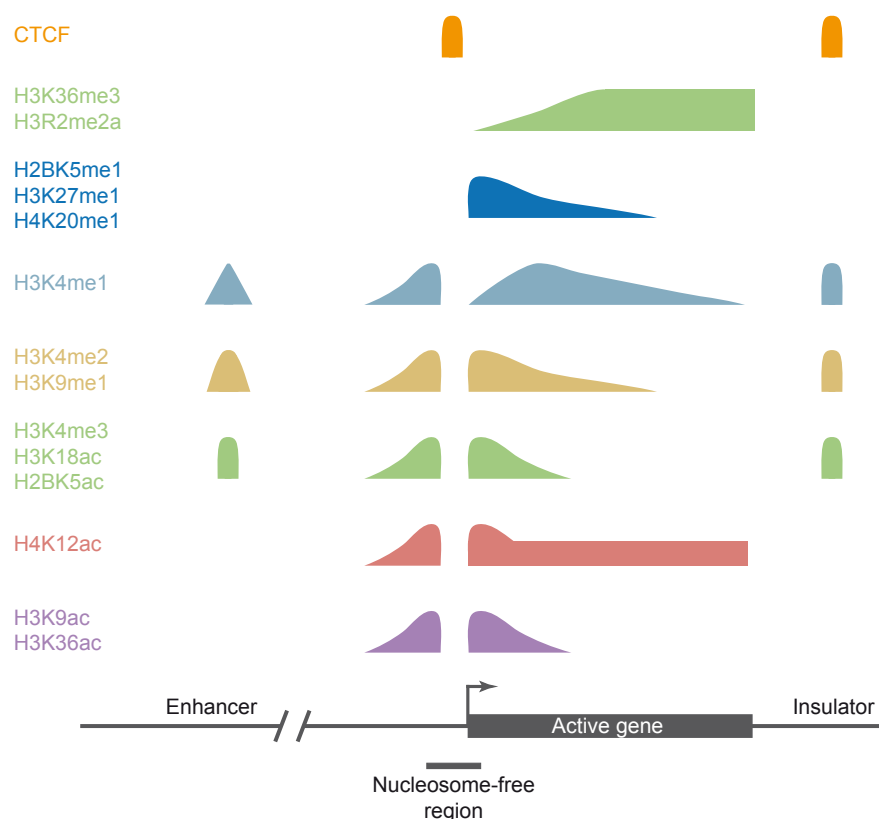
Figure 2.4 Distribution of fourteen histone modifications and CTCF in the region of an active gene.

Distribution of the indicated histone modifications and CTCF in the enhancer, insulator and TSS regions and the gene body of an active gene. This is a cartoon representation and does not show the true complexity and variability of the modification levels within individual genes. Based on data derived from CD4+ T-cells [38, 43] and adapted from [44] with permission from the publisher.

DNA methylation

In addition to the plethora of histone post-translational modifications, DNA can also be chemically modified post-replicationaly. Methylation of the C5' position of cytosines, mainly in CpG dinucleotides, is the most prominent and widely studied DNA modification in mammals [45]. Its existence has been proposed more than 100 years ago, and it was detected and unambiguously identified shortly afterwards [46]. In mammals, this modification is deposited by a class of enzymes – DNA methyl transferases (Dnmts), with differing activities. Dnmt3a and Dnmt3b establish the *de novo* methylation pattern, while Dnmt1 is a maintenance methyl transferase during replication and repair and has a preference for hemimethylated DNA [45]. CpG dinucleotides are found throughout the genome, but are concentrated at gene-rich regions. Moreover, even though 60-80% of the CpG sites in the mouse and human genomes are methylated [47], CpG methylation is found at higher levels in gene bodies, intergenic and repetitive sequences. On the other hand, regulatory

sequences like promoters and enhancers are methylated to a lower level [48]. Sequences with higher than average CpG content are termed CpG islands and are found often unmethylated at promoters of constitutively expressed genes [46]. Methylation is associated with gene silencing, as well as with repression of transposable elements and transcriptional noise in general [49]. Interestingly, DNA methylation at CpG island promoters correlates very well with reduced gene expression, however methylation of promoters with low CpG content does not influence transcription significantly [46, 50, 51]. It is thought that the mechanisms of DNA methylation-regulated gene expression depend on either direct loss of transcription factor binding site on DNA, or on recruitment of proteins binding to methyl-CpG (meCpG) [46]

2.2.2   Functions of chromatin modification-binding proteins

The distribution of histone and DNA modifications throughout the genome and their association with differential gene expression or with particular chromatin domains are well established for a number of model organisms and cell types. Mounting evidence suggests that the main mechanism of localized action of these modifications is by recruitment of specific proteins and protein complexes [25]. It is clear that for the majority of modifications not only the exact chemical group, but also its precise position and surrounding residues are determinant for different functional consequences. While some acetylation positions are interchangeable [52], this is not the case for others [28, 29]. Methylation shows even higher level of specificity, where different degrees of modification, as well as the exact position could have profoundly different impact [43, 53]. The possibility of combinatorial effects of different modifications expands the signaling potential tremendously. It has been postulated that the protein "readers" and "effectors" associated with these marks are part of a "histone code" where specific combinations of PTMs exert specific functions that determine the functional status of the underlying DNA [27, 54]. However, while there are examples of cross-talk between different modifications and combinatorial recruitment, evidence for a true code-like behaviour with high levels of specificity on protein level is scarce.

Recognition of chromatin modifications by "reader" domains

Translation of the signals from individual modifications requires specialized molecular recognition of the recruited proteins. Indeed, chromatin mark-binding proteins contain "reader" domains that bind specifically to modified (or unmodified) residues, as well as DNA. In the majority of cases, they are found in multiple copies or together with other domains, creating specific recognition modular assemblies [53]. Most prominent histone modification-binding domains are bromodomains, tudor domains, plant homeodomains, MBT repeats,

PWWP, chromodomains and 14-3-3 domains. The first identified example of histone modification binding domain is the bromodomain, which binds acetyl-lysine residues with remarkable sequence specificity [55]. It is found in many transcriptional regulators, including HAT enzymes. It is often positioned near another histone-binding domain – the plant homeodomain (PHD) Zn-finger [53]. It was recently demonstrated that PHD fingers can also recognize acetylated lysines [56]. However, they are best known as general protein-protein interaction domains, and in particular - methylated lysine recognition motifs. There are numerous reports, as well as 3D structures, indicating the interaction of PHD domains with methylated H3K4 [57, 58]. Tudor domains are often found in tandem (TTD), where one of the folds interacts with a methylated lysine and the other is free for potential ligands [59, 60]. Tudor domains have also been shown to recruit proteins to methylated arginines [61, 62]. The MBT repeats bind preferentially to monomethyl, or in some cases dimethyl lysines [63], while often having no particular sequence context specificity [64]. PWWP domains have not been investigated as extensively as the other chromatin binding folds. There are several reports demonstrating their binding to mono- [65] and trimethyl lysines [66]. Chromodomains are well characterized as methyl recognition domains both structurally and functionally. They are divided into several subclasses, including the chromobox and HP1-type, the CHD chromodomains, and the chromo shadow and chromo barrel domains [58]. HP1 and chromobox chromodomains have been found to interact with methylated H3K9 and H3K27. CHD-type of chromodomains appear in tandem and several 3D structures including recognition of H3K4me have been determined [67]. The chromoshadow domains, on the other hand, do not interact with histones. In HP1 proteins they act as dimerization domain, and as such, create a platform for protein-protein interactions with factors containing a conserved PxVxL motif [68, 69]. The tudor, PWWP, MBT and chromodomains are structurally related and are part of the so called Royal family of methyl-lysine recogntition domains [53].

Phosphorylation of serine and tyrosine are recognized by distinct protein domains - BRCT and 14-3-3, respectively [70, 71]. Phosphoserine is often found next to another modified residue and can function in disruption or weakening of the recognition of the adjacent modification [72].

Proteins binding to methylated cytosine also contain specific recognition domains. meCpG-interacting proteins can be divided in three different groups – the meCpG-binding domain (MBD), Uhrf and Kaiso families [46]. The MBD group of proteins bind meCpG via their MBD domain and often recruit HDAC and KDM proteins for gene silencing. However, there are examples where these proteins are recruited independently from meCpG and not exclusively linked to gene repression [46]. The two members of the Uhrf family – Uhrf1 and Uhrf2, recognize meCpG via an SRA domain [73, 74]. Uhrf1 can also recruit HDAC and

KDM enzymes, suggesting involvement in gene silencing [73]. The members of the Kaiso family of meCpG binders contain BTB/POZ motif and their recruitment is mediated by a Krüpel-like C_2H_2 zink finger [46].

Functional consequence, modes of chromatin modification binding and cross-talk

There are numerous chromatin-associated proteins that show specific recruitment by histone and DNA modifications. Some have been identified by large-scale approaches [75-78], while others have been targeted following known functional association or by educated guesses [79]. These proteins and their recruitment have been implicated in a wide range of cellular functions. More prominent examples of proteins binding to specific histone modifications, the recognition modules, the influence of other modifications and the known associated functional consequence are summarized in Table 2

From the provided examples it is clear that while there are paradigms where only a single protein is known to recognize a particular modification, in the majority of cases there are multiple different factors, and the associated complexes, binding to a chromatin mark with similar, divergent or opposing functional consequences. Moreover, these interactions can in addition be dependent on the recognition of the surrounding amino acids sequence, as well as the surface of the histone proteins and DNA within the same or adjacent nucleosomes (Figure 2.5 and examples below). Therefore, it is not surprising that the chromatin associated proteins or complexes often are modular assemblies of several functional domains. On one hand, multivalent interaction could enhance specificity, and on the other – multiple weak interactions could contribute to overall stronger binding [6]. For example, the TAFI subunit of TFIID recognizes two acetylation marks on histone H3 (H3K5acK12ac) via double bromodomains [80]. Moreover, another subunit of TFIID, TAF3, specifically recognizes the H3K4me3 mark [81]. The interaction of the complex with H3K4me3 is thus stronger when this modification is flanked by the two acetyl marks [75]. Another level of combinatorial recruitment might be provided by a third TFIID subunit – TBP binds the TATA box DNA sequence and is important for recognition of TSS [82].

Multivalent interaction can also be achieved by a single multidomain chromatin factor. TRIM24 recognizes the nonmethylated H3K4 residue (H3K4me0) and at the same time H3K23ac [83]. UHRF1 is a multidomain protein that interacts with several H3 residues and in addition with DNA [84]. While the contribution of each binding event towards the overall strength of recruitment remains to be quantified accurately, it has been established that the protein recognizes the unmodified H3 tail, methylation of H3K9 and/or DNA, and its binding is antagonized by methylation of H3K4 and H3R2 [84]. Multivalent interaction can also be achieved by binding multiple targets from different histones within the same

Chromatin modifications and their readers

Mark	Histone	Position	Protein	Recognition domain	+/- effect from	Functions
Kme	H3	K4me0	WDR5	WD40		HAT
			UHRF1	PHD	H3R2me1/2	
		K4me	CHD1 (2/3)	Chromo	H3T3ph, H3R2me2a	ATPase
			Tip60 (1)	Chromo		Transcriptional activation
			WDR5 (1/2/3)	WD40		HAT
			ING1,2,4,5 (2/3)	PHD	H3T3ph, H3T6ph, H3R2me2s	HDAC, H3 acetylation
			BPTF (2,3)	PHD	H3K9/14ac, H3R2me2a	Nucleosome remodeling
			TAF3 (2,3)	PHD	H3K9/14ac, H3R2me2a	TFIID
			TRIM24 (2,3)	PHD		Chromatin activation
			PHF2 (3)	PHD		H3K9 demethylation
			YNG1 (3)	PHD		NuA3, histone acetylation
			PHF8 (3)	PHD	H3K9/14ac	Histone demethylation
			JMJD2A (2,3)	Tudor	H3T3ph	Hystone demethylase
			Sgf29 (3)	Tudor	H3K9/14ac	Histone acetylation
		K9	HP1 (2,3)	Chromo	Y41ph, S10ph	Heterochromatin
			Tip60 (2/3)	Chromo		DNA damage
			CDYL	Chromo	S10ph	Represor of REST
			UHRF1 (2,3)	Tudor	H3K4me2/3	DNA methylation
			EED (2/3)	WD40		PRC2 activity
			LRWD1	WD40		DNA replication (ORC binding)
		K23	MPP8 (3)	Chromo		
		K27	EED (2,3)	WD40		PRC mediated repression
			LRWD1	WD40	S28ph	DNA replication (ORC binding)
			PC (2/3)	Chromo		PRC1
			CDYL	Chromo		
			MPP8	Chromo		
		K36	Eaf3 (2/3)	Chromo		Histone deacetylation
			NSD1/2/3	PWWP		Histone methylayion
		K79	53BP1 (2)	Tudor		DSB response
	H4	K20	53BP1 (1/2)	Tudor		DNA damage repair
			ORC1 (1,2,3)	BAH		DNA replication
			EED (2,3)	WD40		PRC mediated repression
			LRWD1	WD40		DNA replication (ORC binding)
			JMJD2A (2,3)	Tudor	H3T3ph	
			PHF20L1	MBT		
Rme	H3	R2me	WDR5 (2s)	WD40		HAT
		R17	TDRD3 (2a)	Tudor		Transcriptional activation
	H4	R3	TDRD3 (2a)	Tudor		Transcriptional activation
			Dnmt3a (2s)			
			PCAF/p300 (2)			H3 acetylation
Sph	H3	S10	14-3-3	14-3-3		Transcriptional activation
			Gcn5			Histone acetylation
		S28ph	14-3-3	14-3-3		
Kub	H2B	K120	WDR82/Cps35			H3K4 methylation
Kac	H3	K14	DPF3b	Tandem PHD	K4me	Remodeling
		K23	TRIM24	Bromo		Chromatin activation
	H4	K5,8	Brdt	Bromo		Chromatin compaction
		K16	BPTF	Bromo		Nucleosome remodeling
			GCN5	Bromo		Histone acetylation

nucleosome. In this sense, BPTF recognizes H3K4me3 via a PHD finger and H4K16ac via a bromodomain, and this trans-histone recruitment shows remarkable selectivity [85]. There is also evidence for proteins and protein complexes interacting in trans-nucleosome fashion [6]. L3MBTL1 can compact nucleosomal arrays dependent on H4K20me1/2 and H1bK26me1/2, and the MBT domains of this protein can interact with at least two nucleosomes simultaneously [86]. Lastly, long non-coding RNAs (lncRNAs) can also function in targeting chromatin proteins. This has been shown for the polycomb repressive complexes PRC1 and PRC2 [6]. lncRNAs have also been implicated in targeting HP1, a prominent H3K9me binder, to pericentromeric heterochromatin [87]. Examples of several possible binding modes are illustrated in Figure 2.5, however, different combinations and variations of these are also feasible.
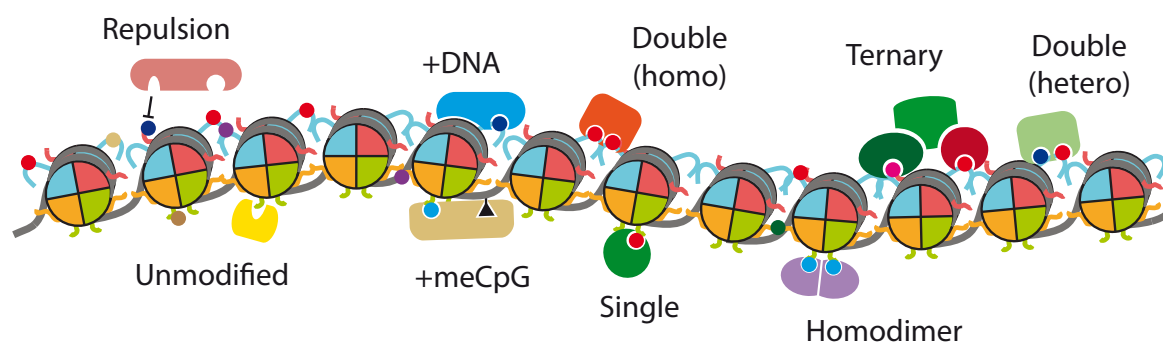
Figure 2.5 Binding modes of proteins recruited to chromatin.

Several scenarios of binding modes depending on histone and/or DNA modifications. Different modifications are represented as coloured circles on the histone tails or triangle on DNA. A protein can interact with one modification and be ejected by another (**Repulsion**); binding can be achieved via histone modification and DNA (**+DNA**) or meCpG (**+meCpG**) interacting domains; recruitment by two identical (**Double homo**) or different (**Double hetero**) modifications; a protein can bind unmodified residue (**Unmodified**) or a single modification (**Single**); a homodimer can be stabilized by two modifications (**Homodimer**); the recruitment of a factor can depend on independent binding of two other proteins (**Ternary**).

← **Table 2 Histone modifications and their binding proteins**

Examples of the most widely studied histone modifications and their binding proteins. The recognition domain, known cross-talk with other modifications and functional relevance are indicated. Based on [6] and references therein and [32, 83-85, 93-99]

Besides in binding of factors to multiple chromatin modifications, a different level of "cross-talk" can be achieved by recruitment of effector proteins via secondary interactions. Importantly, these effector proteins can function as "erasers" or "writers" of additional chromatin modifications. An example of this is the recruitment of H3K4 and H3K79 methyl transferases by recognition of ubiquitylation of H2B [88, 89]. In the case of repressive marks, methylation can recruit enzymatic activities for heterochromatin spreading. In the first example, the EED subunit of the polycomb complex PRC2 is recruited by trimethylated repressive marks and the binding results in allosteric activation of H3K27 methylation activity [90]. Methylation of H3K9, on the other hand, can recruit different HP1-containing complexes and result in DNA methylation and histone deacetylation [91, 92]. Another link between histone modification and DNA methylation are the examples of HDAC recruitment by meCpG binding proteins discussed in the previous section.

Thus, it is clear that the chromatin readout by modifications of histones and DNA is a complex process where multiple synergistic or antagonistic levels of interaction can exist. While there are systematic efforts to catalog the proteins binding to individual or limited sets of modifications [32, 75-78, 81], the examples given above are among the few available where cross-talk between individual modifications has been demonstrated. Examples for recognition of multiple modifications in a code-like manner by individual proteins or protein complexes are even scarcer. There is more and more evidence regarding the transcriptional control of gene-rich areas of the genome, however, it is still far from clear what is the function of the discovered complex patterns of modifications (Figure 2.3 and 2.4). Even though these patterns are thought to be less complex in heterochromatic regions, there is a clear difference in the distribution of individual chromatin marks among separate subdomains. On a more global scale it remains to be revealed how these subdomains function differentially, what is the interactome of a chromatin domain as determined by its modification signature and what is the interplay and connections between the proteomes of multiple modifications and combinations thereof.

Methods for systematic analysis of chromatin modification readout

Several array-based technologies have been used to screen the interaction of modified histone peptides with immobilized reader domains [63, 100], or vice versa using immobilized peptides to screen for a library of interaction partners [101]. However, these methods are restricted to the use of libraries of prey molecules and cannot identify unpredicted interactions. Therefore, affinity approaches for purification of proteins from cellular sources have become the method of choice for *de novo* identification of histone and DNA binding proteins and complexes. Like array-based assays, these methods can function in two ways. In one example, the chromatin protein 53BP1 was used as bait to screen native

purified histones. Mass spectrometric analysis then revealed that H3 carrying methylation at K79 was preferentially bound [102]. On the other hand, affinity approaches using N-terminal modified histone peptides have proven extremely useful for identification of multiple modification-dependent interactions of proteins and protein complexes [103]. In peptide pull-downs a chemically synthesized modified peptide is incubated with isolated nuclear extracts, the bound proteins are analyzed by mass spectrometry, and the results compared to a control experiment using an unmodified counterpart peptide. This approach is straightforward and easy to implement, and has recently been coupled to quantitative mass spectrometry and used for cataloguing the interaction partners of several histone modifications [32, 75, 76]. However, as the peptides in this type of assay can only be synthesized in limited size, they reflect the local environment of the histone modification and cannot replicate additional interactions with histones, DNA or other non-adjacent modifications. Therefore, affinity approaches using recombinant chromatin templates are of great interest. Uniformly and homogeneously modified histones can be created by one of several available approaches [29, 104, 105] and can be reconstituted into mononucleosomes or nucleosomal arrays [106-108]. These complexes not only mimic the native chromatin *in vivo*, but also allow for the incorporation of multiple histone and DNA modifications in combination. Therefore, affinity purification using homogeneously modified *in vitro* reconstituted chromatin species is a promising candidate assay for the elucidation of the complex interactions and cross-talk between different chromatin modifications and their binding partners. Moreover, coupled with state-of-the-art quantitative mass spectrometry, it will allow unbiased and sensitive interrogation of the chromatin modification associated interactomes.

2.3   Mass spectrometry for protein identification and quantification

2.3.1   Protein identification by mass spectrometry

Proteins are one of the major classes of macromolecules in living cells and participate in virtually every molecular process. Unambiguous protein identification is therefore crucial for all areas of life science research. A classical approach for establishing protein identity is immuno-detection with antibodies. However, this method is targeted and therefore biased and also with low-throughput. One of the pioneering methods for unbiased protein sequence determination is Edman degradation [109]. It relies on chemical cleavage of N-terminal amino acid residues in a step-wise manner. The technique has significantly pushed forward protein-centered research, however, it requires a large amount of homogeneous samples, free N-termini and even in an automatized manner is still very time-

consuming. Because of these limitations, protein analytical methods based on mass spectrometry have been developed. The routine exploitation of these methods for large nonvolatile molecules became possible after the advent of the soft ionization techniques of matrix-assisted laser desorption/ionization (MALDI) [110, 111] and electrospray ionization (ESI) [112]. Instrumental for the wide adoption of these approaches was the knowledge of the genome sequence, and hence the proteome sequence, of multiple organisms. The general workflow of a typical proteomics experiment consists of several steps (Figure 2.6): a) hydrolysis of the analyzed proteins into peptides; b) determination of the peptide mass-to-charge ratio (*m/z*) by MS analysis; c) selection of precursor peptide ions and their fragmentation using MS/MS; d) determination of the peptide sequence and identification of the peptides, and consequently the proteins, by comparing the experimentally derived sequences with protein databases using software search engines.
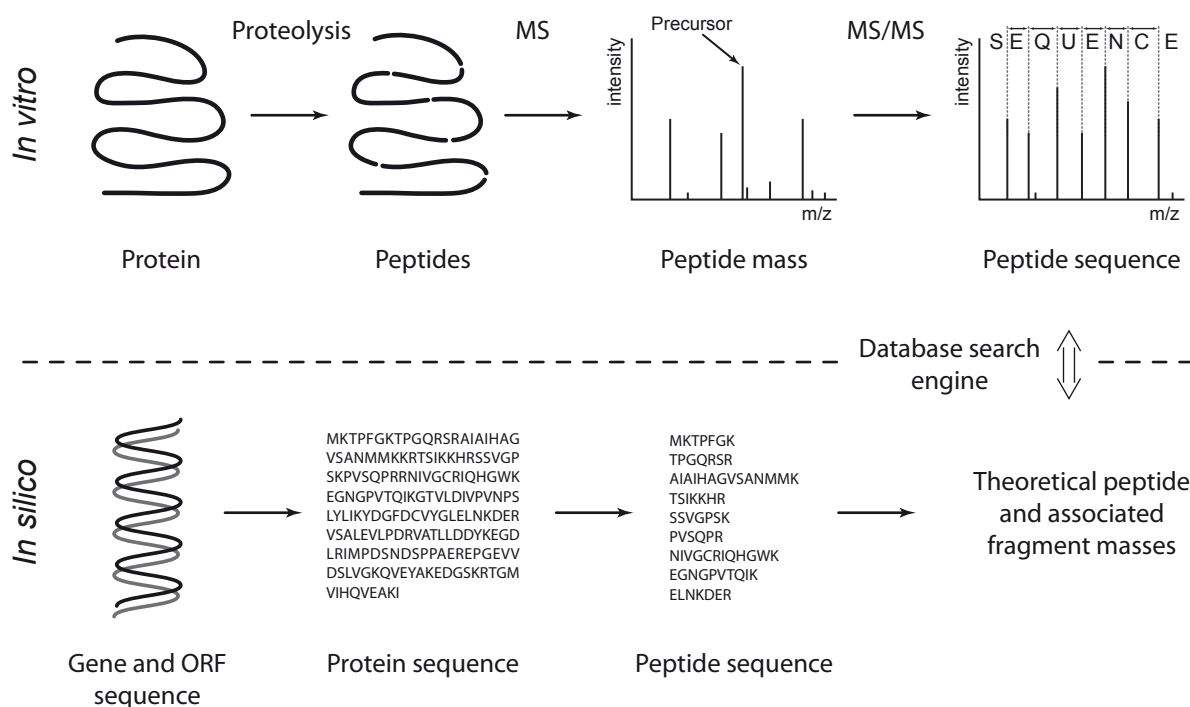
Figure 2.6 MS-based protein identification workflow.

Proteins are hydrolyzed using specific endoproteases, the resulting peptides are analyzed by MS and selected for fragmentation by MS/MS. The precursor and fragment *m/z* values are compared to the theoretical ones from a database by a search engine. Redrawn with permission from [113].

Depending on the size of the protein, the standard proteomics workflow can result in a complex mixture of peptides. As the mass spectrometers can analyze only a limited number of peptides at a time, even for single protein analysis a peptide separation step for reducing the sample complexity is beneficial. For complex proteomics samples with hundreds or thousands of proteins, separation of the peptides is usually not sufficient and reducing the complexity of the sample at protein level is required. Proteins can be separated using standard SDS-PAGE, followed by excision of protein bands and in-gel proteolysis [114]. This approach is very robust and has been the method of choice for numerous proteomics studies. Alternatively, proteins can be separated by isoelectric focusing, size exclusion or ion exchange chromatography followed by in-solution protease digestion.

After protein digestion, the resulting peptides can be separated by different approaches. The most routine one relies on reversed-phase (RP) C18 chromatography and direct elution of the separated peptides into an ESI mass spectrometer. Additionally, peptide pre-fractionation can be done using strong cation exchange (SCX) chromatography. Often the separation at the level of proteins can be circumvented by using high performance peptide separation. This can be achieved by two-dimensional separation, e.g. SCX followed by RP chromatography [115], or peptide isoelectric focusing followed by RP chromatography [116]. Alternatively, novel filter-assisted sample preparation methods [117] and ultra high performance chromatography can be optimized for single step near complete proteome coverage [118].

Regardless of the separation technique used, the resulting peptides are further analyzed by mass spectrometry. In case of MALDI, the peptides are spotted and co-crystalized together with a matrix on a stainless steel target and analyzed "off line". This has the advantage that the peptides can be stored at room temperature and analyzed several times at different time points. For high-throughput identification, however, ESI offers immediate and fast analysis by directly coupling the last chromatography step to the mass spectrometer. Capillary or nano liquid chromatography methods use very low flow rates and sample amounts and are ideally suited for coupling to ESI mass spectrometers.

After elution of the peptides through the ESI source into the mass spectrometer, their *m/z* ratio is measured using different physical approaches depending on the specific mass analyzer. Nevertheless, most mass spectrometers used in routine proteomics research operate by data-dependent-acquisition (DDA). A DDA analysis cycle consists of several steps. First, a complete MS scan over a specific mass range (e.g. 350-1600 *m/z*) detects the *m/z* ratios of all peptides eluting at that particular moment. The most intense precursor ions are then selected consecutively and fragmented independently in a MS/MS experiment. The number of selected peptides per MS scan depends on the speed of the mass analyzer and can range from 2-3 to over 20. Typically, the two stages of analysis are performed one after

another. Therefore, a faster mass analyzer can measure a higher number of peptides during the length of their LC elution profile and detection of peptides eluting later is not disturbed. Apart from speed, important features for high sensitivity are the accuracy and resolution of the mass analyzers. Higher accuracy results in more confident *m/z* and therefore mass assignment and better identification. The resolution is typically measured as the full peak width at half maximum intensity (FWHM). Higher resolution allows better *m/z* accuracy and the distinction between peaks with similar *m/z*. Resolutions of 60,000 and above and accuracy in the low ppm to sub-ppm range are now routine, additionally, accuracy can be further improved by software post-processing [119, 120].

2.3.2 MS-based quantification of peptides and proteins

The application of mass spectrometry for identification of proteins is well established. These approaches allow the analysis of individual proteins or almost entire proteomes within single experiments [118], mapping and localization of post-translational modifications, elucidation of three-dimensional structure and contact sites within macromolecular complexes [121]. The use of mass spectrometry to identify interaction partners of specific baits is also well documented. However, the use of protein identity when comparing bait and control has several disadvantages. First, proteins binding in both experiments but with higher affinity for the bait cannot be identified as specific binders, thus resulting in a high false negative rate. Furthermore, it is difficult to determine whether bait-prey interaction is strong or weak. Therefore, in the last decade, much effort has been invested in developing various MS-based quantitative approaches. The applications of these methods range from focused analysis of the subunit stoichiometry of individual protein complexes [122], to almost entire proteome comparison between cell types [123].

General overview of MS-based quantification

Mass spectrometry is not a quantitative method *per se*. Due to different physicochemical properties of different peptides and proteins, their signals within the mass spectrometer cannot be used for comparison. Quantification can be done only on chemically identical species between measurements or, if they differ in isotope composition, and therefore in mass, within the same measurement. The first group of methods is termed label-free quantification. It relies on comparison of the signals from identical peptides and proteins between different measurements. The second group is based on stable isotope labeling techniques. Here, the peptide or protein species to be compared are differentially labeled with stable heavy isotopes and measured simultaneously. This results in chemically identical molecules that have different mass when measured by MS. In both label-free and label-

based approaches the signals between the investigated molecules are compared. This comparison can be relative – direct measurement of the relative amounts of each species, or absolute – determination of the absolute amount of analyte based on spiked-in standards. The most widely used methods for label-free and label-based relative or absolute quantification are presented in Figure 2.7 [124].

Relative quantification

Absolute quantification

Label-based

Metabolic
N^{15}
SILAC

Chemical
ICAT
ICPL
iTRAQ, TMT
IPTL
DML
mTRAQ

Enzymatic
O^{18}

Label-free
Ion intensities (XIC)
Peptide/spectral counting

Label-based
AQUA
PSAQ, Absolute SILAC
FLEXIQuant
QCONcat

Label-free
(em)PAI
APEX
Top3
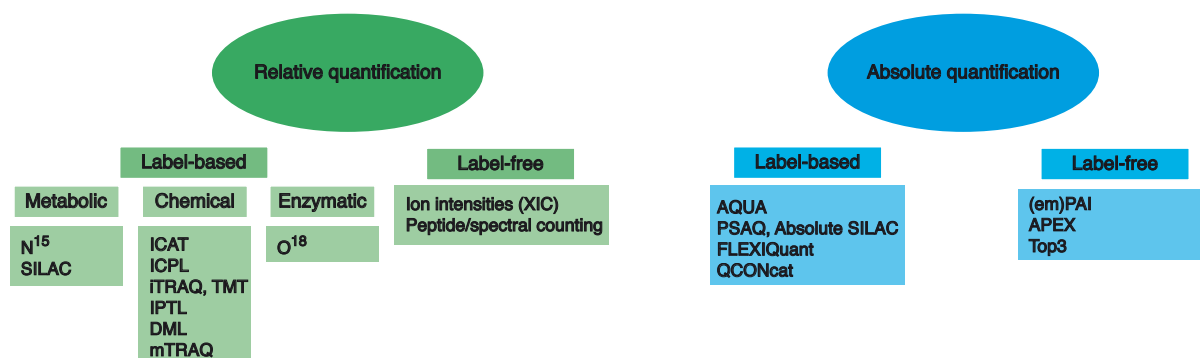
Figure 2.7 Approaches for label-free and label-based relative or absolute MS quantification.

Most widely used methods for relative (left) and absolute (right) MS-based quantification. Adapted from [124] with kind permission from Springer Science and Business Media.

Label-free approaches [124, 125] have the advantage of comparing virtually unlimited number of samples, without additional cost for labeling procedures. They can be based on counting the number of peptides or spectra per peptide in each measurement, or comparing the intensities of precursor peptides on the MS detector. However, these methods rely on highly reproducible sample handling and chromatography separation, and usually require multiple technical replicates. Additionally, quantification of PTMs between different experiments can be very unreliable. These limitations encouraged the development of more accurate methods for direct quantitative comparison within the same mass-spectrometric measurement using stable isotope labeling. Due to their natural occurrence, each peptide/protein contains a certain proportion of stable heavy isotopes of carbon, nitrogen, oxygen and hydrogen (e.g. ^{13}C, ^{15}N, ^{18}O, ^{2}H, respectively) that produce a mass shift and thus – the characteristic isotope envelope seen in MS measurements. The difference in signal intensities between the individual peaks of this envelope reflect accurately the natural abundance of the heavy isotopes contained therein. Therefore, artificial incorporation of such heavy isotopes can be used for protein or peptide labeling and quantification. In this way, a heavy labeled peptide will produce a specific mass shift compared to a light labeled species. The intensities of both isotopomers can be measured within the same spectrum and will be directly proportional to the relative abundance of each

analyte. Stable isotope labeling can be established both *in vivo*, by metabolic labeling [126, 127], or *in vitro*, by chemical [128, 129] or enzymatic [130] labeling. The label-based techniques offer high accuracy but rely on additional sample preparation steps for labeling. This also increases the costs, compared to label-free approaches. Moreover, only a limited number of samples can be compared within one experiment (from two to eight, depending on the method) [124]. However, recent reports describe significant increase in the multiplexing power of combinations of chemical and metabolic labeling [131]. This requires sophisticated computational analysis but seems to be extremely useful for large-scale multiplexing experiments.

Stable isotope labeling by amino acids in cell culture

Stable isotope labeling by amino acids in cell culture (SILAC) [126] is one of the most widely used methods for relative quantification by MS. SILAC, and metabolic labeling in general, introduces heavy isotopes during cellular or organismal growth. This offers a significant advantage over chemical labeling approaches where the isotopes are incorporated later in the experimental workflow. Early labeling and mixing of the samples reduces separate and potentially differential handling steps of the analytes to be compared. On the other hand, the requirement for metabolic labeling limits the type of samples that can be used for comparison. There are hybrid approaches that successfully use SILAC samples for spike-in comparison of samples that cannot be labeled [132]. A typical workflow of a SILAC experiment is shown in Figure 2.8. A cell population is divided in two and cultured in light and heavy SILAC medium, respectively. The two medium preparations are identical, apart from the fact that all arginine and lysine sources are depleted and replaced with amino acids containing a specific number of light (^{12}C, ^{14}N, ^{1}H) or heavy (^{13}C, ^{15}N, ^{2}H) carbon and/or nitrogen and/or hydrogen atoms. After complete incorporation of the heavy isotopes, the SILAC experiment is performed (e.g. differential treatment, overexpression of a protein, immunoprecipitation or pull-down) and the light and heavy samples are mixed and processed for MS analysis. Here, each peptide will be represented as a SILAC pair with a certain mass shift (depending on the number of heavy atoms used for labeling). Comparison of the intensities between the heavy and light peaks of each SILAC pair gives the relative enrichment ratio for each peptide and by extension, each protein.
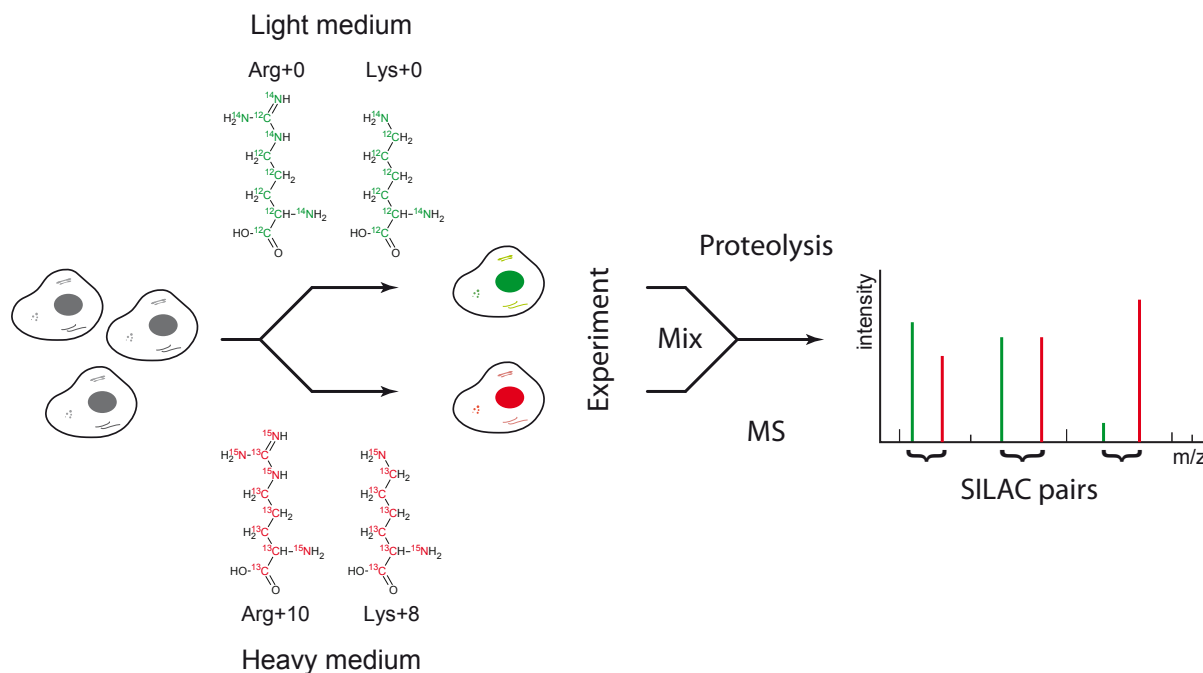
Figure 2.8 Schematic representation of a SILAC experimental workflow.

Workflow of a typical SILAC experiment. First, two cell populations are labeled with light or heavy SILAC medium, then an experiment is performed (e.g. differential treatment of the cells, isolation of extract and immunoprecipitation, etc.), the samples are mixed and measured by MS. Here in the MS spectrum each peptide is represented by SILAC pair peaks with different intensity in light (green) or heavy (red), depending on the relative abundance of the specific protein.

All large-scale MS-based quantification approaches are greatly dependent on software tools for identification and quantification of the proteins within the compared samples [133]. Apart from the straightforward experimental design, another key aspect in the success of the SILAC approach has been the availability of an advanced and semi-automatic software environment for data analysis and interpretation. MaxQuant [119] implements complex algorithms for all steps of data analysis, starting with raw MS data and finishing with statistical evaluation and visualization. The first step in the data analysis is feature detection and processing. This includes 3D peak (intensity, *m/z*, elution time) detection, de-isotoping and SILAC pair identification. Furthermore, this step allows for software peptide mass recalibration improving accuracy. This information, together with the deconvoluted MS/MS spectra, is used for peptide identification with a built-in database search engine [120]. After peptide identification, the protein identities are inferred. Next, the enrichment ratio for each individual SILAC pair is calculated, normalized and used for estimation of the protein enrichment ratio. In cases where only one of the SILAC pairs is detected, a re-quantification step estimates enrichment ratio based on the baseline noise in the spectrum. Furthermore, PTM analysis can also be performed, including localization, occupancy and quantification of

modification sites. The latest versions of MaxQuant allow for data analysis of label-free, as well as chemical label-based experiments.

2.4 Aims and objectives of the presented thesis

The role of histone modifications and their binding partners is undoubtedly at the center of genome biology and chromatin regulation. The plethora of identified modifications and their specific patterns of genome distribution is striking. However, little is known about the proteomes of individual chromatin domains. Are there exclusively unique proteins recruited to each distinct modification? What is the set of proteins specific for functionally related chromatin modifications? Which factors are recruited to chromatin as a function of a combinatorial modification signature? Research in this direction led to cataloguing the interactomes of several individual modifications, however these questions are still open. The majority of the approaches used in the field of chromatin modification proteomics are based on *in vitro* assays using isolated modified or unmodified histone N-terminal peptides. Therefore, the first aim of this study was to establish a novel affinity purification approach for the identification of chromatin modification binding proteins from mammalian cell nuclear extracts using homogeneous *in vitro* reconstituted chromatin templates. False positive hits are a major problem in all pull-down experiments, therefore the assay needed to be set up utilizing SILAC-labeled extracts for quantitative mass spectrometry. In order to validate the novel approach, it had to be compared to classical peptide pull-down using the two most widely studied histone modifications – H3K4me3 and H3K9me3. Following successful validation, the chromatin affinity purification was to be applied to describe the interactomes of a set of heterochromatic modifications, in addition to several euchromatic ones. The binding proteins of each modification were then to be used to deduce novel functional features and relationships within and between the interactomes. Similar analyses were also intended to be applied for identification of proteins binding to combinations of modifications. Lastly, the functional consequences of recruitment of selected factors needed to be interrogated using *in vitro* and *in vivo* approaches.

3 Material and Methods

3.1 Material and reagents

3.1.1 Laboratory equipment and instruments

ÄKTA Explorer/Purifier	GE Healthcare, Buckinghamshire (UK)
Balances	Metler-Toledo, Giesen (DE)
BBD 6220 CO_2 incubator	Heraeus, Hanau (DE)
Bioreactor 5L with ez-Control	Applikon, Schiedam (NL)
Centrifuge Cryofige 6000i	Heraeus, Hanau (DE)
Centrifuge Sorvall Evolution RC	Thermo Scientific, Braunschweig (DE)
Centrifuges tabletop 5415R/5810R	Eppendorf, Hamburg (DE)
Electrophoresis power supplies	Bio-Rad, München (DE)
Gel Doc 2000 gel documentation system	Bio-Rad, München (DE)
HP1100 and HP1200 LC systems	Agilent, Santa Clara (USA)
Kodak X-Omat 2000 processor	Carestream Health, New York (USA)
Laminar flow clean bench	Heraeus, Nahau (DE)
LTQ Orbitral XL	Thermo Fischer Scientific, Bremen (DE)
LTQ Orbitrap Velos	Thermo Fischer Scientific, Bremen (DE)
LTQ XL	Thermo Fischer Scientific, Bremen (DE)
Mini Trans-Blot system	Bio-Rad, München (DE)
Mini-PROTEAN Tetra PAGE cell	Bio-Rad, München (DE)
NanoDrop ND-1000	Peqlab, Erlangen (DE)
Perfection V700 PRO scanner	Epson, Nagano (JP)
Peristaltic pump	Ismatec, Glattburgg (CH)
pH meter	Metler-Toledo, Giesen (DE)
Sonication bath SONOREX Super	BANDELIN Electronic, Berlin (DE)
Sorval SA600 rotor	Thermo Scientific, Braunschweig (DE)
Sorval SS34 rotor	Thermo Scientific, Braunschweig (DE)
SpeedVac Savant SPD121P	Thermo Scientific, Braunschweig (DE)
Sub-Cell-GT agarose gel electrophoresis	Bio-Rad, München (DE)
Thermocycler epgradientS	Eppendorf, Hamburg (DE)
Thermomixer Comfort	Eppendorf, Hamburg (DE)

UV Transiluminator	Bio-Rad, München (DE)
Water bath TW12	Julabo, Selbach (DE)
XCell Sure Lock Mini NuPAGE cell	Invitrogen, Karlsruhe (DE)

3.1.2 Chemicals and reagents

(2-chloroethyl)-methylammonium chloride	Karl Industries, Aurora (USA)
2-mercaptoethanol	Sigma-Aldrich, Steinheim (DE)
4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	VWR, Poole (DE)
Acetic acid	Merck, Darmstadt (DE)
Acetonitrile, LiChrosolv	Merck, Darmstadt (DE)
Acrylamide/Bisacrylamide (37.5:1)	Merck, Darmstadt (DE)
Agarose	Serva, Heidelberg, (DE)
Ammonium hydrogen carbonate	Fluka, Buchs (CH)
Ammonium peroxodisulfate	AppliChem, Darmstadt (DE)
Boric acid	Merck, Darmstadt (DE)
Bovine serum albumin (BSA)	Sigma-Aldrich, Steinheim (DE)
Bromophenol blue	Serva, Heidelberg (DE)
Cleland's reagent (DTT, for MS analysis)	Calbiochem, Darmstadt (DE)
Coomassie Briliant Blue G-250	Fluka, Buchs (CH)
D/L-Methionine	Sigma-Aldrich (DE)
Deoxynucleotide-5'-phosphate (dATP, dCTP, dGTP, dTTP)	Roth, Karlsruhe (DE)
Dithioerythrol (DTE)	Roth, Karlsruhe (DE)
Dithiothreitol (DTT)	Alexis Biochemicals, Farmingdale (USA)
Ethanol	Merck, Darmstadt (DE)
Ethidium bromide (solution 10ml/ml)	Roth, Karlsruhe (DE)
Ethylendiamine tetraacetate (EDTA)	Roth, Karlsruhe (DE)
Formic acid (FA)	Fluka, Buchs (CH)
Glycerol	Merck, Darmstadt (DE)
Guanidine hydrochloride	Sigma-Aldrich, Steinheim (DE)
Hydrochloric acid (37% HCl)	Merck, Darmstadt (DE)
Iodacetamide (IAA)	Sigma-Aldrich, Steinheim (DE)
Magnesium chloride ($MgCl_2$)	Merck, Darmstadt (DE)
Methanol, LiChrosolv	Merck, Darmstadt (DE)
N,N,N',N'-Tetramethylethylendiamid (TEMED)	Sigma-Aldrich, Steinheim (CH)

Non-fat dry milk powder	Regilait, Saint-Martin-Belle-Roche (FR)
Ortho-Phosphoric acid	Merck, Darmstadt (DE)
Phenol:Chlorophorm:Isoamil alcohol (PCI) [25:24:1]	Roth, Karlsruhe (DE)
Phenyl-methylsulfonyl fluoride (PMSF)	Roche, Mannheim (DE)
Poluethylene glycol 6000	Merck, Darmstadt (DE)
Ponceau S	Sigma-Aldrich, Steinheim (DE)
Potassium chloride (KCl)	Merck, Darmstadt (DE)
Potassium dihydrogen phosphate (KH_2PO_4)	Roth, Karlsruhe (DE)
S-(5'-adenosyl)-L-methionine (SAM)	New England Biolabs, Ipswitch (USA)
Sodium acetate	Roth, Karlsruhe (DE)
Sodium azide (NaN_3)	Alfa Aesar, Massachusetts (USA)
Sodium chloride (NaCl)	Merck, Darmstadt (DE)
Sodium dodecyl sulfate	VWR, Poole (DE)
Sodium hydrogen phosphate (Na_2HPO4)	Merck, Darmstadt (DE)
Sodium hydroxide (NaOH)	Merck, Darmstadt (DE)
Triethanolamine	VWR, Poole (DE)
Tris(hydroxymethyl)amino ethane (Tris base)	Roth, Karlsruhe (DE)
Triton X-100	Merck, Darmstadt (DE)
Tween-20	Sigma-Aldrich, Steinheim (DE)
Urea	Merck, Darmstadt (DE)
Water, LiChrosolv	Merck, Darmstadt (DE)

3.1.3 Cell culture media and materials

DMEM High Glucose (w/o Lys, w/o Arg)	PAA Laboratories, Colbe (DE)
Foetal bovine serum, dialyzed	PAA Laboratories, Colbe (DE)
L-Arginine (Arg0)	Sigma-Aldrich, Steinheim (DE)
L-Arginine, $^{13}C_6$ (Arg6)	Euriso-top, Saarbrücken (DE)
L-Arginine, $^{13}C_6$, $^{15}N_4$ (Arg10)	Euriso-top, Saarbrücken (DE)
L-Lysine (Lys0)	Sigma-Aldrich, Steinheim (DE)
L-Lysine, $^{13}C_6$, $^{15}N_2$ (Lys8)	Euriso-top, Saarbrücken (DE)
L-Lysine, $^2D_4$ (Lys4)	Euriso-top, Saarbrücken (DE)
Penicillin/Streptomycin 100x	PAA Laboratories, Colbe (DE)

3.1.4 Cell lines

HeLa S3 (human cervical cancer, Computer cell culture centre, BE) were kindly provided by Thomas Conrad (Facility for Cell Production, Max Planck Institute for Biophysical Chemistry, Göttingen, DE).

3.1.5 Commercial kits, buffers and solutions

1 Kb Plus DNA ladder	Invitrogen, Karlsruhe (DE)
Amersham Hyperfilm ECL	GE Healthcare, Buckinghamshire (UK)
ECL Pus Western Blotting Detection System	GE Healthcare, Buckinghamshire (UK)
ECL Western Blotting Detection System	GE Healthcare, Buckinghamshire (UK)
Hybond ECL nitrocellulose membrane	GE Healthcare, Buckinghamshire (UK)
Hybond N+ membrane	GE Healthcare, Buckinghamshire (UK)
Imperial Protein Stain	Pierce/Thermo Scientific, Rockford (USA)
Mini-PROTEAN 4-12% TGX gels	Bio-Rad, München (DE)
Mini-PROTEAN 4-20% TGX gels	Bio-Rad, München (DE)
NucleoSpin Extract II	Machery&Nagel, Düren (DE)
NuPAGE Antioxidant	Invitrogen, Karlsruhe (DE)
NuPAGE LDS Sample Buffer (4x)	Invitrogen, Karlsruhe (DE)
NuPAGE MOPS SDS Running Buffer (10x)	Invitrogen, Karlsruhe (DE)
NuPAGE Novex 4-12% Bis-Tris gels, 1mm	Invitrogen, Karlsruhe (DE)
NuPAGE Sample Reducing Agent (10x)	Invitrogen, Karlsruhe (DE)
Plasmid Giga Kit	Qiagen, Hilden (DE)
QIAquick PCR purification kit	Qiagen, Hilden (DE)
SeeBlue Plus2 Protein Standard	Invitrogen, Karlsruhe (DE)

3.1.6 Chromatography and affinity media

HiLoad 10/300 Superdex200	GE Healthcare, Buckinghamshire (UK)
HiLoad 16/60 Superdex200	GE Healthcare, Buckinghamshire (UK)
PD-10 columns	GE Healthcare, Buckinghamshire (UK)
Reprosil AQ-3/5µm / 300Å	Dr.Maisch, Ammerbuch (DE)
SilicaTip emmiters	New Objective, Woburn (USA)
Streptavidin MagneSphere particles	Promega, Mahnneim (DE)

3.1.7 Consumables and plasticware

Amicon Ultra centrifugal filter devices (MWCO 3 and 10 kDa)	Millipore, Billerica (USA)
Phase Lock Heavy Tubes (2, 15, 50 ml)	5PRIME, Hamburg (DE)
Slide-A-Lyzer dialysis units and cassettes (MWCO 3500 and 10,000)	Pierce/Thermo Scientific, Rockford (USA)
Spectra/Por dialysis membrane (MWCO	Spectrum Laboratories, Rancho Domingues

3500 and 10,000)   (USA)

3.1.8   Enzymes, proteins and inhibitors

Antarctic phosphatase	New England Biolabs, Ipswitch (USA)
Benzonase	Calbiochem, Darmstadt (DE)
H2Bub1	Moshe Oren (WIS, Rehovot, IL)
M.SssI CpG methyltransferase	New England Biolabs, Ipswitch (USA)
Micrococcal nuclease	Calbiochem, Darmstadt (DE)
Proteinase Inhibitor Cocktail Complete, EDTA free	Roche, Mannheim (DE)
Restriction endonuclease enzymes	New England Biolabs, Ipswitch (USA)
SWI/SNF complex (INI1-FLAG)	Eric Allemand (Institut Pasteur, Paris, FR)
T4 DNA ligase	New England Biolabs, Ipswitch (USA)
Taq polymerase	Winfried Lendeckel (MPIbpc)
Trypsin	Roche, Mannheim (DE)
Trypsin	Promega, Mannheim (DE)
Ubiquitin-biotin	Stefan Kopperman (Max Planck Institute for Biophysical Chemistry, Göttingen, DE)

3.1.9   Peptides

N-terminal histone H3 peptides for pull-down experiments were synthesized in unmodified or modified form using Fmoc-based solid-phase synthesis. C-terminal non-native lysine was used for biotinylation. The peptides obtained from the Department of NMR-based Structural Biology at the Max Planck Institute for Biophysical Chemistry, Göttingen, DE.

H3 unmodified	NH_2 - ARTKQTARKSTGGKAPRKQL - biotin
H3K4me3	NH_2 - ARTK$_{me3}$QTARKSTGGKAPRKQL - biotin
H3K9me3	NH_2 - ARTKQTARK $_{me3}$STGGKAPRKQL - biotin

3.1.10 Antibodies and sera

Reagents and their dilutions used for immunodetection of proteins in Western Blotting:

Name	Host	Supplier, catalog number	Dilution
α-beta-tubulin	mouse, monoclonal	Sigma-Aldrich,	1:20,000
α-BRG1	mouse, monoclonal	Santa Cruz, sc-17796	1:2,000
α-BRM	rabbit, polyclonal	Eric Allemand (Institut Pasteur, Paris, FR)	1:500
α-DYRK1A	rabbit, polyclonal	Cell Signaling, 2771	1:1,000
α-FLAG M2	mouse, monoclonal	Sigma-Aldrich, F1804	1:1,000
α-H2A	rabbit, polyclonal	Abcam, ab13923	1:1,000
α-H3	rabbit, polyclonal	Abcam, ab1791	1:20,000
α-H3K4me3	rabbit, polyclonal	Active Motif, 39159	1:2,000
α-H3K9me3	rabbit, polyclonal	Upstate/Millipore, 07-442	1:1,000
α-H4	rabbit, polyclonal	Upstate/Millipore, 07-108	1:500
α-INI1	rabbit, polyclonal	Eric Allemand (Institut Pasteur, Paris, FR)	1:500
α-INTS3	rabbit, polyclonal	Abcam, ab91419	1:1,000
α-MORC2	rabbit, polyclonal	Bethyl, A300-149A	1:2,000
α-NELFA	rabbit, polyclonal	Yuki Yamaguchi (TIT, Tokyo JP)	1:2,000
α-RCC1	mouse, monoclonal	Abcam, ab54600	1:1,000
α-RECQL5	rabbit, polyclonal	Abcam, ab91419	1:1,000
α-BAF155	rabbit, polyclonal	Abcam, ab22355	1:400
α-SMCHD1	rabbit, polyclonal	Abcam, ab31865	1:1,000
α-SUPT5H	rabbit, polyclonal	Yuki Yamaguchi (TIT, Tokyo JP)	1:2,000

α-rabbit-HRP	swine, polyclonal	Dako, P0399	1:5,000
α-mouse-HRP	goat, polyclonal	Dako, P0447	1:5,000

3.1.11 Plasmids

Name	Promotor	Resistance	Application	Source
pUC18_12x200-601	T7	Ampicillin	Chromatin array template (12x200-601)	Daniela Rhodes (MRC, Cambridge, UK) [107]

3.1.12 DNA oligonucleotides

Oligonucleotides were purchased from Eurofins MWG Operon (Edersberg, DE) and kindly provided by Alexandra Stützer (Chromatin Biochemistry Group, Max Planck Institute for Biophysical Chemistry, Göttingen, DE).

Primers for PCR amplification of scavenger DNA for chromatin array reconstitution:

Forward   5' GTTATCCGCTCACAATTCCACACAACATAC 3'
Reverse   5' TAATGCAGCTGGCACGACAGGTTTC 3'

Oligonucleotides for labeling of 12x200-601 template for chromatin array reconstitution:

EcoRI_3'P   5'-GGGGGGGGGATCCGGGGGGGGp-3'
EcoRI_5'P_3'bio   5'-pAATTCCCCCCCGGATCCCCCCCC-biotin-3'

3.1.13 Commonly used buffers

<u>10x DNA loading dye</u>   30 % [v/v]   Glycerol
                                          10 mM   EDTA
                                            0.25 % [w/v]   Bromophenol blue

<u>10x PBS</u>   1.37 M   NaCl
                                            27 mM   KCl
                                            100 mM Na_2HPO_4
                                            20 mM KH_2PO_4

Material and reagents

10x SDS running buffer	250 mM	Tris base
	1.92 M	Glycine
	1 % [w/v]	SDS

| 10x TB | 890 mM | Tris base |
| | 890 mM | Boric acid |

10x TBE	890 mM	Tris base
	890 mM	Boric acid
	20 mM	EDTA NaOH (pH 8.0)

| 1x PBS-T | 1x | PBS |
| | 0.1 % [v/v] | Tween-20 |

| 2x chromatin loading buffer | 10 % [v/v] | Glycerol |
| | 1x | RB low |

Colloidal Coomassie stain	0.08 % [w/v]	Coomassie Brilliant Blue G 250
	20 % [v/v]	Methanol
	1.6 % [v/v]	Ortho-Phosphoric acid
	8 % [w/v]	Ammonium sulfate

PD150 (pull-down wash buffer)	20 mM	HEPES NaOH (pH 7.9)
	10 % [v/v]	Glycerol
	150 mM	NaCl
	0.1 % [v/v]	Triton X-100

RB High	10 mM	Tris-HCl (pH 7.5)
	1 mM	EDTA NaOH (pH 8.0)
	2 M	NaCl
	1 mM	
		DTT

RB Low	10 mM	Tris-HCl (pH 7.5)
	1 mM	EDTA NaOH (pH 8.0)
	10 mM	NaCl
	1 mM	DTT

Röder C	25 % [v/v]	Glycerol
	20 mM	HEPES
	420 mM	NaCl
	1.5 mM	$MgCl_2$
	0.2 mM	EDTA NaOH (pH 8.0)

Röder D	10 % [v/v]	Glycerol
	20 mM	HEPES
	100 mM	KCl
	1.5 mM	$MgCl_2$
	0.2 mM	EDTA NaOH (pH 8.0)
	0.5 mM	DTT
	0.5 mM	PMSF

| Western Blotting blocking solution | 1x | PBS-T |
| | 5 % [w/v] | Non-fat dry milk powder |

| 1 M Tris.HCl buffer | 1 M | Tris base, desired pH adjusted with 37 % [w/w] HCl |
| 1 M HEPES buffer | 1 M | HEPES, desired pH adjusted with 5 M NaOH or 5 M KOH |

3.1.14 Software

Adobe Creative Suite 4	Adobe Systems, San Hose (USA)
Cytoscape	Cytoscape Consortium [134]
GProX	CEBI [135]
Lasergene 7	DNASTAR, Madison (USA)
MaxQuant and Andromeda	Max Planck Institute for Biochemistry [119, 120]
Microsoft Office	Microsoft Corporation, Redmont (USA)
R language for statistical computing	R Foundation for Statistical Computing [136]

3.2 Molecular biology methods

3.2.1 Analysis of nucleic acids

3.2.1.1 Determination of nucleic acid concentration

DNA concentration was determined by measuring the absorbance of aqueous solutions against a reference at 260 nm using a Nanodrop ND-1000 spectrophotometer. The following equation was used to calculate DNA concentrations [137]:

1 OD_{260nm} = 50 μg/ml double-stranded DNA

3.2.1.2 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out for visualization and small-scale purification of DNA according to standard protocols [137]. Gels were prepared with 0.5 – 2 % [w/v] agarose in 1 x TBE buffer. DNA samples were diluted with 10x DNA loading dye and separated using a horizontal electrophoresis system (Bio-Rad, München, DE) in 1 x TBE buffer at 80-120 V for 40 – 80 min. 1 kb Plus DNA Ladder (Invitrogen, Karlsruhe, DE) was used as size reference. Gels were stained post-run in 0.5 μg/ml ethidium bromide solution for 20 min at RT, and subsequently destained in ddH$_2$O for 30 min. DNA was visualized on a UV transiluminator or gel documentation system (Bio-Rad, München, DE).

Agarose gel solution:
0.5 – 1 % [w/v] agarose
1 x TBE

3.2.2 Preparation of DNA templates for chromatin array assembly

3.2.2.1 Isolation of 12-mer '601' DNA template

The DNA template used for chromatin array reconstitution was a kind gift of Daniela Rhodes (MRC, Cambridge, UK) [107]. The template, consisting of 12 repeats of the '601' nucleosome positioning sequence (147 bp) [138, 139] flanked with a linker sequence (200 bp in total), was within the pUC18 plasmid backbone. The plasmid DNA was amplified in and isolated from transformed dam⁻/dcm⁻ *E.coli* using a plasmid Giga kit (Qiagen, Hilden, DE) (isolation performed by Winfried Lendeckel, Max Planck Institute for Biophysical Chemistry, Göttingen, DE). For excision of the template from the plasmid backbone, the isolated DNA was digested with DdeI, BfuCI, HaeII and EcoRI restriction endonucleases overnight at 37°

(300 U of enzyme per 1 mg of DNA) in NEB4 buffer provided by the manufacturer (New England Biolabs, Ipswitch, USA). The resulting digest, containing the 2.4 kb template and various vector backbone fragments (<400 bp) was fractionated by precipitation with PEG-6000 [140]. The restriction digest was incubated for 10 min at RT with 0.5 M NaCl and 2 % [v/v] PEG-6000. The mixture was centrifuged at 16,000 g for 15 min (4°C). The supernatant was collected and the PEG-6000 concentration stepwise adjusted from 3 % to 12 % [v/v] in 1 % steps. The supernatant was centrifuged at each step as described above. The pellets resulting from each fraction were washed once with 70 % [v/v] ethanol, air-dried and dissolved in Tris-HCl (pH 7.9). Equal amount of DNA from each fraction was analyzed using agarose gel electrophoresis (3.2.1.2). The 12x200-601 template precipitated reproducibly at PEG-6000 concentration of 6-7 % [v/v] with high purity.

3.2.2.2   Biotinylation of 12-mer '601' DNA template

The 12x200-601 template (3.2.2.1) contains a 5' EcoRI overhang sequence (5'-GATC-3') which was used for ligating to a biotinylated oligonucleotide (3.1.12). The "EcoRI_3'P" and "EcoRI_5'P_3'bio" oligonucleotides were mixed in equimolar amounts at 100 μM and denatured at 95°C for 10 min in a thermocycler (Eppendorf). Annealing was achieved by slow cooling from 95°C to RT for 30 min. In order to allow ligation of the oligonucleotide (phosphorylated at the EcoRI overhang) to the 12-mer template, the latter was dephosphorylated using typicaly 1 U Antarctic phosphatase (New England Biolabs, Ipswitch, USA) per 2μg of DNA for 2 hours at 37°C in the corresponding buffer provided by the manufacturer. The phosphatase was deactivated by incubation at 65°C for 10 min. Following dephosphorylation, the DNA was extracted by addition of equal volume of phenol:chlorophorm:isoamyl alcohol (CIP) [25:24:1] in a phase lock heavy tubes (5PRIME, Hamburg, DE). The aqueous phase was precipitated with 0.3 M (final) sodium acetate (pH 5.2) and 3 volumes cold ethanol (99% [v/v]) [137]. The pellet was dissolved in 10 mM Tris buffer (pH 7.9). Ligation of the biotinylated oligonucleotide was performed with T4 DNA ligase. Ten times molar excess of oligonucleotide was ligated for 2 hours at RT using 2.5 U of ligase per 1μg of DNA in buffer supplied by the manufacturer (T4 DNA ligase buffer, New England Biolabs, Ipswitch, USA). The biotinylated DNA template was precipitated by addition of 0.5 M NaCl and 7% [v/v] PEG-6000 as described above (3.2.2.1).

3.2.2.3   Methylation of 12-mer '601' DNA template

The DNA template was methylated specifically at CpG sites using bacterial M.SssI methyltransferase (New England Biolabs, Ipswitch, USA). Typically, 500 μg DNA was

incubated for 4 h at 37°C with 250 U of M.SssI in 500 µl buffer containing 0.32 mM SAM, 1 mM DTT, 1 mM EDTA (pH 8.0), 50 mM NaCl and 10 mM Tris buffer (pH 7.9). To ensure complete methylation, the amount of SAM was doubled and half the original enzyme amount was added to the reaction for additional 4 hours (or overnight). Completion of the methylation was tested by digesting a 0.5 µg DNA aliquot of the reaction mixture with a restriction endonuclease with blocked cleavage at methylated CpG sites (e.g. BstUI, NotI, using buffer supplied by the manufacturer, New England Biolabs, Ipswitch, USA). Cleavage of non-methylated DNA and absence of cleavage of methylated DNA indicated complete methylation.

3.2.2.4 Preparation of scavenger DNA

Scavenger DNA was used to prevent oversaturation of chromatin arrays during reconstitution [107]. The scavenger DNA represents a 147 bp fragment from the pUC18_12x200-601 plasmid backbone without specific nucleosome positioning properties. The fragment was amplified by PCR in 96x100 µl volume using Taq polymerase (purified in-house by Winfried Lendeckel, Max Planck Institute for Biophysical Chemistry, Göttingen, DE). The PCR mix and PCR cycle program were as follows:

PCR mix (100 µl):

100 mM	Tris-HCl pH 8.8
500 mM	KCl
0.8% [v/v]	NP-40
0.2 mM (each)	dNTP mix
0.2 µm (each)	primers
1.5 mM	MgCl$_2$
50 ng	DNA template
20 U/ml	Taq polymerase

PCR cycle program:

Temperature	Time	
95°C	2 min	
95°C	30 s	◄ ¬
68°C	30 s	40x
72°C	30 s	-
72°C	1 min	
10°C	∞	

The PCR product was extracted by CIP and ethanol precipitated (3.2.2.2).

3.3 Protein biochemistry methods

3.3.1 Detection and analysis of proteins

3.3.1.1 Protein concentration determination

Concentration of histone proteins was determined by UV adsorption at 276 nm on a NanoDrop ND-1000 using published molar extinction coefficients [106]. Molar extinction coefficient of ubiquitinated H2B (H2Bub1) was calculated by adding the extinction coefficient of ubiquitin (calculated by sequence using ProtParam [141]) to this of H2B.

Protein	Molecular weight [Da]	Molar extinction coefficient [$M^{-1}cm^{-1}$]
H2A	13960	4050
H2B	13774	6070
H2Bub1	22172	7320
H3	15273	4040
H4	11236	5400

Concentration of nuclear extract was roughly estimated for comparison between preparations by adsorption at 280 nm and assuming $1OD_{280}$ = 1 mg/ml.

3.3.1.2 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

For visualization, histone proteins were separated using discontinuous polyacrylamide gel electrophoresis in the presence of SDS [142] according to standard protocols [143]. Typically, Tris-Glycine gels using 12-15 % resolving gel and 4 % stacking gel were prepared and run using Mini-PROTEAN Tetra system (Bio-Rad, München, DE).

Resolving gel:		Stacking gel	
12-15 % [w/v]	Acrylamide:bisacrylamide (37.1:1)	4 % [w/v]	Acrylamide:bisacrylamide (37.1:1)
0.4 M	Tris-HCl pH 8.8	0.68 M	Tris-HCl pH 68
0.1 % [w/v]	SDS	0.1 % [w/v]	SDS
0.1 % [w/v]	APS	0.1 % [w/v]	APS
0.04 % [v/v]	TEMED	0.1 % [v/v]	TEMED

Prior to loading on gels, protein samples were heated at 70°C for 10 min in 1x NuPAGE LDS sample buffer and 1x NuPAGE reducing agent (Invitrogen, Karlsruhe, DE). SeeBlue Plus2 pre-stained protein standard (Invitrogen, Karlsruhe, DE) was used as reference (6 μl per gel

lane) [144]. Samples were run in 1x SDS running buffer at constant current (30 mA per gel) for the aspired time (typically 50 – 80 min).

Unless otherwise stated, all protein samples for Western Blotting immunodetection were analyzed on 4-12 % or 4-20 % gradient Mini-PROTEAN TGX gels (Bio-Rad, München, DE) using 1x SDS running buffer and constant voltage of 200 V for 30 min.

All samples for mass spectrometry analysis were separated on 4-12 % gradient NuPAGE Novex Bis-Tris gels (Invitrogen, Karlsruhe, DE) using 1 x MOPS SDS running buffer supplemented with NuPAGE antioxidant for 50 min at constant 200 V according to instructions provided by the manufacturer (Invitrogen, Karlsruhe, DE).

3.3.1.3 Staining of protein SDS-PAGE gels

SDS-PAGE gels were stained using Imperial Protein Stain (Pierce/Thermo Scientific, Rockford, USA) according to instructions provided by manufacturer, or alternatively by using Colloidal Coomassie staining solution [145] overnight and destaining by rinsing with water for 2 h.

3.3.1.4 Western Blot transfer and immunodetection

For immunoblot detection, proteins from SDS-PAGE gels were transferred on nitrocellulose membrane using the Mini Trans-Blot system (Bio-Rad, München, DE). Gels were sandwiched between a stack of fiber pad/Whatmann paper/nitrocellulose membrane, equilibrated in 1 x transfer buffer according to instructions by manufacturer. Proteins were blotted onto the membrane electrophoretically in 1 x transfer buffer at 100 V, for 1 h at 4°C. The membrane was washed with ddH$_2$O and for visualization of higher abundant proteins, stained with Ponceau S solution (0.5 % [w/v] in 5 % [v/v] acetic acid) for 10 min at RT. The membrane was destained in 5 % [v/v] acetic acid, washed with 1 x PBS-T and blocked with PBS-T non-fat dry milk blocking solution for 1 h at RT. Then the membrane was rinsed with PBS-T and incubated with primary antibody solution in blocking buffer (3.1.10) overnight at 4°C. After three washes with PBS-T (10 min at RT), the membrane was incubated with secondary antibody solution in blocking buffer (3.1.10). The membrane was washed three times in PBS-T (10 min at RT) and the immunoblot was developed for chemiluminescence detection using ECL or ECL Plus kits (GE Healthcare, Buckinghamshire, UK). Photography films were overlaid on the developed membrane and exposed for 1 s to 20 min.

For reprobing the blots with additional primary antibodies, the membrane was washed with stripping buffer for 30 min at 60°C. After extensive washing with ddH$_2$O and

PBS-T, the membrane was blocked with milk solution and probed again with a different primary antibody using the procedure described above.

3.3.2 Introduction of specific histone chemical modifications

3.3.2.1 Native protein ligation

Native protein ligation for generation of H3K4me3, H3K9me2 and H3K9me3 modified histones was performed as described [104] by Szabolcs Sörös (Max Planck Institute for Biophysical Chemistry, Göttingen, DE). The ligation was performed between N-terminally truncated histone H3 (H3 A21C Δ1-20) and chemically synthesized 20 amino acid-long N-terminal H3 peptide bearing a C-terminal thioester (synthesized by Peter Henklein, Charité, Berlin, DE).

3.3.2.2 Incorporation of methyl lysine analogues.

H3K27me1 modified histone protein was prepared with mono-methyl lysine mimick as described [105, 146]. Histone H3 protein containing K27C and C110A mutations was prepared by Alexandra Stützer (Max Planck Institute for Biophysical Chemistry, Göttingen, DE). 10 mg freeze-dried histone protein was dissolved in alkylation buffer (4 M Guanidine hydrochloride, 1 M HEPES pH 7.8, 10 mM D/L-Methionine, filtered), supplemented with 10 μl fresh 1 M DTT and incubated for 1 h at 37°C in a sealed tube with occasional mixing by inverting the tube. Next, the solution was quantitatively transferred to a new tube and mixed with 100 μl of 3 M (2-chloroethyl)-methylammonium chloride and incubated for 2.5 h at RT in the dark under a fume hood (toxic intermediates!). The reaction was incubated for another 2.5 h after addition of 10 μl fresh 1 M DTT. The reaction was quenched with 50 μl 2-mercaptoethanol for 30 min at RT. The modified histone protein was purified using PD-10 columns (GE Healthcare, Buckinghamshire, UK) equilibrated with 25 ml ddH$_2$O. The reaction mixture (1 ml) was loaded on the column and washed with two consecutive volumes of 750 μl 2 mM 2-mercaptoethanol. The protein was eluted with 3.2 ml 2 mM 2-mercaptoethanol. The identity of the modification was verified after analysis by ESI-MS on LTQ XL (Thermo Fisher Scientific, Bremen, DE) (3.5.3.4) by the specific +58 Da mass shift. H3K27me2 and H3K27me3 proteins were kindly provided by Alexandra Stützer (Max Planck Institute for Biophysical Chemistry, Göttingen, DE).

3.3.3   Recombinant chromatin

3.3.3.1   Histone octamer assembly

Close to equal amounts (typically 1-3 mg) of freeze-dried histone proteins (H2A, H3, H4 and H2B or H2Bub1) were dissolved separately in Unfolding buffer at ~2 mg/ml and incubated at 4°C for 1 h as described [106]. Equimolar amounts of the four proteins (3.3.1.1) were mixed and the total protein concentration was adjusted to 1 mg/ml with Unfolding buffer. The histone proteins were refolded and the octamer complex assembled after dialysis against RB High buffer at 4°C (3 changes of 2 l of buffer each, for minimum of 30 h in total). The assembled complex mixture was concentrated to 500 µl (using Amicon Ultra, Millipore, Billerica, USA) and the complex species (octamers, H3/H4 tetramers and H2A/H2B dimers) were separated using HiLoad 10/300 Superdex200 size exclusion chromatography column connected to ÄKTA Explorer/Purifier FPLC system according to instructions provided by the manufacturer (GE Healthcare, Buckinghamshire, UK) and using RB High buffer as mobile phase. The octamer peak fractions (typically eluting 11-12 ml after injection) were pooled and concentrated to 2-3 mg/ml. For long-term storage at -20°C, the octamer solution was supplemented with 50 % [v/v] glycerol.

3.3.3.2   Chromatin 12-mer array reconstitution

Chromatin array reconstitution was achieved by decreasing salt dialysis as described [106, 107]. Histone octamers were dialyzed against RB High buffer for minimum of 6 h and concentration was determined by UV spectroscopy (0.45 OD_{276} = 1 mg/ml). Octamers and template DNA (3.2.2.2) were mixed in a molar ratio of 1:1 (octamers:DNA) with final buffer concentrations identical to RB High. The reconstitution mixture was dialyzed for 1 h against RB High at 4°C. RB High buffer was then exchanged gradually for RB Low buffer over a period of 36 h using a peristaltic pump, yielding a final NaCl concentration of ~25 mM. Successful reconstitution of chromatin arrays was visualized on native agarose gel electrophoresis (3.2.1.2) using 0.2 x TB running buffer and Chromatin loading buffer. Quality of the reconstitution was also monitored by digestion with micrococcal nuclease (3.3.3.3). The chromatin arrays were stored at 4°C for up to 6 months.

3.3.3.3   Digestion with micrococcal nuclease

Micrococcal nuclease (MNase) digestion was performed as described [147]. 500 µl of free DNA or chromatin array solution at 5µg/ml was incubated with 0.1 U MNase in MNase buffer (20 mM Tris-HCl (pH 7.2), 5 mM $MgCl_2$, 3 mM $CaCl_2$) at RT for 0 s, 10 s, 30 s, 2 min

and 10 min. At each time point a 100 μl aliquot was transferred in PB buffer (QIAquick PCR purification kit, Qiagen, Hilden, DE) to stop the reaction. DNA was purified according to instructions provided by the manufacturer. Samples were analyzed by agarose gel electrophoresis (3.2.1.2).

3.4 Cell culture, metabolic labeling and cell-based assays

3.4.1 Cell culture and metabolic labeling of HeLa S3 cells

HeLa S3 cells were grown in custom High Glucose Dulbecco's Modified Eagle's Medium (DMEM) medium (PAA Labratories, Colbe, DE) lacking Arginine and Lysine. Light (Arg0 and Lys0) or heavy (Lys4 and Arg6, Lys8 and Arg10 or Lys6 and Arg10) amino acids (3.1.3) were supplemented to final concentration of 50 mg/ml. 1/10 volume of foetal bovine serum (FBS) and 1 X Penicillin/Streptomycin were added before culturing of the cells. Typically, a cryostock of 10^8 cells was used for starting culture of 100 ml in spinner flasks. The cells were cultivated at 37°C, 5 % CO_2 and 95 % relative humidity. The culture was expanded to 1.5 L over at least 6 passages and then inoculated in a 5 L bioreactor (Applikon, Schiedam, NL) and grown under standard conditions (2 x 10^6 cells/ml, with barbutation of synthetic air and dissolved oxygen level kept at pO_2 = 20 using a feedback monitoring system).

3.4.2 Transgenic cell lines for microscopy, ChIP and qRT-PCR

Transgenic cell lines (expressing exogenous FLAG-tagged ACTL8, ADNP, FANCF, SPIN1 or ZMYM3) for immunofluorescence microscopy or specific chromatin or peptide pull-down experiments were prepared by Kerstin Mosch (Max Planck Institute for Biophysical Chemistry, Göttingen, DE) [78].

Cell lines used for ChIP and qRT-PCR (siBrg1, siBAF155, siBrm) were prepared by Efrat Shema (Weizmann Institute of Science, Rehovot, IL).

3.4.3 Nuclear extract isolation

Light and Heavy-labeled cells were used for preparation of nuclear extracts (NE) as described [148]. Typically, 10^{10} Cells were harvested from the bioreactor by centrifugation at 2,000 rpm for 5 min (Cryofuge 6000i, Heraeus, Hanau, DE) and washed with ice-cold 1 x PBS. Cells were washed with 1.25 volumes of MC buffer supplemented with 1/500 vol. 0.25 M DTE and EDTA-free protease inhibitor cocktail according to instructions by the

manufacturer (Roche, Mannheim, DE). After incubation on ice for 5 min, the cell suspension was lysed by 50 ml Dounce homogenizer (18 strokes). Nuclei were pelleted by centrifugation for 5 min at 18,000 x g (Sorval SS34 rotor) and then resuspended in 1.3 vol. Röder C buffer (supplemented with 1/500 vol. 0.25 M DTE and 1/200 vol. 0.1 M PMSF) and dounced (20 strokes). Nuclear debris were pelleted by centrifugation for 30 min at 16,000 rpm (Sorval SS34 rotor). The supernatant was flash-frozen in liquid nitrogen (LN_2) and stored at -80°C. The nuclear extract was thawed in water bath (37°C) and dialyzed 3 x 2 hours against 50 vol. Röder D buffer. The dialysate was centrifuged for 2 min at 9,000 x g (Sorval SA600 rotor) and the supernatant aliquoted and flash-frozen in LN_2. The isolated NE was stored at -80°C. Typical protein concentration of the isolated nuclear extracts was 10-12 mg/ml.

3.5 Biochemical binding assays

3.5.1 Peptide affinity purification of histone modification-binding proteins

Peptide pull-down experiments were performed as described [78, 103, 108, 149]. Typically, 10 µg of peptide solution was incubated with 40 µl streptavidin coated paramagnetic beads (Promega, Mannheim, DE) in PD150 buffer overnight at 4°C. Peptide-bound beads were washed three times with PD150 buffer and incubated with 0.2 to 0.5 ml pre-cleared nuclear extracts for 4 hours on head-over-tail shaker at 4°C. Non-specifically bound proteins were removed by three washes with PD150 buffer (5 min each on head-over-tail shaker). Bound proteins were eluted by heating the beads in LDS sample buffer (Invitrogen, Karlsruhe, DE) at 70°C for 10 min. The eluted proteins were separated on SDS-PAGE and transferred on nitrocellulose membrane for immunodetection (3.3.1.4). For SILAC pull-down experiments, unmodified peptides were incubated with light and modified with heavy NE (forward experiment), and unmodified peptides with heavy and modified with light NE (reverse experiment). Protein eluates from unmodified and modified peptides within each forward or reverse experiment were mixed and separated by SDS-PAGE, protein bands were excised and subjected to proteomics analysis (3.6.1, 3.6.2 and 3.6.3).

3.5.2 Chromatin affinity purification

Chromatin affinity purification experiments were performed as described for the peptide pull-down experiments (3.5.1) [78, 108] with some modifications. For SILAC experiments, typically 40-50 µg of chromatin arrays were immobilized on 160-200 µl paramagnetic beads suspension (1 µg chromatin per 4 µl bead suspension) per experimental condition. Following pull-down, bound proteins were eluted with 1 x NuPAGE

LDS sample buffer (containing 2 mM $MgCl_2$, 10 mM Tris-HCl pH 7.9), light and heavy samples were mixed and incubated with 1,000 U benzonase nuclease (Calbiochem, Darmstadt, DE) for 45 min at 37°C. After addition of 1 x NuPAGE reducing agent, the samples were heated at 70°C for 10 min and separated on gradient SDS-PAGE gels as described above (3.3.1.2 and 3.5.1). Chromatin pull-down experiments for Western Blot immunodetection were performed with 2-10 µg chromatin immobilized on 20-40 µl beads suspension.

3.5.3 Affinity purification experiments using internal standard

Chromatin or peptide pull-down experiments were performed as described above (3.5.1 and 3.5.2) with the following modification. For forward experiment, proteins originating from Light nuclear extract pull-downs were eluted in LDS sample buffer and quantitatively mixed. The pooled eluate was divided equally among the Heavy nuclear extract pull-downs and each of them was analyzed separately on SDS-PAGE (3.3.1.2). Reverse experiment was performed alike, but with pooling the Heavy samples and distributing them between the Light pull-down eluates.

3.6 Mass spectrometry methods

3.6.1 In-gel proteolysis of proteins

In-gel digestion of proteins was performed as described [114, 150] with modifications. Unless otherwise stated, all incubation steps were performed at 26°C in thermomixer (Eppendorf, Hamburg, DE) at 1050 rpm for 15 min. All solutions were prepared with LiChrosolv H_2O (Merck, Darmstadt, DE). Solutions were removed after each incubation step. SDS-PAGE gel lanes were cut into 22 equal slices using in-house designed device [129]. Each gel slice was cut into 1 mm x 1 mm cubes, washed with 150 µl water and dehydrated with 150 µl acetonitrile (ACN). Gel pieces were dried and then rehydrated with 100 µl reducing solution (100 mM DTT in 50 mM ammonium hydrogen carbonate pH 8.0). After incubation at 56°C for 50 min, the gel pieces were dehydrated with 150 µl ACN. Reduced protein cysteine residues were alkylated with 100 µl 60 mM iodoacetamide (IAA, in 50 mM ammonium hydrogen carbonate pH 8.0) for 20 min at 26°C in the dark. The gel pieces were washed with 150 µl ammonium hydrogen carbonate (pH 8.0), followed by addition of 150 µl ACN. After dehydration with 150 µl ACN, the gel pieces were dried and

rehydrated with 15-20 μl digestion buffer for 30 min on ice. The rehydrated gel pieces were overlaid with digestion buffer without trypsin and incubated at 37°C ON.

Digestion buffer:

15 μl   Trypsin (0.1 μg/ml)

50 μl   Ammonium bicarbonate (50 mM, pH 8.0

50 μl H_2O

3.6.2   Extraction of peptides

Peptides from in-gel digestion were extracted as described [114, 150] with modifications. All incubation steps were performed using a thermomixer (1050 rpm) at 37°C for 15 min. Gel pieces were incubated with 30 μl water followed by addition of 100 μl ACN. The solution was transferred to new tubes and the dehydrated gel pieces were incubated with 50 μl 5 % [v/v] formic acid (FA), followed by addition of 50 μl ACN. The solution from the second extraction step was pooled with the solution from the first and additional 50μl ACN were added to the gel pieces to ensure complete extraction. The supernatant from the third extraction step was pooled with the solution from steps one and two, dried in vacuum centrifuge (Thermo Scientific, Braunschweig, DE), re-dissolved and analyzed by LC-MS/MS (3.6.3) or stored at -20°C.

3.6.3   LC-MS/MS analysis of peptides

Peptides from in-gel digestion (3.6.1 and 3.6.2) were re-dissolved in 20 μl 5 % [v/v] ACN/1 % [v/v] FA by extensive vortexing and brief sonication (3 min, maximum power). 5 μl aliquots were separated by reverse-phase nanoflow chromatography (HP1100 or HP1200 equipped with autosampler, Agilent, Santa Clara, USA). Peptides were loaded on in-house packed trap (desalting) column (fused silica, 3 cm, 360μm o.d., 150 μm i.d., Reprosil AQ-3/5 μm / 300 Å, Dr.Maisch, Ammerbuch, DE) and separated at flow rate 250 nl/min on C18 analytical column in-house packed in the electrospray emitter (SilicaTip, New Objective, Woburn, USA, 15 cm, 360 μm o.d., 75 μm i.d., Reprosil AQ-3/5 μm / 300 Å, Dr.Maisch, Ammerbuch, DE). The applied gradient was 3 % - 37 % buffer B (buffer A: 0.1 % [v/v] FA; buffer B: 98 % [v/v] ACN / 0.1 % [v/v] FA) for 35 min, followed by column wash (90 % buffer B) and equilibration (3 % buffer B). Peptides eluting from the LC were analyzed by ESI-MS/MS (spray voltage 1.5 kV) on a LTQ-Orbitrap XL mass spectrometer (Thermo Fischer Scientific, Bremen, DE) operating in data-dependent mode. Survey full scan MS spectra were acquired in the orbitrap mass analyzer (m/s range 350-1600, resolution of 30,000 or

60,000 at 400 *m/z* and automatic gain control target at 10^6). The five most intense peaks were selected for collision-induced dissociation (CID) in the LTQ linear ion trap with normalized collision energy of 37.5 %, activation q = 0.25 and activation time 30ms. Precursors with single or unknown charge state were excluded. Previously selected precursors were dynamically excluded for 60 s. Unless otherwise stated, samples were analyzed in triplicate. LC-MS/MS analysis on LTQ-Orbitrap Velos was performed alike, but with the 15 most intense ions selected for CID fragmentation and 30 sec dynamic exclusion time.

3.6.4   Molecular weight determination of intact proteins

Intact proteins were dissolved in 20 % [v/v] ACN / 0.1 % [v/v] FA and manually injected and analyzed on LTQ XL linear ion trap (Thermo Fischer Scientific, Bremen, DE) in full scan mode (350-2000 *m/z*). Spectra were extracted and deconvoluted with ProMass (Thermo Fischer Scientific, Bremen, DE) using Small Protein analysis mode as suggested by the manufacturer.

3.6.5   MS raw data processing

Raw MS data from LTQ-Orbitrap XL and Velos were analyzed using MaxQuant software [119, 120] (version 1.2.2.5). The following settings were used: initial MS mass tolerance 8 ppm, MS/MS tolerance 0.6 Da, false discovery rate (FDR) at both peptide and protein level 1 %, maximum peptide posterior error probability (PEP) 0.05, minimum peptide length 5 amino acids, minimum ratio count 1 (further increased after analysis, 3.7.1), maximum number of modifications per peptide 4, maximum precursor charge 5, "re-quantify" on, "keep low-scoring versions of identified peptides" off, "use razor and unique peptides" on. Processed raw files were searched using International Protein Index (IPI) Human protein database (version 3.87, September 2011, containing 91,464 entries, http://www.ebi.ac.uk/IPI/) supplemented with 179 common contaminants (e.g. keratins, serum albumin). Carbamidomethylation of cysteine and oxidation of methionine were used as variable modifications. Trypsin specificity with no proline restriction and up to 2 missed cleavages was allowed.

3.7 Data analysis and statistics

3.7.1 Data filtering and visualization in R

Tab-delimited text file output from MaxQuant (proteinGroups.txt) was imported in R [136] without pre-processing. All "Reverse" and "Contaminant" entries were deleted. Additionally, for individual chromatin modification experiments within each Forward (F) and Reverse (R) experiments the following ratio count (RC) minima were required: $RC_F > 1$ AND $RC_R > 1$ AND $RC_F + RC_R > 3$.

Total summed peptide intensities were plotted in $\log_{10}$ scale, normalized enrichment ratios in $\log_2$ scale. Hierarchical clustering was performed within the heatmap.2 function using Euclidean distance and combined linkage method. R scripts were previously published in [78].

3.7.2 Protein-protein interaction analysis

For protein-protein interaction network analysis, the STRING database [151] was queried for known protein-protein structural and functional interactions (based on experimental, curated database and text-mining evidence and with confidence level higher than 0.4) using the leading UniProt knowledgebase IDs as reported by MaxQuant. Interaction information was saved as tab-delimited text files and imported in Cytoscape (version 2.8.1, [134]). Proteins without known interactions within the datasets were imported as individual nodes. Nodes were coloured according to experiment origin, with overlapping nodes having multi-colour annotation. The network was clustered using force-directed layout. Network analysis was performed using the NetworkAnalyzer plug-in [152].

3.7.3 GRproX

MaxQuant results from experiments using spiked-in internal standards were imported in GProX (version 1.1.7, [135]), where the enrichment ratios were reversed and plotted for selected proteins as indicated.

4 Results

DNA and histone modifications play a central role in regulating virtually all genome functions. While some chromatin marks exert their function at a structural level, the majority of the studied modifications serve as molecular beacons to recruit various proteins and protein complexes that bring about different functional consequences. Single or multiple interacting partners of several modifications have been described, however the overlap, interplay and correlation between chromatin marks with similar or divergent functional outcomes have not yet been analyzed. To gain more insight into the chromatin modification interactome, a novel affinity purification strategy combined with quantitative mass spectrometry was established. Using recombinant, uniformly modified chromatin, the interaction partners of ten chromatin templates differing in their modification status were described. Apart from global analysis, the particular molecular function of the recruitment of the SWI/SNF complex to chromatin containing monoubiquitylated histone H2B was also elucidated.

4.1 Chromatin affinity purification assay for identification of modification-specific binding proteins

4.1.1 Preparation of recombinant uniformly modified chromatin template

General reconstitution strategy

The first step towards the goal of identification of chromatin modification binding proteins was establishment of a suitable affinity template. Several studies, some of which originating from our research group, successfully report the use of short modified or unmodified peptides that can recruit specific interaction partners [149, 153] including examples in the context of chromatin [81, 103, 108]. In order to bring the approach closer to the native state of chromatin, the novel method was based on homogeneous recombinant chromatin that consists of nucleosome octamers positioned uniformly on a defined DNA template [107, 108]. Recombinant *Xenopus laevis* histone proteins [106] were utilized as they have been widely used for structural and functional characterization of recombinant nucleosome core particles and chromatin [1, 30, 107, 154]. In addition, they have high sequence similarity to their human orthologues (93-100% sequence identity). The DNA template used for in vitro reconstitution was based on the non-natural "601" nucleosome
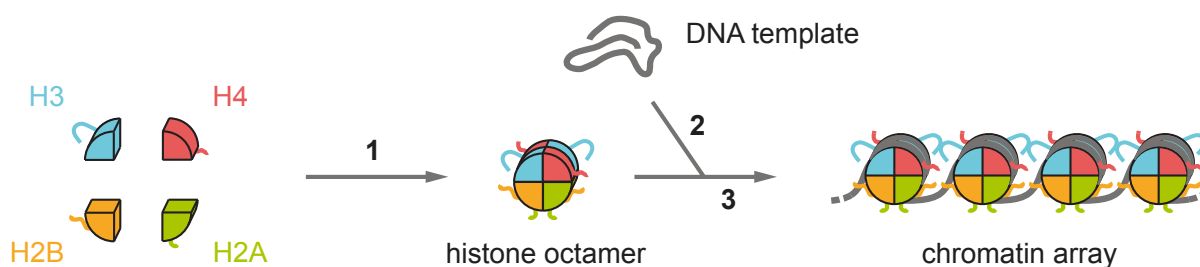
Figure 4.1 General scheme of chromatin array reconstitution.

Unfolded recombinant histone proteins were mixed in equimolar amounts and assembled into a histone octamer via dialysis against 2 M NaCl (step 1). The resulting octamer was mixed with a DNA template in a defined molar ratio (step 2) and gradually dialyzed against a buffer with decreasing NaCl concentration from 2 M to 20 mM (step 3).

positioning sequence [138, 139]. An array of twelve 601 sequences separated by 53 bp of linker DNA was used. This allowed for the reconstitution of uniform and evenly spaced nucleosome core particles along a 2.4 kb long DNA template. Nucleosomal arrays, rather than mononucleosomes, were chosen in order to capture proteins requiring the interaction surfaces offered by multiple or neighbouring nucleosomes. The DNA template and histone octamers were assembled into chromatin arrays using the salt dialysis approach [106] as outlined in Figure 4.1. The basis of the reconstitution approach has been established by Szabolcs Sörös [155].

The salt dialysis method for reconstitution of oligonucleosomes [107] differs from the classical mononucleosome reconstitution method [106] in the addition of a 148 bp DNA fragment with random sequence that prevents oversaturation of the array by scavenging free histone octamers. An octamer to DNA molar ratio of 1.1 was previously found to be optimal for obtaining ~90% saturation of the DNA template [155] as revealed by single molecule atomic force microscopy analysis (Figure 4.2 A, courtesy of Stefan Winter, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany). This molecular ratio resulted in uniformly saturated chromatin arrays as seen by native agarose gel shift (Figure 4.2 B, compare lane 2 with lanes 4 and 7). Regular digestion pattern over time in micrococcal nuclease (MNase) digest assay indicated evenly distributed nucleosomes (Figure 4.2 C).

Incorporation of specific chromatin modifications

Three separate approaches were used for obtaining site-specifically homogeneously modified histones depending on the type of modification and its localization within the protein sequence. The modifications within the first 20 amino acids of the H3 N-terminus were prepared using native chemical ligation [104]. The basis for this approach is a

52
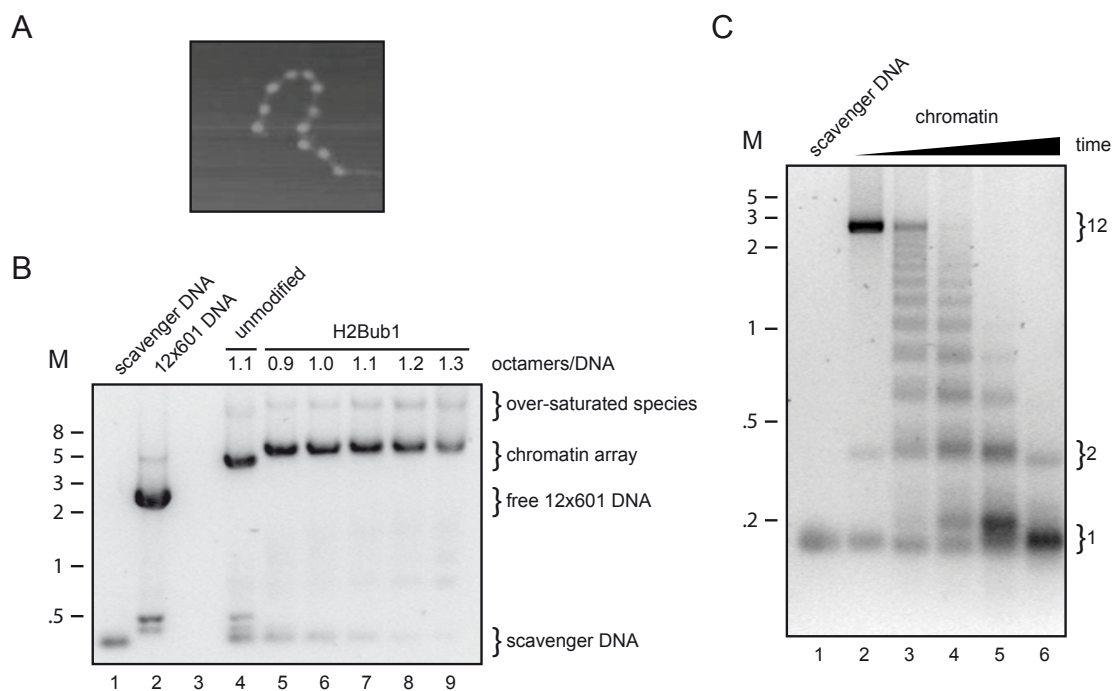
Figure 4.2 Quality control of reconstituted nucleosomal arrays.

A. Atomic force microscopy imaging of a 12x601 nucleosomal array (courtesy of Stefan Winter, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany). **B.** Native agarose gel shift analysis of free DNA and nucleosomal arrays reconstituted using the indicated molar ratios of DNA (calculated per 200 bp repeat) and histone octamer. The samples were separated using 0.2x TB buffered agarose gel electrophoresis and stained with ethidium bromide. **C.** Micrococcal nuclease digestion of nucleosome array reconstituted using a 12x601 DNA template. Samples loaded in lanes 2 - 6 were incubated with the enzyme for 0 s, 10 s, 30 s, 2 min and 10 min, respectively. DNA was extracted and analyzed by agarose gel electrophoresis and stained with ethidium bromide. Ladder-like pattern indicates specifically positioned octamers. Mono-, di- and 12-mer DNA fragments are indicated. M: DNA size marker in kb.

transesterification reaction between a C-terminal thioester on a synthetic histone peptide consisting of amino acids 1-20, and a N-terminal cysteine from a truncated (Δ1-20) recombinant histone H3 protein (Figure 4.3 A). Thus, a ligated histone protein can be created bearing any modification that can be synthesized chemically as part of a short peptide. The native chemical ligation, purification of the products and quality control by SDS-PAGE (Figure 4.3 B) and mass spectrometry (data not shown) were performed by Szabolcs Sörös (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany) for modifications H3K4me3, H3K9me2 and H3K9me3.

A second strategy, used for preparation of K120-ubiquitinylated H2B (H2Bub1), was based on an approach combining orthogonal protection and native chemical ligation [156]. The synthesis of H2Bub1 was done in the group of Ashraf Brink (Ben Gurion University of the Negev, Israel).
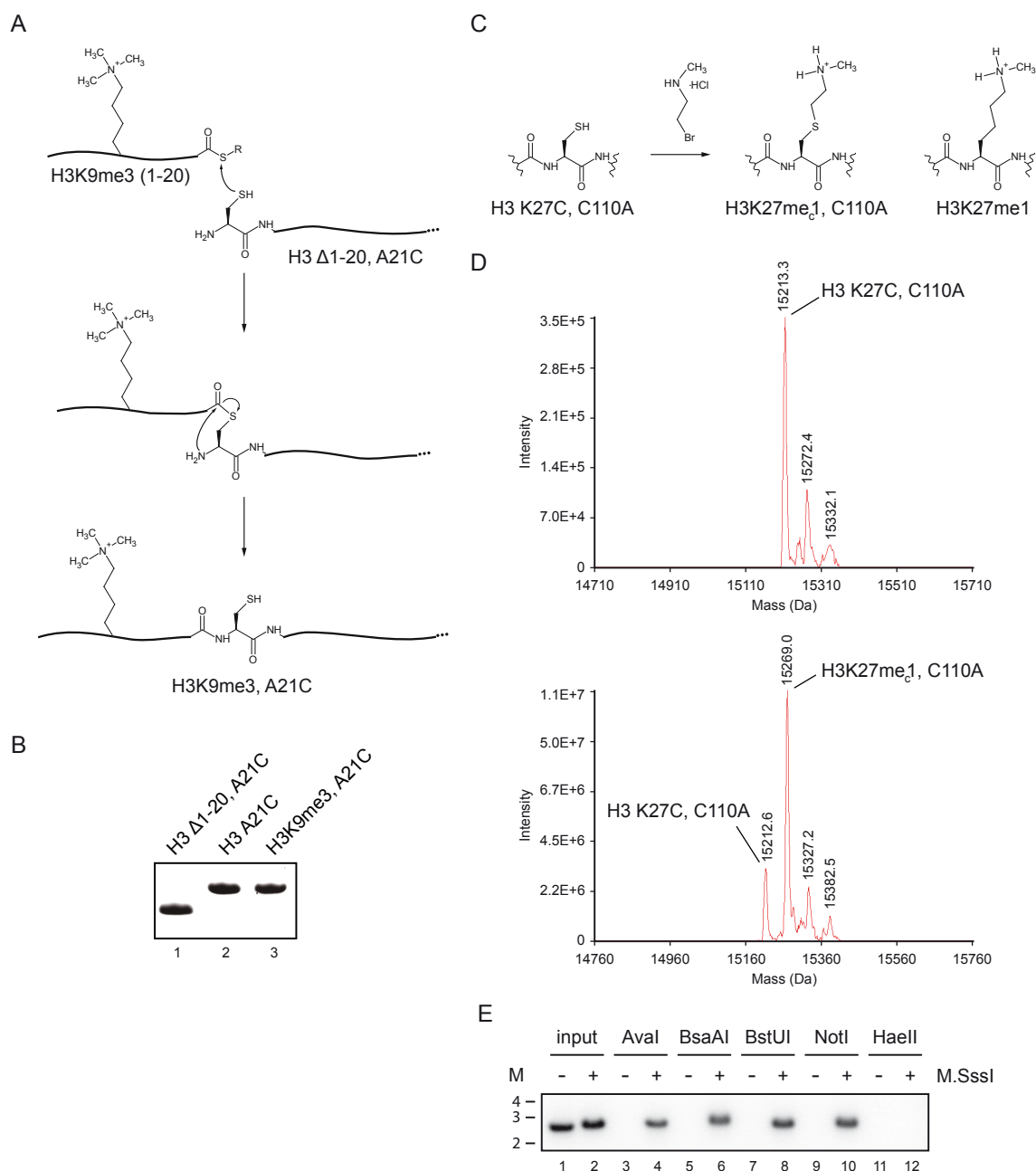
Figure 4.3 *In vitro* **incorporation of specific chromatin modifications.**

A. Schematic representation of the native chemical ligation reaction (H3K9me3 as an example). The synthetic peptide bearing the chemical modification and C-terminal thioester is coupled to the cysteine thiol group of recombinant H3 Δ1-20, A21C. The intermediate product rearranges spontaneously to form a native peptide bond. Adapted from [157]. **B.** SDS-PAGE analysis of recombinant H3 Δ1-20, A21C, as well as mutant full length H3 A21C and ligated H3K9me3 A21 (courtesy of Szabolcs Sörös). **C.** Schematic representation of the incorporation of methyl lysine analog H3K27mec1 on H3 K27C, C110A recombinant protein. The native H3K27me1 residue is shown on the right for comparison. Based on [105]. **D.** ESI-LTQ MS analysis of unmodified H3 K27C, C110A mutant protein (upper panel) and the product of the alkylation reaction H3K27mec1 (lower panel). **E.** Agarose gel electrophoresis analysis of non-methylated and methylated DNA treated with different restriction endonucleases sensitive (AvaI, BsaAI, BstUI, NotI) or non-sensitive towards CpG methylation (HaeII). M: DNA size marker in kb.

The third strategy, used here for methylation of lysine 27 on histone 3, was based on chemical alkylation of a cysteine residue resulting in mono-, di- or trimethyllysine analogs (MLA) [105] (Figure 4.3 C). To this aim, a mutated recombinant histone 3 (H3 K27C, C110A) was used for selective alkylation with 2-chloro-N-mehylethanamine, 2-chrloro-N,N-dimethylethanamine or 2-bromo-N,N,N-trimethylethanaminium for mono-, di or trimethylated product, respectively. The product of the alkylation reaction mimics a native methylated lysine residue [105] with the difference of a thioether group replacing the γ-methylene group of the modified amino acid. In each case, upon completion of the alkylation reaction the mass of the product was measured by ESI-MS (Figure 4.3 D). As shown before [105], the methylation reaction for production of monomethyl analogues is very sensitive and residual unmethylated protein was observed after the reaction (Figure 4.3 D, lower panel). Nevertheless, this strategy is more cost-efficient than native protein ligation and allows for selective methylation of virtually any mutated cysteine residue within a protein sequence. The recombinant protein used for preparation of H3K27me$_c$1, as well as the alkylated H3K27me$_c$2 and H3K27me$_c$3 proteins were a kind gift from Alexandra Stützer (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany).

Commercially available and highly active bacterial CpG methyltransferase M.SssI (New England Biolabs, Ipswitch, USA) was utilized for specific methylation of cytosines in CpG sequences of the DNA template used for reconstitution. Agarose gel electrophoresis analysis showed complete cleavage of non-methylated DNA using all tested restriction endonucleases (Figure 4.3 E), while the M.SssI-treated DNA was not cleaved by the enzymes whose activity is blocked by CpG methylation (AvaI, BsaAI, BstUI and NotI).

4.1.2 Optimization of the chromatin affinity purification approach

General strategy and representation of SILAC results

Several publications have reported the use of affinity purification followed by mass spectrometry for identification of histone modification binding proteins (see for example refs. [108, 158]). Franz et al. used short biotinylated peptides bearing the H3K9me3 modification in pull-down experiments and identified CDYL as a direct H3K9me3-binding partner. The identification of CDYL as an H3K9me3-associated protein was inferred from the notion that it was not identified in parallel pull-down experiments using unmodified biotinylated peptides. However, among the proteins with general chromatin binding properties (or proteins that show high non-specific binding to the matrix used in such experiments) there could be examples that nevertheless are specific for H3K9me3 but would be eliminated from the list of putative H3K9me3 binders using this analysis. While this approach is suitable for the
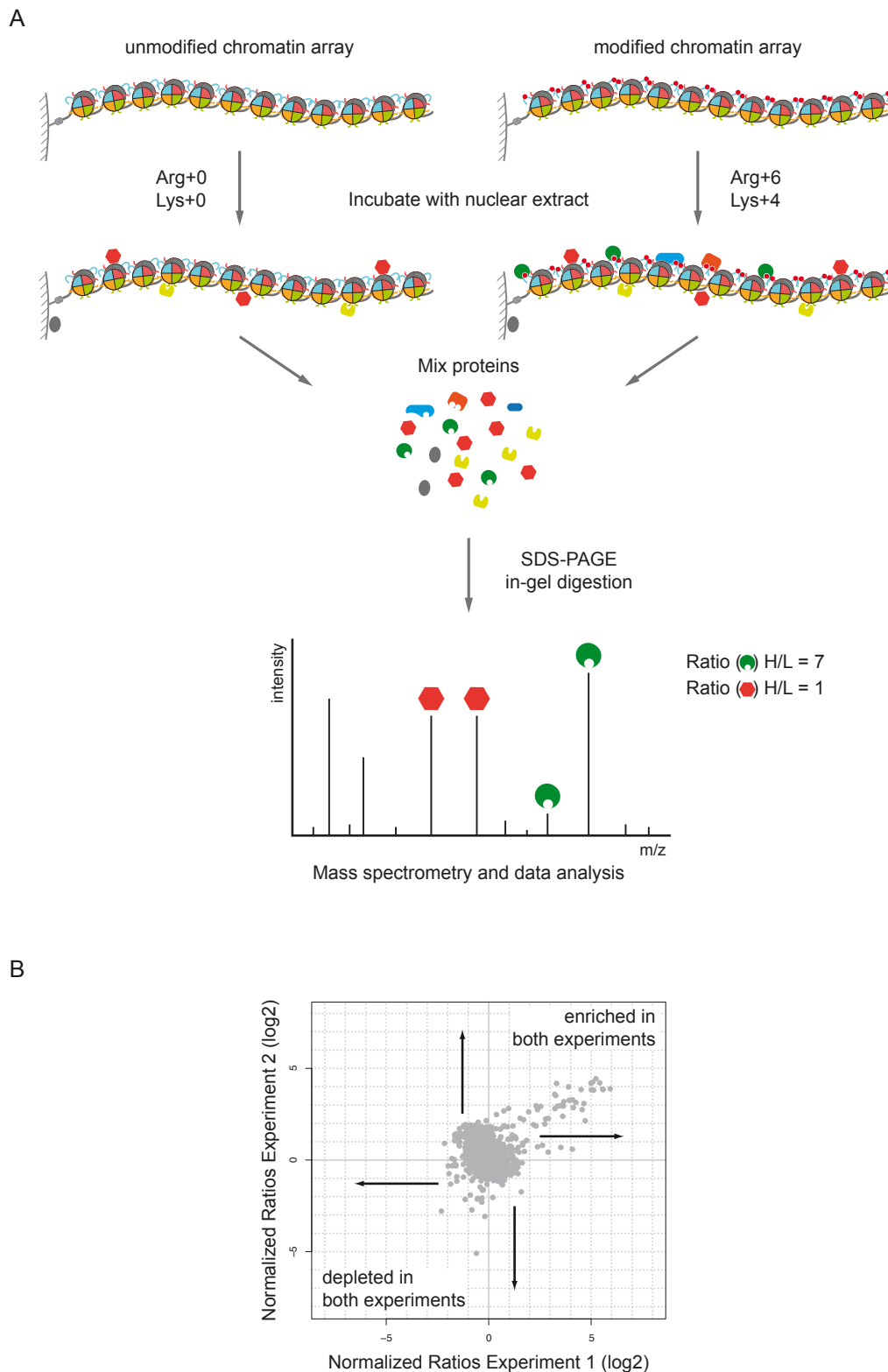
Figure 4.4 Workflow and results representation of the SILAC chromatin affinity purification.

A. Unmodified and modified chromatin templates were immobilized on magnetic beads and incubated with light and heavy SILAC extracts, respectively. The beads were washed and the bound proteins eluted, mixed and separated via SDS-PAGE. The proteins were then in-gel digested and the resulting peptides analyzed by LC-MS/MS. The enrichment ratios of each peptide (and consecutively, each protein) were calculated from the ratio of the light and heavy ➔

dissection of the binding of individual factors, it would restrict the comprehensive large-scale characterization of the chromatin modification interactomes. Therefore, a straightforward and semi-automatic approach for relative quantification of the proteins binding to unmodified and modified chromatin arrays in pull-down experiments was applied here. Stable isotope labeling with amino acids in cell culture (SILAC) [126] was chosen because of its accuracy and readily available software for advanced data processing [119].

A general workflow of the SILAC-based chromatin affinity purification is shown in Figure 4.4 A. Biotinylated chromatin arrays were immobilized on streptavidin-coated paramagnetic beads. This ensured stable interaction and facilitated washing and handling steps. After immobilization, the unmodified and modified chromatin was incubated with light and heavy SILAC nuclear extracts, respectively. Following three mild washing steps, the bound proteins were eluted using SDS sample buffer and mixed. For in-depth proteomics analysis the proteins were separated via SDS-PAGE, the gel lanes were excised into 21-23 equal slices and the proteins therein subjected to in-gel digestion using trypsin. The extracted peptides were analyzed on nano LC-coupled hybrid linear ion trap-orbitrap mass spectrometer. Here, peptides originating from proteins binding to the modified chromatin would show higher intensity heavy peaks, while those originating from background proteins or proteins binding irrespectively from the modification would show close to equal intensity peaks. Large-scale identification and quantification of the analyzed peptides and proteins was done using the MaxQuant software suite [119].

In order to visualize the results, the enrichment ratios were plotted as scatter plots correlating the values from two independent experiments and representing each quantified protein as a circle (Figure 4.4 B). The normalized enrichment ratios were plotted in log2 scale, positive values indicating higher heavy/light (H/L) ratio (higher abundance of the heavy protein), negative values denoting H/L ratios below 1/1 (higher abundance of the light protein). A protein showing similar behavior in both experiments would appear in the upper right or lower left quadrant, while a protein enriched in one and de-enriched (depleted) in the other experiment would appear in the lower right or upper left quadrants.

← peak intensities. **B.** Representative scatterplot showing the correlated enrichment ratio results from two experiments (e.g. forward and reverse experiments or two individual modifications, see main text) in log2 scale. Proteins enriched or excluded (depleted) in both experiment appear in the upper right or lower left quadrants, respectively. Proteins enriched in one and excluded in the other experiment would appear in the upper left or lower right quadrants.

The significance of label swap experiments

For each investigated modification, two pull-down experiments in parallel were performed – one as described above, and the second one identical but with swapping the heavy and light extracts for incubation with unmodified and modified chromatin, respectively. The first experiment was termed "forward" and the second – "reverse". Positive correlation of the enrichment ratios from the forward experiment with the inversed enrichment ratios from the reverse experiment indicates modification-dependent specific binding. Negative correlation implies false positive interaction where high enrichment ratio correlates with the type of extract used (light or heavy) and not with the modification status of the affinity template. Figure 4.5 shows the result from a representative SILAC experiment. The correlation between the enrichment ratios from forward and reverse experiments is plotted in panel A, and in panel B – the enrichment ratios relative to the total summed peptide intensities (indicative of protein abundance). Proteins coloured in orange and blue show average enrichment ratio between forward and reverse experiment higher than 2 (log2 scale). The proteins coloured in orange appear to have opposite enrichment ratios between the two experiments, nevertheless their average ratios exceed the enrichment thresholds that were set. If only one of the label swap experiments was performed, these proteins would have been indistinguishable from the true hits specific for the chromatin modification used. Thus, in order to exclude false positive hits, each SILAC pull-down was performed in duplicate label swap experiments.
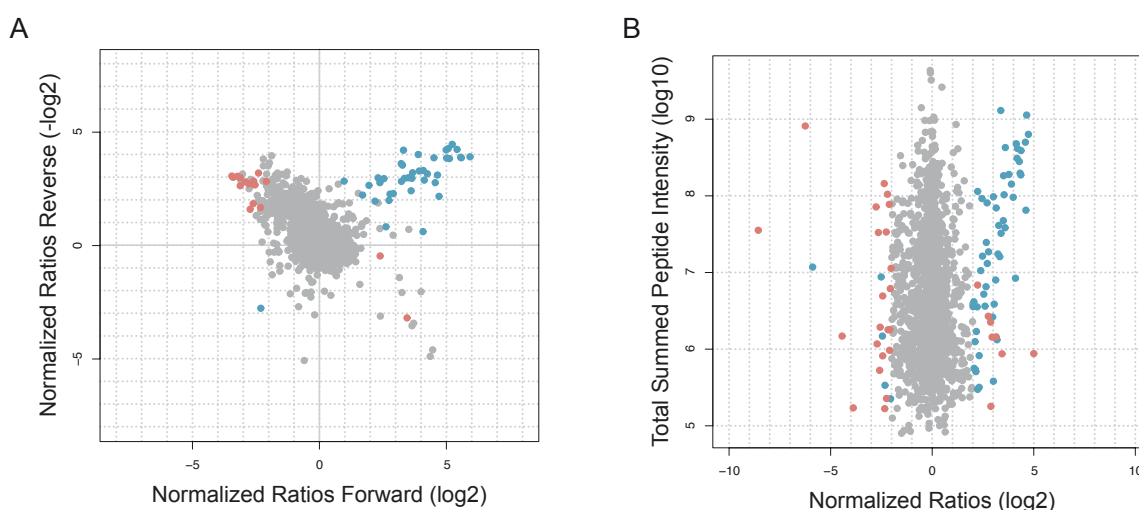
Figure 4.5 Correlation between forward and reverse label swap experiments.

A. Normalized ratios from representative forward and reverse experiments were plotted as scatterplot. All proteins showing average enrichment ratios higher than 2 (log2 scale) were coloured in red or blue. **B.** Average enrichment ratios of the proteins from panel A were plotted against their total summed peptide intensities (y-axis). Protein were coloured as in A.

Nuclear extract preparation and quality control.

SILAC nuclear extracts (NE) used for chromatin affinity purifications were prepared from HeLa S3 human cervical cancer cells. This cell line has modest culturing requirements allowing large yields. The HeLa S3 cells for SILAC labeling were cultured in a 5 L bioreactor with environmental controller (Figure 4.6 A). Near gram amounts of nuclear extract were prepared in a single experiment, with the advantage of having a controlled and reproducible growth environment. Additionally, the optimized culturing conditions within the bioreactor allowed for growing the cells to higher densities (2-3 x 10^6 cells/ml), reducing the time and SILAC medium needed. A single bioreactor run resulted in typically 30-50 g of HeLa cells. The nuclear extracts were prepared at higher protein concentration (typically 10-15 mg/ml), thus mimicking the natural environment in the cell nucleus during the pull-down experiments more closely (typical protein concentration range in the interchromatin compartment of the nucleus is around 100 mg/ml [159]). Mixing equal amounts of each extract, followed by standard SILAC sample preparation and data analysis, was used to compare the difference in abundance of the proteins between the light and heavy extracts. An example of light and heavy extracts prepared by two different lab members is shown in the left panel of Figure 4.6 B, and an example of light and heavy nuclear extracts prepared by the same person in parallel is shown in the left panel of Figure 4.6 C. The enrichment ratios in the case where the extracts were not prepared in parallel showed wider distribution, indicating broader differences in protein abundance. Additionally, a pull-down experiment using the first pair of extracts resulted in even more pronounced spread of the enriched and false positive interactors (Figure 4.6 B and C, middle and right panels). Thus, the performance of each extract preparation was monitored by using control pull-downs and by determining the difference in protein abundance.
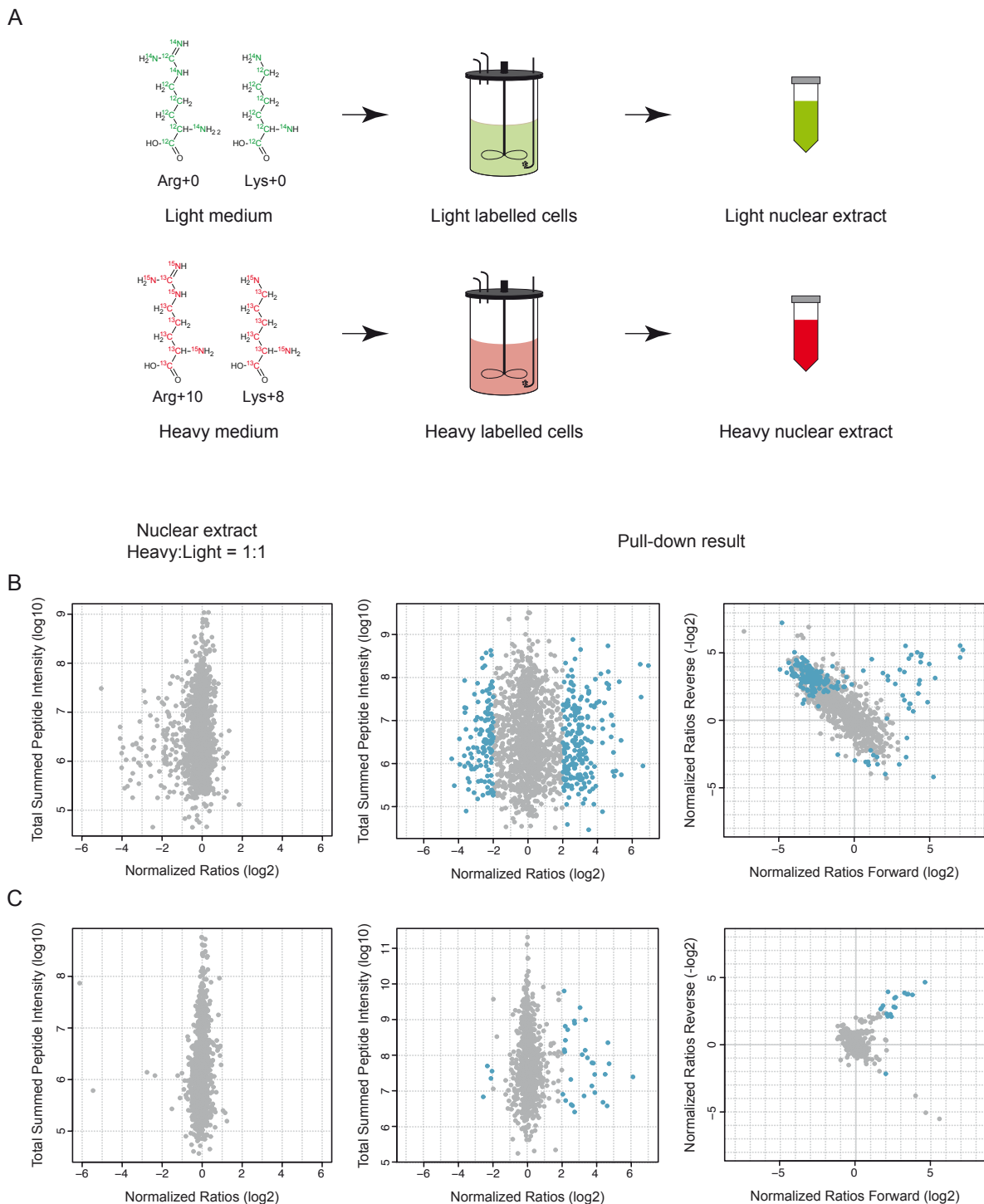
Figure 4.6 Nuclear extract preparation and quality control.

A. HeLa S3 cells used for SILAC labeling and nuclear extract preparation were cultured at high densities in a bioreactor with controlled environment. **B-C.** Comparison of the protein distribution of light and heavy nuclear extracts prepared independently (**A**, left) and in parallel (**B**, left). Results from pull-down experiments using these extracts are shown on the right.

Removal of DNA by benzonase digestion

A typical chromatin affinity purification experiment utilized 40-50 µg of chromatin per condition, resulting in a total of 80-100 µg of chromatin (after mixing light and heavy samples) per gel lane for SDS-PAGE analysis. Such large amount of DNA resulted in blocking the porous structure of the gel [160] and decreased the resolution and protein separation capacity tremendously (Figure 4.7 A). Sonication of a control DNA solution
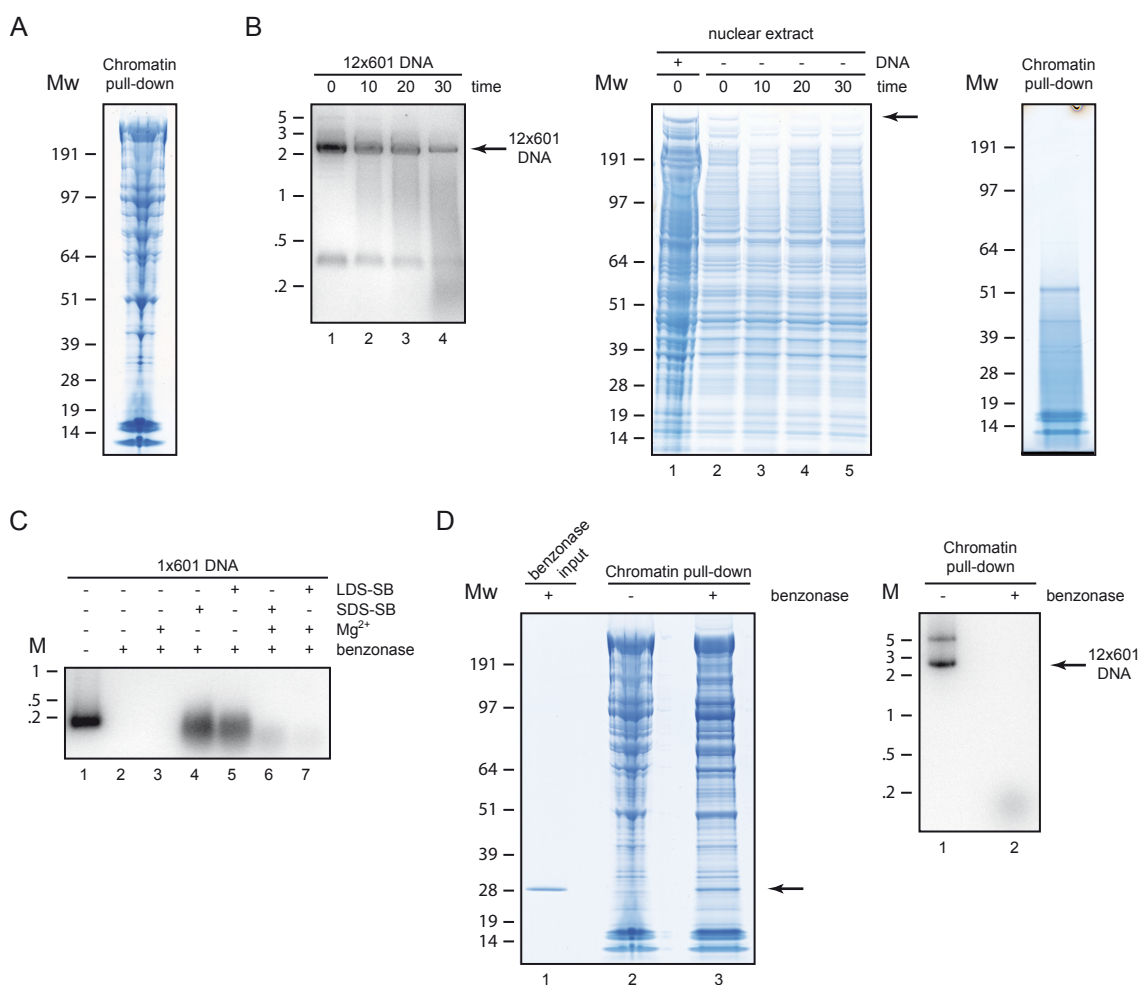
Figure 4.7 Removal of DNA from affinity purification eluates.

A. SDS-PAGE analysis of a pull-down reaction using 100 µg chromatin in total. **B.** 12x601 DNA aliquots were sonicated for 10, 20 and 30 min and analyzed by agarose gel electrophoresis (left panel). The treatment did not remove the large DNA fragment completely. Sonication of nuclear extract aliquots (middle) or chromatin pull-down reaction (right panel) resulted in loss of high molecular weight proteins as observed after SDS-PAGE analysis. **C.** Control DNA samples in SDS or LDS sample buffer were treated with benzonase with or without addition of 4 mM $MgCl_2$ and visualized by agarose gel electrophoresis. The presence of Mg^{2+} clearly improved digestion efficiency. **D.** Chromatin pull-down reactions with or without benzonase treatment were analyzed by SDS-PAGE (left, benzonase band marked with an arrow) and agarose gel electrophoresis (right). Benzonase improved SDS-PAGE resolution tremendously and resulted in near complete loss of DNA. Mw: protein size marker in kDa; M: DNA size marker in kb.

reduced the amount of intact DNA species only moderately, even after 30 min treatment (Figure 4.7 B, left panel, lane 4). Treating a nuclear extract aliquot under the same conditions resulted in loss of large molecular weight proteins (Figure 4.7 B, middle panel, lane 5, see arrow). This effect was even more pronounced after identical treatment (30 min sonication) of a chromatin pull down reaction (Figure 4.7 B, right panel). Thus, other methods for removal of DNA from the samples were tested. Benzonase is a commercially available genetically engineered promiscuous endonuclease [161] that is active in wide range of conditions. The activity of the enzyme in SDS- or LDS- sample buffers (SB) was tested, as the later is used for elution in the final step of the chromatin pull-down. The enzyme was only moderately active under these conditions (Figure 4.7 C, left panel, compare lane 1 to 4 and 5). However, both SB used contain EDTA which could deplete the Mg^{2+} required for optimal enzymatic activity. Indeed, supplementing the reactions with 4 mM $MgCl_2$ improved the cleavage efficiency substantially (Figure 4.7 C, left panel, compare lanes 1, 4 and 5 to 6 and 7). Incubating the chromatin pull-down eluate with 1000 U benzonase for 45 min at 37°C improved significantly the SDS-PAGE resolution (Figure 4.7 D, left panel, compare lanes 2 and 3) and cleaved the 12x601 DNA template completely (Figure 4.7 D, right panel, compare lanes 1 and 2). As the enzyme is sensitive towards higher concentrations of DTT or 2-mercapto ethanol, these components were added to the SB only after incubation with benzonase.

The effect of saturation of the affinity matrix

The amount and immobilization efficiency of the bait complex on affinity matrix beads could be an important parameter in pull-down experiments. Therefore, the immobilization efficiency of the recombinant chromatin on the streptavidin beads used was tested. Varying amounts of 1-11 µg chromatin in 2 µg steps were immobilized on 20 µl beads suspension, then the equivalent of 1 µg chromatin was loaded on SDS-PAGE, transferred on nitrocellulose membrane and detected with polyclonal αH3 antibodies. As shown in the left panel of Figure 4.8 A, the high immobilization efficiency remained similar regardless of the amount of chromatin used (compare lane 1 with lanes 2-7). With higher chromatin per beads ratios, the beads aggregated noticeably (Figure 4.8 A, right panel, e.g. compare tubes 1 and 7). This aggregation is likely due to higher order chromatin inter-bead interactions, promoted by the higher chromatin density on the beads. The influence of the different beads saturation was further tested by comparing two chromatin SILAC pull-downs, one using 20 µg chromatin/200 µl beads (experiment labeled "20 µg") and a second one using 50 µg chromatin/200 µl beads (experiment labeled "50 µg"). The interaction partners of unmodified chromatin and chromatin reconstituted with DNA containing 5-methylcytosine in the context
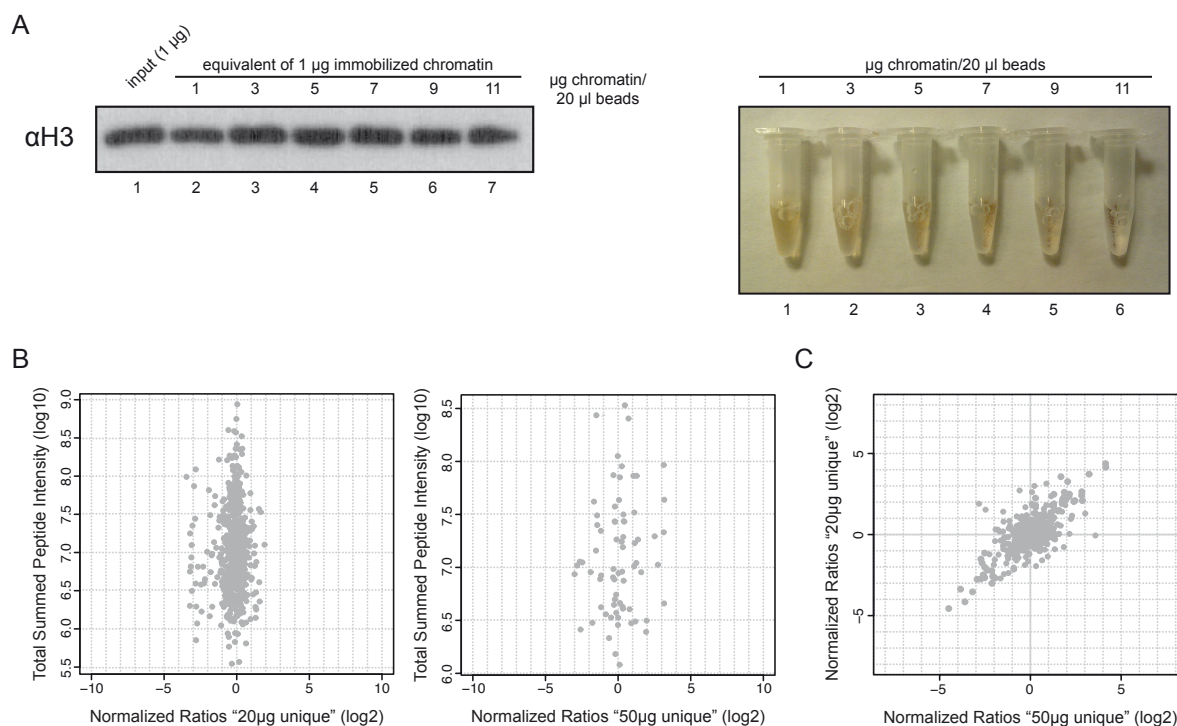
Figure 4.8 Influence of chromatin saturation level.

A. One to eleven µg chromatin was immobilized on 20 µl magnetic beads and the equivalent of 1 µg chromatin was loaded on SDS-PAGE and then transferred on nitrocellulose membrane for immunodetection with polyclonal αH3 antibodies (left panel). The samples of immobilized chromatin showed higher bead aggregation with increasing chromatin amounts (right panel). **B.** 20 µg and 50 µg of chromatin were used for immobilization on 200 µl beads suspension for pull-down experiments. The SILAC results show higher number of background proteins for the "50 µg" experiment (left). **C.** Correlation between the enrichment ratios from the "20 µg" and the "50 µg" experiments.

of CpG sequences (meCpG chromatin) were compared between both experiments. Nearly 1900 proteins were identified in the "20 µg" experiment and 1300 proteins in the "50 µg" experiment, of which 679 and 79 proteins were unique, respectively. Interestingly, the proteins unique for the "20 µg" experiment showed enrichment ratios very close to 1:1 (Figure 4.8 B, left panel, note that only quantified proteins are plotted), indicating that the lower saturation level of the beads promotes more prominent background protein binding. Of the proteins unique for the "50 µg" experiment, a few showed higher enrichment ratios in one or the other directions (Figure 4.8 B, right panel). Notably, two of those proteins appear to be significant interaction partners in the context of this modification (HOMEZ and KDM2B, discussed later). The two experiments reported similar enrichment ratios of the proteins binding to either meCpG or unmodified chromatin (Figure 4.8 C). However, larger number of background binding proteins could lead to lower dynamic range of detection of the MS instrument (in case of very abundant background proteins). Additionally, with higher number

of background proteins and constant sampling rate, the chances of sequencing peptides from true interaction partners become lower. Thus, in order to maximize the sampling rate and avoid identification of excess background proteins, for further chromatin affinity purification experiments a chromatin/beads ratio offering higher saturation level and identical to the "50 µg" experiment was used.

4.1.3   Validation of the chromatin affinity purification approach

As a proof-of-principle the proteins enriched to recombinant modified chromatin were compared with those enriched using modified histone peptides. For these experiments two hallmark histone modifications were used – trimethylation of lysine 4 and lysine 9 on histone 3 (H3K4me3 and H3K9me3, respectively), as there are a number of already known binding proteins [58, 162-164]. Additionally, these modifications have very different functional roles in chromatin biology and an overlap between the proteins they recruit was not likely [27, 43, 54]. In total 16 experiments were performed, four for each modification using both chromatin or peptide as affinity matrix (including label swap, and using two independent nuclear extract preparations). A total of more than 2100 proteins, out of 5000 present in the HeLa nuclear extract (Ilian Atanassov, manuscript in preparation), were identified. The distribution of the enrichment ratios of all quantified proteins is shown in Figure 4.9. In order to exclude proteins behaving non-reproducibly between forward and reverse experiments, a threshold of minimum 1 (log2 scale) in each experiment, and in addition for the average of both experiments minimum 1.5 (log2 scale) was set. Using this threshold, 61 proteins were identified as recruited to the H3K4me3 modification in the context of chromatin, and 67 in the context of the N-terminal histone peptides. For the H3K9me3 modification the number of proteins associated with modified chromatin and peptide settings was 25 and 42, respectively. Interestingly, a number of proteins associated specifically with unmodified chromatin and peptides in each of the experiments (termed here excluded or repelled by the respective modification). 14 proteins were excluded by chromatin-embedded H3K4me3 and 18 in the parallel peptide experiment. The H3K9me3 experiments revealed less proteins excluded by the modification – 3 for chromatin and 5 for peptide. Importantly, the datasets reveal a number of factors that have already been shown to interact directly with H3K4me3 and H3K9me3. TAF3 [81], CHD1 [67], PHF8 [165], WDR5 [98] and ING2 [166] are known to be directly recruited by H3K4me3. The isoforms of HP1 – CBX1, CBX3 and CBX 5 [167], UHRF1 [168] and MPP8 [169] are known H3K9me3 binders. Additionally, proteins known to be recruited to these modifications via secondary interactions though direct binders were also identified. For example, DNMT1 [170, 171] and HAUSP [172] were recruited by UHRF1,

while POGZ is a known binding partner of CBX5 [173]. Several subunits of TFIID were recruited to H3K4me3 by TAF3 [81]. Additionally, UHRF1 bound preferentially unmodified chromatin and histone tail, compared to those bearing the H3K4me3 modification, which is in agreement with recent literature [84].
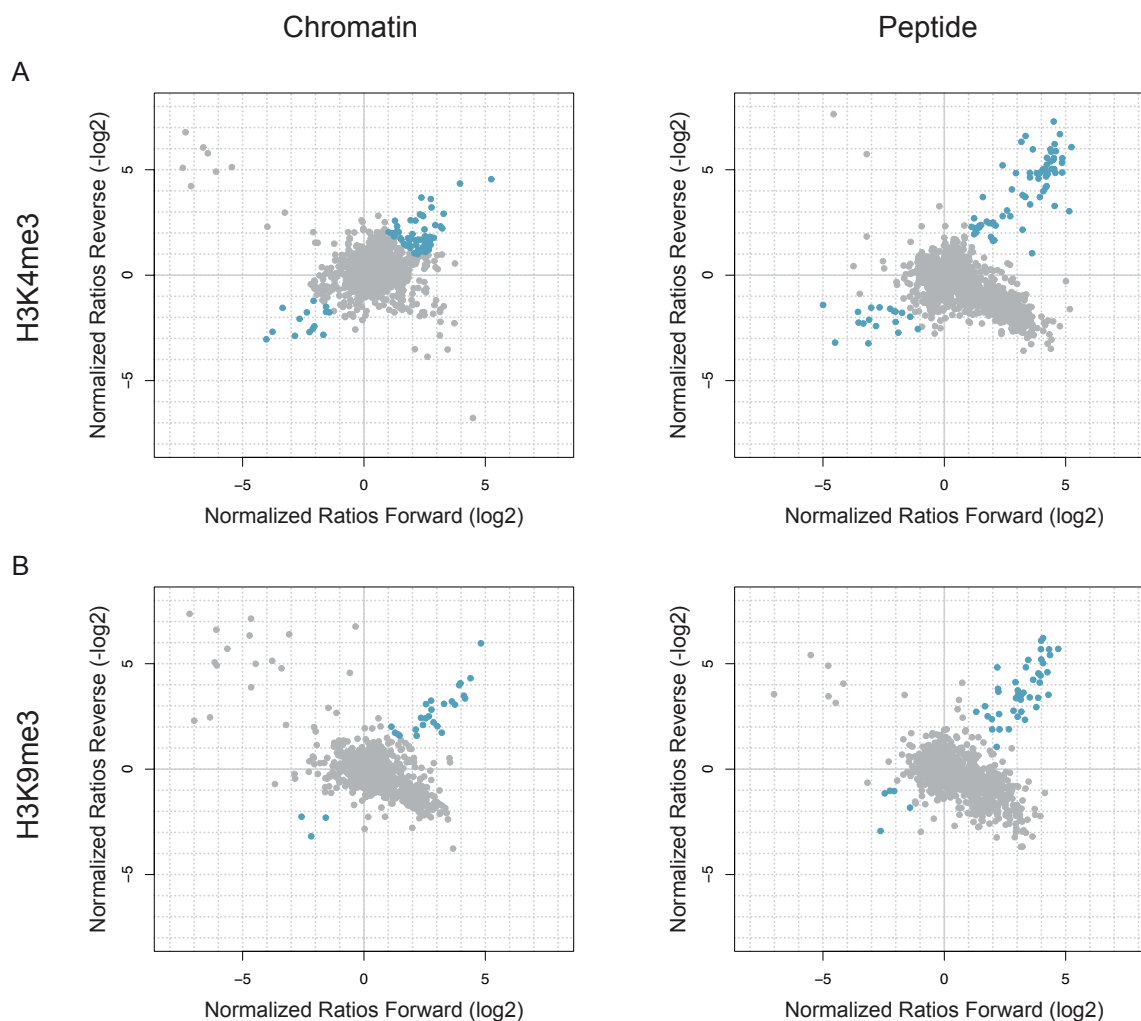
Figure 4.9 Enrichment ratios of proteins recruited by H3K4me3 and H3K9me3 chromatin and peptides.

Scatterplots representing the correlation between the enrichment ratios from forward and reverse experiments for H3K4me3 chromatin (**A**) and peptide (**B**) experiments, and H3K9me3 chromatin (**C**) and peptide (**D**) experiments. Proteins above threshold level (+/-1 in each F and R experiment, +/-1.5 average in log2 scale, see main text) are coloured in blue.

H3K4me3 and H3K9me3 recruit distinct sets of protein factors

Notably, there was a very limited overlap when comparing the results for each modification (Figure 4.10). In the chromatin context the death-associated protein 6 (DAXX) showed moderate enrichment with both modifications, but overall below the set thresholds.

The same was true for PHF8, which showed very strong enrichment with H3K4me3 and only moderate enrichment in the context of H3K9me3. Interestingly, the peptide experiment revealed four proteins excluded by both modifications, three of which were the meCpG associated proteins MTA1, MBD3 and GATAD2B [174]. All three proteins showed moderate exclusion from H3K4me3 chromatin, and near 1:1 binding with H3K9me3 chromatin. As already mentioned above, UHRF1 was enriched with H3K9me3 and excluded by H3K4me3 in both affinity matrices. The ACTL8 protein (discussed later and in the following chapters) followed the same trend in the chromatin experiments. Furthermore, in the chromatin experiment the UHRF1-binding proteins DNMT1 and HAUSP showed similar enrichment behaviour but slightly below the set thresholds. The second protein showing high enrichment with H3K9me3 peptides and exclusion from H3K4me3 peptides was CHD4, a factor involved in nucleosome remodeling and histone deacetylation [175]. Notably, neither the chromatin nor the peptide affinity purification identified proteins that associate with the H3K4me3 modification and were excluded from the H3K9me3 mark.
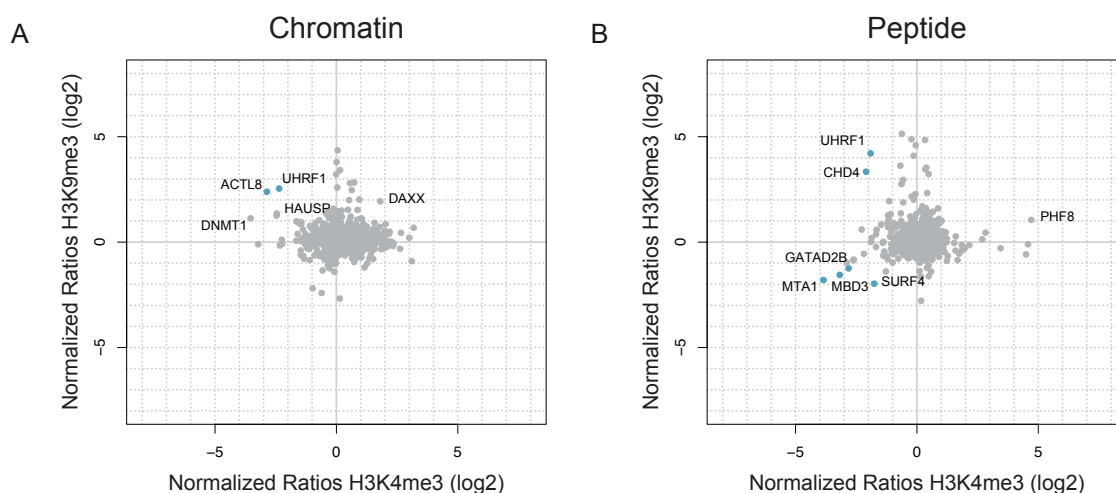
Figure 4.10 Comparison between H3K4me3 and H3K9me3 interactomes.

Scatterplots correlating the average enrichment ratios from H3K4me3 (x-axis) and H3K9me3 (y-axis) in the context of chromatin- (**A**) or peptide- (**B**) based experiments.

Comparison of the chromatin- and peptide-based approaches

As demonstrated in the comparison of the enrichment ratios between the two tested modifications, the chromatin- and peptide-based assays showed only a limited overlap, with several examples of proteins enriched to higher degree with one or the other matrix. This becomes even more evident after direct comparison of the two approaches (Figure 4.11). Both methods produced similar number of enriched factors, slightly higher in the case of the
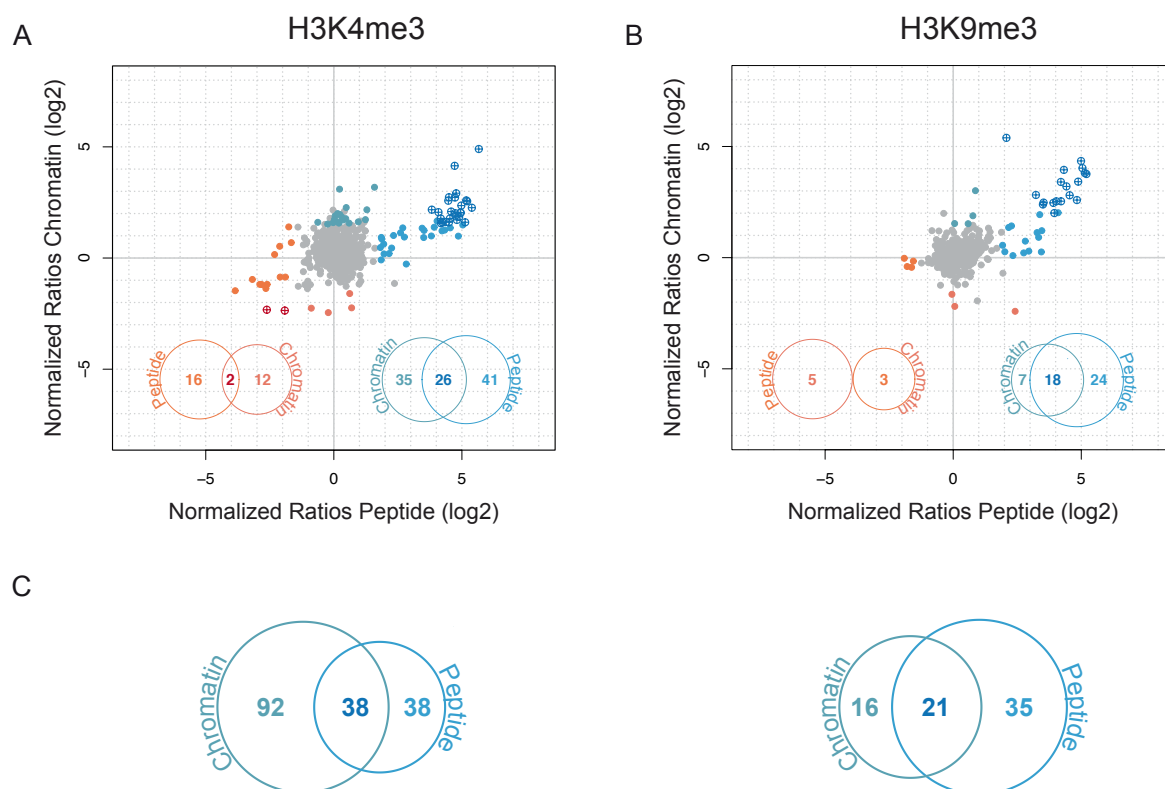
Figure 4.11 Comparison of chromatin- and peptide-based affinity purification approaches.

Correlation between the average enrichment ratios from peptide- (x-axis) and chromatin- (y-axis) based assays for H3K4me3 (**A**) and H3K9me3 (**B**). The overlap between the enriched (blue) or excluded (red) proteins is shown as inlay Venn diagrams (note that the number of dots is lower than the numbers in Venn diagrams as only proteins identified independently in both experiments are plotted). Proteins above threshold level in both approaches are presented as crossed-out circles. **C.** Venn diagrams representing the overlap of the enriched proteins for both approaches in H3K4me3 (left) and H3K9me3 (right) experiments using lower ratio threshold (+/-1 in each experiment, +/-1 average in log2 scale, see main text).

peptide pull-downs (Figure 4.11 A and B, inlay Venn diagrams for enriched (blue) or excluded (red) proteins). Additionally, there was a clear distribution of proteins identified as enriched with one of the affinity templates and as background (1:1) binders with the other (solid blue or orange circles near the x- or y-axes). Interestingly, as seen from the distribution of the proteins identified as enriched in both approaches (crossed out circles), the absolute values of the enrichment ratios for the chromatin experiments were lower compared to the peptide-based assay. Lowering of the ratio threshold (to average from F and R experiments 1, log2 scale) resulted in somewhat higher overlap of enriched proteins (Figure 4.11 C). Strikingly, the number of proteins enriched with the chromatin templates increased more than twofold, while the effect for the peptide-based experiment was only moderate (compare the Venn diagrams in panels A and B to panel C). Thus, a large number of putative novel H3K4me3 and H3K9me3 interaction partners were identified using the two affinity approaches. Both methods showed good agreement in the proteins they identify as

enriched with each modification, in addition to a set of factors unique for each individual affinity matrix.

Validation of novel histone-modification interaction partners

Transfection and harvesting of 293T HEK and NIH3T3 cells, as well as immunofluorescence microscopy, presented in the following paragraph, were performed by Kerstin Mosch (Max Planck Institute for Biophysical Chemistry, Göttingen, Germany).

The analysis of the chromatin and peptide pull-down results for H3K4me3 and H3K9me3 recapitulated the majority of the known interaction partners of both modifications tested. To validate the method, the interaction of several novel factors was further tested using orthogonal approaches. To this end, candidate proteins with engineered FLAG-tag were transiently expressed in 293T HEK cells. Nuclear extracts were isolated and used for peptide or chromatin affinity purification experiments with Western Blot detection using αFLAG antibodies. SPIN1, a protein implicated in regulation of the cell cycle [176] was enriched with the H3K4me3 modification using both pull-down methods (Figure 4.12 A), confirming the results from the SILAC-based experiments. Importantly, SPIN1 was only recently also shown to be a direct H3K4me3 interaction partner using orthogonal approaches [177]. FANCF [178] was recruited by the same modification in the chromatin context, but not in the peptide-based assay, again verifying the MS results. For the H3K9me3 modification the interaction was similarly confirmed for ADNP and ZMYM3 [179] (Figure 4.12 B). The interaction of ADNP with H3K9me3 (mediated by HP1) was further dissected in a recent publication from our group [180]. In contrast to these results, ACTL8 showed no or very little enrichment with H3K9me3. However, the protein contains a PxVxL motif, known to be present in proteins recruited to chromatin by HP1 – a direct H3K9me3 binder [181]. Moreover, the overexpressed protein showed exclusion from the nucleus as seen by immunofluorescence imaging (data not shown). Importantly, ACTL8 was identified as a prominent binder in several other affinity purification experiments using heterochromatic modifications (presented below). Finally, the recruitment of SMCHD1 was also verified using specific antiserum (Figure 4.12 B, lower panel).
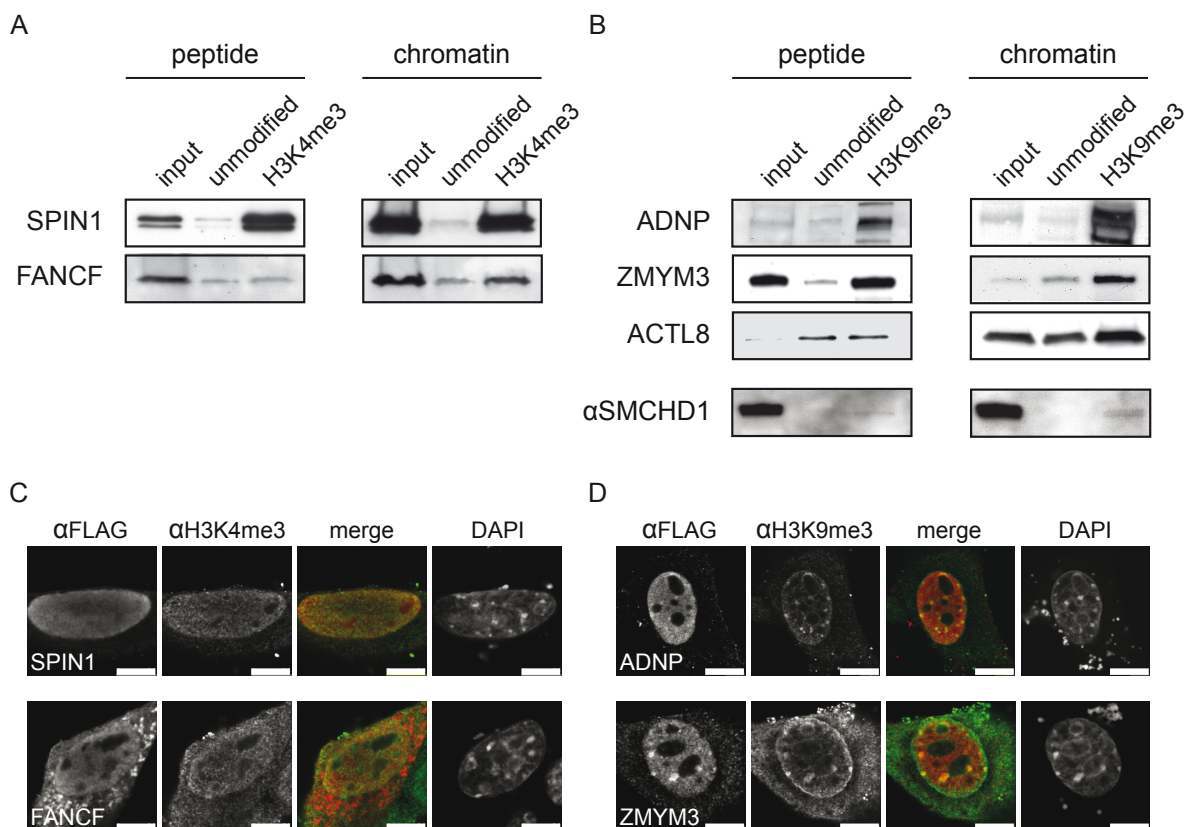
Figure 4.12 Validation of proteins recruited by H3K4me3 and H3K9me3.

A-B. The indicated proteins containing C-terminal FLAG-tag were transiently expressed in 293T HEK cells, which used for preparation of nuclear extracts. Chromatin and peptide affinity purification with H3K4me3 (**A**) and H3K9me3 (**B**) modifications was performed, the bound proteins were separated on SDS-PAGE and transferred on nitrocellulose membrane and immunoblotted using αFLAG antibodies, except for SMCHD1 where wt cells and αSMCHD1 antibodies were used. **C-D.** The indicated proteins containing C-terminal FLAG-tag were transiently expressed in NIH3T3 cells. Immunofluorescence was performed using αFLAG, αH3K4me3 (**C**) and αH3K9me3 (**D**) antibodies. DNA was stained with DAPI. Scale bar corresponds to 7.5 μm. The figure originally appeared in [78].

In addition to the biochemical validation discussed above, the cellular localization of the selected factors was also tested in comparison with H3K4me3 and H3K9me3 nuclear distribution (Figure 4.12) in NIH3T3 cells. Both transiently expressed SPIN1 and FANCF showed diffuse pattern, similar to the distribution of H3K4me3 (Figure 4.12 C). ZMYM3 and ADNP showed enrichment in discrete nuclear foci, which were also marked by H3K9me3 and were DNA-dense (panel D). In summary, the findings from the MS-based screen were recapitulated for 5 out of 6 tested factors.

4.2 Analysis of the interactome of different chromatin modifications

Following the validation of the novel chromatin affinity purification approach, it was used to define the interactomes of a set of histone and DNA modifications. In this chapter the proteins specific for each individual modification will be briefly described, then analyzed comparatively and globally. The focus was on the heterochromatin-associated modifications H3K9me2, H3K9me3, H3K27me2, H3K27me3, meCpG, a combination of two of them (H3K9me3 and meCpG), the euchromatic H3K4me3, and the more poorly understood H3K27me1, which has been implicated both with euchromatin and heterochromatin functions [182]. Furthermore, monoubiquitylation of H2B (H2Bub1) was also tested, in addition to the difference between unmodified and H3Δ1-20 containing chromatin. Lastly, the H2Bub1-dependent recruitment of the SWI/SNF chromatin remodeling complex and its effect on transcription of selected genes was investigated.

4.2.1 Proteins recruited by distinct chromatin modifications

Overview and thresholds

More than 4900 proteins were identified in total in all performed experiments – a value close to the number of proteins identified in the HeLa nuclear extract (Ilian Atanassov, manuscript in preparation). Of these, a varying number of highly enriched factors were specific for each modification. Since there is no consensus in the proteomics community which values reflect clear enrichment in a SILAC experiment, a fixed enrichment ratio threshold within two independent experiments was applied in the following studies. First, the limit of +/- 1 (log2 scale) for each forward and reverse experiments within individual SILAC pull-downs was set. This ensured that false positive proteins showing opposite ratios in both experiments (i.e. being identified as enriched only in light or only in heavy form, e.g. the exclusively light histone proteins part of the affinity matrix) would be excluded. Additionally, this threshold would also discard factors showing generally non-reproducible enrichment between the two experiments. Next, an additional threshold of +/-1.5 (log2 scale) for the average enrichment ratio between the forward and reverse experiments was set, resulting in a flat borderline between the putative background and specifically enriched proteins (Figure 4.13, dark grey boundaries). All proteins above threshold levels for each modification are listed in Table 3 (a detailed list of the proteins with identifiers and enrichment ratios is included in Appendix 1). Nevertheless, as some proteins could be specifically enriched with lower ratios, the threshold level was decreased for some global analyses (as indicated) but enrichment in more than one experiment was required.
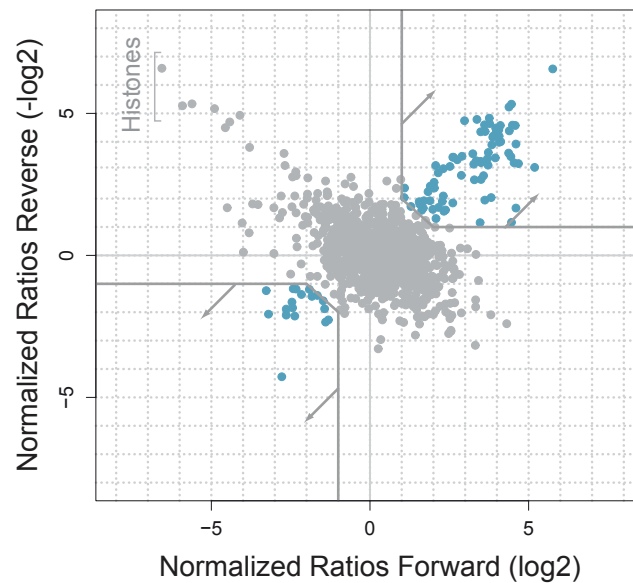
Figure 4.13 Enrichment ratio threshold boundaries.

Scatterplot representing the boundaries of the enrichment ratio threshold used for analysis of the individual interactomes (+/-1 in each F or R experiment, +/-1.5 average, log2 scale). Proteins enriched above the set threshold limits (dark grey lines) are coloured in blue.

➔ **Table 3 Proteins enriched or excluded within each chromatin affinity purification.**

A list of the proteins enriched (green arrow) or depleted (red arrow) in each of the performed experiments sorted according to their enrichment ratios (top – highest enrichment, bottom – highest depletion). Proteins occurring in more than one experiment are colour-coded as indicated in the legend.

H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3 Δ1-20	meCpG H3K9me3

H3K4me3

↑ OCR, ↑ KIAA1111, ↑ CHD1, ↑ C17orf1, ↑ TAF2F, ↑ C6orf61, ↑ KIAA1523, ↑ TAF3, ↑ KIAA0700, ↑ KIAA1227, ↑ GTF2A1, ↑ TAF2C, ↑ TAF2D, ↑ C20orf154, ↑ MBTD1, ↑ hCG_181099, ↑ ESYT1, ↑ TAF2E, ↑ ARID4A, ↑ ARID4B, ↑ KIAA0764, ↑ C11orf30, ↑ JARID1A, ↑ BRMS1L, ↑ FANCF, ↑ KIAA1291, ↑ C17orf53, ↑ GTF2A2, ↑ KIAA0886, ↑ HARP, ↑ PRO2134, ↑ NOC4L, ↑ FACE, ↑ TAF12, ↑ C6orf90, ↑ EPC2, ↑ SAP45, ↑ HSPC301, ↑ TAF9B, ↑ TAF2G, ↑ IPO12, ↑ PPP1R10, ↑ C11orf23, ↑ CHD8, ↑ BA2R, ↑ COPA, ↑ RNF114, ↑ C20orf158, ↑ TAF10, ↑ KIAA1429, ↑ SIN3A, ↑ GTF2D1, ↑ TOP3, ↑ SAP130, ↑ TOX4, ↑ CAGH32, ↑ POLR2B, ↑ TGIF, ↑ ING1, ↑ BRD8, ↑ SPK, ↑ CHET9, ↓ JNKK2, ↓ C17orf96, ↓ HSPC144, ↓ MGMT, ↓ HMG20A, ↓ BRAF35, ↓ KIAA0783, ↓ UHRF1, ↓ HAUSP, ↓ PCL3, ↓ ACTL8, ↓ SCML2, ↓ AIM

H3K9me3

↑ C1orf103, ↑ CBX5, ↑ CBX1, ↑ NIRF, ↑ CBX3, ↑ KIAA0650, ↑ ADNP, ↑ C13orf8, ↑ POGZ, ↑ RNF95, ↑ KIAA0892, ↑ CAF, ↑ CAF1A, ↑ MAFK, ↑ ZMYM3, ↑ NIPBL, ↑ KIAA0425, ↑ ACTL8, ↑ ATF7, ↑ HSPC189, ↑ KIAA1205, ↑ GTF3C1, ↑ AIM, ↑ GTF3C5, ↑ GTF3C3, ↓ BBAP, ↓ ACAC, ↓ BAL2

meCpG

↑ ZHX2, ↑ ZHX1, ↑ HOMEZ, ↑ ZHX3, ↑ KAISO, ↑ ZNF787, ↑ INRF2, ↑ A8K4Q3, ↑ MBD2, ↑ FOSL2, ↑ MAFF, ↑ CDK2AP1, ↑ MTA1L1, ↑ MAFK, ↑ POGZ, ↑ JUND, ↑ ACTL8, ↑ FBL12, ↑ MAFG, ↑ ATF7, ↑ DMAHP, ↑ GATAD2B, ↑ GATAD2A, ↑ RBAP46, ↓ CHD4, ↓ CHD3, ↑ UHRF1, ↓ PCNA, ↓ BAP1, ↓ DPY30, ↓ BRD2, ↓ DP1, ↓ MTF2, ↓ E2F3, ↓ KIAA1798, ↓ KIAA0309, ↓ MEN1, ↓ BHLHB12, ↓ ZFP38, ↓ TIM13B, ↓ SAMD1, ↓ BHLHB2, ↓ BHLHB11, ↓ BCOR, ↓ RBBP5, ↓ EMC19, ↓ ASH2L, ↓ NSPC1, ↓ BHLHB39, ↓ BEND3, ↓ WDR75, ↓ CIRH1A, ↓ UTP15, ↓ L14, ↓ ALL1, ↓ BAP28, ↓ KIAA0007, ↓ HRX2, ↓ ZNF131, ↓ IFI16, ↓ C17orf53, ↓ DEAF1, ↓ CXXC8

H3K9me2

↑ UHRF1, ↑ ACTL8, ↑ CBX5, ↑ ARFIP2, ↑ A2D, ↑ KDP, ↑ EDC4, ↑ NIPBL, ↑ AIM, ↑ GARNL1, ↑ NUP358, ↑ AKAP13, ↑ ATX, ↑ HAUSP, ↑ TRIO, ↑ KIAA1057, ↑ hCG_181214, ↑ GMDS, ↑ HSP75, ↑ EDC3, ↓ SP3, ↓ POLR2, ↓ CREM, ↓ INO80, ↓ E2F3, ↓ POLR3D, ↓ DR1, ↓ ZBTB14, ↑ MAFF, ↓ CIC, ↓ POLR2E, ↓ KIAA1452, ↓ BANP, ↓ ATF1, ↑ DP1, ↓ KIAA0997, ↓ CREM, ↓ ASYIP, ↑ FOSL2, ↓ CEBPB, ↓ ERCC2, ↓ TIGD1, ↓ FOXK2, ↓ BTEB3, ↓ BRD4, ↓ CANPL1, ↓ ELF4, ↓ POLR3F, ↓ WEE1, ↓ AIDD, ↑ KIAA1227, ↓ CREB1, ↑ JUND, ↓ JUN, ↓ IFI16, ↓ ZBP89, ↓ BHLHD3, ↓ IFI16, ↓ BTEB4, ↑ CTCF, ↓ E4F, ↑ TAF2E, ↓ BHLHC41, ↑ HMBOX1, ↑ TFAP2A, ↑ NFIA, ↑ ATF7, ↓ KIAA0414, ↑ NFIB, ↑ NFIC

H3K27me1

↑ DEK, ↑ CTCF, ↑ DDX36, ↑ POLR3D, ↑ POLR3G, ↑ POLR3C, ↑ FBL6, ↑ KIAA1452, ↑ IFI16, ↑ AAG, ↑ NSEP1, ↑ NFIA, ↑ SRBD1, ↑ POLR3A, ↑ C20orf167, ↑ POLR3F, ↑ C14orf117, ↑ H1FX, ↑ POLR3B, ↑ C16orf88, ↑ NFIC, ↑ MUTYH, ↑ HARP, ↑ POLR2E, ↑ CCDC86, ↑ KIAA1649, ↑ PURB, ↑ MUS81, ↑ POLR1C, ↑ CMAS, ↑ ZFP276, ↑ KIAA1665, ↑ PAF53, ↑ ANKT, ↑ DNAJC9, ↑ KIAA1227, ↑ TCF13, ↑ E4F, ↑ ZBTB14, ↑ E2F3, ↑ DP1, ↑ HMBOX1, ↑ TRIP12, ↑ TCF6, ↑ PUR1, ↑ TRIP12, ↑ C20orf1, ↑ TAF1C, ↑ MSH2, ↑ KIAA1196, ↑ POLR2L, ↑ DP2, ↑ ATF7, ↑ RB1, ↑ CSDA, ↑ AUF1, ↑ INO80S, ↑ MBD4, ↑ TFAP2A, ↑ ZBTB9, ↑ ATAD5, ↑ ARID4B, ↑ GLYR1, ↑ POLM, ↑ ZFP64, ↑ MECP2, ↑ AHCTF1, ↑ NR2C2, ↑ UHRF1, ↑ RFX5, ↓ KIAA0886

H3K27me2

↑ PHF1, ↑ POLR3D, ↓ C14orf151, ↓ MYH10, ↓ USP3, ↓ G2E3

H3K27me3

↑ PHF1, ↑ LRWD1, ↑ ORC2, ↑ UHRF1, ↑ ENO1, ↑ ORC5, ↑ FKSG13, ↑ LATHEO, ↑ CBX5, ↑ EIF2G, ↑ POLR3D, ↑ CMAS, ↑ TCF13, ↑ DP1, ↑ PAF53, ↑ PCL3, ↑ TFAP2A, ↑ RBAP48, ↑ DDX47, ↑ RING1, ↑ BHLHB2, ↑ LYT10, ↑ ACTL8, ↑ SAP49, ↑ C21LRP, ↓ TAB3, ↑ KIAA1978, ↑ PSMB2, ↑ LIG1, ↓ C14orf125, ↑ KIAA1797, ↓ COPG, ↓ GRL, ↑ KIAA1741, ↓ RTTN, ↑ AMPK, ↑ PSME3, ↓ ARH, ↑ KIAA0791, ↓ PLEKHG4, ↑ ADTB1, ↑ BAT2D1, ↓ TBC1D15, ↓ FIGNL1, ↓ PI4KA, ↑ NUBP2, ↓ WDR3, ↓ AKAP10, ↓ DNMBP, ↓ CRM1, ↓ CNOT3, ↓ GYG, ↓ GMIP, ↓ EIF4E2, ↓ IRAK2, ↑ STX5, ↑ ESP1, ↑ RAD54B, ↓ C10orf46, ↓ RCD1, ↓ C9JYP6, ↓ CDC36, ↓ PYCR1, ↓ C5orf7, ↓ ARHGAP29, ↓ TMCO4, ↓ C22orf18, ↓ C1orf112, ↓ WDR36, ↓ PKN2, ↓ PPT1, ↓ AD-005, ↓ CAF1, ↓ CALIF, ↓ DGT6, ↓ CGI-120, ↓ MADH5, ↓ AIBP63, ↓ CTNND1, ↓ BTBD14A, ↓ C10, ↓ CNOT10, ↓ ECHDC1, ↓ KIAA0099, ↓ PNO1, ↓ ARHGEF3, ↓ KIAA1728, ↓ PUM1, ↓ KIAA0913, ↓ DIAPH3, ↓ PDE3A, ↓ PATL1, ↓ CVAK104, ↓ HSF1

H2Bub1

↑ SPT4H, ↑ C1orf193, ↑ C9orf80, ↑ LP3587, ↑ PPP2CA, ↑ INTS2, ↑ INTS12, ↑ SPT5, ↑ RSBN1L, ↑ INTS5, ↑ INTS9, ↑ INTS1, ↑ KIAA0852, ↑ A-152E5.7, ↑ INTS4, ↑ OBFC2A, ↑ C1orf73, ↑ C7orf26, ↑ DBI1, ↑ RSBN1, ↑ KIAA0459, ↑ POLR2B, ↑ C12orf11, ↑ CPSF3L, ↑ C8orf35, ↑ C8orf52, ↑ PPP2R1A, ↑ POLR2G, ↑ C19orf62, ↑ PPP2CB, ↓ POLR2, ↑ RECQL5, ↑ PPP2R1B, ↑ ABRA1, ↑ NELFA, ↑ RAP80, ↑ BRCC3, ↑ BTBD12, ↑ C15orf44, ↑ DYRK, ↑ KIAA0211, ↑ POLR2J1, ↑ DCAF7, ↑ KIAA1125, ↑ POLR2D, ↑ POLR2, ↑ HSPC130, ↑ BRCC45, ↑ RNF168, ↑ AD-009, ↑ KIAA1991, ↑ COBRA1, ↑ NELFE, ↑ KIAA1441, ↑ C19orf7, ↑ GPN1, ↑ ERCC1, ↑ C20orf94, ↑ ANP32B, ↑ ERCC11, ↑ LAP2, ↑ ANCO1, ↑ ANCO2, ↑ TOX4, ↑ GRWD, ↑ POLR2L, ↑ RAD18, ↑ WDR82, ↑ POLR2E, ↑ KPNA1, ↑ BAF155, ↑ KIAA0035, ↑ EPC1, ↑ CPSF1, ↑ hCG_29955, ↑ BAF170, ↑ BCL7C, ↑ BAF190B, ↑ KIAA0035, ↑ GPS2, ↑ CPSF100, ↑ IPOA7, ↑ PPP1R10, ↑ CHD1, ↑ ADNP, ↓ HMG17L3, ↓ ANAPC3, ↓ EFO1, ↓ TRIP12, ↓ CATX11, ↓ HMG14, ↓ TRIP12, ↓ CCDC86, ↓ ANX2LG, ↓ ANAPC7, ↓ ANAPC10, ↓ ANAPC8, ↓ RAN, ↓ KIAA0170, ↓ CHC1, ↓ ACAC

H3 Δ1-20

↑ BAZ1B, ↑ UHRF1, ↑ ACAC, ↑ KIAA0783, ↑ JADE3, ↑ BRD2, ↑ C10orf12, ↓ BCLAF1

meCpG H3K9me3

↑ UHRF2, ↑ CBX5, ↑ ZHX2, ↑ HOMEZ, ↑ UHRF1, ↑ MECP2, ↑ KIAA1584, ↑ ZHX1, ↑ CBX1, ↑ FIZ1, ↑ CBX3, ↑ ACTL8, ↑ ZHX3, ↑ ADNP, ↑ UBCH5C, ↑ HAUSP, ↑ HSPC189, ↑ LRWD1, ↑ KIAA0414, ↑ MBD2, ↑ ZNF580, ↑ ORC2, ↑ KIAA1231, ↑ NIPBL, ↑ CDYL, ↑ ZNF787, ↑ POGZ, ↑ LATHEO, ↑ CHD3, ↑ SCML2, ↑ MUTYH, ↑ HMBOX1, ↑ AOF1, ↑ MBD4, ↑ CHD4, ↑ MTA1L1, ↑ A8K4Q3, ↑ C20orf88, ↑ FBL6, ↑ MAFG, ↑ RBAP46, ↑ MAFF, ↑ GATAD2A, ↑ ATF7, ↑ KIAA0892, ↑ NRF1, ↑ KIAA1196, ↑ IFI16, ↑ TAF2D, ↑ GATAD2B, ↑ IFI16, ↑ TFAP2A, ↑ TCHH, ↑ KIAA1266, ↑ PUR1, ↑ ORC5, ↑ NFIB, ↑ LSF, ↑ NFIA, ↑ SRBD1, ↑ PURB, ↑ NFIC, ↑ KIAA1090, ↑ MAFK, ↑ ZMYM3, ↑ LBP1, ↑ IGKJRB, ↑ HDAC2, ↑ AIM, ↑ HDAC1, ↑ C18orf37, ↑ INO80S, ↑ EZF, ↑ SIX4, ↑ JUND, ↑ ACF1, ↑ RBAP48, ↑ DMAHP, ↑ ZKSCAN4, ↑ DUC1, ↑ OCR, ↑ EED, ↑ A8K1E1, ↑ HAP3, ↑ NFYC, ↑ FOSL2, ↑ MSH2, ↑ JUN, ↑ ACTR8, ↑ NFYA, ↑ ZBTB15, ↑ CTCF, ↑ PKNOX1, ↑ CHRAC1, ↑ USP3, ↑ MEIS2, ↑ AYP1, ↑ INO80, ↓ HUSSY-29, ↓ FPGS

Number of experiments where the protein was enriched

Number of experiments
8
6
4
3
2
1

H3K4me3

The affinity purification using H3K4me3 chromatin identified a large number of putative interaction partners (Figure 4.14 A). The major group of proteins of the interactome of the H3K4me3 modification were the multiple subunits of the TFIID complex (e.g. TAF3, TBP). This complex binds directly to H3K4me3 (via TAF3) and the interaction is required for transcriptional activation [81]. Additionally, several subunits of another known H3K4me3 interactor – the SAGA complex (e.g. the chromatin remodeling factor CHD1) [75], were identified. The SAGA subunit binding directly to this modification (CCDC101/SGF29) [183] was also detected, however with a ratio in one of the experiments slightly lower than the set threshold. SPIN1 was another example of a protein known to be directly recruited by H3K4me3 [177]. Other proteins than exert positive transcriptional regulation were the components of the NuA4 histone acetyl-transferase (HAT) complex – EP400, KAT5, ING3 and BRD8 [184]; the mono- and dimethyl lysine-specific demethylase PHF8, known to demethylate e.g. H3K9me1/2 and H4K20me1 [185]; several subunits of the general transcription factor TFIIA [186]. Interestingly, several proteins with transcriptional repression functions were also part of the H3K4me3 interactome – the HDAC-associated BRMS1L, SUDS3 [187] and ING2 - a known H3K4me3 binder [166], and importantly - the H3K4me2/3-specific lysine demethylase KDM5A [188]. Furthermore, several factors were excluded by this modification and bound preferentially to unmodified chromatin. The protein UHRF1 isknown to bind the unmodified N-terminus of H3 and this interaction is abolished when the
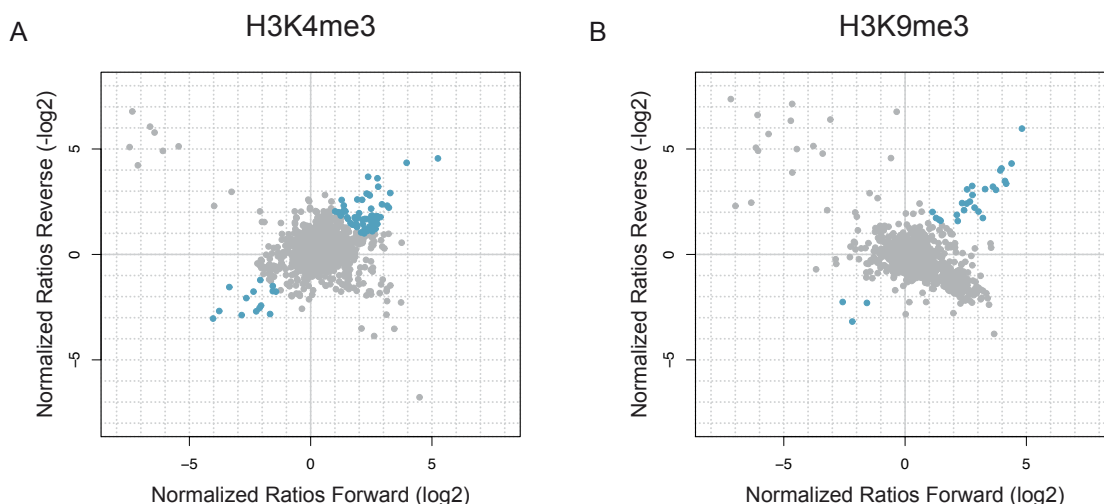
Figure 4.14 Enrichment ratio distribution of the interactomes of H3K4me3 and H3K9me3.

Enrichment ratios (log2 scale) from forward (x-axis) and reverse (y-axis) experiments for H3K4me3 (**A**) and H3K9me3 (**B**) chromatin. Proteins enriched or repelled by the modifications with ratios higher than the set threshold (+/-1 in each F or R experiment, +/-1.5 average, log2 scale) are coloured in blue.

R2 or K4 residues of H3 are modified [84]. Additionally, several UHRF1 interacting partners, including DNMT1/AIM and HAUSP/USP7 [189], were also detected. Another example of a protein from this group was ACTL8, which was already tested in the validation experiments for H3K9me3 (see 4.1.3).

H3K9me3

Among the known interactions partners of H3K9me3 (Figure 4.14 B) identified here were the three isoforms of the hallmark heterochromatin protein HP1 (CBX1/3/5) [167], as well as their interaction partners ADNP [180] and POGZ [173]. UHRF1, ACTL8 and DNMT1, which were excluded by H3K4me3, were strongly enriched with H3K9me3. UHRF2 was another example of a protein that was recently shown to bind H3K9me3 with high affinity [190]. Several zink-finger domain proteins (ZMYM3, ZMYM4, ZNF828, ZNF581) were also identified as specifically enriched, together with three subunits of the general RNA polymerase III transcription factor TFIIIC [191]. Of the proteins repelled by the H3K9me3 modification were the ubiquitin ligase DTX3L, responsible for H4K91 ubiquitylation, and the functionally related poly(ADP-ribose) polymerase PARP14 [192].

H3K9me2

The interactome of the heterochromatic modification H3K9me2 [43] is far less studied than the one of its trimethyl counterpart. The H3K9me2 chromatin pull-down showed asymmetrical enrichment ratio distribution with only a few proteins enriched above the set threshold (Figure 4.15 A). The only known H3K9me2 binder - PGC7, was not identified in the dataset. However, the interaction of PGC7 with H3K9me2 was shown only in mice and in the context of early embryogenesis [193]. Interestingly, ACTL8, UHRF1, UHRF2, CBX5, NIPBL, DNMT1, HAUSP – all of which were already described in the H3K9me3 dataset, were identified with H3K9me2. Additionally, enriched were the ubiquitin ligase HECTD1 [194] and the ubiquitin-specific peptidase USP24 [195]. The proteins excluded by this modification represented a larger set. These included a large number of sequence-specific activating transcription factors - several subunits of the transcription factor nuclear factor I (NFI) [196], E2F3 [197], ELF4 [198], ATF7 [199]. On the other hand, DR1 [200], CTCF [201] and ZNF148 [202] were among the proteins related to transcriptional repression. Two subunits of RNA polymerase II and three subunits of RNA polymerase III were also found.
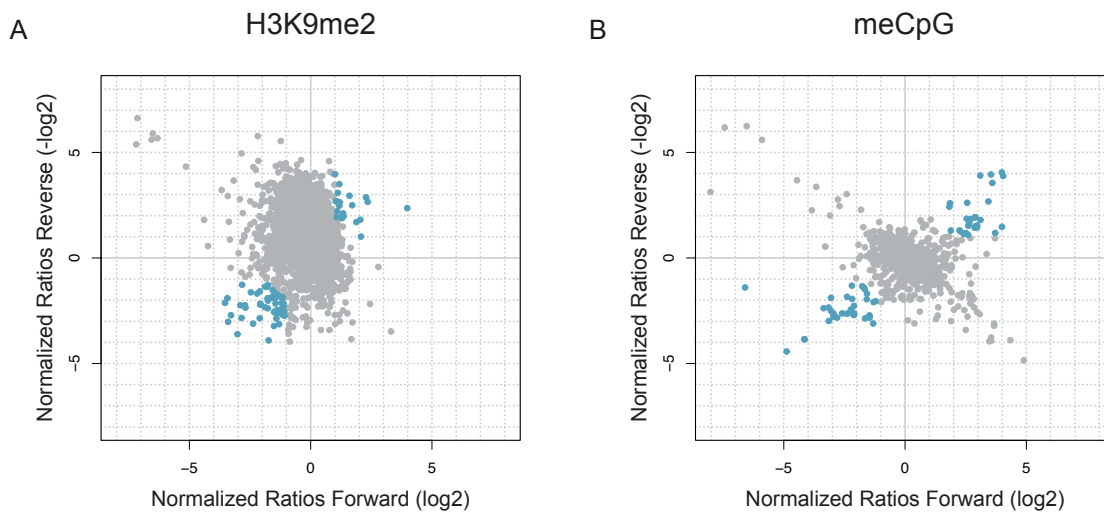
Figure 4.15 Enrichment ratio distribution of the interactomes of H3K9me2 and meCpG.

Enrichment ratios (log2 scale) from forward (x-axis) and reverse (y-axis) experiments for H3K9me2 (**A**) and meCpG (**B**) chromatin. Proteins enriched or repelled by the modifications with ratios higher than the set threshold (+/-1 in each F or R experiment, +/-1.5 average, log2 scale) are coloured in blue.

Methylated DNA (meCpG)

        The interactome of DNA methylated at 5'C in CpG context (meCpG) revealed a similar number of proteins recruited and repelled by this modification (Figure 4.15 B). Among the known meCpG binders [203] MBD2 and KAISO were faithfully identified as enriched, while MBD3, MBD4, ZBTB4 and MECP2 showed ratios slightly lower than the set threshold or were found with very low ratio counts only in forward or reverse experiments. UHRF1, which was described in the context of H3K4me3 and H3K9me3, was also enriched here, supporting previous reports about its binding to methylated DNA [73]. Other proteins enriched in this dataset include ACTL8, POGZ, the homeobox leucine zipper protein HOMEZ [204] and most prominently - the three members of the zink fingers and homeoboxes (ZHX) family – ZHX1, ZHX2, ZHX3, implicated in transcriptional co-repression [205]. Of the proteins repelled by DNA methylation, KDM2A - a H3K36 demethylase, was previously shown to bind non-methylated CpG sequences [206]. Other proteins binding preferentially to unmodified chromatin in this context were WBP7 (or MLL4), MLL and ASH2L - components of H3K4-specific methyltransferase complexes [207-209]; multiple transcription factors, including USF1, USF2 [210], E2F3 [197], MTF2 [211]; the polycomb group member and H2AK119-secific ubiquitin transferase RNF2 [212].

H3K27me1

Monomethylation of Lys27 on H3 has been generally associated with transcriptional repression [213]. However, several reports also indicate roles in actively transcribed genes [182]. The interactome of H3K27me1 (Figure 4.16 A) revealed nearly 70 specifically enriched proteins, including a large number of subunits of RNA polymerases II and III. Additionally, recruited were proteins implicated in both positive and negative transcriptional regulation. Members of the former include E2F3 [197], ATF7 [199], NFIA/C [196], TEAD1 [214]. Proteins with known negative transcriptional functions include for example the meCpG binding protein MECP2 [203], ARID4B and RB1 [215], the chromatin insulator protein CTCF [216]. Additionally, UHRF1 was identified with this modification as well. The only protein found to be excluded by the modification was the reticulon protein RTN4 [217].

H3K27me2

This modification is associated with transcription repression, however not much is known about its direct functions in recruiting chromatin proteins [43]. Surprisingly, only two proteins appeared enriched here (Figure 4.16 B). The first, and highly enriched was PHF1 – a polycomb group protein involved in response to DNA double-strand breaks [218], and the second and with much lower ratio – the RNA polymerase III subunit POLR3D. Interestingly, both proteins were identified in the H3K27me3 dataset, and POLR3D was also present in the H3K27me1 interactome. The proteins repelled by the modification included the E3 ubiquitin ligase G2E3, required for early embryonic development [219], and USP3 – a H2A/H2B deubiquitinating enzyme [220].

H3K27me3

Trimethylation of Lys27 on H3 is a well-studied histone modification implicated in polycomb-mediated transcriptional silencing [221], X-chromosome inactivation [222] and establishment of bivalent chromatin domains, poised for transcriptional activation or silencing [41]. Among the highly enriched factors here (Figure 4.16 C) were LRWD1 and origin recognition complex (ORC) subunits, known to be associated with the repressive H3K9me3 and H4K20me3 modifications [223]. The only polycomb group proteins above the enrichment ratio threshold were the H2AK119-specific ubiquitin ligase RING1 – a subunit of PRC1 [224] and PHF1 [218]. Additionally, several other polycomb group proteins were identified, but either with lower ratios, or not quantified (e.g. BMI1, CBX4, CBX8, EZH1, EZH2, PHC2, SUZ12). Other proteins include the already descibed UHRF1, ACTL8, TEAD1, CBX5, as well as the splicing factor SF3B4 [225] and the transcription factor DP-1 [226].
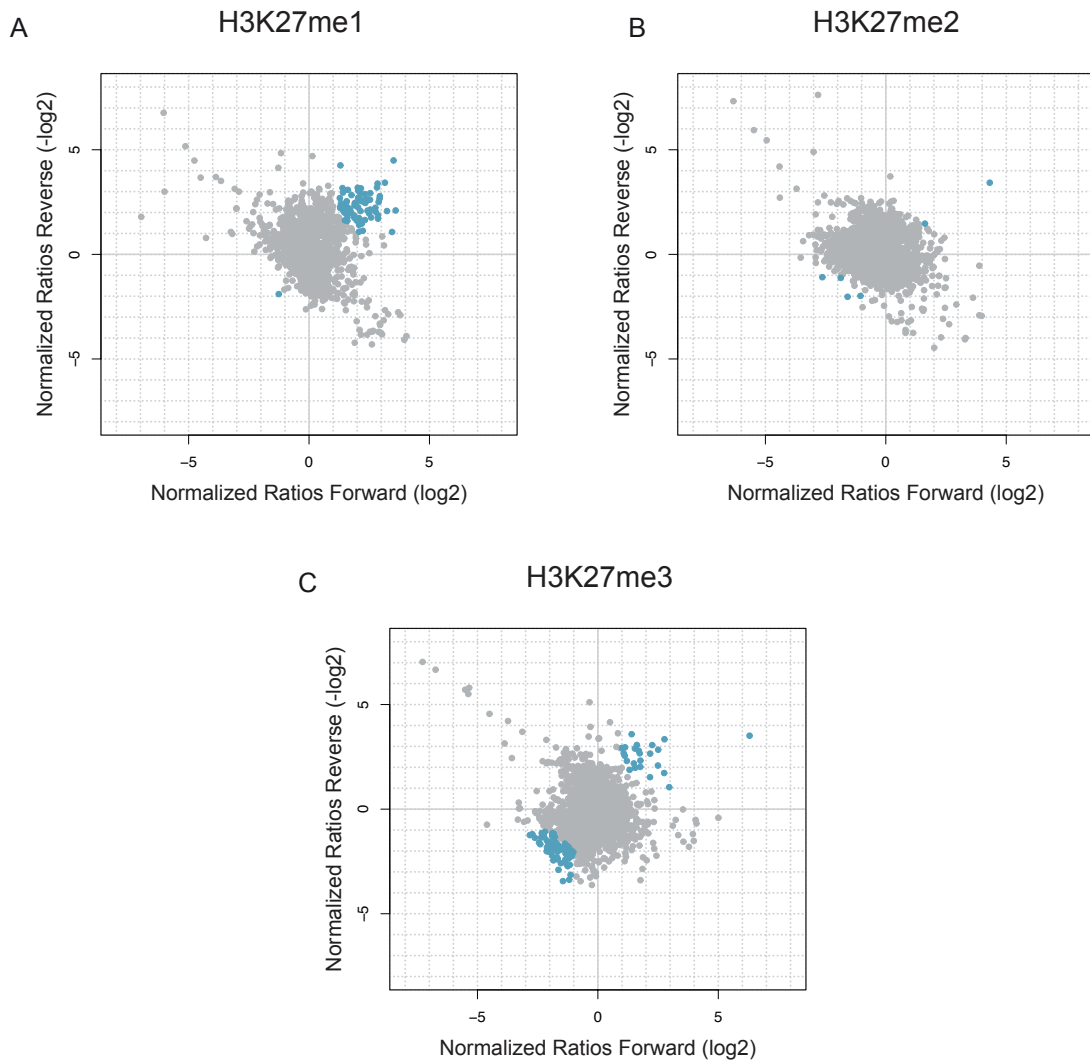
Figure 4.16 Enrichment ratio distribution of the interactomes of H3K27me1, H3K27me2 and H3K27me3.

Enrichment ratios (log2 scale) from forward (x-axis) and reverse (y-axis) experiments for H3K27me1 (**A**), H3K27me2 (**B**) and H3K27me3 (**C**) chromatin. Proteins enriched or repelled by the modifications with ratios higher than the set threshold (+/-1 in each F or R experiment, +/-1.5 average, log2 scale) are coloured in blue.

Interestingly, among the proteins excluded by H3K27me3 were six subunits of CCR4-NOT – a protein complex involved in various aspects of mRNA metabolism, including transcriptional regulation, mRNA decay and transport [227], the H3K9-specific histone demethylase KDM3B [228], the WD repeat-containing proteins WDR3 and WDR36 [229].

H2Bub1

Ubiquitination of H2B on K120 (H2Bub1) has been linked to transcriptional activation/elongation, nucleosome assembly/disassembly and DNA damage [230, 231]. The analysis of the H2Bub1 pull-down revealed more than 90 enriched proteins (Figure 4.17 A), many of which are members of multiprotein complexes. Importantly, one of the identified factors was WDR82, a known H2Bub1-binding partner [89]. Remarkably, the list included all subunits of the negative elongation complex NELF [232], the elongation complex DSIF [233], 11 subunits of the snRNA processing Integrator complex [234], multiple subunits of RNA polymerase II and the SWI/SNF chromatin remodeling complex [235]. Additionally, enriched with H2Bub1 were the DNA helicase RECQL5, a putative tumor suppressor involved in maintaining genome stability [236] and the DYRK1A kinase, implicated in neuronal development [237]. Several of the identified proteins have been implicated in heterochromatin related roles, including the CW zinc finger protein MORC2, involved in histone deacetylation [238] and ADNP - a HP1 interacting partner linked to silencing of major satellite repeats [180]. The most prominent H2Bub1 excluded factors were the regulator of chromosome condensation RCC1 [239] as well as four subunits of the anaphase promoting complex [240].
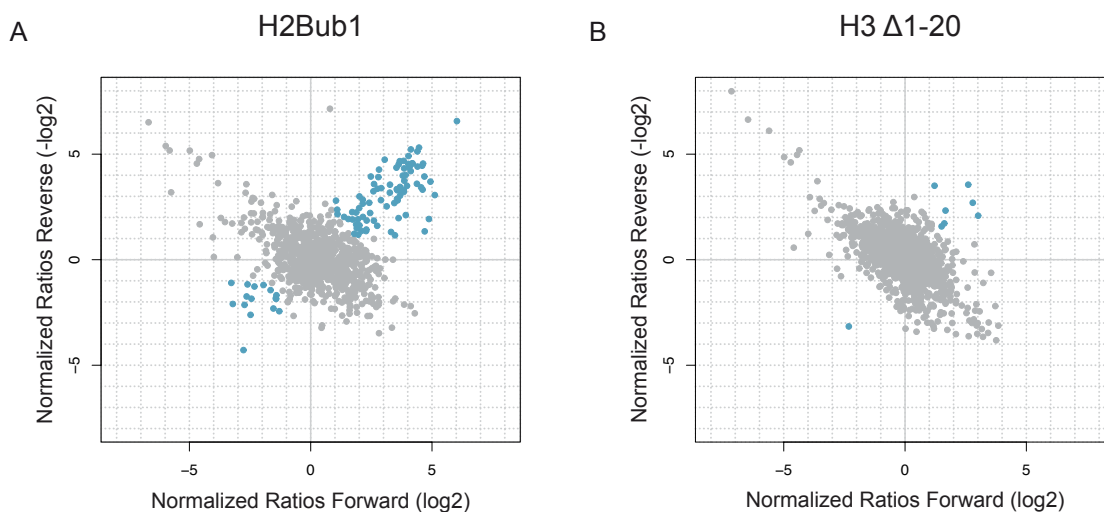
**Figure 4.17 Enrichment ratio distribution of the interactomes of H2Bub1 and H3 Δ1-20.**

Enrichment ratios (log2 scale) from forward (x-axis) and reverse (y-axis) experiments for H2Bub1 (**A**) and H3 Δ1-20 (**B**) chromatin. Proteins enriched or repelled by the modifications with ratios higher than the set threshold (+/-1 in each F or R experiment, +/-1.5 average, log2 scale) are coloured in blue.

H3 Δ1-20

In addition to the presented chromatin modifications, the proteins binding to the unmodified H3 N-terminus *per se* were also investigated. To this aim, an affinity purification experiment comparing the interaction partners of chromatin containing wt H3 to chromatin reconstituted with H3 lacking 20 amino acids from the N-terminus (H3 Δ1-20) was performed. Several proteins indeed bound stronger to the chromatin containing full length H3 (Figure 4.17 B), including the already described UHRF1 [84], the bromodomain-containing proteins BAZ1B [241] and BRD2 [242], and the PHD finger proteins PHF14 and PHF16 [243]. The only protein enriched with the H3 Δ1-20 chromatin was BCLAF2, a transcriptional repressor interacting with members of the BCL-2 family of apoptosis regulatory proteins [244].

H3K9me3 and meCpG, comparison of single against a combination of modifications

In order to study potential synergistic effects of two chromatin modifications, the interactomes of unmodified chromatin and doubly modified H3K9me3/meCpG chromatin (meCpG.H3K9me3) were compared. Nearly 100 proteins were associated specifically with meCpG.H3K9me3 chromatin (Figure 4.18 A). Interestingly, the overlap between this experiment and the interactomes of the individual H3K9me3 (Figure 4.14 B) and meCpG (Figure 4.15 B) modifications was relatively limited. The chromatin carrying both modifications recruited larger number of proteins than the factors from the two single modification experiments taken together (Figure 4.18 B). The proteins enriched in all three experiments were UHRF1, ACTL8 and POGZ. Additionally, several of the factors that were enriched in meCpG.H3K9me3 were also present with high ratios in one of the other two modifications (see crossed out circles in Figure 4.18 C and D) – these represented proteins identified as prominent binders in meCpG (e.g. ZHX1) or H3K9me3 (CBX5). In both comparisons, a number of proteins were enriched with the combination of modifications and present near background level in the other (see solid blue and green circles). The majority of those were the factors specific for the meCpG experiment (in the comparison of meCpG.H3K9me3 with H3K9me3, panel C) or H3K9me3 experiment (in the comparison of meCpG.H3K9me3 with meCpG, panel D). Surprisingly, a large number of proteins excluded by meCpG were not influenced by the combination of meCpG and H3K9me3 (green circles on the negative x-axis in Figure 4.18 D), while only two were uniquely enriched by meCpG. Interestingly, one of the later was the known meCpG binder KAISO [203].

Next, the three datasets were analyzed together in order to identify proteins that were highly enriched in the meCpG.H3K9me3 experiment but remained at background level
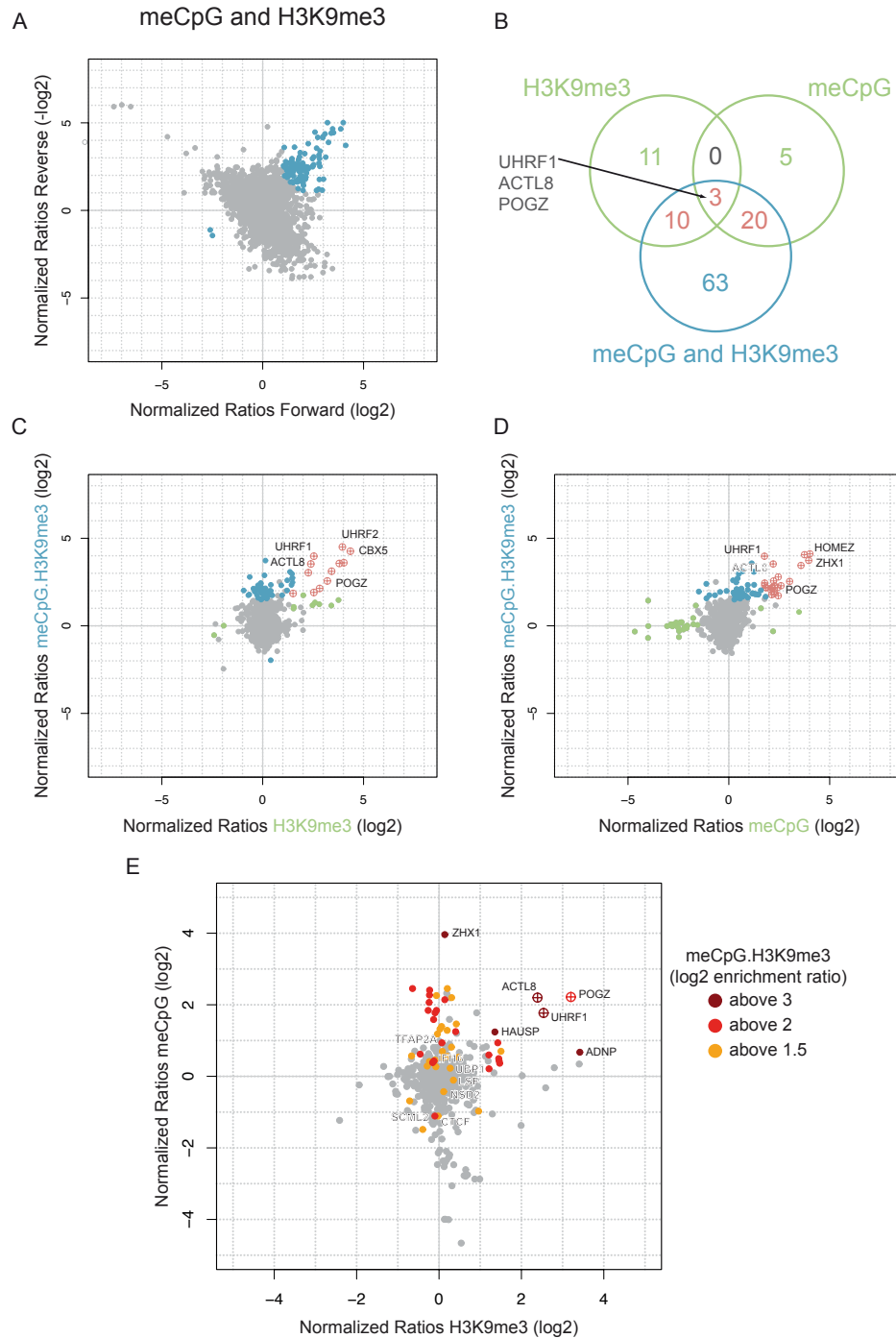
Figure 4.18 The interactome of H3K9me3 in combination with meCpG.

A. Enrichment ratios (log2 scale) from forward (x-axis) and reverse (y-axis) experiments for meCpG.H3K9me3 chromatin. Proteins enriched or repelled by the modifications with ratios higher than the set threshold (+/-1 in each F or R experiment, +/-1.5 average, log2 scale) are coloured in blue. **B.** Venn diagrams representing the overlap of enriched proteins between meCpG, H3K9me3 and meCpG.H3K9me3 experiments. **C-D.** Scatterplots correlating the average enrichment ratios of the meCpG.H3K9me3 experiment with the H3K9me3 (**C**) and meCpG (**D**) experiments. Proteins above enrichment ratio thresholds (see A) are colour-coded as follows: crossed-out red circles - both experiments, meCpG.H3K9me3 – blue circles, H3K9me3 (**C**) or meCpG (**D**) – green circles. **E**. Comparison of the proteins enriched by H3K9me3 (x-axis), meCPG (y-axis) and meCpG.H3K9me3 (colour coded according to enrichment ratio as indicated in the legend).

with both modifications independently. The enrichment ratios of the meCpG and H3K9me3 binders were plotted in a scatter plot and the enrichment ratios from the combined meCpG.H3K9me3 experiment were overlaid as a third dimension using colour code (Figure 4.18 E). Surprisingly, there were several transcription factors that appear to be enriched only when both modifications are present. These included UBP1 – a transcriptional activator that can in addition repress HIV-1 transcription by inhibiting TFIID binding [245]; IFI16, which can function as innate immune sensor for intracellular DNA [246]; several subunits of nuclear factor I (NFI) [196] and the trimeric nuclear transcription factor Y (NFY) [247]; YY1 – a ubiquitous transcription factor that may direct HAT and HDAC enzymes for activation or repression [248, 249]. The histone deacetylases HDAC1 and HDAC2 were also identified in this context. Interestingly, the meCpG-binders MECP2 and MBD4 were among those with high enrichment only in the meCpG.H3K9me3 dataset. Additionally, identified here were the known H3K9me3 binder CDYL [108], the SET, PWWP, HMG and PHD domain-containing probable histone methyltransferase NSD2/WHSC1 [250], the H3K4-specific demethylase KDM1B [228], as well as several origin recognition complex subunits [223]. Strikingly, two of the meCpG.H3K9me3-specific proteins were also identified in the H3K4me3 dataset: SPIN1 – a known H3K4me3 binder [177], and TAF5/TAF2D – a subunit of TFIID [81].

4.2.2 Relationship and overlap between the interactomes of different chromatin modifications

Binary comparison of the interactomes of different modifications

As described above, the analysis of the interactomes of the ten individual chromatin species revealed a set of proteins enriched in multiple chromatin affinity purification experiments. Similar binding partners could indicate functional association; therefore the proteins recruited by selected pairs of modifications were compared next. Illustrating the similarities, differences and, most importantly, the overlap between them was the first step in the global analysis of the investigated interactomes.

Both H3K9me2 and H3K9me3 have been implicated in the context of heterochromatin [43]. The comparison of their interactomes revealed a limited overlap of several highly enriched proteins, including CBX5, UHRF1, ACTL8 and DNMT1 (Figure 4.19 A, crossed-out circles). Some known H3K9me3 binders (e.g. CBX3 and the associated ADNP) showed mild enrichment below threshold levels with H3K9me2, while others (e.g. CHAF1 subunits) were at background level. Additionally, H3K9me2 showed a tendency to repel binding of a large number of factors (green circles along the negative x-axis). These included, for example IFI16, which was enriched with meCpG.H3K9me3, and interestingly – several subunits of RNA polymerase II and TFIID.
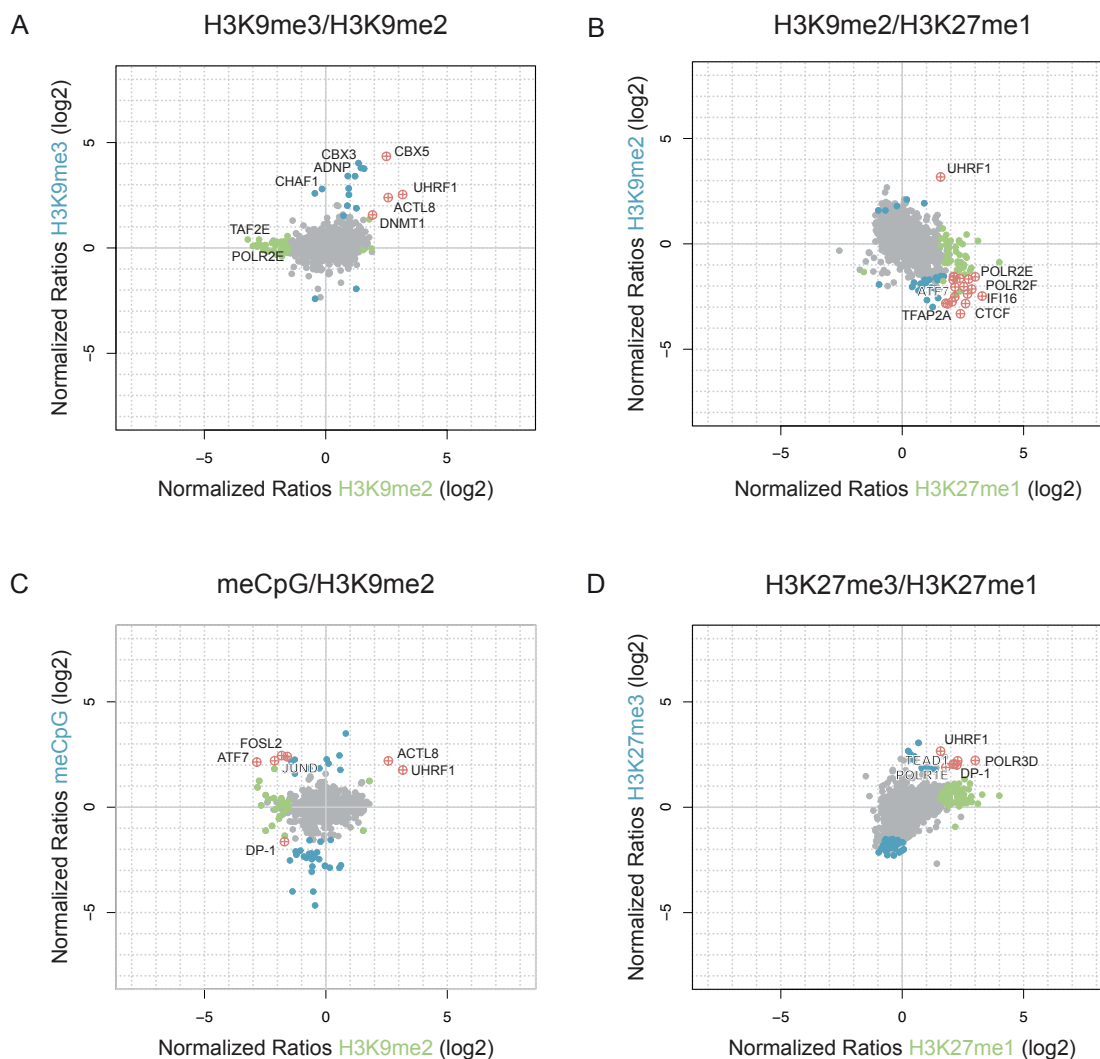
Figure 4.19 Binary comparison of the enrichment ratio distributions of selected chromatin modifications.

Correlation of the average enrichment ratios between H3K9me and H3K9me2 (**A**), H3K9me2 and H3K27me1 (**B**), meCpG and H3K9me2 (**C**) and H3K27me3 and H3K27me1 (**D**). Proteins enriched for only one of the modifications are coloured in green or blue, as indicated. Proteins above threshold levels in both experiments are represented with red crossed-out circles.

The single protein enriched by both H3K9me2 and H3K27me1 (Figure 4.19 B), was UHRF1. Interestingly, a large number of the proteins enriched with H3K27me1 appeared to be repelled by H3K9me2 and the interactomes showed an overall negative correlation of enrichment ratios. These included several RNA polymerase II subunits, the insulator protein CTCF [201], and both activating and repressing transcription factors - ATF7 [199], IFI16 [246], several subunits of nuclear factor I [196], TFAP2A [251].

UHRF1 and ACTL8 were the only two proteins enriched above threshold levels by both H3K9me2 and meCpG (Figure 4.19 C). One protein was excluded by both modifications, albeit mildly – the E2F type transcription factor DP-1 [226], this protein was

enriched by both H3K27me1 and H3K27me3. Several proteins appeared repelled by H3K9me2 and enriched by meCpG – ATF7, as well as two subunits of the transcription factor complex AP-1 (JUND and FOSL2) [252].

While both H3K27me1 and H3K27me3 have been implicated with transcriptional repression [213], recent studies suggest that the monomethyl mark is found in the gene bodies of actively transcribed genes [43, 182]. H3K27me3 can also be found together with H3K4me3 at bivalent chromatin domains poised for transcription [41]. Furthermore, there is a recent report that suggests a role of the trimethyl modification at actively transcribed genes [253]. The comparison of the interactomes of the two marks revealed that the vast majority of proteins enriched with H3K27me1 are at or slightly above background levels in H3K27me3 (Figure 4.19 D, green circles). Conversely, the majority of the proteins repelled by H3K27me3 was near background level or slightly excluded from H3K27me1 (blue circles). Notably, several proteins showed strong enrichment in both experiments. These included UHRF1, subunits of RNA polymerase I (POLR1E) and III (POLR3D), the transcription factors DP-1 [226] and TEAD1 [214].

Several groups have demonstrated that establishing of H2Bub1 is a prerequisite for di- and trimethylation of H3K4, and that the mechanism behind this histone modification cross-talk relies on the recruitment of the specific H3K4 HMT enzymes [254-256]. Interestingly, there was limited overlap between the proteins enriched by H3K4me3 and H2Bub1 (Figure 4.20 A). These included RNA polymerase II subunit POLR2B, two subunits of protein phosphatase complex PTW/PP1 (PPP1R10 and TOX4), known to regulate chromatin structure and cell cycle progression [257, 258], as well as the chromodomain helicase CHD1 [67, 259]. Additionally, there was a group of factors highly enriched with one of the modifications and near background or moderately enriched with the other (blue and green circles near the positive x- and y-axes). H2Bub1 in general also showed limited overlap with the other modifications tested. POLR2E and POLR2L were the common enriched factors with H3K27me1 and ADNP with H3K9me3.

As shown in the previous section, the comparison between meCpG, H3K9me3 and meCpG.H3K9me3 revealed several common interaction partners (Figure 4.18). Next, the overlap between the experiment utilizing meCpG.H3K9me3 chromatin and three single chromatin modifications was investigated. Three of the proteins prominent in almost all modifications screened – UHRF1, ACTL8 and CBX5, were enriched in both meCpG.H3K9me3 and H3K27me3 (Figure 4.20 B). These also included four ORC subunits, known to be associated with the repressive H3K9me3 and H4K20me3 marks [223], the transcription factor TFAP2A [251] and RBAP48 – a subunit of several complexes with functions in chromatin assembly and remodeling and histone deacetylation [260, 261].
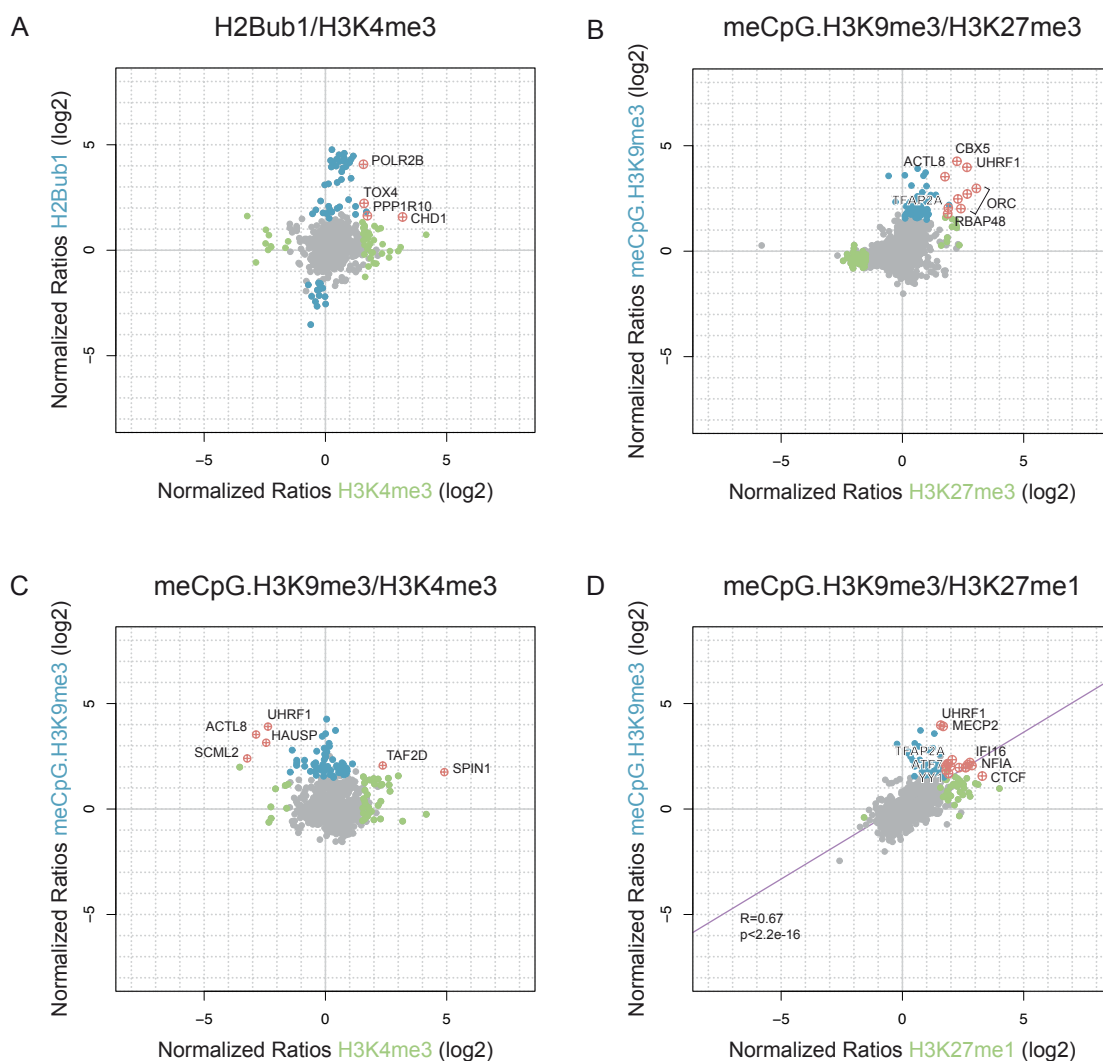
Figure 4.20   Correlation of the enrichment ratio distributions of selected chromatin modifications.

Comparison of the average enrichment ratios between H2Bub1 and H3K4me3 (**A**), meCpG.H3K9me3 and H3K27me3 (**B**), meCpG.H3K9me3 and H3K4me3 (**C**) and meCpG.H3K9me3 and H3K27me1 (**D**). Proteins enriched for only one of the modifications are coloured in green or blue, as indicated. Proteins above threshold levels in both experiments are represented with red crossed-out circles

Two proteins, known to be associated with H3K4me3 (SPIN1 and TAF2D), were also enriched in the meCpG.H3K9me3 pull-down (Figure 4.18 and section 4.2.1). Therefore, further correlation between the two experiments was investigated. The two proteins appeared to be the only enriched above threshold levels in both experiment (Figure 4.20 C), however, a small set of proteins specific for H3K4me3 showed moderate enrichment below threshold level for the double modification as well (green circles below TAF2D). These included several other TFIID subunits. Additionally, as within the comparison between H3K4me3 and H3K9me3 (Figure 4.10), the enrichment ratios of UHRF1, HAUSP and ACTL8 showed negative correlation. The polycomb repressive complex 1 constituent

SCML2 showed similar enrichment pattern. The protein contains two MBT repeats and has been implicated in promiscuous binding to monomethyl lysine, while showing no binding to trimethylated marks [262]. However, it was also enriched with H3K9me2 and more mildly with H3K27me3 and wt H3 in the context of H3 Δ1-20.

Strikingly, the enrichment ratio distributions of meCpG.H3K9me3 and H3K27me1 showed good correlation (Pearson's product correlation coefficient 0.67, p < 2.2e -16) (Figure 4.20 D). A large number of proteins appeared to be enriched above threshold levels in both experiments, including UHRF1, MECP2, IFI16, NFIA, CTCF, TFAP2A, ATF7 and YY1/INO80S. Additionally, the majority of the proteins specific for one of the modifications were enriched just below threshold level with the other (see blue and green circles above the positive x- and y-axes). On the other hand, the two interactomes did not share any proteins repelled by the chromatin modifications.

A protein-protein interaction network of the chromatin modification interactome

Next, a protein-protein interaction network was constructed in order to examine the structural interactions and functional relationships between the proteins recruited by each modification tested. The network was based on known relations inferred from experimental evidence, literature text mining and curated databases (Figure 4.21) [134, 151]. Each protein was represented as individual node, while interacting proteins were linked by edges. Global network analysis [152] revealed relatively good connectedness. 212 out of 324 nodes were connected with an average of 5.154 neighbours. The network diameter (that is, the distance or the path between the two remotest connected nodes) was 13, while the average shortest path length was 4.272. The clustering coefficient of the entire network, which represents the ratio of observed edges to the maximum number of theoretically possible edges, was 0.36.

Visual inspection of the interaction network revealed that the individual modification interactomes were separated well within the network. The H3K27me1 nodes appeared to connect the euchromatic H3K4me3 and H2Bub1 with the heterochromatic modifications. Additionally, a prominent clustering was apparent for the subunits of TFIID (H3K4me3), the RNA polymerase II and the associated complexes NELF, DSIF, Integrator (H2Bub1) and

➔**Figure 4.21 Protein-protein interaction network of the chromatin modification interactome.**

Interactome network constructed based on information from experiments, peer-reviewed publications and curated databases. Known structural and functional relationships are represented as edges connecting individual proteins (nodes). The weight of the edges is indicative of the confidence of the interaction (PPI score, as indicated). Nodes are colour-coded according to identification within separate experiments, and labeled using canonical gene names as reported by STRING database [151].
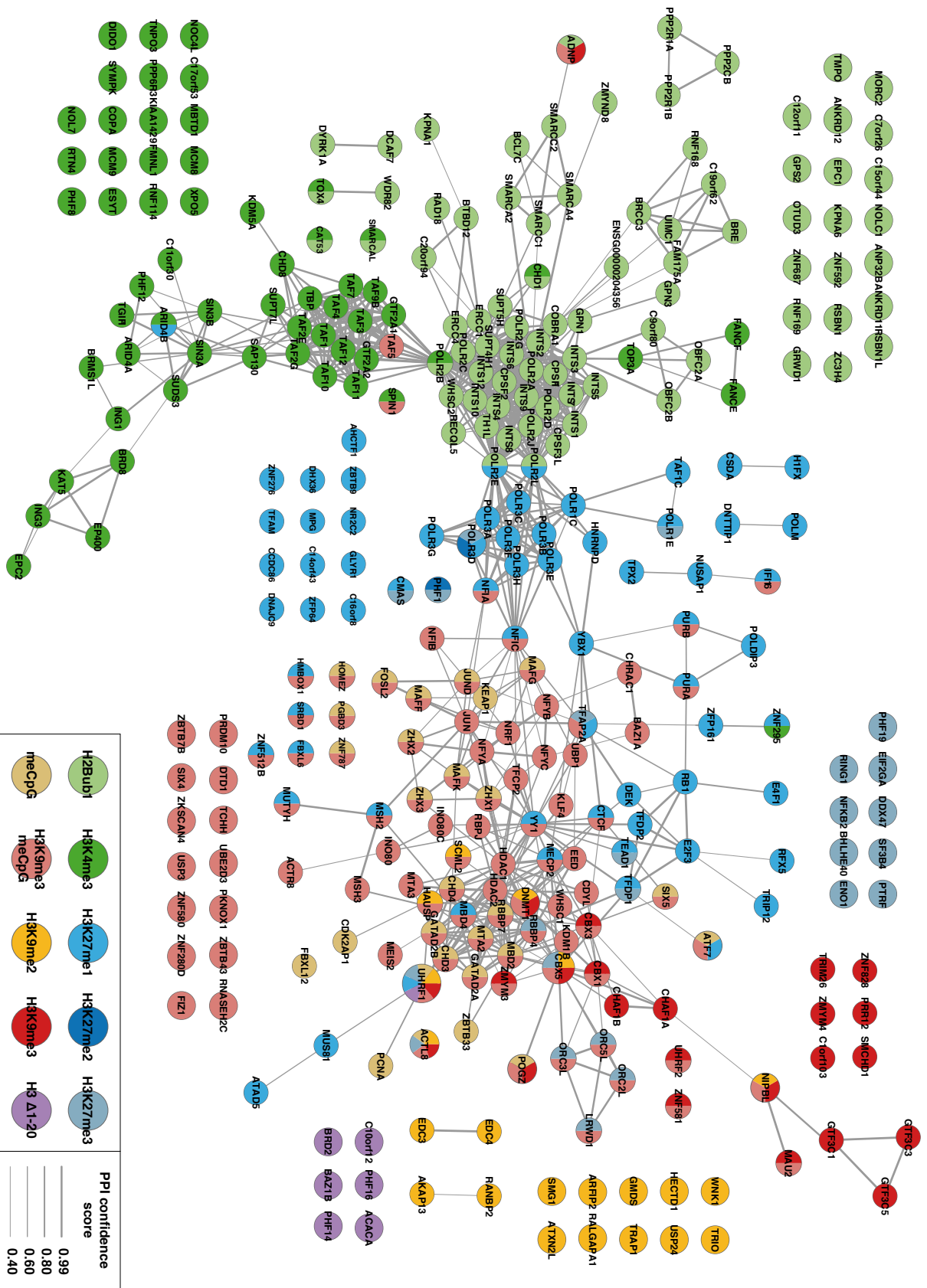
RNA polymerase III (H3K27me1) and one or more subunits common for different datasets connecting them. The heterochromaic interactomes revealed less prominent clustering, with the exception of the highly interconnected nodes around subunits of CHAF1, GATAD2 and several meCpG-binding proteins, part of the meCpG/H3K9me3 related interactomes. However, the majority of those interactions represent individual complexes and not known large macromolecular assemblies. Notably, some of the proteins that were enriched with multiple modifications did not appear to be central members of large multiprotein complexes (e.g. UHRF1, ACTL8, POGZ with 6, 0 and 1 neighbours, respectively), while others showed enhanced connectivity (e.g. DNMT1 and CBX5 with 17 and 18 neighbours, respectively).

The common heterochromatin interactome

A major aim of this study was establishing a chromatin affinity purification approach for identifying a set of proteins that are associated with and conceivably important for the functional heterochromatin domain. The individual or comparative analyses of the interactomes of different chromatin modifications revealed relatively limited overlap. However, several proteins appeared consistently enriched or repelled throughout most of the datasets. In order to compare the enrichment of those prominent interactors, the proteins enriched or excluded in multiple experiments were identified. For this analysis, the enrichment ratio threshold was lowered to +/-0.5 in each individual forward or reverse experiment and +/-1 on average (log2 scale). As this threshold is less stringent than the one used previously, an additional requirement was that a protein was enriched or excluded in at least four out of eight chromatin modifications (excluded from the analysis here were H3K27me2 as it showed only two enriched proteins, and H3 Δ1-20 as it is not directly linked to eu- or heterochromatin).

Strikingly, this analysis resulted in a list of 21 proteins. The enrichment ratios of these factors in different experiments were plotted as a heatmap with proteins as rows and experiments as columns, and both dimensions were clustered using Euclidean distance and complete linkage methods [263] (Figure 4.22 A). The majority of the proteins showed clear enrichment with some of the *bona fide* heterochromatin marks and exclusion with either the euchromatic H3K4me3 or with the heterochromatic H3K9me2. UHRF1, ACTL8, CBX3 and CBX5 were the most prominent heterochromatin modification binders. This analysis also revealed that H3K27me1 and H3K27me3 are most closely related as seen from the column dendrogram. The identical data was also visualized using line plot (Figure 4.22 B). Additionally, interactome network of the 21 most prominent proteins was constructed. While there are known connections between some of them, they do not recapitulate major recognized macromolecular complexes (Figure 4.22 C).
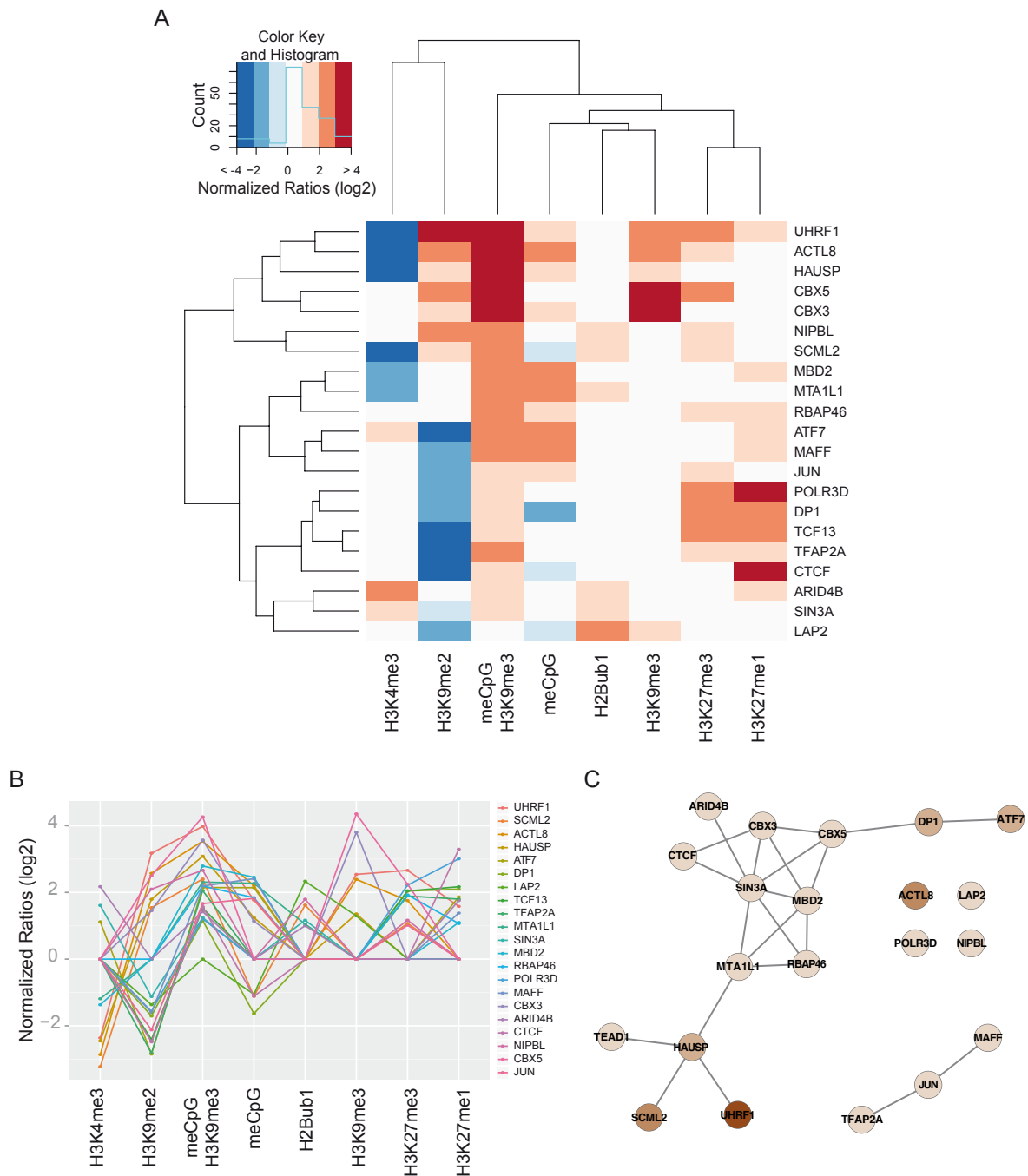
Figure 4.22 The common heterochromatin modification interactome.

A. Heatmap showing the enrichment or exclusion of proteins found in four out of eight of the plotted experiments. Proteins are colour-coded according to their normalized enrichment ratios as indicated in the legend. Columns (modifications) and rows (proteins) were clustered using Euclidean distance and complete linkage methods. **B.** Line plot representing the enrichment ratios of the proteins shown in A. **C.** Interactome network of the proteins shown in A. Darker nodes represent enrichment in more experiments.

In order to further globally compare the different datasets using the approach described in the previous paragraph, the enrichment ratios of all proteins above threshold levels from the ten experiments were plotted as a heatmap. Both the lower threshold used in Figure 4.22 A and the standard for this study threshold (Figure 4.23 A and B, respectively) were applied. Notably, on this global level there wasn't any identifiable close relationship between the interactomes of the different modifications. The column dendrograms did not reveal close distance between the datasets, apart from the few examples with very limited number of data points (e.g. H3K27me2 and H3 Δ1-20 in both heatmaps). Thus on a global scale, the tested chromatin modifications appear to recruit distinct sets of proteins, however, upon closer examination of the datasets, the similar behavior of a small set of factors identified with multiple modifications is apparent (Figure 4.22 A).
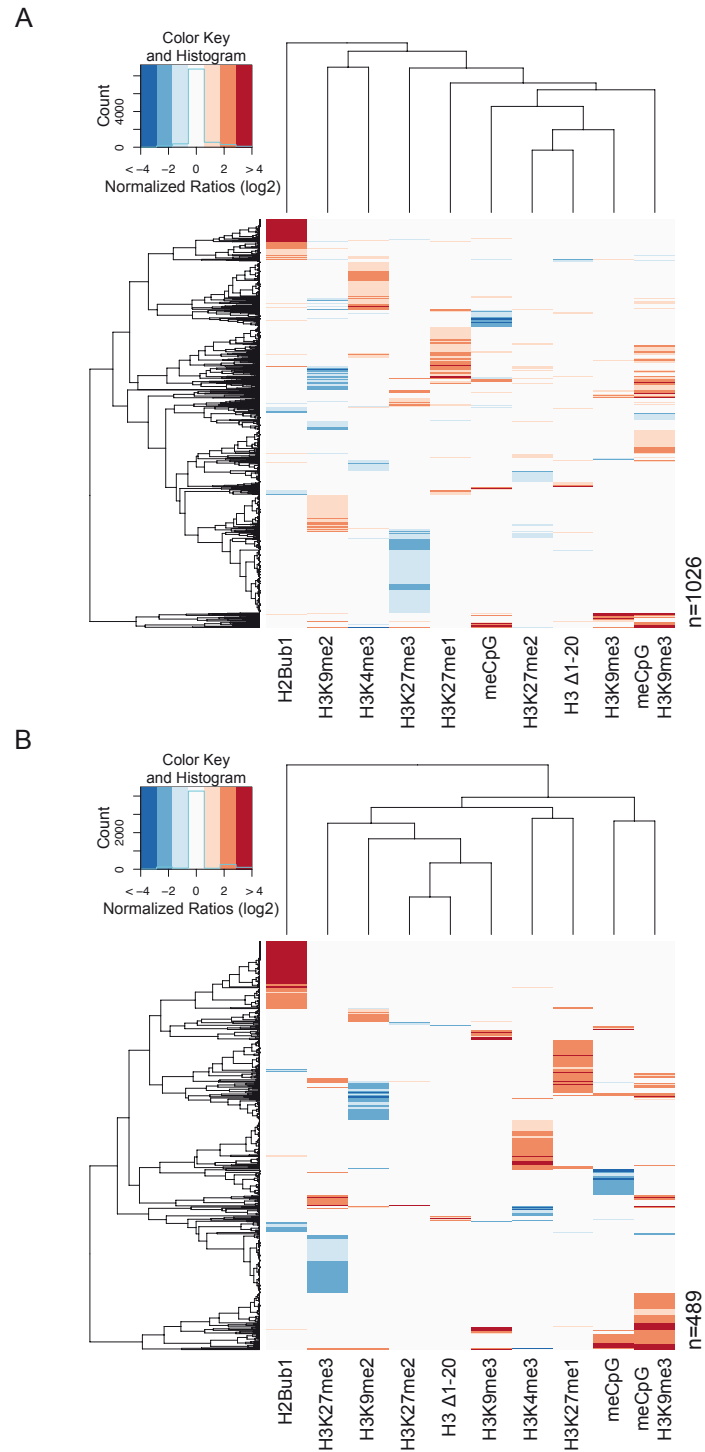
Figure 4.23 Global comparison of the quantified proteins between different chromatin affinity purification experiments.

Heatmap containing all protein identified above threshold levels in at least one experiment. Lower threshold (**A**, as in Figure 4.22) or standard threshold (**B**) was used. The number of clustered proteins is indicated in each case. Proteins are colour-coded according to their normalized enrichment ratios as indicated in the legend.

4.2.3 Recruitment of SWI/SNF to H2Bub1 - investigating the interactome of a single modification and the specific function of a recruited complex

The results presented in Chapter 4.2.3 were obtained in close collaboration with Efrat Shema and Moshe Oren (Weizmann Institute of Sciences, Rehovot, Israel). The experiments involving ChIP, qRT-PCR, Co-IP and transfection were performed by Efrat Shema. The purified SWI/SNF complex was a kind gift from Eric Allemand (Institut Pasteur, Paris, France).

Validation of H2Bub1-dependent interactions

Apart from the global analysis of the H2Bub1 interactome presented in the previous chapter, a deeper insight into the molecular mechanisms that are mediating the cellular functions of this modification was sought out. Multiple subunits of several macromolecular complexes were represented in the list of enriched proteins (colour-coded in Figure 4.24 A). To further validate the SILAC MS results, the recruitment of the different classes of the identified interactors was tested. To this aim, the chromatin affinity purification experiments were repeated using Western Blot detection with specific antibodies. The signals for RECQL5, NELFA, SUPT5H, BAF155, DYRK1A, INTS3 and MORC2 revealed preferential binding to the H2Bub1 containing chromatin (Figure 4.24 B, top panel), essentially verifying the MS results. Two of the proteins identified as non-specific binders, HDAC1 and β-tubulin, showed no preferential binding to either chromatin templates. Furthermore, the Western Blot analysis corroborated the MS results for RCC1 - a member of the small group of proteins showing enrichment for unmodified chromatin (Figure 4.24 B, bottom middle panel). Thus, the SILAC MS enrichment results for H2Bub1 were validated for all 10 proteins tested. Additionally, in order to exclude from further analysis any proteins that could have been recruited by ubiquitin alone, a SILAC pull-down experiment using biotinylated ubiquitin was performed. Importantly, the results showed no overlap of the limited number of identified interaction partners with those recovered when using chromatin reconstituted with H2Bub1 (Figure 4.24 C).

A

B

C
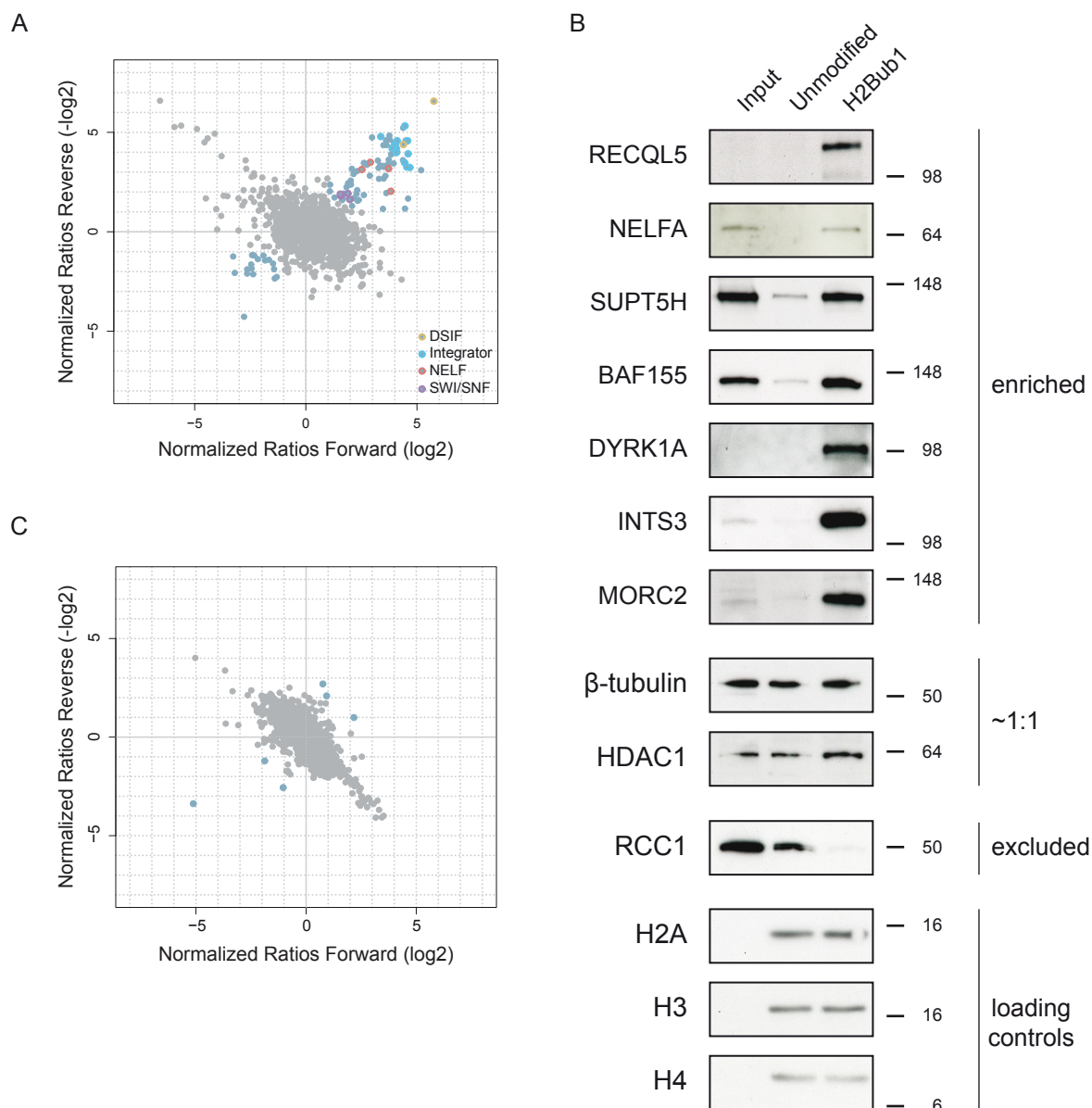
Figure 4.24 Proteins recruited by H2Bub1.

A. Scatterplot representing the ratios from forward and reverse pull-down experiment using H2Bub1 chromatin. Several prominent protein complexes are colour-coded accordingly. **B.** Validation of the recruitment of selected proteins from the H2Bub1 interactome. Affinity purification as in A. was repeated, the bound proteins were separated by SDS-PAGE, transferred on nitrocellulose membrane and immunoblotted using specific antibodies. H2A, H3 and H4 were used as loading controls. **C.** Proteins recruited by ubiquitin alone as revealed by Ub-biotin SILAC pull-down experiment.

The SWI/SNF complex interacts with H2Bub1 in vitro and in vivo

One of the complexes identified as preferential H2Bub1-binder was the chromatin remodeler SWI/SNF. This molecular machine performs ATP-dependent nucleosome remodeling and can either enhance or suppress transcription [14, 235]. Mammalian

SWI/SNF complexes consist of a variety of proteins, including the core subunits BRG1 or BRM, INI1, BAF155 and BAF170. The enrichment of one of the core subunits, BAF155, was already reproduced using Western Blot detection (Figure 4.24 B). In order to further validate the recruitment of the complex, purified mammalian SWI/SNF complex [264] was incubated with unmodified and H2Bub1 chromatin. This revealed moderate, but nevertheless stronger association of four of the core subunits with the modified template (Figure 4.25 A).

Next, this interaction was tested *in vivo*. For this purpose FLAG-tagged WT H2B or H2B resistant to ubiquitylation (mutated at lysines 120 and 125, H2B2KR) [265] were transfected in mammalian cells followed by co-immunoprecipitation (CoIP). Indeed, WT H2B
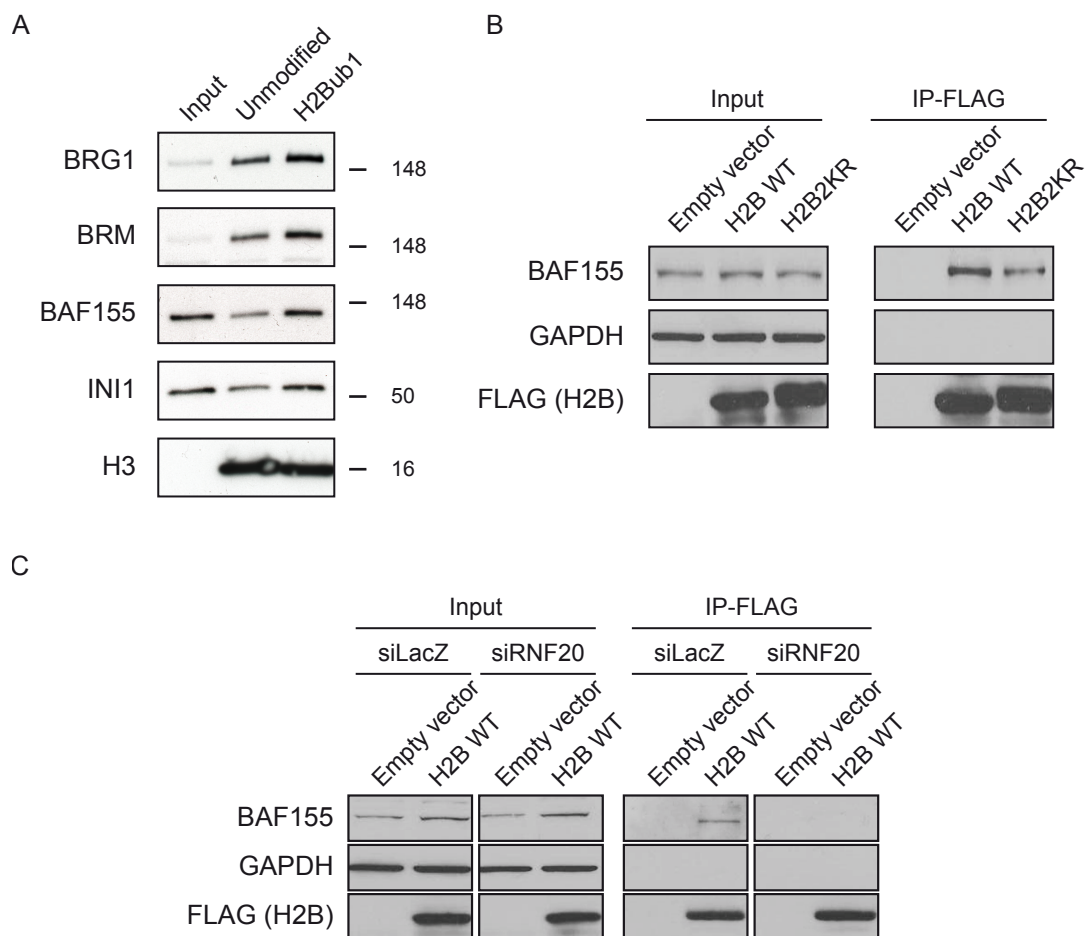
Figure 4.25 H2Bub1-dependent recruitment of SWI/SNF.

A. Unmodified and H2Bub1 chromatin was incubated with purified SWI/SNF complex and the bound proteins were analyzed by SDS-PAGE and immunodetected using specific antibodies. **B.** HeLa cells were transiently transfected with control (empty) vector, or plasmids encoding FLAG-tagged WT H3 or H2B2KR. 24h post-transfection, nuclear extracts were isolated and subjected to immunoprecipitation with αFLAG antibodies, followed by immunodetection by Western Blot using αBAF155 antibodies. **C.** HeLa cells were transfected with control siRNA oligos (siLacZ) or oligos directed against RNF20 (siRNF20). 24h later, the cells were transfected with plasmid carrying FLAG-tagged WT H3 and subjected to co-immunoprecipitation as in B.

precipitated more of the BAF155 subunit, compared to H2B2KR (Figure 4.25 B, right panel). Additionally, WT H2B expressed in cells depleted of RNAF20, the major ubiquitin ligase for H2B [256], resulted in reduced association of BAF155 (Figure 4.25 C, right panel). Taken together, these experiments demonstrated the specific interaction of the SWI/SNF complex both *in vitro* and *in vivo*.

RNF20-dependent genes require SWI/SNF for optimal transcription

It was previously shown that RNF20 and H2B ubiquitylation positively or negatively regulate the transcription of two distinct gene sets, while having no influence on the majority of expressed genes [266]. Therefore, the influence of the recruitment of the SWI/SNF complex to H2Bub1 on the transcription of these gene sets was tested next. Several SWI/SNF subunits were knocked down by siRNA (Figure 4.26 A and B) and the expression of selected genes representative of each class was analyzed. As expected, knock down of RNF20 resulted in significant decrease in expression of genes previously found to be
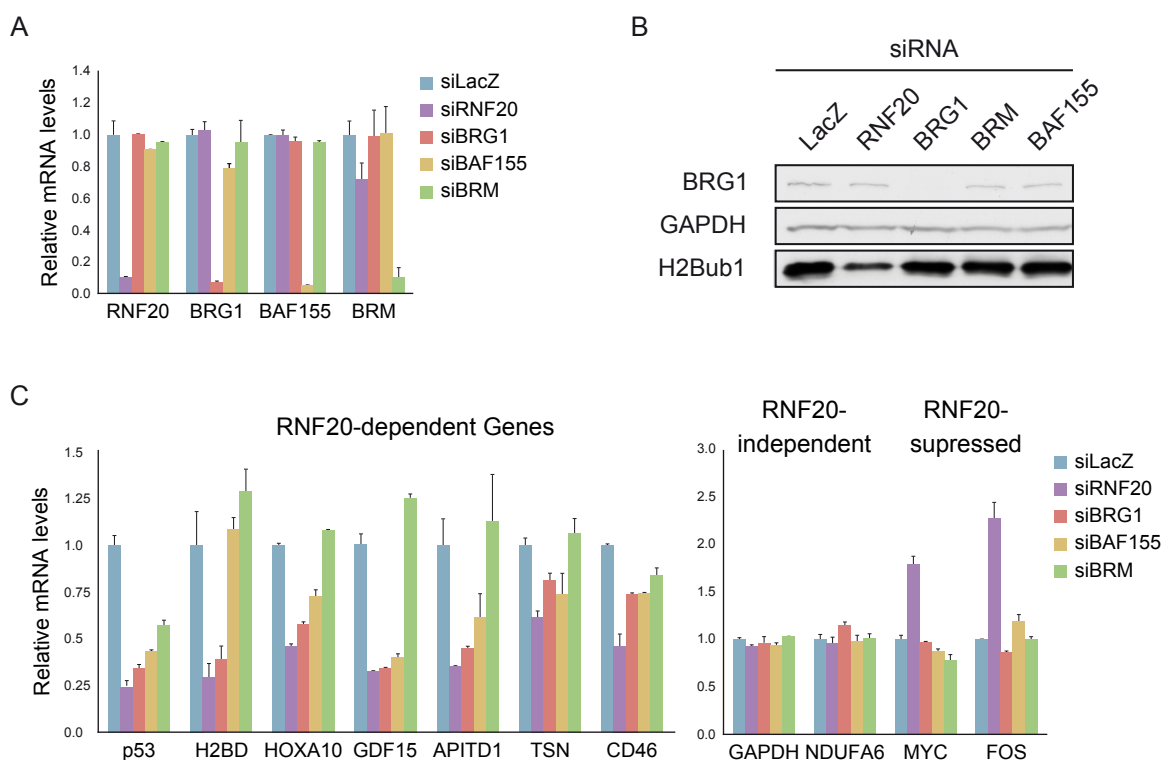
Figure 4.26 RNF20-dependent genes require SWI/SNF for optimal transcription.

A. HeLa cells were transiently transfected with control (LacZ), BRG1, BAF155 and BRM siRNA oligos. 48h later, total RNA was extracted and analyzed by qRT-PCR using primers specific for the targeted genes. mRNA levels were determined after normalization to GAPDH levels within the same sample and as the ratio of the signal from the same transcript in the siLacZ control sample, set as 1. **B.** The experiment presented in A was repeated using Western Blot immunodetection with specific antibodies as indicated. **C.** HeLa cells were transfected and analyzed as in A using primers specific for the genes indicated on the x-axis.

transcriptionally dependent on this ubiquitin transferase (Figure 4.26 C, left panel, compare control blue with violet bars). Interestingly, knock down of BRG1 and to a lesser extent BAF150, resulted in similar decrease in expression of this subset of genes. However, knock down of BRM had little effect in this regard. On the other hand, depletion of the tested SWI/SNF subunits did not affect the expression of RNF20-independent or suppressed genes (Figure 4.26 C, left panel). Thus, the SWI/SNF complex is involved in the positive regulation of genes whose transcription is dependent on RNF20.

SWI/SNF is recruited to RNF-dependent genes

      In order to further validate the recruitment of SWI/SNF by H2Bub1, the association of the complex with the TSS or 5' regions of representative genes was tested using ChIP with antibodies against BRG1. To this aim, the signals from control and RNF20 knock down cells were compared. Depletion of RNF20 resulted in significant decrease in the association of BRG1 with RNF20-dependent genes (Figure 4.27 A, left). In contrast, significant changes in the signal from RNF20-independent or suppressed genes were not observed (panel A, right). As expected, the knock down of RNF20 reduced the H2Bub1 levels at all tested genes (Figure 4.27 B). Thus, H2Bub1 recruits BRG1, and supposedly the SWI/SNF complex, to a subset of genes whose transcription it selectively regulates.

      Next, the rescue of the transcriptional levels of RNF20-dependent genes in cells depleted from RFN20 by BRG1 overexpression was tested. Again, RNF20 knock down led to reduced expression of the subset of RNF20-dependent genes (Figure 4.27 C, compare blue and red bars). Strikingly, BRG1 overexpression partially rescued the transcription of those genes in the RNF20-depleted cells (compare red and yellow bars), while having no significant effect in control cells (compare blue and violet bars).
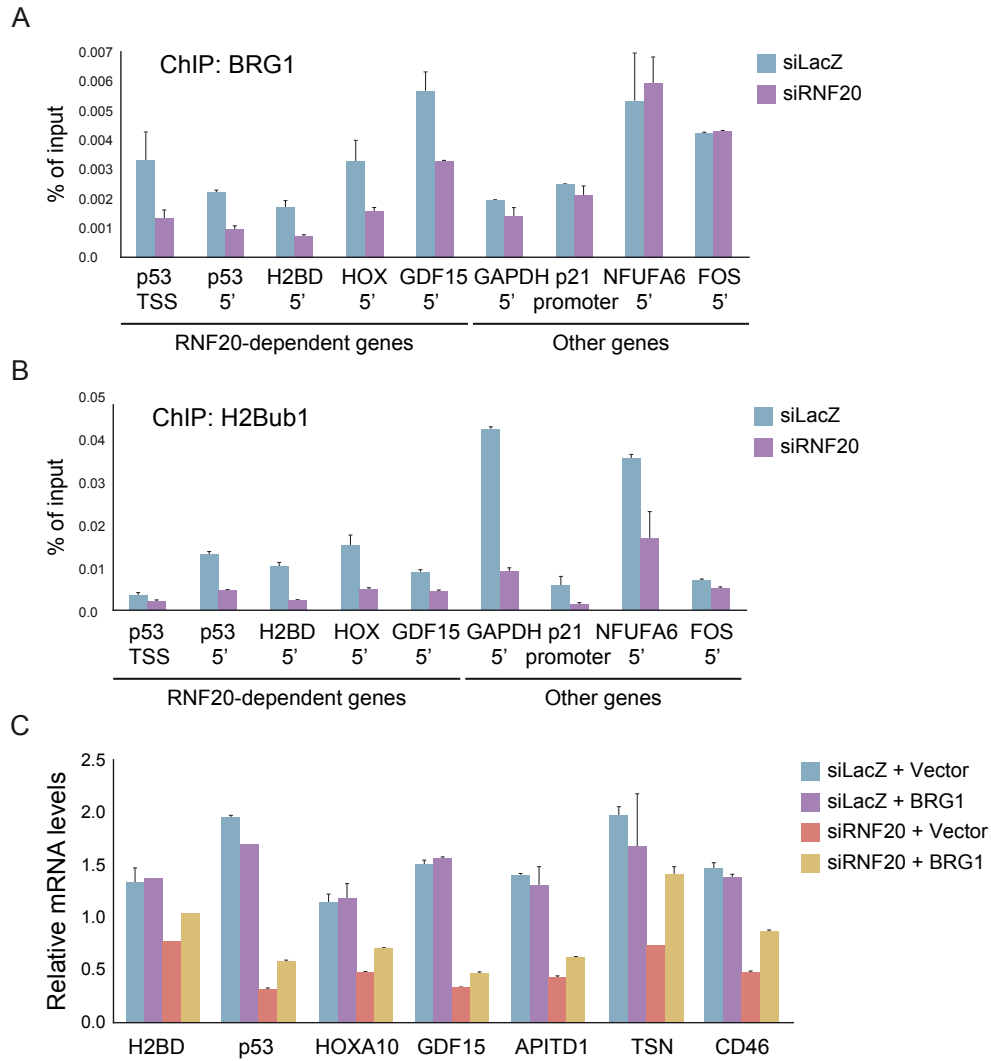
Figure 4.27 H2Bub1-dependent recruitment of SWI/SNF to chromatin *in vivo.*

A-B. Chromatin immunoprecipitation of BRG1 (A) and H2Bub1 (B) from HeLa cells transiently transfected with RNF20 siRNA (siRNF20) or control LacZ siRNA (siLacZ) oligos. Immunoprecipitated and input DNA were quantified by qPCR using primers specific for the indicated gene regions (x-axis). DNA levels are presented as percentage of input. **C.** HeLa cells were transfected with RNF20 siRNA (siRNF20) or LacZ siRNA (siLacZ) as in Figure 4.26. 24h later the cells were transfected with empty vector or plasmid carrying BRG1. After 24h the RNA was extracted and analyzed as in Figure 4.26.

In summary, the SILAC chromatin affinity purification revealed that members of the SWI/SNF complex are associated with H2Bub1-rich chromatin. This interaction was further validated both *in vitro* and *in vivo*. Furthermore, the results demonstrated that it is required for effective transcription of a set of genes whose expression is dependent on RNF20.

4.3   SILAC-based internal standard strategy for relative quantification

Relative quantification by SILAC, or any other label-based or label-free approach, is extremely useful when comparing different samples in a relative manner [124]. However, most approaches are not applicable for comparing relative enrichment between different experiments, especially for peptides and proteins with very high enrichment ratios. The dynamic range limitation in current mass spectrometers restricts accurate enrichment ratio estimation for peptides with high abundant peptides in one of the SILAC pairs. If the intensity ratio of a very high abundant light peak and a very low abundant heavy peak is greater than the measurable dynamic range, the enrichment ratio is calculated based on the noise in the spectrum instead of the heavy peak [119]. Thus, in the example given in Figure 4.28 A, the ratio between the light (blue) peak and the heavy (red) peak in the spectrum from experiment 1 or experiment 2 will result in a high arbitrary L/H ratio but both values would not be comparable. In the specific case of very high enrichment of a protein with two histone modifications and a combination of them, a synergistic effect cannot be quantified by comparing the three individual enrichments ratios. However, if the signals from the three different samples are related to a common standard peptide signal, a comparison of binding strength is possible. Such an internal standard should have identical physicochemical properties as the analytes under investigation. Synthetic heavy-labeled peptides have proven to be a useful standard, for example when determining the subunit stoichiometry of protein complexes [122]. This approach provides very accurate results but is time consuming and costly when applied in large-scale experiments. Therefore, a novel affordable method for estimative enrichment ratio comparison using the SILAC strategy was developed. An internal standard was derived by performing two identical series of experiments in parallel, one using light and the other heavy extract, and then mixing the eluates from one of the series and distributing them equally among the other series (Figure 4.28 B). Thus for signals of the same peptide from different experiments in the light series, there would be equal amount of heavy peptide from the internal standard (Figure 4.28 C), and the L/H ratios should be comparable.
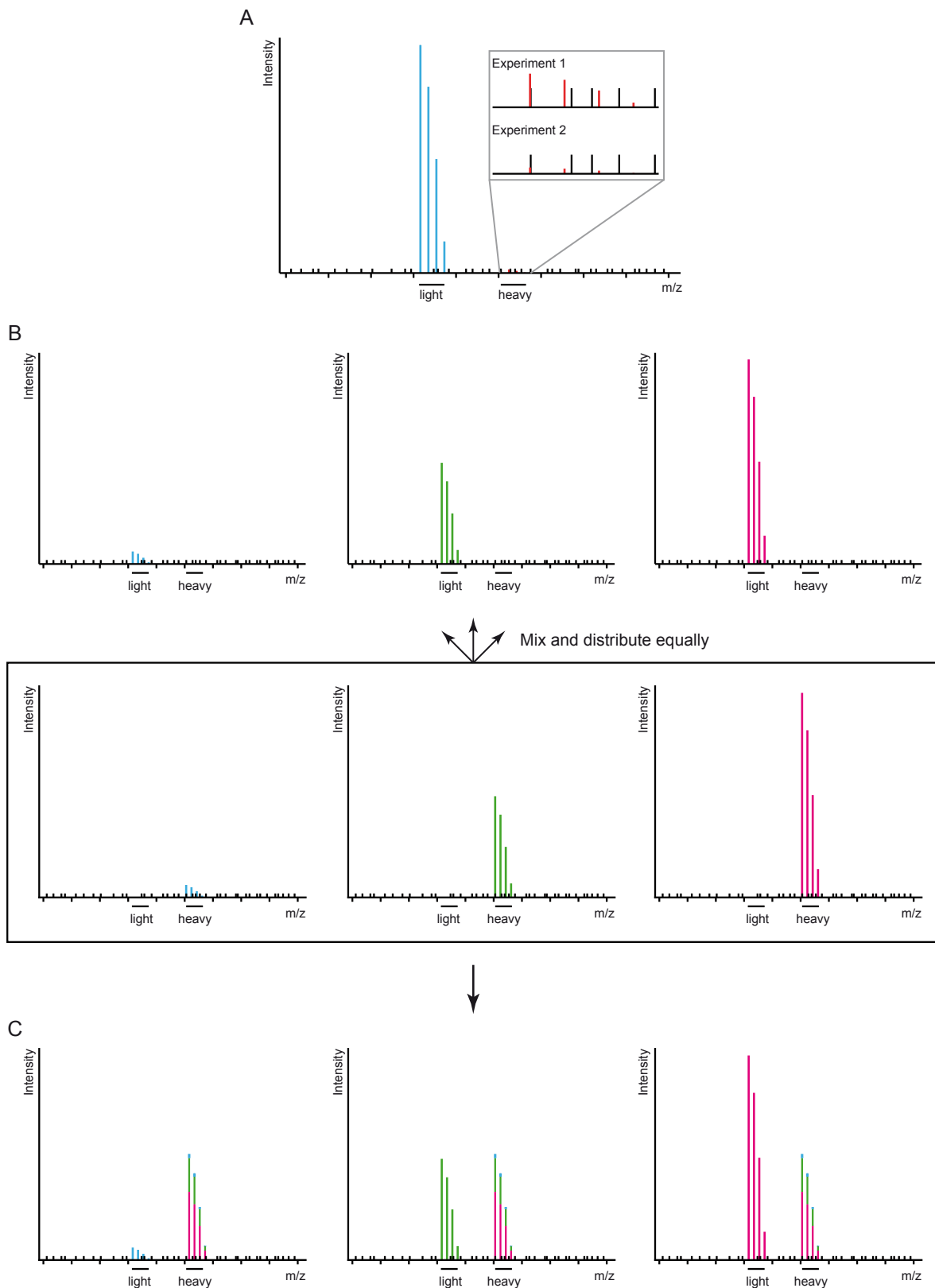
Figure 4.28 Strategy for SILAC quantification using internal standard.

A. A SILAC spectrum of a peptide highly abundant in the light form and near noise levels in the heavy form. **B.** Strategy for mixing and distributing of the internal standard. **C.** A theoretical spectrum showing the light signals from different experiments and the heavy internal standard with equal intensity.

4.3.1 Quantification of proteins binding to meCpG, H3K9me3 and a combination thereof

In order to test the feasibility of the approach for comparing the enrichment to different chromatin modifications using internal standards, an experiment using unmodified, H3K9me3, meCpG and meCpG.H3K9me3 chromatin was set up. Pull-down with each of the chromatin species was performed using light and heavy SILAC NE. The eluates from the light samples were combined and distributed quantitatively between the heavy samples (Figure 4.29). This was termed forward experiment. Additionally, a parallel reverse experiment was performed where the heavy samples were mixed and distributed between the light ones.
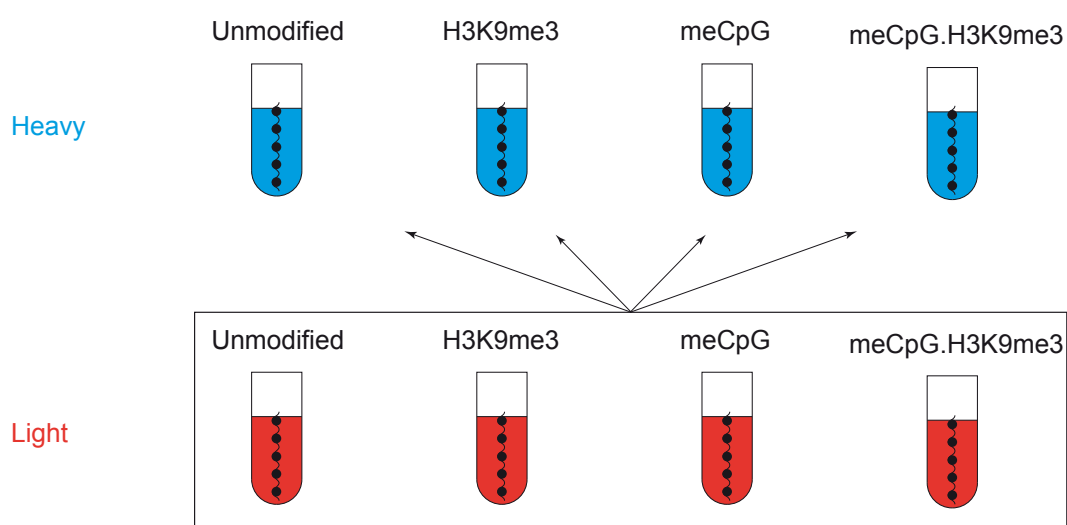
Figure 4.29 Mixing scheme of forward internal standard experiment.

Four chromatin species were used for pull-down from light and heavy nuclear extracts. The eluates from the light sample were mixed quantitatively and distributed equally between the heavy samples. In a reverse experiment (not shown) the heavy samples were mixed and distributed between the light ones.

The enrichment ratios of several prominent meCpG (Figure 4.30 A) and H3K9me3 (Figure 4.30 B) binding proteins identified in previous pull-downs (see section 4.2) were plotted as line plots and their values were compared between the individual experiments. As the distribution of SILAC ratios in this type of experiment with internal standard is different from the distribution in a typical SILAC experiment, the non-normalized ratios [119] were also plotted. Surprisingly, the results from the forward and reverse experiments differed significantly. The ratios in the reverse experiment were on average several fold higher than in the forward, and while the behavior of most proteins appeared similar, UHRF1 and ACTL8 were enriched with much higher ratios in the reverse experiment. Additionally, the normalization changed the profile of some of the proteins mildly (for example ADNP

appeared slightly more enriched in the meCpG.H3K9me3 reverse experiment, compared to H3K9me3, while this was not visible using the non-normalized ratios, panel B).
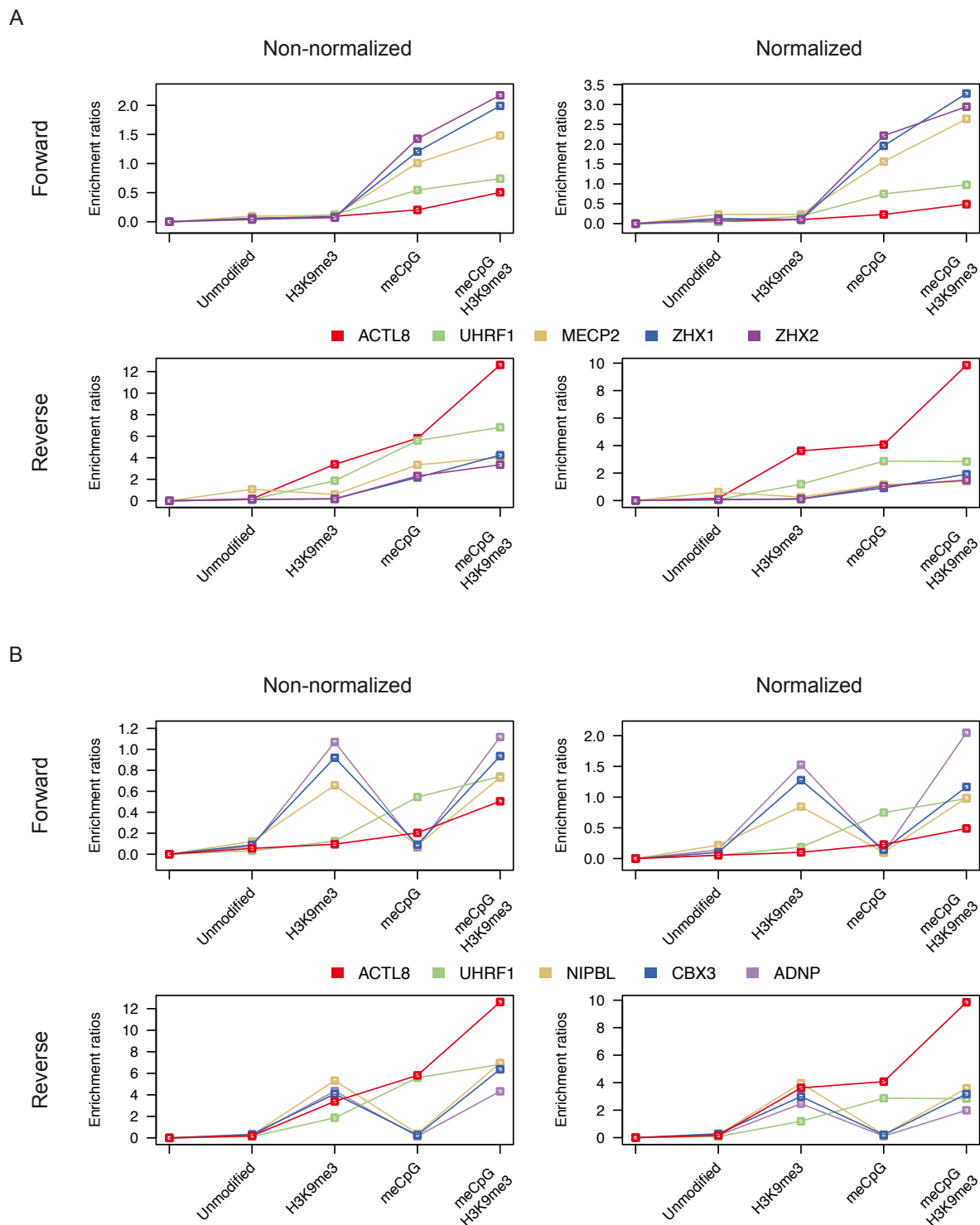
Figure 4.30 Binding profiles of meCpG and H3K9me3-specific interaction partners determined using SILAC internal standard.

The ratio profiles of known meCpG (**A**) or H3K9me3 (**B**) binding proteins were plotted for comparison between unmodified, H3K9me3, meCpG and meCpG.H3K9me3 chromatin. The normalized and non-normalized ratios from forward and reverse experiments are shown. The first data point in each graph is a reference point set to zero.

In order to gain more insight into the behavior of the enrichment ratios in this experiment, the ratio profiles of several ribosomal proteins were plotted (Figure 4.31). Presumably, the ribosomal subunits are not affected by the chromatin modifications and behave as background proteins with constant ratios between the experiments. Interestingly, the ratios, both normalized and non-normalized, showed much higher spread in the forward experiment than in the reverse (note the different y-axis scales). Additionally, the normalization process performed better for the reverse experiment, adjusting the enrichment ratios between experiments closer to equal. Notably, the highest ratio variability for individual proteins in the reverse experiment was less than 0.5 ratio units, well below the effects observed for example for ACTL8 (Figure 4.30). Nevertheless, in order to better describe the system and the observed variability several control experiments were performed.
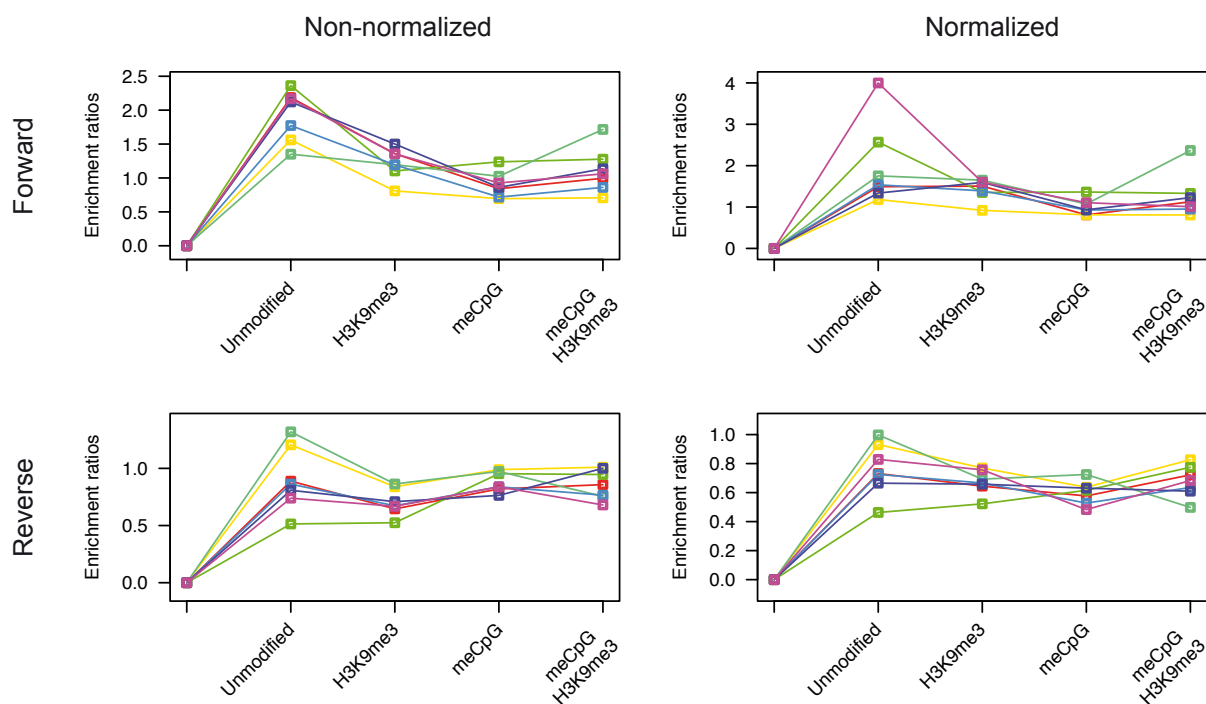
Figure 4.31 Binding profiles ribosomal proteins in internal standard assay using meCpG, H3K9me3 and meCpG.H3K9me3 chromatin.

The normalized and non-normalized ratios of several ribosomal proteins (PRSx and RPLx) quantified with high ratio counts were plotted for the forward end reverse SILAC internal standard experiments. The first data point in each graph is a reference point set to zero.

4.3.2 Control pull-down experiments using SILAC internal standards

In the first control experiment the enrichment of proteins binding to unmodified or different amounts of modified species were compared. Biotinylated N-terminal histone peptides in unmodified or H3K9me3 form were used for simplicity. The pull-downs were performed using unmodified, 50% H3K9me3 and 100% H3K9me3 peptides in forward and reverse (Figure 4.32 A, forward experiment shown). This experiment was done in two replicates: in the first the proteins bound to the peptides were mixed after elution from the magnetic beads ("Experiment 1"), and in the second – during the last washing step of the pull-down ("Experiment 2"). The behavior of several known H3K9me3 binders (Figure 4.32 B) or ribosomal proteins quantified with high ratio counts (Figure 4.32 C) were then compared. Importantly, all but one protein showed higher enrichment with increasing amount of modified peptides (panel B). However, equal and linear behavior of the different proteins in the three samples is not necessary expected as individual factors could have different modes and kinetics of binding. As observed in the experiment using chromatin (Figure 4.30), some proteins showed different enrichment between the forward and reverse experiments. Here again the values of the enrichment ratios were higher for the reverse experiment, however, this effect was minimized after normalization. Furthermore, normalization of the ratios for the ribosomal proteins resulted in near identical behavior between the experiments and ratio variability for individual proteins of less than 0.5 ratio units, again a number smaller than the effect of the recruitment by the modification.
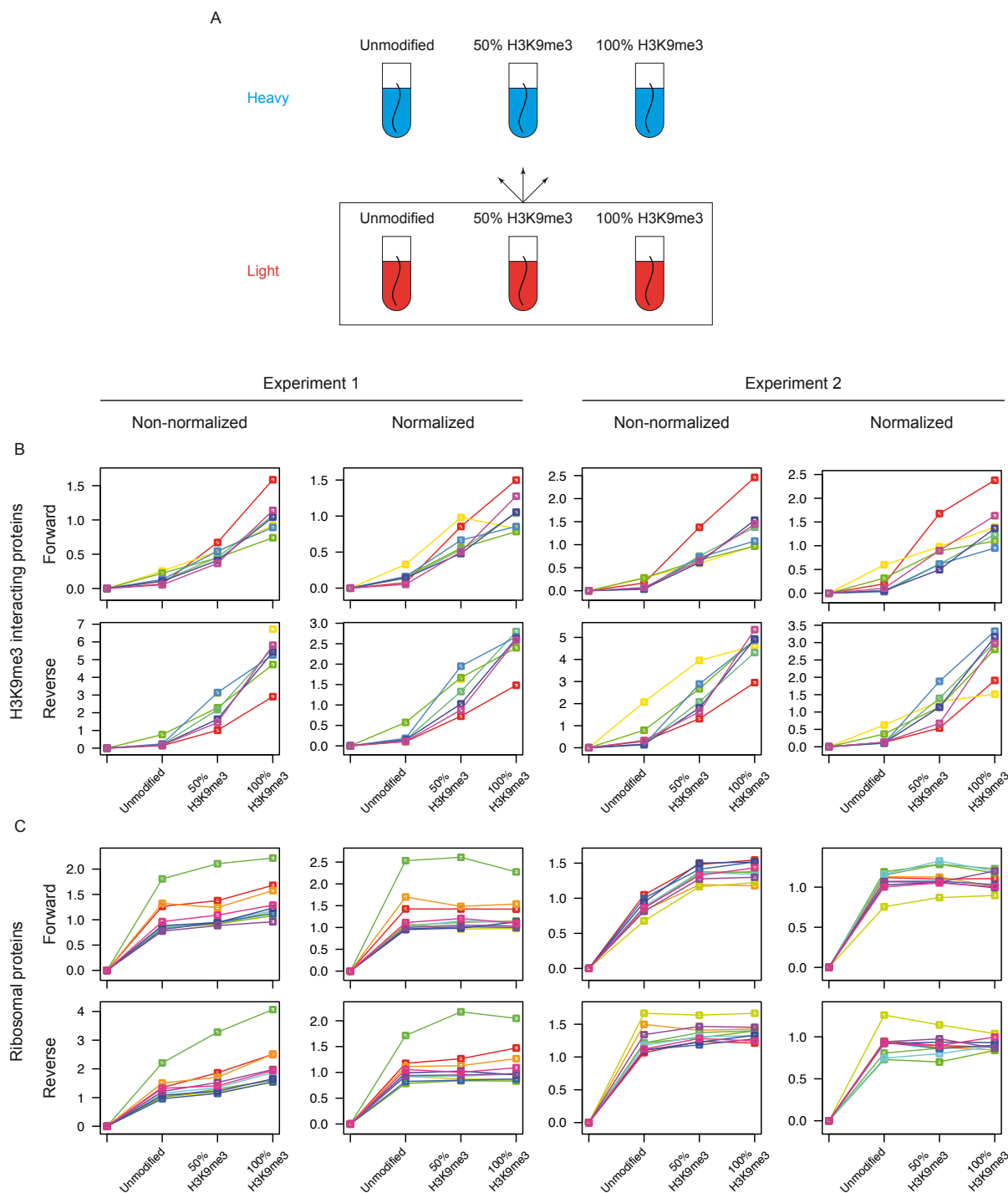
Figure 4.32 Control peptide pull-down using different H3K9me3 modification amount and SILAC internal standard.

A. Unmodified, 50% H3K9me3 and 100% H3K9me3 peptides were used for quantification of known H3K9me3 binders. Shown is the forward experiment where the light samples were mixed and used as internal standard. **B.** Binding profiles of known H3K9me3 interacting partners. In experiment 1 the proteins were mixed after elution, in experiment 2 – at the last washing step before elution. **C.** Binding profiles of ribosomal proteins, as in B. The first data point in each graph is a reference point set to zero.

A second set of control experiments was performed in order to determine the variability of the approach in a more reliable manner. Four parallel pull-downs utilizing only modified H3K9me3 peptides were done. Here, the H3K9me3 binding proteins would be recruited equally in the individual experiments and the variability in their enrichment ratios would represent the variability of the approach. The first three pairs of pull-downs were performed using the internal standard method, the fourth was mixed as a standard SILAC experiment (Figure 4.33 A). The ratios of H3K9me3-specific and ribosomal proteins quantified with high ratio counts (panel B) were plotted as before. The non-normalized ratios of both the H3K9me3 binders and ribosomal proteins showed similar pattern (panel B, left), indicating the usability of ribosomal proteins for internal quality control. The difference in the ratios of experiments "mix 1" was minimized after normalization (panel B, right). After normalization, the variability between different experiments was below 0.3 ratio units. Next, the overall distribution of the normalized protein ratios were compared for experiments "mix 1" and "mix 3", and "mix 2" and "mix 3" (Figure 4.33 C). Interestingly, experiment "mix 1" showed high variability for a set of proteins compared to "mix 3" (left panel), while the ratios of "mix 2" and "mix 3" correlated well (middle panel). Additionally, the normalization factor for mix 1 was higher than for the other experiments (data not shown). A slight bias towards more positive ratios was observed when comparing experiments "mix 2" and "mix 3" (Figure 4.33 C, left panel). In order to exclude the influence of light and heavy extract variability, the ratio distribution of "mix 3" was compared with the ratio distribution of a mixture of equal amounts of light and heavy nuclear extracts. Importantly, there was no correlation between the two distributions. However, in this control experiment, the light samples were mixed and distributed between the heavy ones, therefore the bias towards higher H/L ratios might indicate minute loss of light sample.

In conclusion, these preliminary results demonstrated that a SILAC approach could be used for internal standard-based relative comparison of enrichment between experiments. However, the inherent ratio variability and specific experimental steps (i.e. precise mixing and accurate distribution of the internal standard) require careful optimization and execution of the experiments for optimal performance. Additionally, following the behavior and distribution of the abundant ribosomal proteins is crucial for optimal quality control.
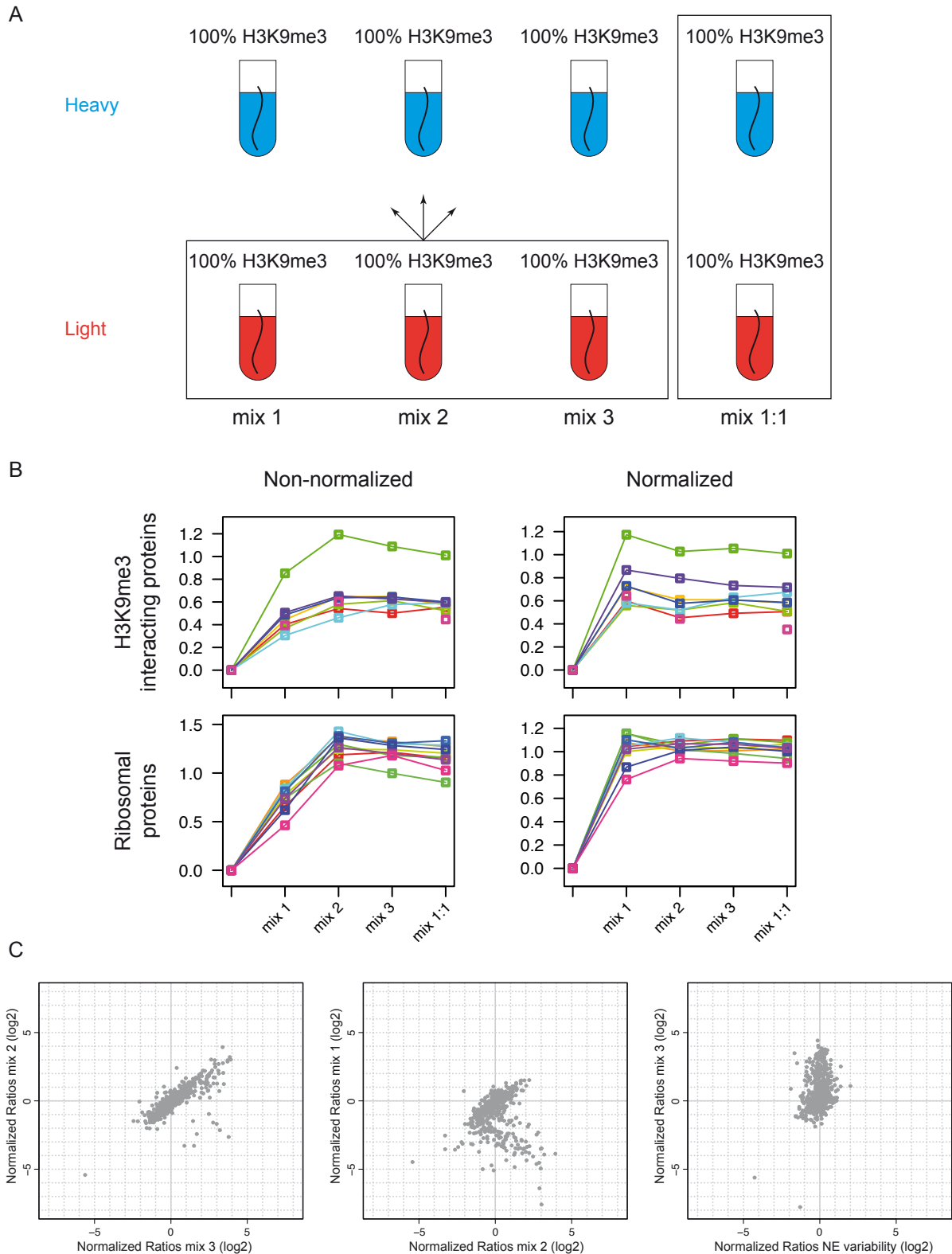
Figure 4.33 Control experiment using H3K9me3 peptides to test variability of the approach.

A. Only modified H3K9me3 peptides were used for pull-down experiments using internal standard (left) or standard SILAC mixing scheme (right). **B.** Binding profiles of known H3K9me3 interacting proteins (top) or ribosomal proteins (bottom). **C.** Differences in enrichment ratio distribution between the separate samples (left and middle) or a 1:1 mixture of light and heavy nuclear extract (right).

5 Discussion

The work presented in this thesis describes the establishment of an *in vitro* affinity purification approach for the analysis of the interactomes of complex chromatin modification patterns based on recombinant and homogeneously modified chromatin template and quantitative mass spectrometry. When using two hallmark histone modifications – H3K4me3 and H3K9me3, the approach performed comparably to a classical peptide-pull down, however each affinity matrix also identified a set of unique binding proteins. The chromatin affinity purification was further applied for the identification of the interactomes of seven other chromatin templates bearing different DNA or histone modifications, as well as a combination of two of them. Analysis of the factors recruited to individual modifications provided new insights into their function. Global analysis revealed moderate overlap between the different datasets. Nevertheless, it identified a set of factors that likely play an important role at the heterochromatin domain. Investigation of the molecular readout of H2Bub1 revealed a novel role of the SWI/SNF remodeling complex for transcriptional regulation of a specific set of genes. Additionally, initial experiments using a novel SILAC-based internal standard quantification strategy showed promising results for its application in the direct comparison of the recruitment of factors to different chromatin species.

5.1 A novel SILAC-based chromatin affinity purification strategy

Identification of chromatin modification-binding proteins has been a focus of research for more than 10 years and has yielded a large number of interacting partners. In addition to several array-based methods [63, 100, 101], affinity purification from cellular extracts combined with mass spectrometry has proven to be the method of choice for *de novo* identification and characterization of these factors. However, the short peptides used in these assays can be synthesized only in limited lengths and as a result reflect a limited local histone tail surface. Therefore, the use of recombinant chromatin template for identification of chromatin modification-binding proteins was established, evaluated and applied in this study. In order to differentiate between specific and background binders, in each experiment the interactomes of modified and unmodified chromatin were compared quantitatively using SILAC labeled nuclear extracts.

Chromatin quality, label-swap experiments and nuclear extract synchronicity are crucial for optimal assay performance

The chromatin affinity purification strategy was setup on the basis of a procedure for *in vitro* reconstitution of defined chromatin templates using salt dialysis [106, 107, 155]. Different histone octamer saturation levels or incomplete modification of the chromatin templates could lead to erroneous results. Therefore each step of the preparation of the template, including comparison of the modified and unmodified species (Figure 4.2 A) and validation of the chemical modifications (Figure 4.3), was monitored carefully to ensure consistent performance.

Additionally, several critical parameters of the assay were identified. First, a label swap experimental design was found to be crucial for eliminating false positive interactors. In this set up, the specific interaction partners showed correlating enrichment ratios between the two experiments (termed "forward" and "reverse"), while the false positive hits from each experiment were identified with irreproducible, often opposite, enrichment ratios (Figure 4.5). There could be several explanations for this behavior. The majority of the false positive hits appear to be only light-labeled (red circles in upper left quadrant, Figure 4.5 A), presumably part of them being contaminants from the experimenter or incompletely labeled proteins with longer half-lives. A fraction of these false positive hits showed low summed peptide intensities, especially in heavy form (red circles on the positive x-axis, Figure 4.5 B), indicating lower abundance, which could indeed lead to false identification or quantification with lower accuracy [119]. A third likely reason for this behaviour are small differences in protein abundance between the independently prepared light and heavy nuclear extracts. A background protein with higher abundance in one of the extracts would appear as enriched with the sample incubated with this extract. Supporting this notion was the finding that highly asynchronous extracts showed even wider spread of enrichment ratio distribution and a significantly increased number of false positive hits (Figure 4.6 B).

Two additional factors could influence the depth of protein identification coverage in the experiments. First, removal of DNA from the affinity purification reaction was necessary for optimal SDS-PAGE resolution. This prevents smear of high abundant proteins along the gel lane (Figure 4.7). Second, using low chromatin-per-beads ratio resulted in the binding of far higher number of background proteins (Figure 4.8). It is likely that this is due to lower saturation of the affinity beads where unspecific binding can occur at unoccupied surfaces. On the other hand, increasing the amount of immobilized chromatin resulted in significant aggregation of the streptavidin beads, presumably owing to higher order chromatin interactions. Therefore, an intermediate saturation level was used here as a compromise between high background binding and irreproducible sample handling.

Comparison to peptide pull-down assay and validation of novel H3K4me3 and H3K9me3 interacting partners

As a proof of principle, the novel chromatin affinity purification approach was compared to a classical peptide pull-down assay. Both approaches showed high specificity in identifying the H3K4me3 and H3K9me3 binding proteins. First, the majority of the known interaction partners for each modification were recapitulated (see section 4.1.3). Second, there was no overlap between the proteins recruited by the two modifications and therefore no general binding to trimethylated lysine was observed (Figure 4.10). Additionally, the interaction of several of the novel binders was validated using orthogonal approaches (Figure 4.12). Here all but one recapitulated the MS results. ACTL8 showed no or low specific enrichment with H3K9me3. However, this is likely due to transfection and overexpression artifacts as the ectopically expressed protein remained in the cytosol and failed to transfer to the nucleus, as shown by immunofluorescence microscopy (data not shown). Moreover, ACTL8 was reproducibly identified as enriched with several heterochromatin-associated modifications (discussed later). Therefore, despite this negative result, it is conceivable that ACTL8 is a true H3K9me3 associated protein.

Both affinity approaches resulted in a similar number of recruited proteins for the two tested modifications, slightly higher for the peptide-based assay. It could be anticipated that all proteins identified with the peptide approach would be present in the set of enriched factors from the chromatin approach, which contains the identical N-terminal peptide in its structure. The chromatin experiments could in addition identify proteins whose recruitment depends on binding to other histone proteins or DNA. Strikingly, the overlap between the two approaches was limited (Figure 4.11). There are several possible causes for this phenomenon. First, the absolute numbers of the enrichment ratios for the chromatin-based experiments were on average lower than for the peptide assay (slope of the distribution of blue circles in Figure 4.11 A and B). Lowering the enrichment ratio threshold indeed resulted in higher overlap, however, the proteins specific for each approach also increased (Figure 4.11 C). Notably, the increase was more significant for the chromatin-specific proteins. Second, due to steric hindrance from the far larger bait molecule in the chromatin experiment, higher streptavidin saturation on the surface of the beads might be achieved in peptide-based assay, thus providing more binding points per bead. This effect can cause the recruitment of higher protein amounts and subsequent better detection and identification in the peptide-based assay. However it does not explain the difference in the absolute values of the enrichment ratios. Moreover, a large proportion of the proteins unique for each approach were faithfully detected and identified in the other approach. Those were slightly above background level in the chromatin assay and around background level in peptide approach (solid blue circles in Figure 4.11 A and B). This, and the fact that the absolute ratio

values were lower in the chromatin experiments, might be the result of the additional surface of the histone proteins, and especially the DNA, acting not only in attractive mode, but also repulsively. Thus, the recruitment of selected factors, whose interaction *in vitro* would not be hindered in the case of isolated histone tails, would be lower in the chromatin approach. Additionally, the histone N-terminal tails might be less accessible in the chromatin context. At physiological salt concentrations they can interact with the surface of neighbouring nucleosomes mediating higher order chromatin conformations [155] and thus limiting the availability for interaction with additional proteins. While these effects have not yet been investigated widely, there is evidence from the comparison of the interaction of the chromodomain of HP1 with H3K9me3-modified chromatin and peptides that is supporting this scenario [155]. Supporting the selectivity of the chromatin approach are also the findings of a recent publication describing the modification-dependent interaction of BPTF with histones [85]. Importantly, a peptide-based approach revealed promiscuous binding of BPTF to three different acetyl marks, in addition to H3K4me3; however, cooperative stronger interaction with H3K4me3 was observed only for one of the acetyl marks (H4K16ac) in a nucleosome context. Finally, two recent publications reported the interactomes of H3K4me3 and H3K9me3 using N-terminal histone peptide [75] and mononucleosome [77] pull-downs. The differences between these two approaches are even more striking, with the mononucleosome-based assay reporting far less enriched factors.

Despite of the differences in the results from the two affinity purification approaches, they both recruited an overall similar number of factors. As this is the first large scale report comparing directly the two experimental platforms, the mechanistic nature of these disparities remains to be confirmed in the future. Nevertheless, the chromatin-based assay has some notable advantages. First, it represents a more native form of chromatin and offers additional histone and DNA interaction surfaces, whose attractive and repulsive roles are relevant *in vivo*. Second, modifications on DNA or in the core histone regions in the context of chromatin can only be investigated using a nucleosome or chromatin-based assay. Third, the recombinant chromatin allows studying of virtually any combination of chemical modifications achievable *in vitro*, including the conjugation of ubiquitin-like proteins. Fourth, in chromatin array-based format, the assay could capture interactions depending on more than one nucleosome (i.e. inter-nucleosome interactions, for example L3MBTL1 [86]), which could be of high relevance in the context of heterochromatin compaction. Fifth, individual nucleosomes with different modifications could be constructed into a chromatin array for the investigation of specific complex modification patterns.

5.2 Analysis of the interactomes of different modifications reveals new insights into chromatin function

Following establishment and validation of the chromatin affinity purification approach, the interactomes of ten distinct chromatin species were further described. The main focus was set on heterochromatic modifications, including H3K9me2, H3K9me3, H3K27me2, H3K27me3, meCpG and a combination of H3K9me3 and meCpG. The other chromatin marks tested were the euchromatic H3K4me3 and H2Bub1, as well as H3K27me1, which has been associated both with euchromatin and heterochromatin. Lastly, the experiment utilizing H3Δ1-20 revealed several factors requiring the N-terminal H3 tail for interaction to chromatin.

The interactomes of individual chromatin modifications

This study expands greatly the list of proteins associated with modified chromatin. While some studies, carried out in parallel to this work [75, 77], have described interactomes of several modifications (H3K4me3, H3K9me3, H3K27me3, meCpG), others have never been investigated at this scale. Moreover, this is the first report utilizing chromatin arrays for affinity purification experiments.

The euchromatic histone marks recruit large number of proteins and macromolecular complexes

Interestingly, the euchromatic H3K4me3 and H2Bub1, as well as the putatively euchromatic H3K27me1, recruited by far the largest number of interaction partners of all single modifications tested. It is likely that this is not a mere coincidence or a technical problem, but a general feature of the interactomes of euchromatic marks. There are multiple protein complexes residing in active euchromatic regions, required for execution of different steps in the gene expression program. These processes require not only large macromolecular assemblies for RNA transcription (including the three RNA polymerase complexes, Pol I, II and III), but also multiple other components for transcriptional regulation and RNA processing. Indeed, these euchromatic marks appear to recruit multiple subunits of several protein complexes, including PolII, NELF, DSIF, SWI/SNF, INTS (H2Bub1), PolI and PolIII (H3K27me1) and TFIID (H3K4me3) (Figure 4.21).

The euchromatic role of monomethylation on H3K27

The fact that H3K27me1 recruits euchromatin-related factors is, to our knowledge, the first evidence at protein level for the involvement of this modification in active chromatin. What is the exact function of this mark at actively transcribed genes? It could be speculated that the large number of RNA polymerase subunits found here and the fact that the

modification is enriched in the gene bodies of actively transcribed genes and depleted from TSS [38, 43] is related to functions in transcriptional elongation. Similarly, it is not yet clear whether monomethylation of H3K27 is simply an intermediate step for di- and tri-methylation [267], or if it has specific functional consequence in the context of transcriptional silencing. Notably, several proteins from the H3K27me1 interactome are linked to negative transcriptional regulation, including the meCpG-binding protein MECP2, UHRF1, RB1 and ARID4B. However, this does not necessarily implicate a heterochromatin-related signaling function of H3K27me1. Supporting this hypothesis is the fact that several binding partners of the euchromatic H3K4me3- and H2Bub1 modifications have also been related to transcriptional repression. For example binding of ING2 links H3K4me3 to histone deacetylation and gene silencing during the DNA damage response [166], while KDM5A binds H3K4me3 for demethylation [188]. On a global scale, the fact that in the protein-protein interaction network (Figure 4.21) the proteins recruited by H3K27me1 are positioned between the euchromatic and heterochromatic interactomes supports the dual nature of the functionality of this modification.

Proteins recruited by H3K27me2, H3K27me3 and H3K9me2

The tested heterochromatic modifications recruited on average lower number of interacting partners, compared to the euchromatic marks. In line with the discussion about the functional requirements of euchromatin above, it is likely that lower number of proteins is required for maintaining a silenced and non-functional heterochromatic state. However, it is possible that the interactomes of H3K9me2, H3K27me2 and H3K27me3 are not complete. The experiment utilizing H3K27me3 did not recapitulate some of the expected polycomb group proteins. The majority of those were identified but with enrichment ratios lower than the threshold or not quantified, presumably due to lower abundance. The H3K27me2 modification recruited two proteins, of which only one was highly enriched. While it is possible that the modification functions only as an intermediate step before trimethylation and has no other signaling roles, binding of factors related to trimethylation of H3K27 might be expected. Additional replicate experiments will show whether the interactomes of H3K27me2, H3K27me3 and H3K9me2, where the enrichment ratios showed skewed distribution, are broader then reported here.

Chromatin array and mononucleosome-based approaches show limited overlap

A recent publication described the interactomes of several of the modifications tested here using a mononucleosome-based approach [77]. The comparison with the results presented here reveals a very limited overlap. For example some of the discussed above polycomb proteins found to bind H3K27me3 by Bartke *et al.* were not identified by the

chromatin array method here. However, it revealed an additional set of factors recruited by H3K27me3 and not detected in the mononucleosome approach. In general a far larger number of proteins were unique for the chromatin array approach, including some known binding partners (e.g. CBX1 for H3K9me3). Some of these differences can be attributed to the different nature of the affinity template used. A chromatin array might stabilize the interaction of selected factors by inter-nucleosome binding. Nevertheless, at this point it cannot be excluded that the disparities are due to the different experimental procedures applied. Bartke *et al.* used nuclear extract preparations with lower protein concentration and analyzed the samples on an instrument with lower sensitivity and in a single technical replicate. Thus, direct comparison with these results is not objective and would require replicate experiments analyzed under identical conditions.

The combinatorial effect of binding to meCpG and H3K9me3

The analysis of the interactome of chromatin containing H3K9me3 and meCpG revealed the largest number of binding factors of all tested chromatin species. These included the majority of the proteins identified in the pull-downs with each individual modification (Figure 4.18 C and D), three of which (UHRF1, ACTL8 and POGZ) were enriched above threshold levels in all three experiments. Strikingly, more than 60 factors were not part of the interactomes of meCpG and H3K9me3. The experiment utilizing the combination of H3K9me3 and meCpG was performed with different extract preparation from the one used in the individual modification pull-downs. Difference in extract composition could explain the finding that certain proteins were identified only in the later experiment. However, more than 15 of the proteins highly enriched only in the combinatorial experiment were faithfully identified near background levels in the other two experiments (Figure 4.18 E). Thus, the recruitment of these factors depends on both meCpG and H3K9me3. Binding to both modifications simultaneously could result in cooperatively stronger enrichment, either by direct bivalent recruitment or by bridging and stabilizing the interaction of individual meCpG- and H3K9me3-specific factors. The genome distributions of the two modifications correlate with each other at multiple heterochromatic regions (Figure 2.3), therefore combinatorial effect between them might be expected. However, to our knowledge this is the first report of such effect on protein level. These findings could provide important evidence supporting the histone code hypothesis, especially considering the large number of meCpG.H3K9me3 specific factors. Most of these proteins are involved in transcriptional regulation – for example the transcription factors UBP1 and YY1, HDAC1 and 2, the mcCpG-binding proteins MECP2 and MBD4, as well as the chromatin insulator CTCF. Member of this group was also the MBT repeat-containing polycomb protein SCML2. MBT

repeat proteins can induce chromatin compaction, however so far they have been shown to have preference for lower methylation states [86].

The interactomes of individual modifications show limited overlap of the recruited factors and both positive and negative correlation trends

In the context of the chromatin landscape, several modifications show different degrees of overlap and correlation in distribution. These patterns are more complex in actively transcribed euchromatic loci and correlate with the different gene regulatory elements (Figure 2.4). However, it is not yet clear what is their importance and exact functional consequence [268]. Heterochromatic loci, on the other hand, show more diffused modification distribution along large areas of silenced chromatin. Additionally, at specific genomic regions some of these marks show redundancy, reduction in the levels of one can be compensated by increase of the others [268]. Thus, recapitulating the exact recruiting potential and functional consequence of protein binding to individual or combinations of modifications is a crucial part in understanding the mechanisms of chromatin regulation. The analysis presented here revealed a striking degree of overlap and different correlation relationships between some of the individual chromatin marks.

Apart from the case of H4K20me1/2 [76], the recruitment of proteins to sites with different degrees of modification has not yet been investigated in detail using large-scale approaches. The interactomes of H3K9me2 and H3K9me3 showed enrichment of several common factors, including the known H3K9me3 binders CBX5 and UHRF1. However, a large number of H3K9me3-specific proteins were not enriched with H3K9me2. Additionally, H3K9me2 repelled a wide range of proteins that were present at background level with the trimethyl mark. While both methylation states are widely present in large areas of silenced heterochromatin, it is possible that the recruitment of different factors by dimethylation of H3K9 is a result of distinct functional roles at other sites. Supporting this notion is the finding that H3K9me2 is specifically enriched in lamina-associated heterochromatic regions, and that the correlation with late replication timing is much stronger for regions with H3K9me2 compared to areas with H3K9me3 [269]. The proteins enriched by H3K27me1 and H3K27me3 also showed limited overlap. However, owing to the distinct functions of the monomethyl mark at euchromatin, this finding is not unexpected. Nevertheless, some proteins showed high enrichment with both marks, including UHRF1, DP-1. Interestingly, POLR1E and POLR3D, subunits of RNA polymerase I and III, respectively, and the positive transcriptional regulator TEAD1, were also among these proteins. While it is surprising to find these factors associated with the heterochromatic H3K27me3, this finding might be explained with the role of this modification at bivalent chromatin domains [41], or with yet

unknown euchromatin-associated function. Indeed, a recent study reported the identification of H3K27me3 at a small fraction of promoters of actively transcribed genes [253].

Discrete function of individual heterochromatic marks is also evidenced by the comparison of meCpG and H3K9me2. While H3K9me3 and meCpG showed correlating enrichment ratios for several factors, including UHRF1 and ACTL8, the interactomes of H3K9me2 and meCpG in addition revealed a set of factors enriched by the DNA methylation and excluded by the histone methylation (e.g. ATF7, FOSL2 and JUND, Figure 4.19 C). More striking, however, is the linear correlation between the enrichment ratios of H3K27me1 and meCpG.H3K9me3 (Figure 4.20 D). While the calculated correlation coefficient is not very high (R = 0.67, p < 2.2e-16), it is unusual considering the random nature of the binding of the background proteins centered around the origin of the graph. Apart from the large number of factors enriched in both experiments (e.g. UHRF1, MECP2, NFIA, CTCF), significant proportion of the proteins specific for one of the chromatin species appeared to be slightly enriched with the other. Furthermore, the interactomes of H3K4me3 and meCpG.H3K9me3 share two enriched proteins - the transcription initiation factor TFIID subunit TAF2D (TAF5) and SPIN1. Additionally, several other H3K4me3-specific TFIID subunits were found enriched slightly below threshold level with meCpG.H3K9me3 (Figure 4.20). Notably, these were not identified as enriched with the individual pull-downs using meCpG or H3K9me3. The nature of these overlaps is unclear and cannot be easily explained. The combination of meCpG and H3K9me3 is a *bona fide* heterochromatin characteristic, and its role in euchromatin is very unlikely. Nevertheless, these observations could hint, for example, at a yet unknown link in euchromatin-heterochromatin transition. Alternatively, it is possible that the euchromatic factors have specific functional roles in hetreochromatin that are not directly related to gene expression regulation. This is supported by the finding that several subunits of another complex generally associated with active transcription - the PolIII transcriptional factor TFIIIC (GTF3C), were identified as highly enriched with H3K9me3, and just below threshold levels with meCpG.H3K9me3. TFIIIC, together with CTCF, have been found at heterochromatic loci unoccupied by PolIII in human, and have been implicated in higher-order chromatin organization in yeast [270]. Interestingly, CTCF was also identified with both H3K9me3 and meCpG.H3K9me3. However, the protein was depleted from meCpG chromatin, and the enrichment of CTCF and TFIIIC subunits was lower in meCpG.H3K9me3 compared to H3K9me3. Thus, it is appears that regarding those factors, these histone and DNA methylations have opposing recruitment roles.

The heterochromatic modifications recruit a limited set of common factors

Individual binary comparison of the different chromatin species tested revealed limited overlap of common factors and a number of proteins unique for each experiment. Taken together, the interactomes did not show strong correlation of the identified factors globally (Figure 4.23 B), indicating in general distinct functional roles at chromatin. However, several proteins (e.g. UHRF1, ACTL8, CBX5) appeared in the majority of the datasets. Comparing the enrichment ratios of all proteins across the experiments further resulted in a list of 21 factors that were present in four out of eight of the interactomes. UHRF1, ACTL8, HAUSP, SCML2, MBD2 and MTA1L1 were enriched with the majority of the heterochromatic modifications and at the same time repelled by the euchromatic H3K4me3. CBX3, CBX5 and NIPBL were not quantified in the H3K4me3 experiment but were enriched with at least four heterochromatic marks. All of them, except ACTL8, have been implicated with transcriptional repression functions. The fact that they were recruited by several distinct heterochromatic marks suggests that they play a central role in the establishment, maintenance or in general the functions of this chromatin domain.

ACTL8, on the other hand, is a poorly studied protein. The only reported evidence about it at protein level originates from two large-scale studies where the factor was found to be ubiquytilated for degradation [271, 272]. The single annotated structural motif within the protein is actin-like domain. Additionally, the sequence of the protein contains a PxVxL motif, known to be present in factors recruited by dimers of CBX (HP1) proteins [273]. It is possible that ACTL8 interacts here with one of the CBX isoforms, however these proteins were not present in all experiments where ACTL8 was identified. Thus, it is likely that this sequence motif is either not involved in interaction, or participates in secondary interactions within a yet unknown protein complex at heterochromatin. UHRF1 was the only protein enriched in all experiments together with ACTL8, so recruitment by this known H3K9me- and meCpG-binding factor is possible. However, ACTL8 was not identified in the experiments using H3K4me3 and H3Δ1-20, where UHRF1 was significantly depleted or enriched, respectively. Thus, direct recruitment of ACTL8 to chromatin is conceivable. Notably, the preliminary results from a SILAC internal standard approach showed remarkably strong recruitment of ACTL8 to the combination of meCpG and H3K9me3 (Figure 4.30).

The rest of the proteins from this group were enriched mainly with H3K27me1, H3K27me3 and meCpG.H3K9me3 and depleted from H3K9me2. Some exhibited more complex recruitment pattern, e.g. CTCF and DP1 in addition being repelled by meCpG. Notably, the majority of these factors were linked to either positive or negative transcriptional regulation and chromatin-related processes in general. Future work will reveal whether these, or additional factors are also present in the interactomes of other heterochromatin

modifications, for example H2AK119ub1, H3R2me2a and H4K20me3 and combinations thereof.

5.3 The interactome of H2Bub1 and the transcriptional effects of SWI/SNF recruitment

The interactome of H2Bub1 chromatin revealed a large number of interacting partners, expanding significantly the list of known associated proteins. Notably, the recruitment of all proteins tested by parallel experiments with Western Blot detection validated the MS results (Figure 4.24 B) thus confirming the role of this modification as a signaling mark. The mode of recruitment to H2Bub1-rich chromatin could be mediated by several, mutually non-exclusive, interaction modes. First, some proteins may bind to the unique H2B-Ub conjugation branched peptide, similar to described antibodies [231]. Additionally, binding can be the result of recognition by one multi-domain factor, or individual subunits of a protein complex, to ubiquitin and histones and/or DNA. Weak individual interactions, for example characteristic of ubiquitin-binding domains [274], can thus be stabilized by cooperative binding. Alternatively, some complexes may recognize different chromatin conformation states, known to be induced by this modification [30]. Importantly, the approach showed high selectivity in identifying proteins recruited by ubiquitin in the context of chromatin. This is supported by the finding that a pull-down with biotinylated ubiquitin alone and using identical experimental conditions did not recapitulate any of the H2Bub1-specific proteins.

H2Bub1 plays a role at sites of DNA damage and is necessary for double strand break (DSB) repair [275, 276], presumably to locally open chromatin structure and increase accessibility of repair proteins. The interactome of H2Bub1 revealed in addition a potential signaling role of this modification in DNA repair by recruitment of various DSB-related protein factors, including the BRCA1-A complex, RNF168, RECQL5 [277, 278]. Additionally, the enrichment of the nucleotide excision repair related proteins ERCC1 and ERCC4 suggests a previously unknown role for H2Bub1 in this pathway [279].

Insight into the role of H2Bub1 in negative transcriptional regulation is provided by the recruitment of two negative elongation complexes – NELF and DSIF. The expression of RNF20-supressed genes depends on the positive elongation factor TFIIS to overcome PolII stalling [266, 280]. Thus, the inhibitory role of H2Bub1 for these genes can be explained by the fact that NELF inhibits the binding of TFIIS to PolII [281].

The association of the SWI/SNF remodeling complex with H2Bub1, shown here for the first time, provides novel insights both for H2Bub1-dependent positive transcriptional regulation as well as for SWI/SNF mode of recruitment. While the activity of chromatin

remodeling complexes is generally indifferent to the modification status of the template, different histone PTMs have been implicated in recruiting/retaining them locally [282, 283]. The H2Bub1-mediated recruitment of SWI/SNF to the transcribed regions of RNF20-dependent genes was required for their efficient transcription. It is conceivable that chromatin remodeling, resulting in a specific nucleosome pattern, is necessary for this process. Alternatively, SWI/SNF can exert another function at these loci, not directly linked to nucleosome remodeling. It is not yet clear what are the exact molecular details for the mechanism of SWI/SNF function here. Pull-downs using nuclear extract showed higher specificity for the H2Bub1 template (BAF155, Figure 4.24 B) compared to experiments utilizing purified SWI/SNF complex (Figure 4.25A). SWI/SNF has general affinity for chromatin and it is likely that additional subunits, not present in the purified complex, might result in higher selectivity. Furthermore, pull-downs using several individual SWI/SNF subunits didn't show significant enrichment for H2Bub1 (data not shown). It is possible that the specific recruitment is a result of multisubunit multivalent interaction and single proteins do not show high affinity binding. Therefore, pull-down experiments using combinations of SWI/SNF subunits might reveal more in this direction. Additionally, *in vitro* chromatin remodeling assays comparing unmodified and H2Bub1 templates will show whether the ubiquitylation of H2B facilitates SWI/SNF remodeling action.

5.4 Comparative quantification of binding to chromatin modifications using a novel SILAC internal standard strategy

Absolute comparison of enrichment ratios between individual SILAC affinity purification pull-downs can be very beneficial, for example in determining cooperative binding for two chromatin modifications in combination. However, in the case of highly enriched proteins this is not always possible as the enrichment ratios are calculated on the basis of the detector noise in the mass spectrometer [119]. The SILAC spike-in strategy presented here can help alleviate this problem and allow direct enrichment ratio comparison between individual pull-down experiments. When the approach was applied to compare the enrichment to H3K9me3, meCpG and a combination thereof it revealed striking differences between the absolute enrichment ratios of the forward and reverse experiments. While these disparities were reduced after normalization of the enrichment ratios, their values remained on average higher in the reverse experiment (Figure 4.30). A possible explanation is that this is a result of unequal nuclear extract amount and unbalanced mixing of the internal standard samples. Normalization would indeed at least partially reduce these differences. However, the same trend was observed when a control experiment using different amounts of modified

H3K9me3 peptide was performed (Figure 4.32). In this case normalization again reduced the difference significantly. The binding profiles of the highly abundant ribosomal proteins in all experiments, and in addition the H3K9me3-binding proteins in the peptide control experiment (Figure 4.33), demonstrated that normalization of the enrichment ratios results in near-equal values between the individual samples. The variability in each case, except for the forward experiment shown in Figure 4.30 and 4.31, was less than the effect of specific enrichment to the tested modifications. The variability shown in Figures 4.30, 4.31 and 4.33 C is nevertheless significant. It is likely that this is a consequence of differences in the mixing and precise distribution of small volumes of the internal standard samples while diluted in SDS-PAGE sample buffer with high density. Indeed, when the internal standard samples were mixed in larger volumes of PD150 buffer during the last washing step, the variability appeared less prominent (experiment 2 in Figure 4.32 B and C). Importantly, given that the enrichment ratio of individual background proteins is constant between the compared samples, its absolute value is of lower importance (e.g. green line compared to the rest, Experiment 1 in Figure 4.32 C). Furthermore, all proteins with specific binding to the modified species are characterized with very low ratio values in the unmodified samples. These values are the result of the ratio between a very low background binding signal in the unmodified sample and relatively high signal from the internal standard. Thus, specific interactors could be identified based on this feature. Lastly, even with some inherent variability, the approach can be used in a more restricted "targeted" manner – independent pull-down experiments can determine the specific interactomes of individual modifications and the enrichment of only those proteins can be followed in and internal standard SILAC experiment, thus disregarding potential false positive hits.

Following further optimization and estimation of the variability, the approach could be used for comparison of enrichment ratios from different SILAC pull-down experiments. Due to the inherent variability, the mixing and distribution of the internal standard samples is the most critical step of the workflow. Multiple replicates, conceivably in semi-automated fashion using microtiter plates, and monitoring the binding profiles of abundant background proteins are thus important for obtaining reliable results and optimal quality control.

5.5 Conclusions and future perspectives

Given the fundamental role of chromatin modifications in genome regulation and cell signaling, deciphering the interaction and functional protein networks associated with them is of utmost importance. The chromatin affinity approach presented here is promising for elucidation of the interactomes of single and complex patterns of chromatin modifications. On one hand, it provides a straightforward and practical first step in the functional characterization of individual modifications using a native chromatin template. Identification of associated proteins is an essential part of dissecting the signaling roles of modification-specific recruitment. This was demonstrated with the example of the H2Bub1-dependent binding of the SWI/SNF complex for regulation of transcription. These results provided novel molecular insights into the function of both the H2Bub1 modification and the chromatin remodeling complex. Individual modification interactomes can in addition hint at novel general functional features of specific marks. Furthermore, the possibility to test combinations of different modifications in the context of chromatin can reveal important insights into the nature and complexity of the "histone code". Indeed, the combinatorial experiment presented here showed large number of factors whose recruitment conceivably depends on both meCpG and H3K9me3. This result has significant implications for the histone code hypothesis and underlines the importance of testing further combinations of modifications, including additional marks not investigated here. The analysis of multiple interactomes on global level identified novel relationships between individual modifications and indicated distinct roles of functionally related marks. Additionally, taken together, this analysis resulted in a condensed list of proteins with potentially important roles in the function of the heterochromatin domain. In this respect, the detailed functional characterization of some of the most prominent factors, for example ACTL8, is needed. Lastly, the presented SILAC internal standard strategy provides a straightforward way of testing and quantification of combinatorial recruitment by multiple chromatin modifications. Upon optimization, it can be used for large-scale analysis of multiple chromatin species in parallel.

References

1. Luger, K*., et al.*, "Crystal structure of the nucleosome core particle at 2.8A resolution," *Nature,* vol. 389, pp. 251-260, 1997.
2. Davey, C. A*., et al.*, "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution," *J Mol Biol,* vol. 319, pp. 1097-113, Jun 21 2002.
3. Dechassa, M. L. and Luger, K., "Nucleosomes as Control Elements for Accessing the Genome," in *Genome Organization and Function in the Cell Nucleus*, ed: Wiley-VCH Verlag GmbH & Co. KGaA, 2011, pp. 55-87.
4. Luger, K. and Richmond, T. J., "DNA binding within the nucleosome core," *Curr Opin Struct Biol,* vol. 8, pp. 33-40, Feb 1998.
5. Fletcher, T. M. and Hansen, J. C., "Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms," *J Biol Chem,* vol. 270, pp. 25359-62, Oct 27 1995.
6. Yun, M*., et al.*, "Readers of histone modifications," *Cell Res,* vol. 21, pp. 564-78, Apr 2011.
7. Holde, K. E. v., *Chromatin*, First ed.: Springer, 1988.
8. Bassett, A*., et al.*, "The folding and unfolding of eukaryotic chromatin," *Curr Opin Genet Dev,* vol. 19, pp. 159-65, Apr 2009.
9. Nishino, Y*., et al.*, "Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure," *EMBO J,* vol. 31, pp. 1644-53, Apr 4 2012.
10. Flemming, W., *Zell Substanz, Kern und Zellheilung*. Leipzig: Vogel, 1882.
11. Heitz, E., " Das Heterochromatin der Moose," *Jahrb Wiss Botanik,* vol. 69, pp. 762–818, 1928.
12. Filion, G. J*., et al.*, "Systematic protein location mapping reveals five principal chromatin types in Drosophila cells," *Cell,* vol. 143, pp. 212-24, Oct 15 2010.
13. Felsenfeld, G. and Groudine, M., "Controlling the double helix," *Nature,* vol. 421, pp. 448-53, Jan 23 2003.
14. Wilson, B. G. and Roberts, C. W., "SWI/SNF nucleosome remodellers and cancer," *Nat Rev Cancer,* vol. 11, pp. 481-92, Jul 2011.
15. Nightingale, K., "Histone Modifications and Their Role as Epigenetic Marks," in *Genome Organization and Function in the Cell Nucleus*, ed: Wiley-VCH Verlag GmbH & Co. KGaA, 2011, pp. 89-110.
16. Allfrey, V. G. and Mirsky, A. E., "Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis," *Science,* vol. 144, p. 559, May 1 1964.
17. Sidoli, S*., et al.*, "Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry," *J Proteomics,* vol. 75, pp. 3419-33, Jun 27 2012.
18. Britton, L. M*., et al.*, "Breaking the histone code with quantitative mass spectrometry," *Expert Rev Proteomics,* vol. 8, pp. 631-43, Oct 2011.
19. Villar-Garea, A*., et al.*, "Analysis of histone modifications by mass spectrometry," *Curr Protoc Protein Sci,* vol. Chapter 14, p. Unit 14 10, Feb 2008.
20. Tan, M*., et al.*, "Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification," *Cell,* vol. 146, pp. 1016-28, Sep 16 2011.
21. Peterson, C. L. and Laniel, M. A., "Histones and histone modifications," *Curr Biol,* vol. 14, pp. R546-51, Jul 27 2004.
22. Kurdistani, S. K. and Grunstein, M., "Histone acetylation and deacetylation in yeast," *Nat Rev Mol Cell Biol,* vol. 4, pp. 276-84, Apr 2003.
23. Struhl, K., "Histone acetylation and transcriptional regulatory mechanisms," *Genes Dev,* vol. 12, pp. 599-606, Mar 1 1998.

24.   Grunstein, M., "Histone acetylation in chromatin structure and transcription," *Nature,* vol. 389, pp. 349-52, Sep 25 1997.

25.   Kouzarides, T., "Chromatin modifications and their function," *Cell,* vol. 128, pp. 693-705, Feb 23 2007.

26.   Izzo, A. and Schneider, R., "Chatting histone modifications in mammals," *Brief Funct Genomics,* vol. 9, pp. 429-43, Dec 2010.

27.   Jenuwein, T. and Allis, C. D., "Translating the histone code," *Science,* vol. 293, pp. 1074-80, Aug 10 2001.

28.   Shogren-Knaak, M*., et al.*, "Histone H4-K16 acetylation controls chromatin structure and protein interactions," *Science,* vol. 311, pp. 844-7, Feb 10 2006.

29.   Neumann, H*., et al.*, "A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation," *Mol Cell,* vol. 36, pp. 153-63, Oct 9 2009.

30.   Fierz, B*., et al.*, "Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction," *Nat Chem Biol,* vol. 7, pp. 113-9, Feb 2011.

31.   Khare, S. P*., et al.*, "HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes," *Nucleic Acids Res,* vol. 40, pp. D337-42, Jan 2012.

32.   Migliori, V*., et al.*, "Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance," *Nat Struct Mol Biol,* vol. 19, pp. 136-44, Feb 2012.

33.   Bannister, A. J. and Kouzarides, T., "Regulation of chromatin by histone modifications," *Cell Res,* vol. 21, pp. 381-95, Mar 2011.

34.   Rosenfeld, J. A*., et al.*, "Determination of enriched histone modifications in non-genic portions of the human genome," *BMC Genomics,* vol. 10, p. 143, 2009.

35.   Black, J. C. and Whetstine, J. R., "Chromatin landscape: methylation beyond transcription," *Epigenetics,* vol. 6, pp. 9-15, Jan 2011.

36.   Trojer, P. and Reinberg, D., "Facultative heterochromatin: is there a distinctive molecular signature?," *Mol Cell,* vol. 28, pp. 1-13, Oct 12 2007.

37.   Ong, C. T. and Corces, V. G., "Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme," *J Biol,* vol. 8, p. 73, 2009.

38.   Barski, A*., et al.*, "High-resolution profiling of histone methylations in the human genome," *Cell,* vol. 129, pp. 823-37, May 18 2007.

39.   Phillips, J. E. and Corces, V. G., "CTCF: master weaver of the genome," *Cell,* vol. 137, pp. 1194-211, Jun 26 2009.

40.   Wang, Z*., et al.*, "Characterization of human epigenomes," *Curr Opin Genet Dev,* vol. 19, pp. 127-34, Apr 2009.

41.   Bernstein, B. E*., et al.*, "A bivalent chromatin structure marks key developmental genes in embryonic stem cells," *Cell,* vol. 125, pp. 315-26, Apr 21 2006.

42.   Pan, G*., et al.*, "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells," *Cell Stem Cell,* vol. 1, pp. 299-312, Sep 13 2007.

43.   Wang, Z*., et al.*, "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nat Genet,* vol. 40, pp. 897-903, Jul 2008.

44.   Rando, O. J. and Chang, H. Y., "Genome-wide views of chromatin structure," *Annu Rev Biochem,* vol. 78, pp. 245-71, 2009.

45.   Cheng, X*., et al.*, "Chapter 2 - Mechanisms of DNA Methylation, Methyl-CpG Recognition, and Demethylation in Mammals," in *Handbook of Epigenetics*, Trygve, T., Ed., ed San Diego: Academic Press, 2011, pp. 9-24.

46.   Frauer, C*., et al.*, "DNA Methylation," in *Genome Organization and Function in the Cell Nucleus*, ed: Wiley-VCH Verlag GmbH & Co. KGaA, 2011, pp. 21-54.

47.   Gruenbaum, Y*., et al.*, "Methylation of CpG sequences in eukaryotic DNA," *FEBS Lett,* vol. 124, pp. 67-71, Feb 9 1981.

48.   Lister, R*., et al.*, "Human DNA methylomes at base resolution show widespread epigenomic differences," *Nature,* vol. 462, pp. 315-22, Nov 19 2009.

49.   Bird, A. P., "Gene number, noise reduction and biological complexity," *Trends Genet,* vol. 11, pp. 94-100, Mar 1995.

50.   Bird, A., "DNA methylation patterns and epigenetic memory," *Genes Dev,* vol. 16, pp. 6-21, Jan 1 2002.

51.   Weber, M*., et al.*, "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome," *Nat Genet,* vol. 39, pp. 457-66, Apr 2007.

52.   Dion, M. F*., et al.*, "Genomic characterization reveals a simple histone H4 acetylation code," *Proc Natl Acad Sci U S A,* vol. 102, pp. 5501-6, Apr 12 2005.

53.   Yap, K. L. and Zhou, M. M., "Keeping it in the family: diverse histone recognition by conserved structural folds," *Crit Rev Biochem Mol Biol,* vol. 45, pp. 488-505, Dec 2010.

54.   Strahl, B. D. and Allis, C. D., "The language of covalent histone modifications," *Nature,* vol. 403, pp. 41-5, Jan 6 2000.

55.   Mujtaba, S*., et al.*, "Structure and acetyl-lysine recognition of the bromodomain," *Oncogene,* vol. 26, pp. 5521-7, Aug 13 2007.

56.   Lange, M*., et al.*, "Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex," *Genes Dev,* vol. 22, pp. 2370-84, Sep 1 2008.

57.   Khorasanizadeh, S., "Recognition of methylated histones: new twists and variations," *Curr Opin Struct Biol,* vol. 21, pp. 744-9, Dec 2011.

58.   Yap, K. L. and Zhou, M. M., "Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription," *Biochemistry,* vol. 50, pp. 1966-80, Mar 29 2011.

59.   Charier, G*., et al.*, "The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding," *Structure,* vol. 12, pp. 1551-62, Sep 2004.

60.   Zhao, Q*., et al.*, "Structure of human spindlin1. Tandem tudor-like domains for cell cycle regulation," *J Biol Chem,* vol. 282, pp. 647-56, Jan 5 2007.

61.   Cote, J. and Richard, S., "Tudor domains bind symmetrical dimethylated arginines," *J Biol Chem,* vol. 280, pp. 28476-83, Aug 5 2005.

62.   Cheng, D*., et al.*, "The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing," *Mol Cell,* vol. 25, pp. 71-83, Jan 12 2007.

63.   Kim, J*., et al.*, "Tudor, MBT and chromo domains gauge the degree of lysine methylation," *EMBO Rep,* vol. 7, pp. 397-403, Apr 2006.

64.   Taverna, S. D*., et al.*, "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers," *Nat Struct Mol Biol,* vol. 14, pp. 1025-40, Nov 2007.

65.   Wang, Y*., et al.*, "Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein," *Mol Cell,* vol. 33, pp. 428-37, Feb 27 2009.

66.   Laue, K*., et al.*, "The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity," *Development,* vol. 135, pp. 1935-46, Jun 2008.

67.   Flanagan, J. F*., et al.*, "Double chromodomains cooperate to recognize the methylated histone H3 tail," *Nature,* vol. 438, pp. 1181-5, Dec 22 2005.

68.   Brasher, S. V*., et al.*, "The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer," *EMBO J,* vol. 19, pp. 1587-97, Apr 3 2000.

69.   Cowieson, N. P*., et al.*, "Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis," *Curr Biol,* vol. 10, pp. 517-25, May 4 2000.

70.   Macdonald, N*., et al.*, "Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3," *Mol Cell,* vol. 20, pp. 199-211, Oct 28 2005.

71.   Stucki, M*., et al.*, "MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks," *Cell,* vol. 123, pp. 1213-26, Dec 29 2005.

72.   Fischle, W*., et al.*, "Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation," *Nature,* vol. 438, pp. 1116-1122, 2005.

73. Unoki, M*., et al.*, "ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain," *Oncogene,* vol. 23, pp. 7601-10, Oct 7 2004.

74. Mori, T*., et al.*, "NIRF/UHRF2 occupies a central position in the cell cycle network and allows coupling with the epigenetic landscape," *FEBS Lett,* vol. 586, pp. 1570-83, Jun 4 2012.

75. Vermeulen, M*., et al.*, "Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers," *Cell,* vol. 142, pp. 967-80, Sep 17 2010.

76. Oda, H*., et al.*, "Regulation of the histone H4 monomethylase PR-Set7 by CRL4(Cdt2)-mediated PCNA-dependent degradation during DNA damage," *Mol Cell,* vol. 40, pp. 364-76, Nov 12 2010.

77. Bartke, T*., et al.*, "Nucleosome-interacting proteins regulated by DNA and histone methylation," *Cell,* vol. 143, pp. 470-84, Oct 29 2010.

78. Nikolov, M*., et al.*, "Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns," *Mol Cell Proteomics,* vol. 10, p. M110 005371, Nov 2011.

79. Shi, X*., et al.*, "Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36," *J Biol Chem,* vol. 282, pp. 2450-5, Jan 26 2007.

80. Jacobson, R. H*., et al.*, "Structure and function of a human TAFII250 double bromodomain module," *Science,* vol. 288, pp. 1422-5, May 26 2000.

81. Vermeulen, M*., et al.*, "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4," *Cell,* vol. 131, pp. 58-69, Oct 5 2007.

82. Li, B*., et al.*, "The role of chromatin during transcription," *Cell,* vol. 128, pp. 707-19, Feb 23 2007.

83. Tsai, W. W*., et al.*, "TRIM24 links a non-canonical histone signature to breast cancer," *Nature,* vol. 468, pp. 927-32, Dec 16 2010.

84. Rajakumara, E*., et al.*, "PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression," *Mol Cell,* vol. 43, pp. 275-84, Jul 22 2011.

85. Ruthenburg, A. J*., et al.*, "Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions," *Cell,* vol. 145, pp. 692-706, May 27 2011.

86. Trojer, P*., et al.*, "L3MBTL1, a histone-methylation-dependent chromatin lock," *Cell,* vol. 129, pp. 915-28, Jun 1 2007.

87. Maison, C*., et al.*, "SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin," *Nat Genet,* vol. 43, pp. 220-7, Mar 2011.

88. Lee, J. S*., et al.*, "Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS," *Cell,* vol. 131, pp. 1084-96, Dec 14 2007.

89. Wu, M*., et al.*, "Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS," *Mol Cell Biol,* vol. 28, pp. 7337-44, Dec 2008.

90. Margueron, R*., et al.*, "Role of the polycomb protein EED in the propagation of repressive histone marks," *Nature,* vol. 461, pp. 762-7, Oct 8 2009.

91. Honda, S. and Selker, E. U., "Direct interaction between DNA methyltransferase DIM-2 and HP1 is required for DNA methylation in Neurospora crassa," *Mol Cell Biol,* vol. 28, pp. 6044-55, Oct 2008.

92. Honda, S*., et al.*, "Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation," *Nat Struct Mol Biol,* vol. 19, pp. 471-7, S1, May 2012.

93. Xu, C*., et al.*, "Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2)," *Proc Natl Acad Sci U S A,* vol. 107, pp. 19266-71, Nov 9 2010.

94. Sun, Y*., et al.*, "Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60," *Nat Cell Biol,* vol. 11, pp. 1376-82, Nov 2009.

95. Jeong, K. W*., et al.*, "Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation," *Nat Struct Mol Biol,* vol. 18, pp. 1358-65, Dec 2011.

96. Wang, M*., et al.*, "Structural genomics of histone tail recognition," *Bioinformatics,* vol. 26, pp. 2629-30, Oct 15 2010.

97. Kuo, A. J*., et al.*, "The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome," *Nature,* vol. 484, pp. 115-9, Apr 5 2012.

98. Wysocka, J*., et al.*, "WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development," *Cell,* vol. 121, pp. 859-72, Jun 17 2005.

99. Winter, S*., et al.*, "14-3-3 proteins recognize a histone code at histone H3 and are required for transcriptional activation," *EMBO J,* vol. 27, pp. 88-99, Jan 9 2008.

100. Espejo, A*., et al.*, "A protein-domain microarray identifies novel protein-protein interactions," *Biochem J,* vol. 367, pp. 697-702, Nov 1 2002.

101. Bua, D. J*., et al.*, "Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks," *PLoS One,* vol. 4, p. e6789, 2009.

102. Huyen, Y*., et al.*, "Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks," *Nature,* vol. 432, pp. 406-11, Nov 18 2004.

103. Wysocka, J., "Identifying novel proteins recognizing histone modifications using peptide pull-down assay," *Methods,* vol. 40, pp. 339-43, Dec 2006.

104. Shogren-Knaak, M. A. and Peterson, C. L., "Creating designer histones by native chemical ligation," *Methods Enzymol,* vol. 375, pp. 62-76, 2004.

105. Simon, M. D*., et al.*, "The site-specific installation of methyl-lysine analogs into recombinant histones," *Cell,* vol. 128, pp. 1003-1012, Mar 2007.

106. Luger, K*., et al.*, "Expression and purification of recombinant histones and nucleosome reconstitution," *Methods Mol Biol,* vol. 119, pp. 1-16, 1999.

107. Huynh, V. A*., et al.*, "A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone," *J Mol Biol,* vol. 345, pp. 957-68, Feb 4 2005.

108. Franz, H*., et al.*, "Multimerization and H3K9me3 binding are required for CDYL1b heterochromatin association," *J Biol Chem,* vol. 284, pp. 35049-59, Dec 11 2009.

109. Edman, P., "A method for the determination of amino acid sequence in peptides," *Arch Biochem,* vol. 22, p. 475, Jul 1949.

110. Karas, M. and Hillenkamp, F., "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons," *Analytical Chemistry,* vol. 60, pp. 2299-2301, 1988.

111. Tanaka, K*., et al.*, "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry," *Rapid Communications in Mass Spectrometry,* vol. 2, pp. 151-153, 1988.

112. Fenn, J. B*., et al.*, "Electrospray ionization for mass spectrometry of large biomolecules," *Science,* vol. 246, pp. 64-71, October 6, 1989 1989.

113. Schmidt, C., "Absolute and relative quantification of proteins in large protein-RNA assemblies by mass spectrometry," Dr. rer. nat., Bioanalytical Mass Spectrometry Group, Georg August University, Goettingen, 2010.

114. Shevchenko, A*., et al.*, "Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels," *Anal Chem,* vol. 68, pp. 850-8, Mar 1 1996.

115. Washburn, M. P., "Utilisation of proteomics datasets generated via multidimensional protein identification technology (MudPIT)," *Brief Funct Genomic Proteomic,* vol. 3, pp. 280-6, Nov 2004.

116. Hubner, N. C*., et al.*, "Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis," *Proteomics,* vol. 8, pp. 4862-72, Dec 2008.

117. Wisniewski, J. R*., et al.*, "Universal sample preparation method for proteome analysis," *Nat Methods,* vol. 6, pp. 359-62, May 2009.

118. Nagaraj, N*., et al.*, "System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap," *Mol Cell Proteomics,* vol. 11, p. M111 013722, Mar 2012.

119. Cox, J. and Mann, M., "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nat Biotechnol,* vol. 26, pp. 1367-72, Dec 2008.

120. Cox, J*., et al.*, "Andromeda: a peptide search engine integrated into the MaxQuant environment," *J Proteome Res,* vol. 10, pp. 1794-805, Apr 1 2011.

121. Schmidt, C*., et al.*, "Investigation of protein-RNA interactions by mass spectrometry--Techniques and applications," *J Proteomics,* vol. 75, pp. 3478-94, Jun 27 2012.

122. Schmidt, C*., et al.*, "Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring," *Anal Chem,* vol. 82, pp. 2784-96, Apr 1 2010.

123. de Godoy, L. M*., et al.*, "Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast," *Nature,* vol. 455, pp. 1251-4, Oct 30 2008.

124. Nikolov, M*., et al.*, "Quantitative mass spectrometry-based proteomics: an overview," *Methods Mol Biol,* vol. 893, pp. 85-100, 2012.

125. Liu, H*., et al.*, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics," *Anal Chem,* vol. 76, pp. 4193-201, Jul 15 2004.

126. Ong, S. E*., et al.*, "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics," *Mol Cell Proteomics,* vol. 1, pp. 376-86, May 2002.

127. Conrads, T. P*., et al.*, "Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling," *Anal Chem,* vol. 73, pp. 2132-9, May 1 2001.

128. Gygi, S. P*., et al.*, "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags," *Nat Biotechnol,* vol. 17, pp. 994-9, Oct 1999.

129. Schmidt, C. and Urlaub, H., "iTRAQ-labeling of in-gel digested proteins for relative quantification," *Methods Mol Biol,* vol. 564, pp. 207-26, 2009.

130. Desiderio, D. M. and Kai, M., "Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue," *Biomed Mass Spectrom,* vol. 10, pp. 471-9, Aug 1983.

131. McAlister, G. C*., et al.*, "Increasing the multiplexing capacity of TMT using reporter ion isotopologues with isobaric masses," *Anal Chem,* Aug 2 2012.

132. Geiger, T*., et al.*, "Super-SILAC mix for quantitative proteomics of human tumor tissue," *Nat Methods,* vol. 7, pp. 383-5, May 2010.

133. Bantscheff, M*., et al.*, "Quantitative mass spectrometry in proteomics: a critical review," *Anal Bioanal Chem,* vol. 389, pp. 1017-31, Oct 2007.

134. Shannon, P*., et al.*, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Res,* vol. 13, pp. 2498-504, Nov 2003.

135. Rigbolt, K. T*., et al.*, "GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data," *Mol Cell Proteomics,* vol. 10, p. O110 007450, Aug 2011.

136. R-Development-Core-Team, "R: a Language and Environment for Statistical Computing," *R Foundation for Statistical Computing, Vienna, Austria,* 2008.

137. Sambrook, J. and Russell, D. W., *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor Laboratory Press, 2001.

138. Lowary, P. T. and Widom, J., "New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning," *J Mol Biol,* vol. 276, pp. 19-42, Feb 13 1998.

139. Thastrom, A*., et al.*, "Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences," *J Mol Biol,* vol. 288, pp. 213-29, Apr 30 1999.

140. Lis, J. T. and Schleif, R., "Size fractionation of double-stranded DNA by precipitation with polyethylene glycol," *Nucleic Acids Res,* vol. 2, pp. 383-9, Mar 1975.

141. Gasteiger, E*., et al.*, "ExPASy: The proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Res,* vol. 31, pp. 3784-8, Jul 1 2003.

142. Laemmli, U. K., "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature,* vol. 227, pp. 680-5, Aug 15 1970.

143. Gallagher, S. R., "One-dimensional SDS gel electrophoresis of proteins," *Curr Protoc Mol Biol,* vol. Chapter 10, p. Unit 10.2A, Aug 2006.

144. Weber, K. and Osborn, M., "The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis," *J Biol Chem,* vol. 244, pp. 4406-12, Aug 25 1969.

145. Neuhoff, V*., et al.*, "Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250," *Electrophoresis,* vol. 9, pp. 255-62, Jun 1988.

146. Simon, M. D., "Installation of site-specific methylation into histones using methyl lysine analogs," *Curr Protoc Mol Biol,* vol. Chapter 21, pp. Unit 21 18 1-10, Apr 2010.

147. Sandaltzopoulos, R*., et al.*, "Transcriptional repression by nucleosomes but not H1 in reconstituted preblastoderm Drosophila chromatin," *EMBO J,* vol. 13, pp. 373-9, Jan 15 1994.

148. Dignam, J. D*., et al.*, "Eukaryotic gene transcription with purified components," *Methods Enzymol,* vol. 101, pp. 582-98, 1983.

149. Schulze, W. X. and Mann, M., "A Novel Proteomic Screen for Peptide-Protein Interactions," *J. Biol. Chem.,* vol. 279, pp. 10756-10764, March 12, 2004 2004.

150. Shevchenko, A*., et al.*, "In-gel digestion for mass spectrometric characterization of proteins and proteomes," *Nat Protoc,* vol. 1, pp. 2856-60, 2006.

151. Jensen, L. J*., et al.*, "STRING 8--a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Res,* vol. 37, pp. D412-6, Jan 2009.

152. Assenov, Y*., et al.*, "Computing topological parameters of biological networks," *Bioinformatics,* vol. 24, pp. 282-4, Jan 15 2008.

153. Oellerich, T*., et al.*, "SLP-65 phosphorylation dynamics reveals a functional basis for signal integration by receptor-proximal adaptor proteins," *Mol Cell Proteomics,* vol. 8, pp. 1738-50, Jul 2009.

154. Gordon, F*., et al.*, "The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays," *J Biol Chem,* vol. 280, pp. 33701-6, Oct 7 2005.

155. Sörös, S., "H3K9me3 dependent HP1 chromatin association and its consequences for chromatin packaging," Doctor rerum naturalium, Max Planck Institute for Biophysical Chemistry, Georg August University, Göttingen, 2009.

156. Kumar, K. S*., et al.*, "Expeditious chemical synthesis of ubiquitinated peptides employing orthogonal protection and native chemical ligation," *Bioconjug Chem,* vol. 22, pp. 137-43, Feb 16 2011.

157. Dawson, P. E*., et al.*, "Synthesis of proteins by native chemical ligation," *Science,* vol. 266, pp. 776-9, Nov 4 1994.

158. Fujimoto, S*., et al.*, "Proteome analysis of protein partners to nucleosomes containing canonical H2A or the variant histones H2A.Z or H2A.X," *Biol Chem,* vol. 393, pp. 47-61, Jan 1 2012.

159. Hancock, R., "Packing of the polynucleosome chain in interphase chromosomes: evidence for a contribution of crowding and entropic forces," *Semin Cell Dev Biol,* vol. 18, pp. 668-75, Oct 2007.

160. Hagen, J. v., *Proteomics Sample Preparation*. Weinheim, Germany: Wiley-VCH, 2008.

161. Eaves, G. N. and Jeffries, C. D., "Isolation and Properties of an Exocellular Nuclease of Serratia Marcescens," *J Bacteriol,* vol. 85, pp. 273-8, Feb 1963.

162. Adams-Cioaba, M. A. and Min, J., "Structure and function of histone methylation binding proteins," *Biochem Cell Biol,* vol. 87, pp. 93-105, Feb 2009.

163. Torok, M. S. and Grant, P. A., "The generation and recognition of histone methylation," *Results Probl Cell Differ,* vol. 41, pp. 25-46, 2006.

164. Daniel, J. A*., et al.*, "Effector proteins for methylated histones: an expanding family," *Cell Cycle,* vol. 4, pp. 919-26, Jul 2005.

165. Feng, W*., et al.*, "PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation," *Nat Struct Mol Biol,* vol. 17, pp. 445-50, Apr 2010.

166. Shi, X*., et al.*, "ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression," *Nature,* vol. 442, pp. 96-9, Jul 6 2006.

167. Lachner, M*., et al.*, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins," *Nature,* vol. 410, pp. 116-20, Mar 1 2001.

168. Karagianni, P*., et al.*, "ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation," *Mol Cell Biol,* vol. 28, pp. 705-17, Jan 2008.

169. Kokura, K*., et al.*, "Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion," *EMBO J,* vol. 29, pp. 3673-87, Nov 3 2010.

170. Bostick, M*., et al.*, "UHRF1 plays a role in maintaining DNA methylation in mammalian cells," *Science,* vol. 317, pp. 1760-4, Sep 21 2007.

171. Sharif, J*., et al.*, "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA," *Nature,* vol. 450, pp. 908-12, Dec 6 2007.

172. Qin, W*., et al.*, "Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1," *J Cell Biochem,* vol. 112, pp. 439-44, Feb 2011.

173. Nozawa, R. S*., et al.*, "Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation," *Nat Cell Biol,* vol. 12, pp. 719-27, Jul 2010.

174. Spruijt, C. G*., et al.*, "CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex," *Mol Biosyst,* vol. 6, pp. 1700-6, Sep 2010.

175. Tong, J. K*., et al.*, "Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex," *Nature,* vol. 395, pp. 917-21, Oct 29 1998.

176. Zhang, K. M*., et al.*, "Characterization of Spindlin1 isoform2 in mouse testis," *Asian J Androl,* vol. 10, pp. 741-8, Sep 2008.

177. Wang, W*., et al.*, "Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes," *EMBO Rep,* vol. 12, pp. 1160-6, Nov 2011.

178. D'Andrea, A. D., "Susceptibility pathways in Fanconi's anemia and breast cancer," *N Engl J Med,* vol. 362, pp. 1909-19, May 20 2010.

179. Beever, C*., et al.*, "Methylation of ZNF261 as an assay for determining X chromosome inactivation patterns," *Am J Med Genet A,* vol. 120A, pp. 439-41, Jul 30 2003.

180. Mosch, K*., et al.*, "HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats," *PLoS One,* vol. 6, p. e15894, 2011.

181. Smothers, J. F. and Henikoff, S., "The HP1 chromo shadow domain binds a consensus peptide pentamer," *Curr Biol,* vol. 10, pp. 27-30, Jan 13 2000.

182. Steiner, L. A*., et al.*, "Patterns of histone H3 lysine 27 monomethylation and erythroid cell type-specific gene expression," *J Biol Chem,* vol. 286, pp. 39457-65, Nov 11 2011.

183. Bian, C*., et al.*, "Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation," *EMBO J,* vol. 30, pp. 2829-42, Jul 20 2011.

184. Doyon, Y. and Cote, J., "The highly conserved and multifunctional NuA4 HAT complex," *Curr Opin Genet Dev,* vol. 14, pp. 147-54, Apr 2004.

185. Qi, H. H*., et al.*, "Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development," *Nature,* vol. 466, pp. 503-7, Jul 22 2010.

186. Tang, H*., et al.*, "Protein-protein interactions in eukaryotic transcription initiation: structure of the preinitiation complex," *Proc Natl Acad Sci U S A,* vol. 93, pp. 1119-24, Feb 6 1996.

187. Silveira, A. C*., et al.*, "Over-expression of the BRMS1 family member SUDS3 does not suppress metastasis of human cancer cells," *Cancer Lett,* vol. 276, pp. 32-7, Apr 8 2009.

188. Klose, R. J*., et al.*, "The retinoblastoma binding protein RBP2 is an H3K4 demethylase," *Cell,* vol. 128, pp. 889-900, Mar 9 2007.

189. Ma, H*., et al.*, "M phase phosphorylation of the epigenetic regulator UHRF1 regulates its physical association with the deubiquitylase USP7 and stability," *Proc Natl Acad Sci U S A,* vol. 109, pp. 4828-33, Mar 27 2012.

190. Pichler, G*., et al.*, "Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways," *J Cell Biochem,* vol. 112, pp. 2585-93, Sep 2011.

191. Oettel, S*., et al.*, "The activity binding to the termination region of several pol III genes represents a separate entity and is distinct from a novel component enhancing U6 snRNA transcription," *Nucleic Acids Res,* vol. 26, pp. 4324-31, Oct 1 1998.

192. Hakme, A*., et al.*, "The macroPARP genes Parp-9 and Parp-14 are developmentally and differentially regulated in mouse tissues," *Dev Dyn,* vol. 237, pp. 209-15, Jan 2008.

193. Nakamura, T*., et al.*, "PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos," *Nature,* vol. 486, pp. 415-9, Jun 21 2012.

194. Zohn, I. E*., et al.*, "The Hectd1 ubiquitin ligase is required for development of the head mesenchyme and neural tube closure," *Dev Biol,* vol. 306, pp. 208-21, Jun 1 2007.

195. Li, Y*., et al.*, "Genetic evidence for ubiquitin-specific proteases USP24 and USP40 as candidate genes for late-onset Parkinson disease," *Hum Mutat,* vol. 27, pp. 1017-23, Oct 2006.

196. Blomquist, P*., et al.*, "Increased nuclear factor 1 binding to its nucleosomal site mediated by sequence-dependent DNA structure," *Nucleic Acids Res,* vol. 27, pp. 517-25, Jan 15 1999.

197. Lees, J. A*., et al.*, "The retinoblastoma protein binds to a family of E2F transcription factors," *Mol Cell Biol,* vol. 13, pp. 7813-25, Dec 1993.

198. Miyazaki, Y*., et al.*, "MEF, a novel transcription factor with an Elf-1 like DNA binding domain but distinct transcriptional activating properties," *Oncogene,* vol. 13, pp. 1721-9, Oct 17 1996.

199. Gaire, M*., et al.*, "Isolation and characterization of two novel, closely related ATF cDNA clones from HeLa cells," *Nucleic Acids Res,* vol. 18, pp. 3467-73, Jun 25 1990.

200. Inostroza, J. A*., et al.*, "Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription," *Cell,* vol. 70, pp. 477-89, Aug 7 1992.

201. Cuddapah, S*., et al.*, "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains," *Genome Res,* vol. 19, pp. 24-32, Jan 2009.

202. Tommerup, N. and Vissing, H., "Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders," *Genomics,* vol. 27, pp. 259-64, May 20 1995.

203. Bogdanovic, O. and Veenstra, G. J., "DNA methylation and methyl-CpG binding proteins: developmental requirements and function," *Chromosoma,* vol. 118, pp. 549-65, Oct 2009.

204. Ghimouz, R*., et al.*, "The homeobox leucine zipper gene Homez plays a role in Xenopus laevis neurogenesis," *Biochem Biophys Res Commun,* vol. 415, pp. 11-6, Nov 11 2011.

205. Yamada, K*., et al.*, "Analysis of zinc-fingers and homeoboxes (ZHX)-1-interacting proteins: molecular cloning and characterization of a member of the ZHX family, ZHX3," *Biochem J,* vol. 373, pp. 167-78, Jul 1 2003.

206. Blackledge, N. P*., et al.*, "CpG islands recruit a histone H3 lysine 36 demethylase," *Mol Cell,* vol. 38, pp. 179-90, Apr 23 2010.

207. Ikegawa, S*., et al.*, "Cloning and characterization of ASH2L and Ash2l, human and mouse homologs of the Drosophila ash2 gene," *Cytogenet Cell Genet,* vol. 84, pp. 167-72, 1999.

208. Lee, S., *et al.*, "Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases," *Proc Natl Acad Sci U S A,* vol. 103, pp. 15392-7, Oct 17 2006.

209. Ziemin-van der Poel, S., *et al.*, "Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias," *Proc Natl Acad Sci U S A,* vol. 88, pp. 10735-9, Dec 1 1991.

210. Shieh, B. H., *et al.*, "Localization of the gene-encoding upstream stimulatory factor (USF) to human chromosome 1q22-q23," *Genomics,* vol. 16, pp. 266-8, Apr 1993.

211. Li, X., *et al.*, "Mammalian polycomb-like Pcl2/Mtf2 is a novel regulatory component of PRC2 that can differentially modulate polycomb activity both at the Hox gene cluster and at Cdkn2a genes," *Mol Cell Biol,* vol. 31, pp. 351-64, Jan 2011.

212. Lee, S. J., *et al.*, "E3 ligase activity of RING finger proteins that interact with Hip-2, a human ubiquitin-conjugating enzyme," *FEBS Lett,* vol. 503, pp. 61-4, Aug 10 2001.

213. Peters, A. H., *et al.*, "Partitioning and plasticity of repressive histone methylation states in mammalian chromatin," *Mol Cell,* vol. 12, pp. 1577-89, Dec 2003.

214. Maeda, T., *et al.*, "TEF-1 and MEF2 transcription factors interact to regulate muscle-specific promoters," *Biochem Biophys Res Commun,* vol. 294, pp. 791-7, Jun 21 2002.

215. Cao, J., *et al.*, "RBP1L1, a retinoblastoma-binding protein-related gene encoding an antigenic epitope abundantly expressed in human carcinomas and normal testis," *J Natl Cancer Inst,* vol. 93, pp. 1159-65, Aug 1 2001.

216. Filippova, G. N., *et al.*, "An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes," *Mol Cell Biol,* vol. 16, pp. 2802-13, Jun 1996.

217. GrandPre, T., *et al.*, "Identification of the Nogo inhibitor of axon regeneration as a Reticulon protein," *Nature,* vol. 403, pp. 439-44, Jan 27 2000.

218. Hong, Z., *et al.*, "A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell," *Nucleic Acids Res,* vol. 36, pp. 2939-47, May 2008.

219. Brooks, W. S., *et al.*, "G2E3 is a dual function ubiquitin ligase required for early embryonic development," *J Biol Chem,* vol. 283, pp. 22304-15, Aug 8 2008.

220. Nicassio, F., *et al.*, "Human USP3 is a chromatin modifier required for S phase progression and genome stability," *Curr Biol,* vol. 17, pp. 1972-7, Nov 20 2007.

221. Cao, R., *et al.*, "Role of histone H3 lysine 27 methylation in Polycomb-group silencing," *Science,* vol. 298, pp. 1039-43, Nov 1 2002.

222. Plath, K., *et al.*, "Role of histone H3 lysine 27 methylation in X inactivation," *Science,* vol. 300, pp. 131-5, Apr 4 2003.

223. Chan, K. M. and Zhang, Z., "Leucine-rich repeat and WD repeat-containing protein 1 is recruited to pericentric heterochromatin by trimethylated lysine 9 of histone H3 and maintains heterochromatin silencing," *J Biol Chem,* vol. 287, pp. 15024-33, Apr 27 2012.

224. Bardos, J. I., *et al.*, "HPC3 is a new human polycomb orthologue that interacts and associates with RING1 and Bmi1 and has transcriptional repression properties," *J Biol Chem,* vol. 275, pp. 28785-92, Sep 15 2000.

225. Champion-Arnaud, P. and Reed, R., "The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site," *Genes Dev,* vol. 8, pp. 1974-83, Aug 15 1994.

226. Girling, R., *et al.*, "A new component of the transcription factor DRTF1/E2F," *Nature,* vol. 365, p. 468, Sep 30 1993.

227. Miller, J. E. and Reese, J. C., "Ccr4-Not complex: the control freak of eukaryotic cells," *Crit Rev Biochem Mol Biol,* vol. 47, pp. 315-33, Jul 2012.

228. Kim, J. Y., *et al.*, "KDM3B Is the H3K9 Demethylase Involved in Transcriptional Activation of lmo2 in Leukemia," *Mol Cell Biol,* vol. 32, pp. 2917-33, Jul 2012.

229. Neer, E. J., *et al.*, "The ancient regulatory-protein family of WD-repeat proteins," *Nature,* vol. 371, pp. 297-300, Sep 22 1994.

230. Weake, V. M. and Workman, J. L., "Histone ubiquitination: triggering gene activity," *Mol Cell,* vol. 29, pp. 653-63, Mar 28 2008.

231. Minsky, N*., et al.*, "Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells," *Nat Cell Biol,* vol. 10, pp. 483-8, Apr 2008.

232. Yamaguchi, Y*., et al.*, "NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation," *Cell,* vol. 97, pp. 41-51, Apr 2 1999.

233. Wenzel, S*., et al.*, "The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with alpha/beta type topology," *Biochem Biophys Res Commun,* vol. 370, pp. 414-8, Jun 6 2008.

234. Baillat, D*., et al.*, "Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II," *Cell,* vol. 123, pp. 265-76, Oct 21 2005.

235. Wu, J. I*., et al.*, "Understanding the words of chromatin regulation," *Cell,* vol. 136, pp. 200-6, Jan 23 2009.

236. Islam, M. N*., et al.*, "RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase," *Mol Cell Biol,* vol. 30, pp. 2460-72, May 2010.

237. Wegiel, J*., et al.*, "The role of DYRK1A in neurodegenerative diseases," *FEBS J,* vol. 278, pp. 236-45, Jan 2011.

238. Shao, Y*., et al.*, "Involvement of histone deacetylation in MORC2-mediated down-regulation of carbonic anhydrase IX," *Nucleic Acids Res,* vol. 38, pp. 2813-24, May 2010.

239. Makde, R. D*., et al.*, "Structure of RCC1 chromatin factor bound to the nucleosome core particle," *Nature,* vol. 467, pp. 562-6, Sep 30 2010.

240. Morgan, D., *The Cell Cycle Principles of Control*: OUP/New Science Press, 2007.

241. Peoples, R. J*., et al.*, "Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, in the Williams-Beuren syndrome deletion at 7q11.23," *Cytogenet Cell Genet,* vol. 82, pp. 238-46, 1998.

242. Thorpe, K. L*., et al.*, "Phylogeny and structure of the RING3 gene," *Immunogenetics,* vol. 44, pp. 391-6, 1996.

243. Doyon, Y*., et al.*, "ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation," *Mol Cell,* vol. 21, pp. 51-64, Jan 6 2006.

244. Kasof, G. M*., et al.*, "Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins," *Mol Cell Biol,* vol. 19, pp. 4390-404, Jun 1999.

245. Yoon, J. B*., et al.*, "Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro," *Mol Cell Biol,* vol. 14, pp. 1776-85, Mar 1994.

246. Unterholzner, L*., et al.*, "IFI16 is an innate immune sensor for intracellular DNA," *Nat Immunol,* vol. 11, pp. 997-1004, Nov 2010.

247. Mantovani, R., "The molecular biology of the CCAAT-binding factor NF-Y," *Gene,* vol. 239, pp. 15-27, Oct 18 1999.

248. Shi, Y*., et al.*, "Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein," *Cell,* vol. 67, pp. 377-88, Oct 18 1991.

249. Yao, Y. L*., et al.*, "Regulation of transcription factor YY1 by acetylation and deacetylation," *Mol Cell Biol,* vol. 21, pp. 5979-91, Sep 2001.

250. Stec, I*., et al.*, "WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma," *Hum Mol Genet,* vol. 7, pp. 1071-82, Jul 1998.

251. Williams, T. and Tjian, R., "Analysis of the DNA-binding and activation properties of the human transcription factor AP-2," *Genes Dev,* vol. 5, pp. 670-82, Apr 1991.

252. Hess, J*., et al.*, "AP-1 subunits: quarrel and harmony among siblings," *J Cell Sci,* vol. 117, pp. 5965-73, Dec 1 2004.

253. Young, M. D*., et al.*, "ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity," *Nucleic Acids Res,* vol. 39, pp. 7415-27, Sep 1 2011.

254. Sun, Z. W. and Allis, C. D., "Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast," *Nature,* vol. 418, pp. 104-8, Jul 4 2002.

255. Kim, J*., et al.*, "RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells," *Cell,* vol. 137, pp. 459-71, May 1 2009.

256. Zhu, B*., et al.*, "Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation," *Mol Cell,* vol. 20, pp. 601-11, Nov 23 2005.

257. Allen, P. B*., et al.*, "Isolation and characterization of PNUTS, a putative protein phosphatase 1 nuclear targeting subunit," *J Biol Chem,* vol. 273, pp. 4089-95, Feb 13 1998.

258. Lee, J. H*., et al.*, "Identification and characterization of a novel human PP1 phosphatase complex," *J Biol Chem,* vol. 285, pp. 24466-76, Aug 6 2010.

259. Woodage, T*., et al.*, "Characterization of the CHD family of proteins," *Proc Natl Acad Sci U S A,* vol. 94, pp. 11472-7, Oct 14 1997.

260. Zhang, Q*., et al.*, "Histone binding protein RbAp48 interacts with a complex of CREB binding protein and phosphorylated CREB," *Mol Cell Biol,* vol. 20, pp. 4970-8, Jul 2000.

261. Nicolas, E*., et al.*, "RbAp48 belongs to the histone deacetylase complex that associates with the retinoblastoma protein," *J Biol Chem,* vol. 275, pp. 9797-804, Mar 31 2000.

262. Santiveri, C. M*., et al.*, "The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine," *J Mol Biol,* vol. 382, pp. 1107-12, Oct 24 2008.

263. Everitt, B. S*., et al., Cluster Analysis*: Hodder Arnold, 2001.

264. Sif, S*., et al.*, "Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes," *Genes Dev,* vol. 15, pp. 603-18, Mar 1 2001.

265. Minsky, N. and Oren, M., "The RING domain of Mdm2 mediates histone ubiquitylation and transcriptional repression," *Mol Cell,* vol. 16, pp. 631-9, Nov 19 2004.

266. Shema, E*., et al.*, "The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression," *Genes Dev,* vol. 22, pp. 2664-76, Oct 1 2008.

267. Shen, X*., et al.*, "EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency," *Mol Cell,* vol. 32, pp. 491-502, Nov 21 2008.

268. Campos, E. I. and Reinberg, D., "Histones: annotating chromatin," *Annu Rev Genet,* vol. 43, pp. 559-99, 2009.

269. Zhou, V. W*., et al.*, "Charting histone modifications and the functional organization of mammalian genomes," *Nat Rev Genet,* vol. 12, pp. 7-18, Jan 2011.

270. White, R. J., "Transcription by RNA polymerase III: more complex than we thought," *Nat Rev Genet,* vol. 12, pp. 459-63, Jul 2011.

271. Emanuele, M. J*., et al.*, "Global identification of modular cullin-RING ligase substrates," *Cell,* vol. 147, pp. 459-74, Oct 14 2011.

272. Falsone, S. F*., et al.*, "A proteomic approach towards the Hsp90-dependent ubiquitinylated proteome," *Proteomics,* vol. 7, pp. 2375-83, Jul 2007.

273. Thiru, A*., et al.*, "Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin," *EMBO J,* vol. 23, pp. 489-99, Feb 11 2004.

274. Hicke, L*., et al.*, "Ubiquitin-binding domains," *Nat Rev Mol Cell Biol,* vol. 6, pp. 610-21, Aug 2005.

275. Moyal, L*., et al.*, "Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks," *Mol Cell,* vol. 41, pp. 529-42, Mar 4 2011.

276. Nakamura, K*., et al.*, "Regulation of homologous recombination by RNF20-dependent H2B ubiquitination," *Mol Cell,* vol. 41, pp. 515-28, Mar 4 2011.

277. Doil, C*., et al.*, "RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins," *Cell,* vol. 136, pp. 435-46, Feb 6 2009.

278. Popuri, V*., et al.*, "Recruitment and retention dynamics of RECQL5 at DNA double strand break sites," *DNA Repair (Amst),* vol. 11, pp. 624-35, Jul 1 2012.

279. Fagbemi, A. F*., et al.*, "Regulation of endonuclease activity in human nucleotide excision repair," *DNA Repair (Amst),* vol. 10, pp. 722-9, Jul 15 2011.

280. Shema, E*., et al.*, "RNF20 inhibits TFIIS-facilitated transcriptional elongation to suppress pro-oncogenic gene expression," *Mol Cell,* vol. 42, pp. 477-88, May 20 2011.

281. Palangat, M*., et al.*, "A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS," *Proc Natl Acad Sci U S A,* vol. 102, pp. 15036-41, Oct 18 2005.

282. Hassan, A. H*., et al.*, "Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes," *Cell,* vol. 104, pp. 817-27, Mar 23 2001.

283. Hassan, A. H*., et al.*, "Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes," *Cell,* vol. 111, pp. 369-79, Nov 1 2002.

Appendix

Appendix 1 Proteins enriched or excluded within at least one of the chromatin affinity purification experiments. The leading official gene names, IPI and UniProt IDs are included as reported by MaxQuant. Enrichment ratios are shown in log2 scale.

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
99D8.1	IPI00297779.7	P78371	-0.21	0.27	0.29	0.75	0.12	-0.50	-1.15	-0.93	-0.25	0.05
A-152E5.7	IPI00018288.1	P19387	0.80	-0.36	-	-0.42	1.14	0.56	-0.35	4.26	-0.18	0.13
A18HNRNP	IPI00641579.1	Q14011	1.01	-0.17	-0.20	-	-	0.18	-	-	-0.53	1.17
A2D	IPI00456363.1	Q8WWM7-3	-	-	-	2.34	-	-	-	-	-	-0.22
AAG	IPI00218495.3	P29372-2	-0.17	-0.21	0.27	-0.65	2.66	0.16	0.19	-0.04	0.26	-
ABBP1	IPI00334713.1	Q99729-3	1.03	-0.36	-0.28	0.55	1.28	0.26	0.17	0.82	-0.59	1.30
ABC1	IPI00464999.2	Q6AI08	0.16	0.44	0.23	1.39	-0.59	-1.33	-1.31	-	0.16	-0.77
ABRA1	IPI00030384.4	Q6UWZ7	-	-	-	-	-	-	-	3.54	-	-
ACAC	IPI00396015.5	Q13085-4	-0.61	-2.41	-1.23	-0.44	-1.36	0.05	0.53	-3.52	2.55	-0.53
ACF1	IPI00412415.2	Q9NRL2-1	-0.62	-0.67	0.57	-0.44	1.50	0.29	0.72	-0.41	0.18	1.78
ACTL8	IPI00329820.2	Q9H568	-2.86	2.39	2.20	2.57	-	-1.10	1.76	-0.58	-	3.53
ACTR3	IPI00028091.3	P61158	-	-	-	-	-	-	-	-	-	1.29
ACTR8	IPI00025646.5	Q9H981-1	0.19	0.06	-	-1.15	1.53	1.12	0.59	0.94	-	1.64
AD-005	IPI00166010.6	A5YKK6-1	-0.03	0.72	-0.19	1.54	-0.47	-1.09	-1.92	0.40	0.41	-0.50
AD-005	IPI00880052.1	A5YKK6-2	-	-	-	-	-	-	-1.70	-	-	-0.45
AD-009	IPI00856080.1	Q9UHW5-2	-	-	-	0.81	-	-	-	3.06	-	-
ADNP	IPI00022215.1	Q9H2P0	0.15	3.42	0.67	0.91	0.51	0.30	0.37	1.51	0.11	3.11
ADPRTL1	IPI00296909.4	Q9UKK3	-	-	-	1.46	-0.31	-0.83	-1.29	-	0.05	-0.25
ADTB1	IPI00940292.1	Q10567-3	-0.13	0.56	0.07	1.13	-	-1.33	-1.63	0.60	0.40	-0.60
AFR1	IPI00007256.1	Q9Y6X8	0.10	-	4.02	-	-	-	-	-	-	4.11
AHCTF1	IPI00979341.1	Q8WYP5-2	-	-	-	0.73	1.67	0.24	0.70	-1.31	-0.25	0.57
AIBP63	IPI00010486.2	Q9UPZ3-1	-	-	-	1.17	-0.87	-0.59	-1.99	-	-0.19	-0.52
AIDD	IPI00027547.2	P81605	-	-	-	-2.04	-	-	-	-	0.25	-
AIM	IPI01011182.1	F5GX68	-	1.57	-	1.93	0.91	-0.53	1.35	-0.78	1.47	-
AIM	IPI00031519.4	P26358-1	-3.53	1.13	0.56	-	-	-	-	-	-	1.98
AIM	IPI00220918.1	P26358-2	-	1.50	0.70	-	-	-	-	-	-	1.85
AKAP10	IPI00306265.2	O43572	-	-	-	1.31	-	-1.03	-1.68	-	0.55	-0.42
AKAP13	IPI00065931.4	Q12802-2	-	-	-	1.86	-	-	-	-	-	-
ALL1	IPI00009286.4	Q03164-1	0.02	-0.11	-2.81	-0.56	1.65	0.70	0.45	0.29	0.63	-
AMPK	IPI00220409.3	Q9Y478	-	-	-	1.27	-	-1.01	-1.57	-	0.35	-0.74
AMPK1	IPI00410287.3	Q13131-2	-	-	-	1.37	-0.35	-0.89	-1.10	-	0.70	-0.37
ANAPC10	IPI00007088.6	Q9UM13	-0.57	-1.34	0.28	-0.70	-	0.06	0.13	-2.18	-0.33	0.99
ANAPC3	IPI00794278.1	P30260	-0.28	-1.24	-0.01	0.12	0.96	-0.10	0.03	-1.55	-0.07	1.02
ANAPC4	IPI00966303.1	B3KN47	0.15	-0.45	0.40	-	0.27	-0.15	-0.52	-2.19	-0.40	0.48
ANAPC7	IPI00008248.4	Q9UJX3-1	-0.26	-0.77	-0.04	0.51	0.90	-0.09	0.49	-2.15	0.15	0.82
ANAPC8	IPI00005822.2	Q9UJX2-1	-0.02	-0.98	0.17	0.93	-	0.02	0.23	-2.19	-0.51	0.41
ANCO1	IPI00914930.2	Q6UB99	-	-	-	-	-	-0.13	0.46	2.32	-	0.81
ANCO2	IPI00448465.5	Q6UB98-1	-	-	-	-	-	-	-	2.31	-	-
ANKRA1	IPI00022520.1	O14593-1	1.10	0.80	-	0.05	1.19	0.63	0.15	-0.14	0.72	1.39
ANKRD25	IPI00470559.2	Q63ZY3-1	-	-	-	1.31	0.01	-0.44	-1.21	-	-0.08	-0.21
ANKRD52	IPI00398957.6	Q8NB46	1.28	1.65	1.44	-	-	-	-	0.50	-0.84	-
ANKT	IPI00020484.4	Q9BXS6-2	-0.05	-1.05	-0.49	-0.37	2.22	0.06	0.87	-0.76	0.53	1.16
ANP32B	IPI00007423.1	Q92688-1	-	-	-	-0.10	-	0.31	0.08	2.50	0.31	-1.14
ANX2LG	IPI00183695.9	P60903	-	-0.08	-	-0.21	-	0.11	-	-1.93	-	-
AOF1	IPI00784464.1	Q8NB78-1	-0.05	1.42	0.94	0.06	0.71	0.57	-0.28	0.55	-0.21	2.33
APE	IPI00215911.3	P27695	-	0.15	-	-1.00	-0.04	-0.50	1.31	-	-0.53	0.72
ARAP3	IPI00103380.1	Q8WWN8	-	-	-	1.52	-	-0.77	-1.50	-	0.15	-0.26
ARC240	IPI00913988.1	Q93074-2	0.26	0.07	-	1.20	0.43	-0.24	-1.01	0.58	-0.17	0.15
ARFIP2	IPI00910034.1	B4DUZ3	-	-	-	2.49	-	-	-	-	-0.12	-0.77
ARG134	IPI00033143.1	Q9UBQ5	-0.10	0.16	0.46	1.13	-0.09	-0.65	-1.45	-	-0.28	-0.28
ARH	IPI00004758.5	Q5SW96	-0.70	0.47	0.16	1.33	-0.71	-1.15	-1.59	0.32	0.69	-0.66
ARHGAP29	IPI00152011.4	Q52LW3-1	-	-	-	-	-	-	-1.81	-	0.42	-0.58
ARHGEF3	IPI00385169.1	Q9NR81-2	-	-	-	1.24	-	-0.84	-2.04	-	-	-0.42
ARHGEF7	IPI00900260.1	Q14155-5	-	-	-	1.34	-	-	-	-	-	-0.36
ARID4A	IPI00296069.2	P29374-1	2.19	0.49	-0.35	-	-	-	-	1.25	-	-
ARID4B	IPI00328828.5	Q4LE39-1	2.17	0.40	0.02	-	1.75	0.99	0.67	1.00	-0.11	1.42
ARL1	IPI00219518.7	P40616	-	-	-	1.20	-	-0.57	-1.43	-	0.22	-0.40
ASH2L	IPI00328658.3	Q9UBL3-1	-0.03	0.15	-2.41	-0.80	0.80	0.61	0.06	0.45	0.14	0.20

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
ASYIP	IPI00028946.2	O95197-3	1.48	-0.40	0.09	-1.80	-	-0.69	-	-	-0.75	0.04
ATAD5	IPI00102575.5	Q96QE3-1	-0.15	-0.21	-0.84	-	1.76	0.42	0.24	-0.52	1.08	0.65
ATF1	IPI00002501.1	P18846	0.26	0.31	-0.22	-1.70	1.12	0.77	0.41	0.34	-	1.32
ATF3	IPI00002502.1	P18847-1	-	-	1.77	-	-	-	-	-	-	-
ATF7	IPI00015147.1	P17544-1	1.12	0.14	2.14	-2.85	1.86	1.24	0.75	-0.31	0.24	2.14
ATG2B	IPI00872410.3	Q96BY7	-	-	-	-	-	-	-1.45	-	-	-0.46
ATX	IPI00556369.3	Q96Q15-1	-	-	-	1.83	-	-0.94	-	-	-	-0.19
ATXN1L	IPI00900336.1	P0C7T5	-	-	-	-1.97	-	-	-	-	-	-
AUF1	IPI00028888.1	Q14103-1	1.01	-0.51	-0.34	0.43	1.81	0.50	0.31	0.35	-0.65	1.38
AUF1	IPI00220683.1	Q14103-2	-	-	-	-	-	-	-	-	-	1.35
AYP1	IPI00984287.1	E9PN81	-	-	-	-	-	-	-	-	-	1.50
BA2R	IPI00645793.3	P21675-4	1.71	-	-	-	1.75	0.80	0.64	0.66	-	-
BAF155	IPI00234252.3	Q92922	0.17	0.24	0.81	0.17	0.53	0.34	0.02	1.93	0.48	0.72
BAF170	IPI00216047.3	Q8TAQ2-1	0.26	0.19	0.58	-0.14	0.54	0.34	0.70	1.81	0.62	0.67
BAF190B	IPI00514648.1	P51531-1	0.04	0.26	0.90	0.37	-	0.66	-0.32	1.74	0.70	1.43
BAF45D	IPI00023322.2	Q92785	-0.36	0.42	1.20	-0.07	-	0.44	0.15	1.17	0.00	0.76
BAL2	IPI00291215.6	Q460N5-6	0.13	-2.68	-0.19	-	-	-	-	-	-	-
BANP	IPI00956625.1	F5GZM0	0.12	0.22	-1.36	-1.69	0.92	1.02	0.83	-0.01	0.45	-
BAP1	IPI00026993.1	Q99496	0.22	0.46	-1.55	0.20	1.05	0.24	0.45	0.89	-0.31	0.05
BAP28	IPI00024279.4	Q9H583	-0.02	0.99	-2.87	0.16	0.61	0.11	0.04	1.04	-0.50	0.02
BAT1	IPI00848161.1	Q13838-1	-	-	-	-0.09	0.15	0.24	1.43	-0.34	-0.96	-0.47
BAT2D1	IPI00083708.5	Q9Y520-7	-	-	-	1.80	-0.33	-0.94	-1.63	-	-0.14	0.00
BAZ1B	IPI00069817.2	Q9UIG0-1	0.03	0.59	-0.10	-0.41	0.76	0.45	0.66	0.45	3.08	-0.06
BBAP	IPI00152503.1	Q8TDB6-1	-	-1.94	-0.24	1.26	-	0.00	-	-	-	0.01
BCL7C	IPI00447051.3	Q8WUZ0-2	-0.10	0.39	0.51	-0.47	-	0.73	-	1.77	0.26	0.96
BCLAF1	IPI00006079.1	Q9NYF8-1	1.12	-	-	0.62	-0.42	0.31	0.58	0.24	-2.74	0.13
BCOR	IPI00439548.1	Q6W2J9-1	0.69	0.19	-2.34	-0.91	1.48	0.72	0.28	0.76	-0.03	0.21
BEND3	IPI00741958.1	Q5T5X7	0.64	0.13	-2.52	-1.48	0.86	0.77	0.77	0.16	0.43	-0.08
BHLHB11	IPI00026559.1	P22415	0.17	0.24	-2.25	-1.22	0.72	0.75	0.23	0.46	0.30	-0.21
BHLHB12	IPI00020037.1	Q15853-1	0.36	0.18	-2.10	-1.24	0.77	0.87	0.67	0.42	0.59	-0.14
BHLHB2	IPI00021143.1	O14503	0.50	0.11	-2.21	-0.69	0.93	0.84	1.84	0.00	-0.04	-0.18
BHLHB39	IPI00030964.1	Q14469	0.82	0.36	-2.47	-0.58	0.96	0.80	0.12	0.29	0.47	-0.64
BHLHC41	IPI00009005.1	Q01664	0.61	0.13	0.09	-2.67	1.02	0.79	0.25	0.58	-0.67	1.57
BHLHD3	IPI00016177.1	Q99583	1.37	0.24	-0.89	-2.22	0.67	1.09	0.70	0.18	0.77	0.75
BHLHD4	IPI00018214.1	P61244-1	0.80	0.30	-0.67	-1.49	1.47	0.91	0.22	0.29	0.45	0.75
BHLHE8	IPI00007284.1	O15516	1.47	-	-	-	-	-	-	-	-	-
BIG3	IPI00005492.2	P61964	-0.27	-0.38	-0.55	-0.73	1.10	0.05	0.74	-0.26	1.30	0.04
BN51	IPI00215978.4	P05423	-	-	-	-1.57	3.01	1.55	2.22	0.12	-	1.21
BNIP2	IPI00030399.2	Q12982-1	-	-	-	1.25	-	-1.07	-1.35	-	-	-0.57
BOZF1	IPI00292615.3	Q96BR9-1	-	-	-	-	2.80	0.88	-	-	-	-
BP75	IPI00647008.1	Q9NPI1-2	0.46	-0.14	-	0.02	-	0.60	0.64	1.25	0.96	0.67
BRAF35	IPI00464951.5	Q9P0W2-1	-2.25	0.11	-0.37	-0.27	0.72	0.22	0.26	0.19	0.54	0.10
BRCC3	IPI00939471.6	P46736-1	-	-	-	-	-	-	-	3.44	-0.69	0.13
BRCC45	IPI00149276.1	Q9NXR7-1	-0.03	-	-	0.86	-	-0.30	0.05	3.10	-0.39	0.08
BRD2	IPI00797929.3	A2AAU1	-0.32	0.16	-1.57	-0.68	0.59	-0.01	-0.03	0.74	1.67	-
BRD3	IPI00014266.1	Q15059-1	1.06	-0.04	-0.02	-1.46	0.68	0.42	0.31	0.45	0.46	1.24
BRD4	IPI00440727.1	O60885-1	0.58	0.26	0.00	-1.93	0.84	0.95	0.31	1.22	0.31	1.20
BRD8	IPI00019226.2	Q9H0E9-1	1.54	-	-	-	-	-	-	-	-	0.05
BRMS1L	IPI00031653.9	Q5PSV4	2.02	0.41	-0.02	-0.46	-	0.69	0.92	1.15	0.24	1.19
BTBD12	IPI00291796.2	Q8IY92-1	-	-	-	-	-	-	-	3.43	-	-
BTBD14A	IPI00059930.1	Q96BF6	-	-	-	1.00	-	-0.93	-2.01	-	0.18	-0.43
BTBD14B	IPI00045207.2	Q96RE7	0.29	0.71	-	0.99	-0.33	-0.46	-1.15	-	0.21	-0.26
BTEB3	IPI00219691.2	Q9Y2Y9	0.11	-	-	-1.91	-	-0.10	0.65	-	0.18	1.08
BTEB4	IPI00011115.1	Q9BXK1	0.37	0.15	0.58	-2.48	-	0.28	0.19	-0.23	-0.18	1.21
C10	IPI00016925.1	Q99622	-0.03	0.12	-	1.01	-0.44	-	-2.02	0.35	-0.13	-0.26
C10orf12	IPI00166933.1	Q8N655	-	-	-	0.23	0.29	0.18	0.20	-	1.54	0.65
C10orf46	IPI00410319.1	Q86Y37-1	-	-	-	1.87	-	-	-1.77	-	0.14	-0.57
C11orf23	IPI00719725.1	Q5H9R7-5	1.73	1.45	1.22	-	-	-	-	0.54	-	-
C11orf30	IPI00981684.1	Q17RM7	2.09	-0.22	-0.09	0.58	1.29	0.66	0.77	-0.63	0.58	-
C12orf11	IPI00550986.4	Q9NVM9	-	-	-	-0.65	0.80	0.25	-0.33	3.99	-0.22	0.19
C12orf14	IPI00639887.3	B7Z287	1.24	0.17	-0.05	-	-	-	-0.10	0.81	0.32	0.56
C13orf8	IPI00064212.2	Q96JM3	-	3.40	0.34	1.21	0.83	0.27	-0.22	0.96	-0.38	1.15
C14orf117	IPI00784739.1	Q6PJG2	0.36	-0.63	0.18	-1.12	2.52	0.41	0.20	-0.81	-0.37	1.16
C14orf125	IPI01013086.1	Q86XA9-1	-0.15	0.68	0.01	1.36	-0.39	-0.77	-1.52	-	-0.12	-0.44
C14orf151	IPI00895800.3	Q27J81-1	0.33	0.43	-	1.33	-0.60	-1.50	-1.44	0.11	0.48	-0.74
C14orf46	IPI01025516.1	Q52LA3	0.54	-	-	-	-	-	-	-	-	1.21
C14orf92	IPI00006586.1	O94842	1.59	-	-	-1.59	0.25	-0.19	-	2.23	-0.29	0.07
C15orf25	IPI00018198.3	Q9NVX0-1	-	-	-	1.10	-	-0.83	-1.38	-	-0.09	-0.16
C15orf44	IPI00306017.3	Q96SY0-1	1.06	0.07	0.21	-0.12	0.78	0.10	-0.27	3.41	-0.42	0.21
C16orf88	IPI00554560.4	Q1ED39	-	-0.87	-0.30	-	2.40	0.23	0.77	-0.71	0.80	-
C17orf1	IPI00025202.6	O95466-2	3.10	-0.90	-0.25	-1.31	-1.19	-0.56	-0.16	0.14	-1.10	-
C17orf53	IPI00386891.4	Q8N3J3-1	1.99	0.15	-4.00	-	-	1.96	-	-0.61	-	1.44

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
C17orf96	IPI00246649.7	A6NHQ4	-1.65	0.99	-	0.35	-	0.21	1.19	0.11	1.20	-
C18orf21	IPI00845453.1	Q8TBS0	1.39	-0.15	-0.11	-	-	-	-	-	-	-
C18orf37	IPI00413363.1	Q6PI98-1	0.22	0.09	0.52	-1.25	-	1.19	0.56	0.74	-0.42	1.84
C19orf62	IPI00101987.6	Q9NWV8-1	-	-	-	-	-	-	-	3.89	-0.42	0.00
C19orf7	IPI00187011.4	Q9UPT8	-	-	-	-0.39	-	-	-	2.83	-	-0.09
C1orf103	IPI00479789.2	Q5T3J3-1	-	5.39	-	-	-	-	-	-	-	-
C1orf112	IPI00178512.7	Q9NSG2-1	-	-	-	0.84	-0.55	-1.09	-1.85	-	0.02	-0.22
C1orf149	IPI00009373.3	Q9HAF1-3	0.93	-0.03	-1.02	0.21	-	0.15	-0.25	0.66	1.33	-0.07
C1orf174	IPI00844507.2	Q8IYL3	0.10	-0.22	-0.05	-0.32	1.13	-0.15	-	-1.31	-	-
C1orf193	IPI00418336.3	Q68E01-2	-	-	-	-0.21	0.39	0.12	-0.06	4.89	0.07	0.01
C1orf24	IPI00328350.6	Q9BZQ8	-	-	-	1.29	-	-0.55	-1.12	-	-0.07	-0.65
C1orf73	IPI00743871.2	Q9NVH2-1	0.58	-	0.56	-0.49	0.74	0.21	-0.43	4.21	-0.18	0.09
C1orf77	IPI00719040.1	Q9Y3Y2-3	0.44	-0.29	-0.17	0.88	-	0.15	-0.04	-0.38	-1.17	-1.18
C20orf1	IPI00008477.1	Q9ULW0	0.07	-0.30	0.20	-0.20	1.93	0.08	0.07	-1.06	-0.20	0.65
C20orf154	IPI00032496.2	Q9UJA3-1	2.33	-0.19	-0.06	-	-	1.37	0.00	-0.16	-0.52	1.17
C20orf158	IPI00619921.3	Q9BTC0-4	1.63	0.03	-1.32	0.32	0.67	0.34	0.74	-1.26	-0.85	0.24
C20orf167	IPI00057097.3	Q9H147	0.14	-0.41	0.00	-0.99	2.54	0.54	0.61	-0.83	-0.56	0.98
C20orf20	IPI00019451.3	Q9NV56	1.31	0.35	0.37	-0.09	-	0.47	0.21	0.79	0.25	0.83
C20orf88	IPI00152692.2	Q8TEA8	-	-	-	-	-	-	-	-	-	2.20
C20orf94	IPI00046188.4	Q5VYV7	-	-	-	-	-	-	-	2.50	-	-
C21LRP	IPI00030770.1	O95456-1	-0.34	-	-	1.58	-	-0.75	-1.51	-	0.12	-0.16
C22orf18	IPI00031566.3	Q9NSP4-1	-0.17	0.07	-	0.91	-	-	-1.85	-	0.43	-0.41
C2orf3	IPI00418340.6	P16383-1	-	-	-	1.11	-	-1.04	-1.48	-	0.01	-0.48
C5orf7	IPI00298935.5	Q7LBC6-1	-	-	-	1.37	-0.55	-0.57	-1.81	-	-0.19	-0.22
C6orf61	IPI00014977.5	Q9NXL9-1	3.00	0.21	0.32	-	-	1.44	0.30	-0.05	-0.57	1.57
C6orf90	IPI00007729.1	Q9UMY1-1	1.84	-0.32	-0.72	0.02	1.36	0.48	0.21	0.30	-0.25	-0.49
C7orf26	IPI00043294.1	Q96N11-1	0.29	0.27	0.11	-0.46	-	-0.12	-0.72	4.17	-0.36	0.14
C8orf33	IPI00867735.1	Q9H7E9-2	-	-	-	-	-	-	-	2.37	-	-0.16
C8orf35	IPI00550272.3	Q9NVR2	0.84	-	-	-0.52	0.08	0.16	-0.52	3.97	-0.30	-0.06
C8orf52	IPI00375653.5	Q75QN2-1	-	-	-	0.43	0.32	-0.22	-0.96	3.96	-0.51	-0.17
C8orf55	IPI00171421.3	Q8WUY1	0.81	-	-	-1.30	-	-0.85	-	-	-	-0.12
C9orf142	IPI00030968.4	Q9BUH6-1	0.19	0.01	0.60	-1.44	1.21	-	-	0.46	-	-0.20
C9orf80	IPI00000330.5	Q9NRY2-1	0.27	-0.04	-	-0.41	-	-	-0.32	4.76	-	0.04
CAF	IPI00023177.4	Q13111-1	0.56	2.80	0.24	-0.14	-0.43	0.17	0.11	-	0.61	1.23
CAF1	IPI00006552.2	Q9UIV1	-0.38	0.35	0.18	1.67	-0.46	-1.23	-1.92	0.28	0.21	-0.49
CAF1A	IPI00011857.1	Q13112	0.04	2.59	-0.32	-0.45	-0.25	0.33	0.61	-0.13	0.18	1.32
CAGF28	IPI00973545.1	Q6ZW49-6	-	-	-	1.07	0.30	-0.32	-1.37	1.41	0.50	-0.32
CAGH32	IPI00167535.7	Q96L91-1	1.57	0.11	-0.24	0.90	0.66	-0.20	-0.61	0.47	-0.07	0.49
CALIF	IPI00295501.1	Q9UFF9	-0.69	0.13	-	1.63	-	-	-1.96	-	0.35	-0.58
CAMSAP3	IPI00176702.4	Q9P1Y5-1	-	-	-	1.04	0.24	-0.29	-1.13	0.63	0.41	-0.58
CANPL1	IPI00011285.1	P07384	-	-	-0.54	-1.93	-0.96	-0.39	-	-0.40	-	0.46
CASKIN2	IPI00103516.2	Q8WXE0-1	-	-	-	-	-	-0.87	-1.39	-	0.39	-0.23
CAT53	IPI00298731.2	Q96QC0	1.73	0.20	-	-0.58	0.14	-0.41	-1.32	1.63	-0.40	-0.34
CATX11	IPI00008708.5	O76021	-0.71	-0.65	0.24	0.96	0.52	-0.76	0.74	-1.64	1.14	0.26
CBX	IPI00878669.2	P83916	-	4.03	0.97	1.35	-	-	0.10	-	-	3.60
CBX3	IPI00297579.4	Q13185	0.01	3.80	1.14	1.45	-	0.44	-0.57	0.81	-0.16	3.57
CBX5	IPI00024662.1	P45973	0.05	4.35	-	2.50	-	-	2.25	-	-	4.26
CCDC111	IPI00065356.2	Q96LW4	2.20	-0.36	-	-	-	-	-	-0.08	-	-
CCDC120	IPI00181265.2	B4DFC1	-	-	-	0.60	-	-	-1.33	-	0.67	-0.37
CCDC86	IPI00012199.1	Q9H6F5	-0.34	-1.01	0.05	-0.52	2.36	0.05	0.40	-1.89	0.27	1.17
CDABP0017	IPI00879210.1	Q9Y5Q8-3	0.48	1.53	-	-	-	-	-	-	-	1.08
CDC36	IPI00005011.1	Q9NZN8-1	-	0.47	-	0.81	-0.46	-0.93	-1.78	0.10	0.42	-0.41
CDCA5	IPI00879287.1	B5MBX0	-0.25	0.08	0.04	0.36	1.35	-0.06	0.15	-1.48	0.41	0.73
CDK2AP1	IPI00022340.1	O14519	-1.22	-0.25	2.40	-	-	-	0.14	0.97	-	-
CDYL	IPI00293963.4	Q9Y232-1	-0.03	1.21	0.21	0.95	1.52	0.70	0.62	0.05	-0.31	2.59
CEBPB	IPI00289773.3	P17676	0.63	0.30	0.42	-1.84	0.49	1.02	0.69	0.08	-0.36	1.04
CENPE	IPI00296365.4	Q02224-1	-	-	-	-	-0.64	-	-1.23	-	0.70	-0.25
CENPI	IPI00552142.1	Q92674-1	-	-	-	-	-	-	-1.36	-	0.64	-0.35
CGI-120	IPI01020803.1	F8VVA7	0.07	0.22	0.54	1.35	-0.30	-0.82	-1.97	-0.06	0.19	-0.38
CHC1	IPI00001661.3	P18754-2	-0.35	-0.86	0.10	-0.26	1.49	-0.09	0.83	-2.65	0.05	0.72
CHD1	IPI00297851.1	O14646-1	3.18	0.68	-0.77	0.83	0.38	0.51	0.61	1.57	-0.70	-0.58
CHD3	IPI00465222.2	E9PG89	-	-0.10	1.77	0.59	-	0.09	0.53	1.34	0.05	2.44
CHD4	IPI01012026.1	F5H2G1	-0.86	0.92	1.78	-	-	-	-	-	-	2.31
CHD8	IPI00398992.7	Q9HCK8-1	1.72	-	-	0.66	0.90	0.40	-0.18	0.07	-0.55	0.69
CHEK1	IPI00909845.2	B4DT73	-	-	-	0.25	-	-1.32	-1.10	-	-	0.25
CHET9	IPI00299526.7	Q15022	-1.53	0.89	-0.95	-0.48	0.56	0.24	0.59	0.42	1.27	1.17
CHRAC1	IPI00010158.3	Q9NRG0	-0.08	0.20	0.56	-0.48	0.49	0.03	0.11	-0.55	-	1.55
CIC	IPI00045360.4	Q96RK0	-	-	-	-1.62	1.44	0.83	-	-	-	-
CIP150	IPI00164719.11	Q4ADV7-1	-	-	-	1.34	-	-1.17	-1.44	-	-	-0.62
CIRH1A	IPI00239815.9	Q969X6-1	-0.42	0.69	-2.75	0.62	0.46	0.17	0.37	0.59	-0.39	0.03
CLH17	IPI00024067.4	Q00610-1	1.41	-0.61	-0.27	0.12	-0.27	-0.92	1.21	0.04	0.17	-0.66

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
CMAS	IPI00303158.3	Q8NFW8-1	-0.01	-0.47	0.27	-0.85	2.29	-0.02	2.20	-0.61	-0.59	1.44
CNBP	IPI00430812.6	P62633-1	1.07	-0.31	-0.85	-0.70	1.36	0.44	0.30	0.05	-0.12	0.99
CNOT10	IPI01014565.1	Q9H9A5-1	-0.23	0.25	-0.26	1.00	-0.71	-1.00	-2.02	0.67	0.24	-0.33
CNOT3	IPI00005015.1	O75175-1	-	-	-	0.61	-0.43	-1.22	-1.68	-	0.49	-0.46
COBRA1	IPI00103483.1	Q8WX92	-	0.51	-	1.09	-0.74	-0.53	-0.70	3.02	0.08	-0.21
COPA	IPI00295857.7	P53621-1	1.71	0.32	-0.02	-	-	-	-	-	-	-
COPG	IPI00783982.1	Q9Y678	0.57	0.25	0.27	0.95	-0.43	-0.56	-1.54	-0.06	-0.18	-0.14
COPG2	IPI00002557.1	Q9UBF2	-	0.40	-	0.59	-0.45	-0.64	-1.38	-	0.01	-0.30
CPSF1	IPI00026219.4	Q10570	1.69	0.12	-0.75	-0.11	-0.27	-0.17	0.75	1.82	-0.55	-0.60
CPSF100	IPI00419531.2	Q9P2I0	1.24	0.00	-0.54	-0.28	-	-0.23	0.29	1.70	-0.42	-0.19
CPSF3L	IPI00063404.8	Q5TA45-1	0.44	-	-	-0.81	1.13	0.44	0.50	3.99	0.07	0.14
CREB1	IPI00027713.1	P16220-1	0.42	0.13	-0.10	-2.08	1.21	1.31	0.77	0.00	0.24	1.19
CREM	IPI00337364.4	Q03060-9	-1.10	-0.01	-0.06	-1.54	-	-	-	-	-	1.38
CREM	IPI00748922.3	Q03060-6	0.46	0.17	0.22	-1.76	-	0.90	0.18	-0.10	-0.05	1.26
CRM1	IPI00298961.3	O14980	-0.06	0.73	-0.05	1.43	-0.27	-1.40	-1.68	0.64	0.12	-0.49
CSDA	IPI00031801.4	P16989-1	-	-	-0.02	0.10	1.83	0.54	0.37	-	-0.44	1.81
CTCF	IPI00027988.1	P49711	0.32	-0.03	-1.11	-2.48	3.29	0.72	0.60	-0.62	-0.35	1.57
CTNND1	IPI00182540.7	O60716-3	-	-	-	0.38	0.07	-0.27	-2.00	0.35	-0.61	0.13
CTTN	IPI00029601.6	Q14247	-	-	-	0.71	-	-0.41	-1.33	-	0.50	-0.31
CUTL1	IPI00419894.1	P39880-3	-0.06	-0.26	0.63	-1.46	0.93	0.88	0.71	0.88	0.64	1.20
CVAK104	IPI00396218.2	Q6P3W7	-0.20	0.80	-	1.27	-0.35	-0.66	-2.28	-	-0.04	-0.15
CXXC2	IPI00185326.1	Q8NHM5-1	0.15	-0.17	-	-	1.41	0.59	0.47	0.48	-0.23	-
CXXC8	IPI00166009.2	Q9Y2K7-1	-0.32	0.54	-4.66	-0.44	0.19	0.01	0.21	0.25	0.19	-0.32
DADB-100D22.1-002	IPI00607575.2	B7Z7J5	0.40	0.64	-	1.21	0.00	-0.42	-1.13	-0.11	0.03	-0.06
DBI1	IPI00015922.2	Q9UL03-1	0.99	0.48	0.30	-0.05	0.74	0.08	-0.42	4.16	-0.18	-0.02
DCAF7	IPI00006754.1	P61962	-	-	-	0.65	0.04	-0.31	-0.80	3.22	-0.16	-0.46
DCC1	IPI00299884.4	Q9BVC3	-	-	-	1.29	-	-0.57	-1.21	-	-	-
DDX36	IPI00027415.3	Q9H2U1-1	0.15	-0.06	-0.51	0.14	3.12	0.30	0.17	0.32	-0.28	1.16
DDX47	IPI00023972.5	Q9H0S4	0.01	-0.04	0.03	0.37	-0.29	0.59	1.84	-0.43	-	0.64
DDX52	IPI00032423.3	Q9Y2R4	-0.18	-0.08	-0.40	0.86	0.44	-0.64	-1.27	0.03	1.22	-0.43
DEAF1	IPI00216889.5	O75398-1	-0.22	0.23	-4.01	-0.52	-0.24	0.57	0.61	0.93	0.15	-0.69
DEK	IPI00020021.3	P35659	-0.13	-0.03	-0.02	-0.87	4.00	0.24	0.55	-0.31	-0.18	0.97
DFS70	IPI00028122.1	O75475-1	0.17	0.06	-0.17	-0.40	1.47	-0.23	0.93	-0.35	-0.09	1.08
DGCR1	IPI00217560.3	P54198-1	1.15	-	-	-0.26	0.73	0.61	-0.15	0.19	0.32	-
DGT6	IPI00719210.1	Q7Z4H7-1	-	-	-	1.75	-	-0.82	-1.97	-	0.09	-0.53
DIA	IPI01015143.1	O60879-1	-	-	-	1.63	-	-0.86	-1.45	-	0.14	-0.52
DIAP1	IPI00936125.3	Q6URC4	-	-	-	1.41	-0.18	-0.46	-	-	-0.09	-0.37
DIAPH3	IPI00942390.1	C9JDG0	-	0.71	-	1.42	-0.97	-1.02	-2.14	-	0.34	-0.57
DMAHP	IPI00410575.2	Q8N196-1	-	-	2.13	-	-	-	-	-	-	1.78
DNAJC9	IPI00154975.3	Q8WXX5	-0.21	-0.06	0.22	-0.16	2.20	0.27	0.23	-0.38	0.00	1.17
DNM2	IPI00218889.3	P50570-2	-	-	-	1.38	-	-0.57	-	0.27	-	-
DP1	IPI00029095.1	Q14186	0.71	-0.09	-1.63	-1.70	2.10	1.07	2.04	-0.30	0.31	1.16
DP2	IPI00029096.1	Q14188-1	-	-	-	-	1.87	-	0.97	-0.09	-	-
DPY30	IPI00028109.1	Q9C005	0.35	0.07	-1.55	-	-	-	0.00	-0.10	-	-
DR1	IPI00008991.1	Q01658	0.82	0.00	-	-1.57	-	-	0.10	-	-0.05	-
DUC1	IPI00329605.6	P20585	0.80	-	-	-	1.41	0.76	0.38	-0.13	0.16	1.76
DXS254E	IPI00005658.3	P11441	-	-	-	1.46	-0.72	-0.74	-	-	-1.13	-0.21
DXS9928E	IPI00030098.5	Q14320	-	-	-	1.40	-0.50	0.12	0.18	-	0.03	-0.02
DYRK	IPI00014344.1	Q13627-1	-	-	-	-	-	-0.18	-	3.40	-	-
E1AF	IPI00017382.4	P43268	-	-	-	-	-	-	-1.33	-	-	0.07
E2F3	IPI00743509.2	O00716	-	-	-	-1.55	2.11	0.78	0.74	-0.15	0.48	1.18
E2F3	IPI01011332.1	Q24JQ3	0.58	-0.26	-1.70	-	-	-	-	-	-	-
E2F6	IPI00289560.3	O75461-1	1.67	1.55	-	0.41	-	-	-	0.61	-	1.34
E4BP4	IPI00290562.4	Q16649	0.43	-0.04	-0.10	-0.83	1.34	1.18	1.29	-0.03	-0.09	-
E4F	IPI00184135.6	Q66K89	-	0.26	-	-2.52	2.17	-	-	-	-	-
ECHDC1	IPI00302688.7	Q9NTX5-1	-	-	-	0.97	-	-0.30	-2.03	-	0.32	-0.07
EDC3	IPI00018009.2	Q96F86	-	-	-	1.51	-	-	-	-	-	0.10
EDC4	IPI00376317.4	Q6P2E9-1	-	-	-	2.11	-	-	-	-	-0.21	-0.75
EED	IPI00432040.1	O75530-2	-1.45	0.96	-0.97	0.14	0.87	0.35	1.07	-0.07	1.35	1.75
EFO1	IPI00552279.5	Q5FWF5-1	-0.21	-	0.16	-0.52	1.60	0.10	0.01	-1.57	0.04	0.88
EIF2B2	IPI00028083.1	P49770	-	-	-	1.19	-	-0.55	-	-	-0.33	-0.52
EIF2G	IPI00297982.7	P41091	-	-0.07	-	0.12	0.50	-0.43	2.24	-0.18	0.42	1.11
EIF4E2	IPI00744211.2	O60573	0.39	-0.03	0.12	1.61	-0.73	-0.43	-1.74	0.38	0.09	-0.54
ELF4	IPI00170744.1	Q99607	-	-	-	-1.94	-	0.49	-	-	-0.26	0.91
ELP1	IPI00293735.2	O95163	-	0.39	-	1.10	-0.40	-0.64	-0.77	-	0.10	-0.64
EMC19	IPI00301364.3	P63208-1	-0.25	0.38	-2.40	-	-	-	-	-	-	-
ENO1	IPI00465248.5	P06733-1	-1.20	-0.30	-	-	-	0.35	-0.04	2.49	-	-0.71
EPC1	IPI00018823.1	Q9H2F5-1	1.85	-	-	-	-	0.64	-	1.86	-	-
EPC2	IPI00141118.6	Q52LR7	1.82	-	-	-	-	-	-0.58	-	-	-
ERCC1	IPI00014040.1	P07992	-	-	-	-	-	-	-	2.55	-0.37	-

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
ERCC11	IPI00219179.6	Q92889	0.96	-	-	1.11	-	1.12	-0.42	2.40	-0.13	-
ERCC2	IPI00029728.3	P18074	0.36	-0.09	0.21	-1.85	0.99	1.19	0.81	-0.45	-0.06	0.33
ERF	IPI00032936.4	P50548	0.61	0.32	-0.53	-0.10	1.21	0.56	-0.10	-0.16	-0.01	0.35
ESP1	IPI00397474.2	Q14674-1	-	-	-	-	-	-0.31	-1.75	-	-	0.00
ESYT1	IPI00746655.1	Q9BSJ8-2	2.27	0.02	-	-0.10	-0.85	-0.59	-	-	-0.43	-0.47
EXP	IPI00021692.2	Q9NR56-1	-	-	-	1.32	-	-	-1.49	-	-0.46	-0.53
EYA3	IPI00015990.2	Q99504-1	-	-	-	-	-	-	-	0.33	-0.18	1.35
EZF	IPI00167069.2	O43474-3	-	0.20	1.29	-	1.09	0.98	0.63	-0.07	-	1.81
EZH2	IPI00947357.1	Q15910-1	-	-	-	0.01	0.54	0.26	0.76	0.58	1.35	-
EZH2	IPI00376787.4	Q15910-2	-1.45	0.85	-1.01	-	-	-	-	-	-	1.26
FACE	IPI00030252.1	Q9HB96	1.87	0.39	-0.23	0.18	0.22	-	-1.39	-0.38	-	-0.10
FAM91A1	IPI00152671.7	Q658Y4	-	0.45	-	0.57	0.05	-0.80	-1.17	-	-0.23	0.08
FAM96A	IPI00030985.1	Q9H5X1	-	-	-0.43	-1.25	-0.76	-0.62	-	-	-	-
FANCF	IPI00009290.3	Q9NPI8	2.02	0.85	0.02	0.78	-	-	-0.61	-	-0.24	-0.95
FBI1	IPI00026317.3	O95365	-0.17	-0.02	0.70	-0.57	0.66	0.43	0.53	0.26	0.00	1.43
FBL12	IPI00015011.1	Q9NXK8-1	0.68	0.26	2.20	-	-	-	-	-1.28	-	-0.31
FBL6	IPI00166767.3	Q8N531-1	-	-	-1.02	-	2.78	0.84	0.30	-	-0.81	2.20
FBX3	IPI00296789.2	Q9UK99-1	-0.23	-	-0.75	-0.29	0.09	0.49	1.15	-0.34	0.21	0.05
FEN1	IPI00026215.1	P39748	0.01	-0.21	-	0.46	-0.11	0.28	0.43	-0.46	0.06	1.42
FHOD1	IPI00001730.3	Q9Y613	-	-	-	1.28	-	-0.44	-1.29	-	-	-0.17
FIGNL1	IPI00335421.4	Q6PIW4-1	-	-	-	0.69	-0.47	-0.83	-1.66	-	0.25	-0.46
FIZ1	IPI00045801.4	Q96SL8	-	-	-	-1.28	1.31	0.75	0.87	-0.57	-0.59	3.57
FKHL16	IPI00401044.2	Q08050-3	-	-	-	-	-	-	-1.42	-	0.43	-
FKSG13	IPI00176903.2	Q6NZI2-1	-	-	-	-0.78	0.37	0.70	2.34	-	-0.01	0.28
FOSL2	IPI00011593.1	P15408-1	0.57	0.20	2.45	-1.82	-	-	-	0.78	-	1.72
FOXK1	IPI00556645.3	P85037-1	0.85	0.21	-0.34	-1.77	1.38	1.01	0.69	0.38	0.26	1.02
FOXK2	IPI00006029.2	Q01167-1	1.13	0.21	-0.43	-1.89	0.75	1.16	0.67	0.17	0.45	1.04
FPGS	IPI00016745.1	Q05932-1	-	0.40	-	0.77	-	-	-	-	-	-1.96
G2E3	IPI01026109.1	Q7L622	-0.79	-0.12	-0.64	0.19	-	-1.86	-0.64	0.54	0.10	-0.06
GARNL1	IPI00456722.3	Q6GYQ0-2	-	-	-	1.92	-	-	-	-	0.36	-
GATAD1	IPI00171123.4	Q8WUU5	1.36	-0.06	0.21	-0.09	1.42	0.23	0.31	-0.90	0.45	0.81
GATAD2A	IPI00478128.2	B5MC40	-1.18	-0.26	1.84	-0.27	0.55	0.57	0.41	0.92	0.68	2.18
GATAD2B	IPI00103554.1	Q8WXI9	-1.19	-0.24	2.07	0.11	0.55	0.41	0.40	1.14	0.55	2.07
GIGYF2	IPI00784233.2	Q6Y7W6-1	0.41	0.38	0.32	1.23	-0.08	-0.21	-1.13	0.60	0.25	-0.45
GLYR1	IPI00647648.3	Q49A26-2	-	-	-	-0.90	1.74	0.33	0.56	-1.35	0.66	-
GMDS	IPI00030207.1	O60547	-	-	-	1.59	-0.69	-0.87	-1.13	-	0.12	-0.56
GMIP	IPI00292376.4	Q9P107	-	-	-	1.57	-0.44	-0.50	-1.73	-	-0.03	-0.42
GPN1	IPI00027035.2	Q9HCN4	-	-	-	1.49	-0.54	-	-1.35	2.79	0.17	-0.52
GPS2	IPI00012301.1	Q13227	-	-	-	0.58	-	-	-	1.73	-	-0.11
GRL	IPI00022282.3	P04150-1	-	-	-	0.99	-0.51	-0.88	-1.54	-	0.48	-0.28
GRWD	IPI00027831.1	Q9BQ67	0.79	-	-	0.14	-	-0.29	0.20	2.22	-0.15	-0.14
GTBP	IPI00384456.4	P52701-1	1.00	-0.02	0.57	-0.28	1.33	0.80	1.09	0.09	-0.20	1.34
GTF2A1	IPI00004350.1	P52655-1	2.62	-	-	-	-	-	-	-	-	-
GTF2A2	IPI00004353.1	P52657	1.97	0.00	-	-	-	-	-	-	-	-
GTF2D1	IPI00022831.1	P20226	1.61	0.58	0.35	-0.25	1.13	0.68	0.39	0.25	0.02	0.37
GTF2E1	IPI00019977.2	P29083	2.70	-	-	-0.38	1.48	0.51	0.22	-	0.13	0.66
GTF2H4	IPI00016839.1	Q92759	0.09	-0.16	0.09	-0.81	1.40	1.09	1.88	-0.40	-0.48	0.65
GTF3C1	IPI00414482.3	Q12789-2	-	1.88	-	1.26	0.67	0.05	0.02	-	-0.27	1.06
GTF3C3	IPI00015806.3	Q9Y5Q9-1	0.87	1.52	-	0.73	0.48	0.15	-0.22	0.70	-0.39	0.98
GTF3C4	IPI00016725.2	Q9UKN8	-	1.29	-	0.65	0.68	0.16	0.21	0.44	-0.27	1.20
GYG	IPI00180386.5	P46976-1	-	-	-	-	-	-	-1.72	-	-	-
H1FX	IPI00021924.1	Q92522	-0.25	-0.12	0.02	-0.02	2.49	0.25	0.20	-0.41	0.37	1.21
HAP3	IPI00013217.1	P25208	0.54	0.47	0.53	-0.78	-	0.79	0.35	0.09	0.81	1.74
HARP	IPI00386311.2	Q9NZC9	1.94	0.39	-0.68	0.08	2.38	1.22	0.32	-0.12	-	1.34
HARS	IPI00021808.3	P12081	-	-	-	-	-	-	-	-	-	1.29
HAUSP	IPI00003965.5	Q93009	-2.45	1.35	1.24	1.79	-0.22	-0.70	0.99	0.97	1.39	3.08
HCFC2	IPI00002469.1	Q9Y5Z7	-0.05	-	-2.07	-	1.22	-	-	0.62	-	-
hCG_1742080	IPI00470473.2	Q5VW26	-	0.40	-	-3.21	-	1.90	-	-	-	2.00
hCG_1744585	IPI00796462.1	B5MDF5	-0.41	-0.16	0.18	0.22	1.13	-0.09	0.32	-2.43	0.08	0.45
hCG_1810992	IPI00953051.1	B2R8A7	2.28	-	-	0.08	-	0.38	-	1.04	-	0.36
hCG_1811093	IPI00165068.1	Q96AN2	-	-	-	-	-	-	-	-	-	2.13
hCG_1812148	IPI00871372.2	D3DS86	0.03	-	-	1.65	-	0.12	-	-	-	0.01
hCG_19946	IPI00032355.3	Q5T1Z8	-0.36	0.77	-	1.63	-0.67	-1.47	-2.07	0.79	1.15	-0.72
hCG_29955	IPI00900285.1	B1A8Z5	0.24	0.45	0.40	0.20	0.84	0.55	0.27	1.82	0.31	0.81
hCG_32740	IPI00941474.3	A8K4Q3	-	-	2.59	-	-	-	-	1.40	-	2.29
hCG_37170	IPI00009453.2	Q96SH1	0.46	0.07	0.94	-2.82	1.80	1.48	1.89	-0.72	-0.20	2.05
hCG_37966	IPI00027454.5	F5GXF0	-	-	-	0.55	-0.55	-	-1.45	0.27	0.75	-0.33
HD	IPI00002335.1	P42858	-	-	-	1.34	-	-0.30	-	-	-	-
HDAC1	IPI00013774.1	Q13547	-0.02	0.00	1.17	-0.25	1.38	0.65	0.81	0.24	0.22	1.84
HDAC2	IPI00289601.10	Q92769	-0.22	-0.04	1.18	-0.18	1.39	0.66	0.77	0.33	0.17	1.86
HDAC3	IPI00217965.1	O15379-2	0.88	0.41	0.34	0.19	-	0.71	-	1.24	-0.47	0.67

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
HELLS	IPI00010590.2	Q9NRZ9-1	-	-	-	1.19	0.08	-0.72	-1.40	-	0.09	-0.20
HIP12	IPI00024417.1	O75146	-0.52	0.40	-	1.61	-0.48	-1.32	-0.76	-	0.44	-0.73
HMBOX1	IPI00956180.1	Q6NT76-5	0.30	0.40	1.25	-2.74	2.05	1.46	0.99	-0.40	0.41	2.34
HMG14	IPI00554761.2	P05114	-0.11	-0.35	0.08	0.07	1.17	-0.05	-0.06	-1.80	0.20	0.33
HMG17L3	IPI00220484.3	O00479	-	0.20	-	0.18	-	-	-	-1.55	0.06	-
HMG20A	IPI00018924.3	Q9NP66-1	-2.24	-0.08	-0.02	-0.25	-0.47	0.36	0.28	0.15	1.67	-0.44
HMG2L1	IPI00455982.1	Q9UGU5	1.03	-0.73	-1.03	-0.85	1.89	0.66	1.07	-0.74	1.22	1.05
HNRNPU	IPI00479217.1	Q00839-2	1.29	0.10	-0.18	-	-	-	-	-	-	-
HOMEZ	IPI00941497.2	B4DX80	-	-	3.75	-	-	-	-	-	-	4.06
HRX2	IPI00218823.4	Q9UMN6-1	-0.15	0.31	-3.06	-0.59	0.90	0.72	0.76	-0.43	-0.23	-0.02
HSF1	IPI00024071.1	Q00613-1	-	-	-	2.12	-	-	-2.44	-	0.26	-0.46
HSP60	IPI00784154.1	P10809	-	-	-	1.55	-0.12	-0.56	-1.15	-	-0.12	-0.16
HSP75	IPI00030275.5	Q12931	-	-	-	1.59	-0.99	-1.04	-1.59	-	0.40	-0.40
HSPC130	IPI00164949.4	Q8IXH7-1	0.12	0.23	-	1.50	-0.10	-0.53	-0.83	3.15	-0.04	-0.04
HSPC144	IPI00383163.5	Q9P016-1	-1.65	0.06	-	-	0.39	0.05	0.42	-0.82	0.03	1.11
HSPC189	IPI00009777.3	Q9P0T4	-	2.26	-	-	-	-	-	-	-	3.03
HSPC301	IPI00387159.3	Q9NXR8-1	1.79	-	-	-	-	0.61	-	0.82	-	-
HUSSY-29	IPI00007089.1	O95478	-	-	-	-	-	-	-	-	-	-1.86
IDH3B	IPI00304417.7	O43837-1	-	-	-	1.52	-	-	-	-	-	-
IDN3	IPI00026466.9	Q6KC79-2	-	-	-	2.10	0.19	-0.14	1.17	1.79	-0.21	2.66
IDN3	IPI00436632.1	Q6KC79-1	0.64	2.48	-	-	-	-	-	-	-	-
IFI16	IPI00003443.3	Q16666-1	-0.11	-0.15	0.38	-2.37	2.69	-0.01	0.27	-0.25	1.33	2.10
IFI16	IPI00217474.7	Q16666-2	-0.43	-0.12	0.43	-2.14	2.86	0.02	0.24	-0.22	1.40	2.06
IGKJRB	IPI00030177.3	Q06330-1	-	-	1.01	-	-	-	-	-	-	1.89
ING1	IPI00099385.1	Q9UK53-2	1.56	0.35	-0.25	-	0.90	0.95	0.80	1.06	-	-
ING1	IPI00014324.3	Q9UK53-1	-	-	-	-	-	-	-	-	-	1.19
INO80	IPI00008091.2	Q9ULG1	0.34	-0.07	0.27	-1.54	1.71	1.25	1.00	0.97	-0.29	1.50
INO80S	IPI00014513.1	P25490	0.60	0.02	1.32	-0.70	1.81	0.24	0.18	0.00	-0.16	1.82
INRF2	IPI00106502.5	Q14145	-	-	2.73	-	-	-	-	-	-	-
INTS1	IPI00876931.1	Q8N201	1.05	0.47	-0.16	-0.09	0.48	0.22	-0.68	4.31	-0.22	0.02
INTS12	IPI00060379.1	Q96CB8	1.15	0.33	-0.04	-0.26	0.69	0.30	-0.51	4.47	-0.14	0.01
INTS2	IPI00477759.2	Q9H0H0	0.56	-	-	0.45	-0.21	-0.45	-0.82	4.52	-0.01	-0.47
INTS4	IPI00446765.2	Q96HW7-1	0.21	0.75	0.53	0.24	0.30	-0.36	-1.35	4.25	0.52	-0.19
INTS5	IPI00304676.5	Q6P9B9	0.77	0.68	0.33	0.40	0.13	-0.34	-0.98	4.38	-0.39	-0.18
INTS9	IPI00290514.1	Q9NV88	0.46	-	-	-0.52	-	0.44	0.09	4.31	0.10	0.24
IPO12	IPI00395694.1	Q9Y5L0-2	1.77	0.07	-0.77	0.08	0.00	-0.53	-0.66	-0.77	-0.44	-0.23
IPOA7	IPI00747764.3	B4DWX3	-	-	-	0.52	0.32	-0.19	-0.50	1.69	0.01	0.07
IRA1	IPI00002922.5	Q9BZK7	1.27	0.20	0.11	-0.05	1.04	0.45	0.08	1.20	-0.39	0.67
IRAK2	IPI00304986.5	O43187	-	-	-	-	-	-0.83	-1.74	-	0.81	-0.34
JADE3	IPI00006077.1	Q92613	0.63	-	-	0.38	0.82	0.12	0.21	0.49	2.00	-0.50
JARID1A	IPI00021363.3	P29375-1	2.06	-0.03	-	-	-	0.91	0.48	-	-	-
JNKK2	IPI00745806.3	O14733-3	-1.60	0.38	-	-	-	-	-	-	-	0.04
JUN	IPI00008965.1	P05412	0.65	-	1.82	-2.12	-	-	1.02	-	-0.47	1.66
JUND	IPI00289547.4	P17535	0.89	0.31	2.20	-2.11	1.35	-	-	-	-	1.79
KAISO	IPI00465140.3	Q86T24	-	-	3.50	0.82	-	-	-0.78	-	0.06	0.80
KBRAS2	IPI00021124.3	Q9NYR9-1	-	-	-	0.79	-	-0.58	-1.30	-0.07	-0.09	-
KDP	IPI01012075.1	F5H2M7	-	-	-	2.28	-	-	-	-	-0.39	-0.03
KHNYN	IPI00829596.3	O15037	-	-	-	0.75	-	-	-1.44	-	0.81	-0.54
KIAA0007	IPI00937477.1	Q15061	-0.50	0.87	-2.87	0.56	0.45	0.31	0.42	0.86	-0.47	0.08
KIAA0017	IPI00300371.5	Q15393-1	1.48	-0.06	0.02	0.10	-0.55	-0.09	-0.10	0.61	-0.22	-0.77
KIAA0035	IPI00216654.2	Q14978-2	-0.40	0.57	-0.34	0.16	-0.05	-0.31	0.02	1.92	0.02	-
KIAA0035	IPI00908873.1	Q14978-3	-0.53	0.56	-0.31	-	-	-	0.34	1.73	-	-
KIAA0042	IPI00299554.3	Q15058	1.08	0.27	-	1.19	-0.22	-0.89	-1.22	0.21	0.71	-0.44
KIAA0099	IPI00551014.1	Q14671-1	-	-	-	0.72	-	-	-2.04	-	1.24	-0.81
KIAA0170	IPI00552897.2	Q14676-1	0.01	-0.68	-0.20	0.44	0.93	0.59	0.88	-2.55	-0.55	0.38
KIAA0211	IPI00022460.2	Q92610	-	-	-	-	-	-	-	3.36	-	-
KIAA0295	IPI00307591.5	O15014	-	-	-	-	-	-	-	4.03	-0.39	-
KIAA0309	IPI00444046.4	Q6ZRS2-1	-0.87	-0.18	-1.77	-	-	-	-	-	-	0.44
KIAA0370	IPI00465296.2	Q96QU8	-	-	-	1.24	0.08	-0.82	-1.33	-	-0.04	-0.77
KIAA0395	IPI00301844.2	Q9H4I2	-0.13	-	3.58	-	-	-	-	0.82	-	3.45
KIAA0406	IPI00011702.4	O43156	0.04	0.46	0.16	1.04	-0.82	-1.13	-1.06	-	-0.20	-0.70
KIAA0414	IPI00022264.1	O43298	-0.05	0.09	-	-3.00	1.25	-0.06	-	-0.53	-	2.88
KIAA0425	IPI00477949.1	Q5VZL5-1	-	2.47	-	-	-	-	-	-	-0.14	1.17
KIAA0459	IPI00847831.1	Q5T2D3	-	-	-	-	-	-	-	4.09	-	-
KIAA0461	IPI00410717.2	Q7Z3K3-1	-	3.20	2.22	-	-	-	-0.69	-	-	2.57
KIAA0570	IPI00297593.4	Q70CQ2-1	-	-	-	1.44	-	-1.29	-	-	-	-0.19
KIAA0650	IPI00890837.1	A6NHR9-1	-	3.76	-	1.58	0.32	0.03	0.37	-0.29	-	1.47
KIAA0663	IPI00328306.9	O75152	-	-	-	-	-	-	-	-	-1.29	-1.12
KIAA0700	IPI00607589.1	O75182-2	2.66	0.44	-	-0.55	-	0.53	0.60	-0.05	-	0.84
KIAA0764	IPI00006651.1	O94864-1	2.17	-	-	0.30	-	-	-	-	-	0.18
KIAA0775	IPI00006064.4	A7MCY6-1	-	-	-	-	-	-	4.04	-	-	-

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG.H3K9me3
KIAA0783	IPI00472782.1	O94880-1	-2.33	-0.15	-0.07	-0.46	-0.11	-0.08	-0.19	0.72	2.36	-0.62
KIAA0791	IPI00026625.1	O75694-1	-0.25	0.50	-0.49	1.22	-0.49	-1.00	-1.60	0.02	0.38	-0.69
KIAA0852	IPI00455518.5	Q9Y6X9-1	-	-	-	0.72	-	-	0.29	4.28	0.29	-0.32
KIAA0886	IPI00298289.1	Q9NQC3-2	-	-	-	-1.33	-1.58	-0.43	-	-	-	-0.40
KIAA0886	IPI00478442.3	Q9NQC3-5	1.96	-0.30	-0.58	-	-	-	-	-	-	-
KIAA0892	IPI00000656.2	Q9Y6X3-1	0.73	2.82	0.33	0.95	-	0.10	0.39	1.61	-0.33	2.14
KIAA0913	IPI00877106.1	A7E2V4-4	-	-	-	1.18	-0.65	-1.26	-2.07	-	0.85	-0.39
KIAA0997	IPI00007273.3	Q9Y2K1-1	0.60	-	-	-1.72	1.45	0.99	0.82	-0.13	0.74	-
KIAA1057	IPI00902614.2	Q9UPU5	-	-	-	1.66	-0.16	-0.55	-0.59	-	-0.21	-0.16
KIAA1090	IPI00218240.3	O96028-1	-0.28	0.11	-0.43	-	-	0.56	1.02	-0.42	0.49	1.93
KIAA1111	IPI00480187.2	Q9UPP1-1	4.15	-	-0.07	-	1.04	-0.17	-	0.73	0.33	-0.25
KIAA1115	IPI00873586.3	Q9UPN7	1.15	1.47	1.19	1.49	-	-0.66	-	-	-0.57	-0.04
KIAA1125	IPI01008745.1	Q9ULU4-13	-	-	1.42	-	0.90	1.15	-	3.22	-	-
KIAA1148	IPI00868787.1	Q9HCD6-1	-	-	-	0.69	0.06	-0.30	-1.37	1.13	-0.03	-0.25
KIAA1196	IPI00074893.1	Q96KM6	-0.48	-0.45	0.62	-	1.88	0.16	0.15	-0.40	-0.24	2.13
KIAA1197	IPI00008200.2	Q9ULM3	1.46	0.05	-1.02	-0.55	0.67	0.70	0.80	0.12	0.31	0.41
KIAA1205	IPI00855833.1	Q9ULL5-3	0.04	2.01	-	0.90	0.81	-	-	-	-	1.74
KIAA1227	IPI00008137.1	Q9ULJ3-1	2.62	-0.31	-	-2.05	2.18	1.39	-0.92	-0.34	0.36	1.15
KIAA1231	IPI00398772.4	Q9NQV6-6	-	-	-	-	-	-	1.38	-	-	2.67
KIAA1266	IPI00165357.4	Q9BTC8-1	-0.85	-0.13	1.59	-0.01	0.89	0.54	0.80	0.53	0.66	2.03
KIAA1291	IPI00640703.3	Q9HAV4	2.00	0.54	-0.13	0.27	-0.43	-0.68	-0.71	0.48	-0.22	-0.51
KIAA1429	IPI00036742.6	Q69YN4-1	1.62	-0.02	-0.08	0.65	0.07	-0.07	0.01	1.12	-1.01	-0.48
KIAA1441	IPI00395813.1	Q8N1G0-1	-	0.03	-	-0.11	-	-	-	2.86	-	0.78
KIAA1452	IPI00220045.3	Q9NVU0-1	-	-	-	-1.67	2.74	1.26	0.85	0.51	-0.07	0.77
KIAA1523	IPI00292750.3	Q96QT6-1	2.74	-	-	-	-	0.84	-	-	0.91	1.10
KIAA1584	IPI00149044.9	Q6N043-1	-	-	-	-	-	-	-	-	-	3.90
KIAA1649	IPI00440688.4	Q9BY77-1	1.15	-1.20	0.00	0.16	2.34	0.80	0.04	-0.93	-0.21	-0.33
KIAA1665	IPI00289667.3	Q9Y535-1	-	-	-	-1.22	2.26	1.26	0.55	-	-0.62	1.07
KIAA1728	IPI00173946.8	Q9C0D5-1	0.21	0.52	-	1.37	-0.01	-0.59	-2.06	-	0.15	-0.30
KIAA1741	IPI00304589.5	Q9C0C2-1	-	-	-	1.52	-0.18	-1.18	-1.55	-	0.37	-0.57
KIAA1797	IPI00748360.2	Q5VW36	-	-	-	1.66	-	-0.90	-1.53	-	-0.18	-0.41
KIAA1798	IPI00306511.2	Q96JM7-1	-0.13	0.07	-1.75	-	-	-	-	-	-	0.09
KIAA1978	IPI00217661.4	Q8IY67-2	0.45	-0.38	0.40	0.51	-0.68	-0.15	-1.51	-0.14	0.39	-0.46
KIAA1991	IPI00456899.2	Q8NCN4	-	-	-	0.04	0.68	0.08	-1.13	3.02	0.18	0.07
KIAA2030	IPI00785015.3	Q6ZU65-1	1.71	-	0.79	-	0.50	0.94	0.14	0.27	0.29	0.27
KIF4	IPI00178150.5	O95239-1	0.52	0.50	-	-0.43	1.14	-0.37	0.85	0.21	0.91	0.96
KPNA1	IPI00303292.2	P52294	0.48	0.52	0.44	1.02	0.38	-0.25	-	1.98	0.07	0.41
L14	IPI00303813.5	Q9H8H0	-0.57	0.65	-2.79	-0.03	0.60	0.36	0.42	1.09	-0.41	-0.03
LAP2	IPI00216230.3	P42166	-0.19	1.29	-1.05	-1.37	-0.61	0.36	-0.09	2.33	-1.72	0.28
LATHEO	IPI00294402.1	Q9UBD5-1	0.01	1.47	0.36	0.55	0.34	0.52	2.29	-0.38	0.02	2.48
LATHEO	IPI00374747.5	Q9UBD5-2	-	-	-	1.38	-	0.34	2.02	-	-	2.49
LBP1	IPI00005018.3	Q9NZI7-1	0.89	0.27	0.23	-2.03	1.42	1.12	0.92	0.05	-	1.90
LIG1	IPI00219841.6	P18858	-	-	-	1.38	-	1.71	-1.52	-	-0.17	-0.36
LP3587	IPI00031633.4	Q9BQ15-1	-	-	-	-0.45	-	-	-	4.68	-	-
LRWD1	IPI00069309.6	Q9UFC0	0.16	1.45	0.50	0.87	0.67	0.73	3.05	-0.73	-0.05	2.97
LSF	IPI00037599.3	Q12800-1	0.74	0.35	-0.11	-2.16	1.52	1.25	1.01	0.03	-0.04	1.99
LYT10	IPI00845373.1	Q00653-1	-	-	-	1.16	1.36	1.18	1.77	-0.40	-0.44	1.59
MADH5	IPI00017730.1	Q99717	-	-	-	0.87	-	-0.27	-1.99	-	-	0.28
MAFF	IPI00099497.2	Q9ULX9	-	-0.23	2.41	-1.60	1.38	0.71	-	-	-	2.19
MAFG	IPI00007311.1	O15525	0.69	0.70	2.15	-1.43	-	-	-	-	-	2.19
MAFK	IPI00031018.2	O60675	0.09	-0.06	2.26	-1.28	-	1.48	-	-	-	1.93
MAP3K7IP3	IPI00166840.4	Q8N5C8-1	-	-	-	-	-	-	-1.51	-	0.10	-0.29
MBD2	IPI00434623.1	Q9UBB5-1	-1.37	-0.64	2.45	0.56	1.10	0.64	0.82	0.85	0.64	2.79
MBD3	IPI00439194.1	O95983-1	-0.96	-0.15	0.37	0.51	0.94	0.49	-0.08	1.49	0.48	0.55
MBD4	IPI00426728.3	O95243-2	-0.39	-0.47	1.16	-0.08	1.81	0.04	0.68	-0.68	-0.11	-
MBD4	IPI00426727.1	O95243-1	-	-	-	-	-	-	-	-	-	2.33
MBTD1	IPI00784023.1	Q05BQ5-1	2.32	-	-	0.10	-	-	-	-	-	-
MECP2	IPI00645192.4	P51608-2	-	-	-	-1.77	1.68	0.25	0.63	-0.64	0.01	3.91
MEIS2	IPI00023696.1	O14770-1	-	-	-	-	-	-	-	-	-	1.53
MEN1	IPI00182106.5	O00255-2	-0.12	0.02	-2.06	-1.05	0.86	0.57	0.53	-0.26	0.16	0.02
MGMT	IPI00028618.2	P16455	-2.06	-0.65	-	0.26	-	-	0.34	-	0.30	0.96
MIZ1	IPI00556634.1	Q13105-2	0.46	0.04	-	-	1.55	1.11	-	0.29	-	-
MLN70	IPI00013895.1	P31949	-1.31	0.24	-	-	-1.09	-	-	-	-	-
MORF4L1	IPI00550968.2	A5D8W6	1.36	0.38	0.26	0.39	-	-0.03	0.39	0.11	0.15	0.40
MSH2	IPI00017303.1	P43246	0.86	0.08	0.69	-0.42	1.90	0.53	0.10	-0.16	-0.25	1.67
MTA1L1	IPI00171798.1	O94776	-1.19	-0.23	2.27	0.03	0.42	0.49	0.29	1.05	0.61	2.31
MTF2	IPI00291983.8	Q9Y483-1	-1.01	0.71	-1.63	-0.22	0.92	0.13	0.80	0.07	1.15	-
MUS81	IPI00289454.3	Q96NY9	-	-	-	0.14	2.34	0.39	1.12	-0.42	-	-
MUTYH	IPI00414235.1	Q9UIF7-3	-	-	-	-2.25	2.38	0.46	0.80	-0.93	0.06	-
MUTYH	IPI00414236.7	Q9UIF7-2	-	-	-	-	-	-	-	-	-	2.35
MYH10	IPI00790503.3	P35580-3	-	-	-	-0.07	-0.43	-1.52	-	-	-0.03	-0.92

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG.H3K9me3
NAK	IPI00293613.2	Q9UHD2	-	-	-	0.59	-0.29	-0.83	-1.27	-	0.47	-0.31
NELFA	IPI00394679.3	B3KM78	-	-	-	1.15	-0.19	-0.33	-0.59	3.52	0.03	-0.08
NELFE	IPI00000858.2	B4DYX9	-	-	-	-0.44	0.14	0.21	0.11	2.91	0.06	0.42
NFI	IPI00029795.1	P08651-1	-	-	-	-3.32	2.39	0.93	0.30	-0.22	-0.19	-
NFI	IPI00218041.1	P08651-5	-	-	-	-	-	-	-	-	-	1.95
NFIA	IPI00163319.3	B4DS53	0.47	0.07	-	-2.83	2.60	-	-	-	-0.33	1.97
NFIX	IPI00218260.3	Q14938-1	-	0.00	-	-	-	1.31	-	-	-	-
NFYA	IPI00333568.1	P23511-1	0.75	-	0.94	-	1.11	0.63	0.70	0.36	0.67	1.61
NFYC	IPI00071697.1	Q13952-1	0.69	0.30	0.82	-0.97	1.41	0.75	0.80	0.28	0.50	1.74
NIRF	IPI00044681.1	Q96PU4-1	-	3.95	-	-	-	-	-	-	-	4.50
-	IPI00980612.1	No_ID10	0.23	-0.27	1.25	-	-	-	-	-	-	-
-	IPI00181352.2	A8K1B6	-	-0.50	-	1.34	-0.23	-0.55	-0.73	0.27	0.00	-
-	IPI01011704.1	A8K1E1	0.82	-0.71	-0.69	-0.93	1.38	0.54	0.14	-0.06	0.28	1.75
-	IPI00979371.2	B4E3V7	-	-	-	-	-	-	-	3.20	-	-
-	IPI00945820.1	C9JYP6	-	-	-	1.21	-	-	-1.78	-	0.05	-0.50
-	IPI00514550.1	E9PEQ4	0.31	-0.27	1.24	-	-	-	-	0.24	-	-
-	IPI00879166.1	E9PHA7	-1.47	-0.40	1.03	-0.32	0.70	0.71	0.34	1.25	1.05	-
-	IPI00925034.4	F2Z2U4	1.26	0.25	-0.43	0.83	0.26	-0.37	-0.21	0.39	-0.17	0.53
NOC4L	IPI00031661.1	Q9BVI4	1.89	-0.09	-0.36	0.54	1.05	0.78	-0.47	-0.10	0.16	-0.12
NR2B1	IPI00418394.2	P19793	-	-	-	1.93	0.13	-0.79	-1.49	-	0.04	-0.02
NR2C2	IPI00749258.1	P49116-1	0.86	0.11	-0.65	0.69	1.50	0.09	0.37	-0.38	-	-0.20
NR2C2	IPI00332451.3	P49116-2	-	-	-	0.25	1.59	0.17	0.53	-0.67	0.44	-
NSEP1	IPI00031812.3	P67809	0.40	-0.14	-0.41	0.44	2.65	0.69	0.35	0.06	-0.36	1.46
NSPC1	IPI00902987.2	Q9BSM1-1	-0.10	-0.04	-2.47	-0.28	0.41	0.60	0.19	0.54	-0.32	-0.30
NUBP2	IPI00644674.1	Q9Y5Y2	-1.04	0.24	-	1.71	-0.65	-0.91	-1.66	-	0.05	-0.42
NUP358	IPI00221325.3	P49792	-0.03	-0.02	-	1.88	-	-1.07	-	0.03	-0.16	0.03
OBFC2A	IPI00059434.2	Q96AH0-1	0.73	0.08	0.34	-1.35	-	0.44	0.37	4.24	0.12	0.38
OCR	IPI00550655.4	Q9Y657	4.90	-	-	-	-	-	-	-	-	1.75
OK/KNS-cl.5	IPI00885108.1	Q86TB9-1	-	-	-	1.45	-0.61	-1.07	-2.27	-	0.08	-0.31
ORC1	IPI00013215.2	Q13415	-0.09	0.87	0.43	-	-	1.16	-	-0.54	-	-
ORC2	IPI00013216.1	Q13416	0.06	1.46	0.44	1.00	0.25	0.58	2.67	-0.48	0.07	2.71
ORC5	IPI00015143.1	O43913	0.09	1.21	0.60	0.47	0.50	0.11	2.41	-0.76	0.18	2.02
ORP10	IPI00032971.2	Q9BXB5	-	-	-	1.47	-	-0.68	-1.40	-	0.54	-0.36
PAF53	IPI00550638.2	Q9GZS1-2	-	-	-	-	2.26	1.12	2.01	-1.57	-	-
PAGA	IPI00000874.1	Q06830	-1.29	-0.09	-0.26	-0.47	0.29	0.20	0.72	-	-0.47	0.53
PCL1	IPI00973445.1	B4DI51	-	-	-	-	-	3.87	4.91	-	-	-
PCL3	IPI00157837.5	Q5T6S3-1	-2.47	1.26	-1.12	-	-	0.80	1.97	0.32	-	-
PCNA	IPI00021700.3	P12004	-0.13	0.25	1.59	-1.28	0.21	0.49	0.39	0.05	0.39	1.00
PDE3A	IPI00291205.4	Q14432	-	-	-	1.58	-0.12	-	-2.15	-	0.16	-0.40
PGDH3	IPI00011200.5	O43175	1.32	-0.16	-0.05	-0.13	-0.06	0.03	-0.59	-0.14	-0.46	0.09
PHF23	IPI00063434.2	Q9BUL5-1	3.05	-	-	-	-	-	-	-	-	-
PI4KA	IPI00070943.4	P42356-1	-	-	-	-	-	-	-1.66	-	-	-0.35
PIR51	IPI00093253.3	Q96B01-1	-	-	-	-	-	-	-	2.59	-	0.10
PKN2	IPI00002804.1	Q16513	-	-	-	1.76	-	-	-1.88	-	0.30	-0.43
PKNOX1	IPI00220929.4	P55347-1	-	-	-	-	-	-	-	-	-	1.56
PLEKHG4	IPI00025442.2	Q58EX7-1	-	-	-	-	-	-	-1.63	-	-	-0.78
PMS2	IPI00746337.2	P54278-1	-	-	-	1.06	0.22	-0.63	-1.26	-	-0.15	0.28
PNO1	IPI00024524.4	Q9NRX1	-	0.21	-	1.72	-	-0.68	-2.04	-	0.48	-0.47
POLD	IPI00655631.2	E7EVW0	-	-	-	0.81	-0.42	-0.45	-1.50	-	0.05	-0.40
POLM	IPI00002325.1	Q9NP87	-	-	-	-	1.74	-	0.73	-	-	-
POLR1C	IPI00005179.1	O15160-1	-	-	0.08	-0.95	2.30	1.05	0.86	-0.68	-0.20	0.46
POLR2	IPI00031627.4	P24928	0.67	-0.09	-	-1.52	1.59	0.82	0.45	3.73	-	-
POLR2B	IPI00027808.1	P30876	1.57	0.25	-	-0.65	0.25	0.17	-0.32	4.07	0.21	-0.19
POLR2D	IPI00007283.1	O15514	0.49	-	-	-	-	-	-	3.21	-	-
POLR2E	IPI00291093.3	P19388	0.58	-0.06	0.13	-1.64	2.37	1.11	0.72	2.05	-0.02	0.81
POLR2G	IPI00218895.6	P62487	0.72	-0.25	-	-1.15	-	0.86	-	3.89	-0.02	0.74
POLR2J1	IPI00003310.2	Q9H1A6	0.66	-0.38	-	-0.49	-	-	-	3.35	-	-
POLR2L	IPI00003311.1	P62875	-	-	-	-0.91	1.88	0.86	0.39	2.17	-0.10	0.41
POLR3A	IPI01015730.1	F5H7E9	-	-	-	-1.46	2.57	1.12	1.33	0.22	-0.10	-
POLR3B	IPI00301346.3	Q9NW08	-	-	-	-1.10	2.49	1.20	0.54	-	-	0.67
POLR3C	IPI00007948.2	Q9BUI4	-	-	-	-1.28	2.79	0.94	1.13	-0.16	-0.06	0.60
POLR3F	IPI00100245.1	Q9H1D9	-	-	-	-2.03	2.54	1.41	0.88	-	-0.45	1.19
POLR3G	IPI00968202.1	O15318	-	-	-	-1.14	2.85	1.23	0.51	-	-0.38	-
PPP2CA	IPI00008380.1	P67775	0.78	-	0.09	-0.27	0.69	0.04	-0.24	4.59	-0.12	-0.11
PPP2CB	IPI00429689.3	P62714	-	-	-	-	-	-	-	3.80	-	-
PPP2R1A	IPI00554737.3	P30153	0.24	0.31	0.03	0.21	0.34	-0.26	-0.48	3.95	-0.39	-0.04
PPP2R1B	IPI00335449.3	P30154-2	-	-	-	0.31	-0.08	-0.26	-0.26	3.57	-0.39	-0.15
PPT1	IPI00514424.4	E9PES1	-	-	-	1.55	-	-0.62	-1.91	-	0.38	-0.39
PRO2134	IPI00018110.1	Q15544	1.93	-	-	-	-	-	-	-	-	-
PSEC0006	IPI00328985.1	Q86W42-1	0.63	0.09	-0.14	0.33	-0.55	-0.22	0.43	0.93	-1.20	-1.49
PSMB2	IPI00028006.1	P49721	-	-	0.67	0.99	-	-0.55	-1.52	0.09	-0.03	-0.33

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG.H3K9me3
PSME3	IPI00219445.1	P61289-2	-0.47	0.65	0.53	1.50	-0.51	-0.66	-1.59	-0.18	-0.08	-0.51
PUR1	IPI00023591.1	Q00577	0.74	-0.94	-	-0.79	1.98	1.21	0.65	0.49	-0.51	2.02
PURB	IPI00045051.3	Q96QR8	0.85	-0.40	-1.49	-1.26	2.34	0.80	0.47	0.64	-0.39	1.96
PYCR1	IPI00550882.3	B4DMU0	-	-	-	1.59	-	-	-1.81	-	0.07	-0.46
RAD18	IPI00024579.2	Q9NS91	0.13	-	-0.40	0.16	-	0.26	-0.02	2.09	-0.34	0.18
RAD54B	IPI00018281.1	Q9Y620	-0.26	0.74	-	0.80	-0.18	-1.12	-1.76	-	0.80	-0.51
RAP80	IPI00384342.3	F8VQY2	-	-	-	-	0.91	-	-	3.49	-	0.56
RB1	IPI00302829.5	P06400	0.30	0.02	-1.54	-0.52	1.85	0.27	0.25	0.09	0.30	0.85
RBAP46	IPI00395865.4	Q16576	-0.40	-0.06	1.84	-0.68	1.07	0.78	1.92	0.62	0.50	2.19
RBAP48	IPI00328319.8	Q09028-1	-0.43	0.42	1.46	-0.65	0.80	0.75	1.87	0.36	0.35	1.78
RBBP5	IPI00478230.2	Q15291-1	-0.07	0.10	-2.37	-0.46	1.04	0.57	0.56	0.10	0.10	0.18
RCD1	IPI00023101.1	Q92600	-0.22	0.50	0.16	1.79	-0.62	-1.04	-1.78	0.29	0.04	-0.61
RECQ1	IPI00178431.12	P46063	-	-	-	-0.88	1.50	0.32	0.43	-	0.44	1.29
RECQL5	IPI00185769.3	A5YM55	-	-	-	-	-	-	-	3.57	-	-
REPA1	IPI00020127.1	P27694	0.64	-0.11	-1.07	-0.78	1.06	1.17	0.65	-0.02	-0.14	1.05
REPA2	IPI00646500.1	P15927-3	0.49	-0.01	-0.68	-0.25	1.23	1.19	0.17	-0.13	-0.27	1.01
RFX5	IPI00009303.1	P48382	1.25	0.74	0.92	-0.10	1.56	0.66	0.40	-0.25	0.64	1.02
RFXAP	IPI00010890.1	O00287	1.12	0.85	0.84	-0.70	1.26	0.65	-0.20	0.19	0.63	1.25
RING1	IPI00641330.1	Q06587-1	0.48	0.31	-1.07	-0.41	1.16	0.93	1.84	0.29	-0.11	0.44
RNF114	IPI00032955.1	Q9Y508-1	1.66	-0.03	0.21	-	-	-	-	-	-	-
RNF168	IPI00217899.1	Q8IYW5	0.06	-	-	-	-	-	-	3.06	-	-
RNF95	IPI00010948.2	Q12899	-	3.01	-	-	-	-	-	-	-	-
RP11-101E13.2-006	IPI00642374.1	Q5LJA9	0.39	0.11	0.23	-1.27	1.14	0.91	0.45	1.27	-0.54	1.21
RP11-114F7.3	IPI01011375.1	Q5SVK8	-	-	-	1.34	-	-0.93	-1.68	-	0.39	-0.48
RPL34	IPI00219160.3	P49207	-	-	-	0.64	-	-	-	-	0.20	-1.37
RSBN1	IPI00019999.2	Q5VWQ0-1	-	-	-	-	-	-	-	4.09	-	-
RSBN1L	IPI00925255.1	Q6PCB5-1	-	-	-	-	-	-	-	4.38	-	-
RTTN	IPI00645947.4	Q86VV8-1	-	-	-	1.14	-0.42	-0.97	-1.56	-	-0.07	-0.51
SAMD1	IPI00395474.1	Q6SPF0	-0.16	0.26	-2.19	-0.59	-0.18	0.38	0.31	-0.33	0.04	-0.09
SAP130	IPI00002220.4	Q9H0E3-2	1.60	0.38	0.14	-	-	0.92	0.45	0.96	0.59	-
SAP30	IPI00022019.1	O75446	1.22	0.33	0.04	-0.25	0.75	0.90	0.34	0.98	0.10	1.30
SAP45	IPI00607645.5	Q9H7L9	1.82	0.27	-0.04	-1.16	0.83	0.88	0.83	0.72	0.17	1.38
SAP49	IPI00017339.1	Q15427	0.81	-	-0.30	-0.42	-	0.44	1.60	0.08	-0.50	0.27
SCML2	IPI00328688.3	Q9UQR0-1	-3.22	-0.10	-1.11	1.54	-	-0.18	1.07	1.62	0.93	2.40
SEP1gene	IPI00657805.1	Q8IZH2-1	-	-	-	1.26	-0.08	-0.25	-1.26	-	0.63	-0.38
SIN3A	IPI00170596.1	Q96ST3	1.61	0.34	-0.07	-1.12	1.00	0.77	1.17	1.17	0.43	1.24
SIX4	IPI00006525.3	Q9UIU6	0.64	0.06	1.39	-0.65	1.29	0.74	0.27	0.67	-0.54	1.80
SP3	IPI00025807.2	Q02447-1	-	-	-	-1.51	-	-	-	-	-	-
SPK	IPI00023344.2	Q92797-1	1.52	0.32	-0.03	0.61	-0.39	-0.73	-0.88	1.33	-0.27	-0.56
SPT4H	IPI00002895.1	P63272	-	-	-	0.28	-	-	-	6.30	-	-
SPT5	IPI00298058.1	O00267-1	-	-	-	-	-	0.39	-	4.41	-	-0.04
SRBD1	IPI00171087.7	Q8N5C6-1	-0.02	-0.10	0.46	-0.57	2.60	-0.24	0.35	-0.57	0.64	1.96
STX5	IPI00744834.2	Q13190-1	-	-	-	1.82	-0.88	-1.37	-1.74	-	0.48	-0.49
TADA1	IPI00060046.1	Q96BN2	1.22	-0.05	-	0.44	-	-	-0.35	-	-0.03	-0.05
TAF10	IPI00030364.1	Q12962	1.63	0.25	-	-0.62	-	1.47	0.06	-	-	0.89
TAF12	IPI00002806.1	Q16514-1	1.85	-	-	-1.22	-	-	-	-	-	-
TAF1A	IPI00024263.1	Q15573-1	-	-	-	-	-	1.45	-	-	-	-
TAF1B	IPI00291416.6	Q53T94-1	-	-	-	-	-	1.47	0.74	-	-	-
TAF1C	IPI00246842.5	Q15572-1	-	-	-	-	-	1.90	1.32	0.61	-	-
TAF2C	IPI00413755.1	O00268	2.60	-	-	-	-	0.99	0.33	-	-	1.46
TAF2D	IPI00298925.2	Q15542-1	2.36	-0.06	-	-1.77	2.43	1.02	0.65	0.28	0.90	2.07
TAF2E	IPI00944951.1	B4DT11	2.27	-0.10	-	-2.57	1.47	1.23	0.52	0.13	0.84	1.13
TAF2F	IPI00018111.1	Q15545	3.03	-	-	-	-	-	-	-	-	-
TAF2G	IPI00002993.1	Q16594	1.78	0.15	0.03	0.09	0.51	0.37	-0.62	0.18	0.23	0.07
TAF3	IPI00853240.1	Q5VWG9	2.72	-	-	-	-	-	-	-	-	-
TAF8	IPI00065313.2	Q7Z7C8-2	1.26	-0.10	-0.09	-1.05	0.49	0.65	0.25	0.20	0.21	-
TAF9B	IPI00642105.1	Q9HBM6	1.78	0.12	-	0.28	-	0.23	-	-	0.18	0.02
TARDBP	IPI00025815.2	Q13148-2	1.12	-0.32	0.05	0.81	-0.60	-0.18	-0.03	0.16	-0.45	0.45
TBC1D15	IPI00793999.2	E9PH93	-	-	-	1.11	-0.52	-0.30	-1.64	-	-0.33	-0.12
TCF13	IPI00002901.5	P28347	-0.10	-0.19	0.23	-2.41	2.17	1.06	2.04	-0.22	-0.64	1.53
TCF4	IPI00221009.1	Q9NQB0-7	-	-	-	-1.78	1.34	0.93	0.46	-0.65	-	-
TCF6	IPI00020928.1	Q00059	-0.58	0.28	-0.13	-0.81	1.98	0.16	0.32	-0.46	0.53	1.06
TCHH	IPI00015869.2	Q07283	-	-	-	-	-	-	-	-	-	2.04
TCOF1	IPI00815944.2	Q13428-3	1.05	1.11	-1.11	-	-	-	-	-	-	-
TEM6	IPI00658152.1	Q68CZ2-1	-	-	-	0.66	-	-0.44	-1.34	-	0.27	-0.41
TGIF	IPI00297138.2	Q15583-1	1.57	0.91	1.77	-	-	-	-	0.79	-	1.53
THOC3	IPI00063729.4	Q96J01	0.58	-	0.41	0.67	-	-0.15	-0.08	-	-0.97	-1.07
TIGD1	IPI00043467.4	Q96MW7	-	-	-	-1.85	-	-	-	-	-	1.96
TIM13B	IPI00001589.1	Q9Y5L4	0.50	-0.02	-2.14	-0.37	0.25	0.43	-0.02	-	0.14	-0.05
TMCO4	IPI00299169.4	Q5TGY1-1	-	-	-	0.89	-0.79	-0.69	-1.82	0.40	0.02	-0.36

Gene Names	IPI	Uniprot	H3K4me3	H3K9me3	meCpG	H3K9me2	H3K27me1	H3K27me2	H3K27me3	H2Bub1	H3Δ1-20	meCpG. H3K9me3
TOP3	IPI00013378.1	Q13472-1	1.61	0.31	-0.32	-0.47	1.48	0.60	0.25	-0.20	0.00	0.89
TRIO	IPI00657953.2	O75962-1	-	-	-	1.73	-	-0.61	-1.32	-	-	0.00
TRIP12	IPI00783250.1	Q14CF1	-0.26	-1.06	-0.42	-0.19	1.93	0.34	0.30	-1.87	-0.07	0.67
TRIP12	IPI00032342.4	D4HL82	-0.23	-0.78	-0.28	0.15	2.00	0.25	0.29	-1.63	-0.04	0.62
TTF2	IPI00290812.2	Q9UNY4-1	0.39	-0.32	0.54	0.07	1.32	0.04	-0.37	-0.09	0.34	-0.16
UBCH5C	IPI00749013.2	P61077-2	-	-	-	-	-	-	-	-	-	3.11
UFD1L	IPI00873326.3	C9JS35	-	-	-	1.14	-	-	-	-	-0.45	-0.12
UHRF1	IPI00797279.1	A8K024	-2.36	2.54	1.77	3.17	1.58	-0.32	2.66	0.02	2.75	3.98
UNQ9342/PRO34047	IPI00152695.2	Q6UXN9	1.24	-0.01	-0.06	-0.38	-0.55	-0.11	-1.18	2.08	-0.38	-0.43
USP3	IPI00002330.2	Q9Y6I4	-	-	-	0.10	-	-1.81	-	-	-	1.55
UTP15	IPI00152708.4	Q8TED0	-0.09	0.61	-2.78	-0.04	0.86	0.33	0.59	0.60	-0.32	0.09
WDR18	IPI00032533.3	Q9BV38	0.81	0.20	0.04	0.17	-0.35	0.03	0.48	0.40	-0.13	-1.13
WDR3	IPI00009471.1	Q9UNX4	-	-	-	-	0.01	-	-1.66	-	-	-0.46
WDR36	IPI00169325.1	Q8NI36	-	-	-	0.84	-0.34	-0.36	-1.86	-	0.10	-0.14
WDR75	IPI00217240.1	Q8IWA0	-0.53	0.65	-2.61	-	0.91	0.34	0.49	1.94	-0.47	0.19
WEE1	IPI00025830.1	P30291	-	-	-	-2.04	0.41	0.33	-0.84	0.60	-	-0.16
XRCC1	IPI00002564.4	P18887	-0.09	0.03	-0.28	-1.57	0.38	1.10	0.72	0.12	-0.16	-0.47
ZBP89	IPI00010833.1	Q9UQR1	0.81	-	-	-2.14	0.79	0.41	-0.05	0.39	-0.61	1.19
ZBTB14	IPI00307325.2	O43829	0.52	0.56	-	-1.57	2.14	0.69	1.12	-0.39	0.68	0.19
ZBTB15	IPI00026277.8	O15156	-0.38	-0.23	0.41	-0.82	0.96	0.53	0.54	0.37	0.10	1.60
ZBTB9	IPI00060141.1	Q96C00	-	-	-	0.22	1.77	-	-	-	-	-
ZFP276	IPI00847391.2	Q8N554-1	-0.59	-	-1.04	-	2.28	0.38	0.41	-	-0.15	-
ZFP38	IPI00001344.4	Q9Y5A6	0.38	0.10	-2.12	-	1.64	0.58	0.79	-0.54	-0.59	-
ZFP64	IPI00018906.6	Q9NPA5-1	0.02	-	-	-	1.73	0.88	0.74	-0.15	-0.01	-
ZHX1	IPI00293086.2	Q9UKY1	0.41	0.14	3.96	-	0.74	0.40	0.96	1.45	0.45	3.73
ZKSCAN4	IPI00005191.2	Q969J2	-0.07	-0.29	0.28	-1.45	2.64	0.20	1.07	-0.61	0.22	1.76
ZMYM3	IPI00641109.1	A6NHB5	-	2.53	-	0.95	-	0.28	0.72	0.60	-0.17	1.91
ZNF131	IPI00658062.1	P52739-1	0.22	0.13	-3.99	-1.38	0.76	0.93	0.51	0.34	0.76	-0.02
ZNF580	IPI00478837.1	Q9UK33	-0.36	1.47	-	-	-	-	-	0.20	-	2.74
ZNF691	IPI00301583.3	Q5VV52-1	-0.05	0.16	0.30	-1.51	1.26	0.16	0.57	-0.17	-0.28	1.41
ZNF787	IPI00056880.7	Q6DD87	-	-	3.06	-	-	-	-	-	-	2.58

Curriculum Vitae

Personal details

Name: Miroslav Nikolov

Date of birth: 1984

Place of birth: Pleven, Bulgaria

Nationality: Bulgarian

Education

04.2009 – 09.2012   IMPRS Molecular Biology, Göttingen

                     PhD thesis in Bioanalytical Mass Spectrometry Group,

                     Max Planck Institute for Biophysical Chemistry, Göttingen

                     "Systematic Analysis of the Interactome of Modified Chromatin"

10.2009   Master of Science in Molecular Biology

10.2008 – 03.2009   IMPRS Molecular Biology, Göttingen

                     MSc thesis (integrated) in Bioanalytical Mass Spectrometry Group,

                     Max Planck Institute for Biophysical Chemistry, Göttingen

                     "Analysis of HP1-Chromatin Interactions by Mass Spectrometry"

10.2007 – 09.2008   IMPRS Molecular Biology, Göttingen

                     First segment of theoretical and practical MSc studies

08.2007   Bachelor of Science in Molecular Biology

09.2003 – 06.2007   BSc studies in Molecular Biology at Sofia University, Bulgaria

09.1997 – 05.2003   High School of Mathematics "Geo Milev", Pleven, Bulgaria

Awards and scholarships

10.2008 – 09.2012   IMPRS doctoral stipend

09.2007 – 08.2008   Stipend of the Excellence Foundation for the Promotion of the MPS

03.2004 – 06.2007   Scholarship for academic excellence, Sofia University, Sofia, Bulgaria

10.2003 – 06.2007   Scholarship from Evrkia Foundation, Sofia, Bulgaria

07.2003   Bronze medal from the 13th International Biology Olympiad

Publications

Jaspers MH, Nolde K, Behr M, Joo S, Plessmann U, **Nikolov M**, Urlaub H, Schuh R. "The Claudin Megatrachea Protein Complex." *J Biol Chem.* 2012 Aug 28. (Epub ahead of print)

Nikolov M, Schmidt C, Urlaub H; "Quantitative Mass Spectrometry-based proteomics: an overview" *Methods Mol Biol.* 2012; 893:85-100

Alkhaja AK, Jans DC, **Nikolov M**, Vukotic M, Lytovchenko O, Ludewig F, Schliebs W, Riedel D, Urlaub H, Jakobs S, Deckers M. "MINOS1 is a conserved component of mitofilin complexes and required for mitochondrial function and cristae organization." *Mol Biol Cell.* 2012 Jan; 23(2): 247-57. Epub 2011 Nov 23.

Nikolov M, Stützer A, Mosch K, Krasauskas A, Soeroes S, Stark H, Urlaub H, Fischle W. "Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns." *Mol Cell Proteomics.* 2011 Nov; 10(11): M110.005371. Epub 2011 Aug 11.

Ivanov I, Tsacheva I, Stoyanova V, **Nikolov M**, Tchorbadjieva MI, Petrova S, Christov L, Georgieva V, Georgiev G. "Chaperone-like effect of polyzwitterions on the interaction of C1q with IgG." *Z Naturforsch C.* 2009 Jan-Feb; 64(1-2):149-54.

Shema E*****, **Nikolov M***, Haj-Yahya M, Allemand E, Yamaguchi Y, Muchardt C, Urlaub H, Brik A, Oren M, Fischle W. "Systematic identification of proteins binding to chromatin-embedded ubiquitylated H2B reveals recruitment of the SWI/SNF complex to regulate transcription." under review in *Mol Cell*, *****equal contribution

Nikolov M, Fischle W. "Systematic analysis of histone modification readout." under review in *Mol Biosyst*, invited publication

Tirard M, Hsiao HH, **Nikolov M**, Urlaub H, Melchior F, Brose N. "In vivo identification of SUMOylated proteins in the brain of His$_6$-HA-SUMO1 knock-in mice." under review in *Proc Natl Acad Sci USA*