Quantitative Proteomic Analysis of Chromatin Associated Protein Complexes

Hans Christian Eberl

aus
Eggenfelden

2012

Summary

Nucleosomes are the basic building blocks of chromatin. They consist of an octamer of histone proteins, around which the DNA is wrapped. Rather than only packaging and compacting the DNA, they also play an active role in the regulation of many processes such as transcription, DNA repair or development. The histone tails, which protrude out of the nucleosome as unstructured polypeptide chains, are subject to a variety of post-translational modifications. These modifications are believed to form a "histone code" that extends the information contained in the genetic code. Histone modifications can recruit so-called chromatin readers, which in turn modify nearby chromatin or influence chromatin-asssociated processes. Despite tremendous research efforts during the last years, the precise function of most histone modifications remains unclear. A first step towards understanding the molecular mechanisms of the histone code is to elucidate the repertoire of chromatin readers. Mass spectrometry-based proteomics offers a uniquely suited tool to uncover chromatin readers and their associated interaction partners.

The aim of this thesis was to develop workflows to study protein-protein interactions of chromatin associated complexes by quantitative mass spectrometry. These technologies are applied to discover novel chromatin readers and their associated complexes.

In the first project, a SILAC-based quantitative proteomics screen using peptide pull-downs was performed in HeLa cells to discover readers for the major lysine trimethylation marks on histone H3 and H4. The study analyzed the activating H3K4me3 and H3K36me3 as well as the repressive H3K9me3, H3K27me3 and H4K20me3 marks. Many known chromatin readers and associated proteins could be retrieved, as well as several novel putative readers. The SAGA complex was shown to be recruited to the H3K4me3 mark via the double tudor domain of its subunit SGF29. The PWWP domain of the H3K36me3 associated protein NPAC was demonstrated to be necessary for chromatin binding. GFP pull-downs using stable cell lines generated by BAC TransgenOmics allowed the assignment of putative readers into protein complexes. Genome wide profiling of histone modifications and their readers show a good overlap, which verified the peptide pull-down approach *in vivo*. H3K4me3 readers, which are found on the promoters of actively transcribed genes, could be clustered into distinct sub-

groups.

In the second project, a high-throughput label-free interaction pipeline was established, enabling chromatin reader interaction screens from unlabeled protein extracts. A proof-of-concept study applied this technology to screen for readers of the activating H3-K4me3 and the repressive H3K9me3 mark from four different mouse tissues – brain, liver, kidney and testis. This screen generated the currently most comprehensive list of chromatin readers for these marks. Screening from different tissue extracts provided the unique opportunity to discover chromatin readers, which are only present in very specialized cell types and are thus not accessible in standard cell line-based assays. CHD5 is a brain specific NuRD complex subunit, which replaces CHD3/CHD4 and directly binds to the H3 tail via its two PHD finger domains. The largest number of tissue-specific chromatin readers was found in testis, most likely due to its specialized chromatin. Known testis-specific readers like MBD3L and DNMT3A and putative novel readers like SSTY1 and SSTY2 were retrieved.

The replacement of canonical histones with histone variants is an alternative possibility to index chromatin. In a collaborative project we investigated a novel histone variant splice isoform. Histone variants are incorporated into chromatin in a highly controlled fashion by histone chaperones. A novel splice variant for H2A.Z with a distinct C-terminus was discovered and termed H2A.Z.2.2. Quantitative proteomics was applied to investigate which proteins associate with H2AZ.2.2 and H2A.Z.2.1 outside of chromatin. Both splice variants were found to interact with the TIP60 and SRCAP histone chaperone complexes.

In summary, generic workflows were established to screen for protein-protein interactions of chromatin associated protein complexes and to discover chromatin readers from SILAC labeled as well as unlabeled protein extracts. These technologies were succesfully applied to uncover novel chromatin associated complexes and to describe general and tissue-specific chromatin readers.

Contents

1 Introduction

1.1 Chromatin biology

The genetic information of an organism is encoded in the DNA i a one dimensional sequence of the four bases adenine, guanine, cytosine and thymine. In eukaryots, DNA is stored in the nucleus in a highly ordered manner in the form of chromatin. The term chromatin describes the structure formed by DNA, the small basic histone proteins and other associated proteins. As every human cell contains about 2 m of DNA which has to be packaged into a nucleus with a diameter of around 6 μm, one major function of chromatin is to compact and store the DNA. Nucleosomes, the basic building blocks of chromatin are formed by wrapping the DNA around an octamer of core histone proteins. A three dimensional arrangement of nucleosomes can generate a higher order structure to achieve further compaction. Histones, however, are not only compacting DNA, but they are a carrier of information themselves. An additional layer of information is added to the underlying genetic code by modifying histones or replacing the canonical histones with histone variants.

In the 1930s, Emil Heitz, who studied mitotic chromosomes in moss, described two different chromatin compaction states. Certain parts of chromosomes remained condensed throughout interphase which he termed heterochromatin. He further suggested the term euchromatin for those parts of chromosomes which become invisible during late telophase [82, 178]. The macroscopically observable chromatin compaction states were later associated with inactive and active transcription. The human genome consists of around 3.2 billion base pairs, of which only a small fraction encodes genes. Gene rich regions, in which active transcription takes place, are euchromatic. Euchromatin adopts an open state to allow access for the transcriptional machinery. Because of its reduced density these regions are not observable by light microscopy. Conversely, heterochromatin is highly compacted and generally not accessible to the transcriptional machinery. Heterochromatic regions span large parts of the genome which are either not protein coding or code for genes whose expression is not beneficial or even detrimental for the current status of the cell. Heterochromatin, for instance, can be found on the inactive X chromosome, on telomeres or centromers.

Histones, the most abundant proteins in chromatin, can be found in all eukaryots and in archea [235], but not in prokaryots. The formation of a chromatin structure could be a way for organisms with a more complex genome to organize and structure their genetic information. It is beyond debate that processes on chromatin play crucial roles in regulating almost all aspects of the life of a cell. The importance of chromatin can also be seen by the large number of diseases and developmental defects that are associated with malfunctioning chromatin processes [34, 58, 182]. Despite enormous research efforts over the last years, understanding of many mechanisms in chromatin biology is still lacking. Moreover, for many processes all of the players involved are still not known.

1.1.1 Epigenetics

The term epigenetics was first used by Conrad Waddington in the context of developmental biology. To him, epigenetics was "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" [255].

He generated the metaphor of the "epigenetic landscape" (Figure 1.1.1), in which a cell, represented by a ball, roles down a landscape, and can take several decision which lead to specific cell fates. With the growing knowledge in chromatin biology, the term epigenetics received a more chromatin centric meaning. Currently, epigenetics is defined as the "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" [15]. However, as the fact of heredity poses a source of discussion, Adrian Bird suggested to define epigenetics as "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" [19]. Regardless of

Figure 1.1.1: The Epigenetic landscape after Waddington Phenotypic decision are illustrated as a ball rolling down a complex landscape (from [69]).

whether heredity is a essential feature of epigenetics, it describes processes on chromatin which influence gene expression, phenotype and cell fate decisions without changing the underlying DNA sequence.

Most epigenetic phenomena can be attributed to three major mechanisms: DNA methylation, histone modifications and regulation by non-coding RNAs. The trans-
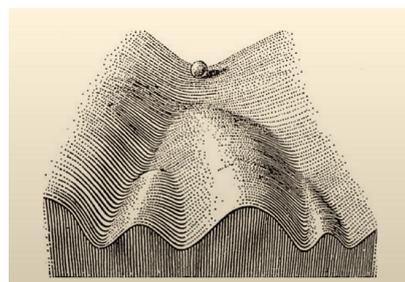
fer of a methyl group from S-adenosyl methionine to the 5' position of the pyrimidine ring of cytosine is referred to as DNA methylation. Most DNA methylation in mammals occurs on CpG dinucleotides; a high density of CpG dinucleotides is referred to as a CpG island. Methylation of these islands correlates with transcriptional repression [71]. Histone modifications are far more complex than DNA methylation, as a wide variety of modifications can occur on histones. Moreover, histone modifications are associated with gene control, but also other chromatin based processes like DNA repair or mitosis (see Section 1.1.3). The importance of non-coding RNAs for epigenetic and chromatin related processes only emerged recently [17]. Non-coding RNAs play a crucial role in controlling processes as diverse as X chromosome inactivation or silencing of repetitive DNA sequences by acting in concert with the cellular chromatin modification and DNA methylation machinery.

A striking example of a heritable epigenetic phenomenon is an epimutation observed in the flower *Linaria vulgaris*. Already more than 250 years ago, a mutant with a characteristic symmetric (peloric) flower instead of the asymmetric wild type (Figure 1.1.2) was described by Linnaeus. This phenotype is not based on a DNA mutation, but on DNA methylation and thus repression of the LCYC gene which controls flower asymmetry [43]. DNA methylation is transmitted through the germ line and propagated similar to a DNA sequence mutation. X chromosome inactivation in mammals is another

Figure 1.1.2: Epimutation in *Linaria vulgaris* Wild type flower is dorsoventrally asymmetrical, whereas the mutant peloric flower is radially symmetrical with all petals resembling the ventral petal of the wildtyp flower (from [43]).

example of a complex developmental process that is regulated by epigenetics. Female cells have to inactivate one of their two X chromosomes to achieve the correct gene dosage [139]. Early during development, one of the two X chromosomes is chosen randomly and silenced. Once one of the X chromosomes is silenced, this state is stably propagated over all following cell divisions. X chromosome silencing is a complex process which involves the combination of various repressive epigenetic mechanisms [36]. The long non-coding RNA XIST (X inactive specific transcript) is exclusively expressed from the inactive X chromosome [21] and plays a crucial role in silencing [264]. Furthermore, methylation of CpG islands [83] and modification of histones, e.g. trimethylation of lysine 27 on histone H3 contribute to X chromosome silencing [190].

1.1.2 The structure of chromatin

Nucleosomes, which form the basic buildings blocks of chromatin, consist of two copies of each of the core histones H2A, H2B, H3 and H4. Isolated histones in solution form heterodimers of H2A and H2B, as well as H3 and H4. 146 base pairs of DNA are wrapped in 1.65 turns around the nucleosome and specific interactions can be observed between the outer surface of the nucleosome and the DNA bases. Seminal work was performed by Luger et al., who described, for the first time, a high resolution crystal structure of the nucleosome (Figure 1.1.3 A) [137]. In the structure, a globular histone octamer directly contacts the DNA and forces it into a bent conformation. The histone tails, especially the amino-termini of H3 and H2B are protruding out of the nucleosome. They could not be resolved in the crystal structure and are believed to form unstructured extensions. The individual nucleosomes are connected by a short stretch of interconnecting DNA. In electron micrographs, a "beads on a string" structure of regularly interspaced nucleosomes can be observed (Figure 1.1.3 B). This fiber has a diameter of 11 nm and represents the first level of chromatin organization.
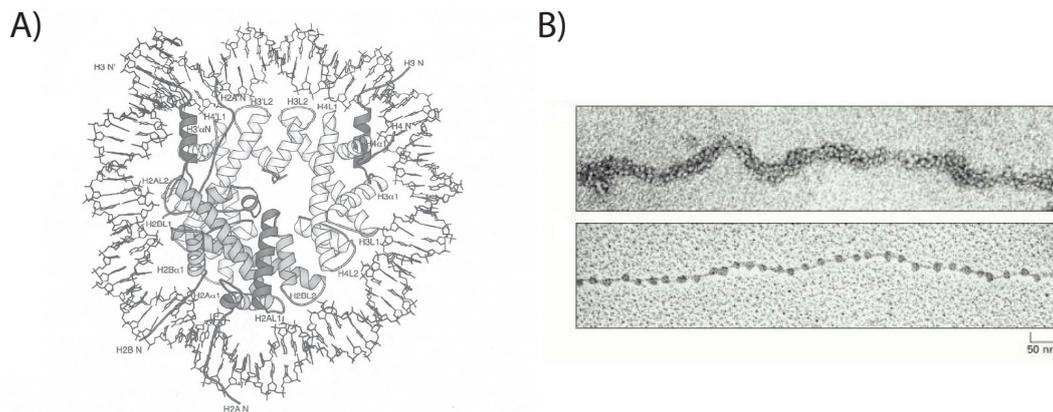
Figure 1.1.3: The nucleosome is the basic building block of chromatin A)The crystal structure of the nucleosome (from [138]). B) Different chromatin compaction states observed by electron microscopy. Upper panel: 30 nm thick fiber from interphase chromatin. Lower panel: Preparation of loose chromatin shows nucleosomes, which are spaced on the DNA like "beads on a string" (from [3]).

Binding of the linker histone H1 generates a more condensed fiber of 30 nm diameter, which is considered the next level of structural organization of chromatin [200]. However, due to different preparation methods and analysis techniques, the exact structure of this fiber *in vivo* [73] and even its existence [141] are still a matter of intense debate. Two basic structural models are suggested for the 30 nm fiber based on EM studies:

In a one-start-helix (solenoid), individual nucleosomes form a superhelix with about six to eight nucleosomes per turn [261]. The linker DNA is bent to follow the helical path and each nucleosome interacts with its fifth or sixth neighboring nucleosome. The two-start-helix forms a zigzag structure in which the linker DNA is straight and each nucleosome interacts with the second neighbour nucleosome [50]. Higher order structures are thought to be generated by fiber-fiber interactions and further loop formation [129].

1.1.3 Histone modifications and the histone code hypothesis

A characteristic feature of histones is their high content of the basic amino acids arginine and lysine. These amino acids not only serve to provide a positive charge that facilitates interaction with the negatively charged DNA, but are also subject to a variety of post-translational modifications (PTMs). The highest diversity can be observed on lysines, which can be mono- (me1), di- (me2) and trimethylated (me3), acetylated (ac) as well as ubiquitinated and sumoylated. Recently, lysine crotonylation was described as a novel histone modification [236]. Arginines can be mono- (me1) and dimethylated. If both methyl groups are added onto the same amino group it is called asymmetric dimethylation (Rme2a), if both amino groups of the guanidinium are monomethylated, it is called symmetric dimethylation (Rme2s) (Figure 1.1.4). Moreover, the classical signal transduction modification phosphorylation can be observed on serines (Sph), threonines (Tph) and tyrosines (Yph). Histone modifications can be associated with a wide array of cellular processes by influencing transcriptional rates and chromatin structure [115].

Modifications on the core histone fold domains mostly influence the biochemical and biophysical properties of the nucleosome. PTMs can change the chemical properties of amino acids by either affecting the charge of the side chain or by introducing a bulky group (Figure 1.1.4 A). This can disrupt the contacts between adjacent nucleosomes, or between the modified histone and the DNA. The basic charge of lysines is neutralized by acetylation, which can lead to a less compacted chromatin structure. For example, acetylation of lysine 16 on histone H4 (H4K16ac) interferes with formation of higher order chromatin structure and also prevents chromatin remodeling by the chromatin remodeler ACF [223]. Modifications on the unstructured histone tails mostly serve as recruiting platforms for proteins. These proteins, which specifically recognize histone modifications in an amino acid sequence context, are generally referred to as "chromatin readers". The recognition of the modified amino acid is accomplished by specialized protein domains, which are capable of distinguishing modified
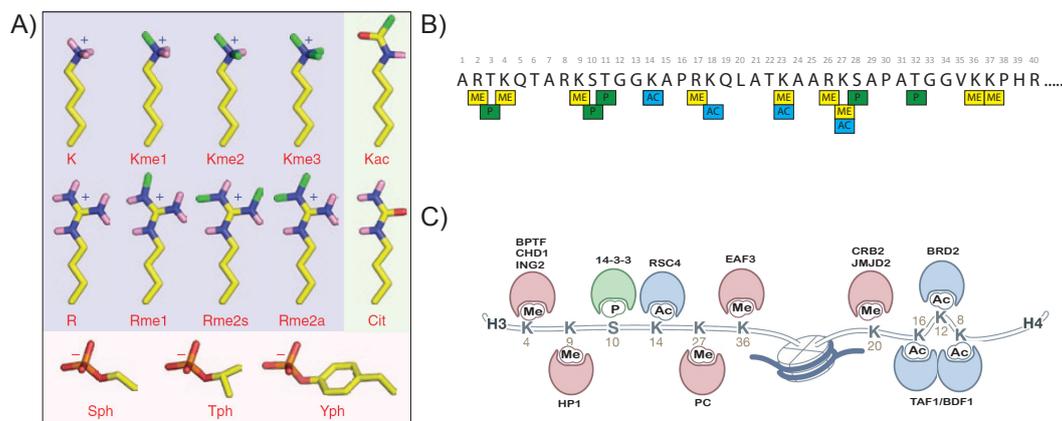
Figure 1.1.4: Histones are subject to a variety of post-translational modifications A) Chemical structure of the major histone modifications. Pink: hydrogen, green: methyl group, red: oxygen, blue: nitrogen (from [237]). B) The N-terminal tail of histone H3 and its modifications. C) Histone modifications on the unstructured tails recruit proteins which contain specific binding domains (modified from [115]).

and unmodified histone tails. Methylated lysines, for example, can be recognized by chromodomains [9, 122], PHD fingers [130, 181], double tudor domains [87], MBT domains [243], PWWP domains [254] and ZW type zinc fingers [80]. Figure 1.1.5 shows the PWWP domain of BRPF1 binding to a trimethylated histone peptide. The PWWP domain recognizes the amino acid sequence surrounding the modification site (Figure 1.1.5 A). A hydrophobic pocket (Figure 1.1.5 B) allows the distinction between the unmodified and the trimethylated peptide.

Bromo domains specifically bind to acetylated lysines [47]. Phosphorylated serines can be recognized by 14-3-3 proteins [140]. All these modification dependent interactions are very specific due to highly specialized binding pockets [237]. Many proteins or protein complexes contain more than one histone modification binding module to achieve combinatorial binding. BPTF, for example, binds via its PHD finger to H3K4me2/3 and via its bromodomain to H4K16ac on the same nucleosome [205]. The CHD4 subunit of the NuRD complex interacts with both histone H3 tails on the same nucleosome, which are either unmodified or K9 trimethylated [161]. Recruitment of the general transcription factor TFIID to H3K4me3 is further augmented by acetylation of adjacent lysines (K9 and K14) [252]. In this case, the binding modules are placed on two separate complex subunits. The TAF3 PHD finger binds to H3K4me3 and the TAF1 subunit contains a tandem bromodomain module which can bind to acetylated lysines [95].

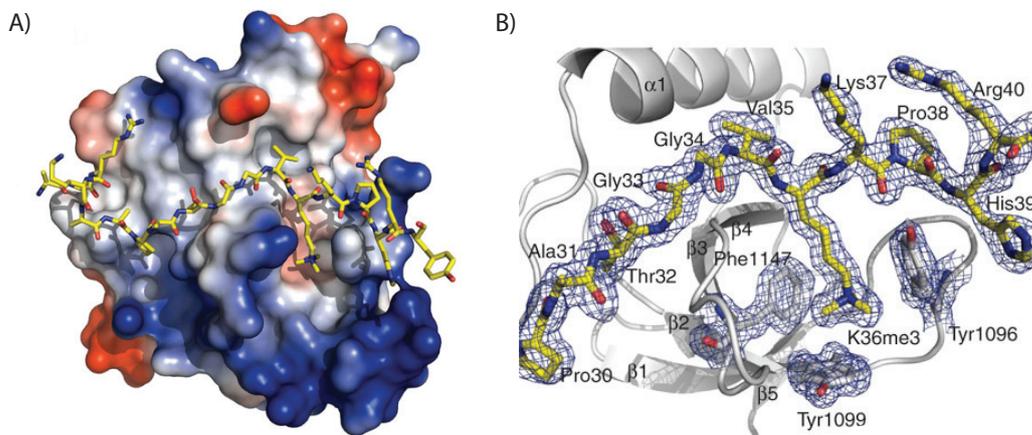**Figure 1.1.5: The PWWP domain as an example for a trimethyl lysine binding domain** A) The crystal structure of the PWWP domain of BRPF1 shows an extended interaction surface, which contacts the modified and its adjacent amino acids. B) Stick and ribbon representation of the peptide binding domain. The side chains of tyrosine and phenylalanine form a hydrophobic cage, which accommodates the trimethyl lysine moiety (from [254]).

The distribution of histone modifications can be correlated to the transcriptional state of the neighboring chromatin. Active genes, for example, are marked by trimethylation of H3K4 [208], which peaks at the transcription start site [106], while the gene body is covered by trimethylation of H3K36 [119]. Lysine acetylation is almost exclusively associated with active genes, whereas repressive chromatin is usually marked by trimethylation of H3K9 and H3K27 [115]. Whereas most modifications on their own could be linked to specific processes in gene regulation or cellular differentiation, their combinatorial patterns could only be analyzed due to improvements in sequencing technologies. Chromatin immunoprecipitation in combination with next generation sequencing was used to study the combinatorics of chromatin marks in human T cells [55]. A total of 51 distinct chromatin states have been described that possess a unique chromatin signature, for example promoter-associated, active intergenic or large-scale repressed regions. Using the DamID technology [249], the van Steensel group analyzed 53 broadly selected chromatin components (histone marks and chromatin associated proteins) in *Drosophila*. In contrast to the abovementioned studies, they describe only five principle chromatin types, each of which was assigned a color [59]. Yellow and red are both actively transcribed chromatin regions, but show a different composition of histone marks and associated proteins and also differences in the timing of DNA replication. Repressive chromatin can be grouped into three types: classical HP1 and H3K9me3 heterochromatin (green), polycomb and H3K27me3 marked heterochro-

matin (blue) and a very abundant additional type, which is marked by histone H1, D1, IAL and SUUR (black).

The correlation between histone modifications and the transcriptional state of the neighboring genes prompted David Allis and co-workers to suggest the histone code hypothesis [96, 231]. This hypothesis states that the combinations of histone modifications provide a code that can be read by the cellular machinery and regulates almost all chromatin associated processes. Proteins involved in the histone code can be separated into three groups. "Chromatin writers" modify histones, "chromatin readers" bind to modified histones and "chromatin erasers" remove histone modifications. Although this theory explains the importance of histone modifications and relates them to a direct output, it is controversially debated [84]. The major point of criticism towards a histone code is, that, according to the theory, histone modifications would dictate the transcriptional outcome. However, so far histone modifications could only be correlated to a transcriptional outcome, and experimental data showing a causal relation between a modification and an outcome are still missing. In addition, for some histone modifications there is even no causal relation between them and a specific output. For example, knockout of SET1, the only methyltransferase for H3K4 in yeast, generates strains without K4 trimethylation. However, despite the clear correlation of H3K4me3 with active transcription, these strains are viable. Surprisingly, a very specific phenotype – defects in mating type loci silencing and telomeres – was observed [168]. Furthermore, yeast lacking the amino-termini of histone H3 or H4 are viable [133]. These yeast strains are devoid of several histone modifications at once and can still accomplish transcriptional regulation.

Nucleosome occupancy is another mechanism which plays an important role in regulating chromatin associated processes. In addition to a combination of histone modifications on active genes, a clearly defined nucleosome occupancy pattern also exists [124]. As the presence of histones inhibits transcription, a nucleosome pattern – maybe defined by the underlying DNA sequence – could contain enough information to guide transcription. Furthermore, nucleosome patterns can be modified by chromatin remodelers, thus allowing regulation. Taken together, two phenomena — histone modification patterns and nucleosome occupancy — can be correlated to a transcriptional output. It remains to be discovered which mechanism is cause and which consequence, or if they are both equally important. Nevertheless, irrespective of whether histone modifications generate a code that defines the transcriptional output or only act by fine tuning gene expression, their importance is beyond debate. Loss of proteins associated with histone modifications often leads to severe developmental defects and loss or misregulation of these proteins can also frequently be observed in malignancies

[7, 18, 34, 182].

1.1.4 Lysine methylation of histone proteins

Methylation of lysine residues on histone proteins is one of the most prevalent and versatile histone modifications (Figure 1.1.4). Up to three methyl groups can be transferred to a lysine side chain. Although lysine methylation has so far only been studied in detail on histone proteins, other cellular proteins are also methylated [270]. Among the well described lysine methylation sites on histones, five of them can be found on histone tails: H3K4, H3K9, H3K27, H3K36 and H4K20 and one site in the histone core: H3K79 [144, 271]. At physiological pH, all methylated forms are believed to be cationic and trimethyl lysines carry a positive charge. Despite their charge, the addition of methyl groups increases the hydrophobicity of the side chain. Proteins specifically recognizing methylated lysines contain a hydrophobic pocket that accommodates the methylated side chain. Especially trimethyl lysine binders evolved very specialized binding domains, that not only provide a hydrophobic environment but also make contacts with the charged nitrogen atom [237].

Lysines are methylated by protein lysine methyl transferases (PKMTs). PKMTs catalyze the transfer of a methyl group from S-adenosyl methionine to the epsilon amino group of lysine [226]. Currently at least 27 PKMTs are known in human [4]. For a long time, lysine methylation was considered to be irreversible. First of all, no protein with the necessary enzymatic capability to remove methyl groups from amino side chains was known. And, more importantly, studies observed that half-lives of histones and total histone methyl groups were comparable [27, 240], which led to the interpretation that no turnover of the methyl group takes place. A revision of this paradigm happened when the first lysine demethylase LSD1 was found [221]. LSD1 demethylates mono-
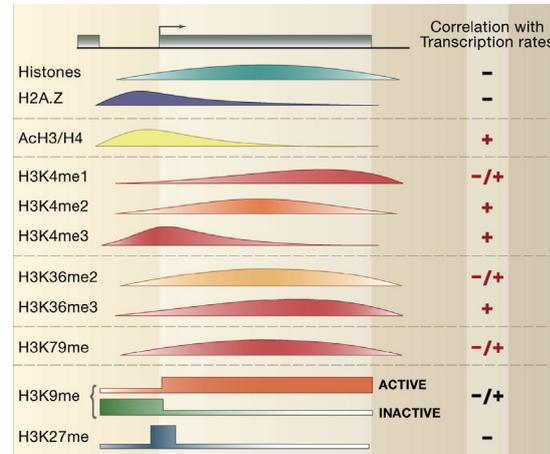

Figure 1.1.6: Histone modification pattern on a representative gene Schematic representation of the distribution of histones and histone modifications in relation to a gene. The sidebar indicates the correlation of the respective modification with transcription rates (modified from [128]).

9

and di- but not trimethylated H3K4 by an FAD dependent oxidative reaction that generates formaldehyde. An additional family of proteins capable of performing lysine demethylation reactions is formed by proteins containing a Jumonji (JmjC) domain. The Jumonji domain of JHDM1 specifically demethylates H3K36me2 via an oxidative reaction [244]. JMJD2A, which demethylates H3K9me3 and H3K36me3, was the first protein described to remove lysine trimethylation [260]. In the same year the H3K9me2/3 demethylase GASC-1 (JMJD2C) was discovered [38]. Recently, an additional mechanism to remove lysine trimethylation was described: LOXL2, a lysyl oxidase, specifically deaminates H3K4me3, but not the mono- and dimethylated lysine, and generates a deaminated lysine (allysine) [85]. More than 20 lysine demethylases have been described so far in human. As they play a crucial role in maintaining the homeostasis of lysine methylation, it is not surprising that deregulation or loss is often associated with diseases or developmental defects [182].

Lysine methylation can – depending on the methylation state and on the position of the lysine on the histone – be associated with active as well as repressive chromatin. Some methylation sites are linked to active transcription, such as H3K4, H3K36, H3K79, whereas others, including H3K9, H3K27, H4K20 are linked to transcriptional repression [115] (Figure 1.1.6).

Lysine methylations associated with transcriptional activation

H3K4: Methylation of lysine 4 of histone H3 is highly conserved and is associated with the initiation of transcription [232]. Lysine trimethylation marks active genes [208] and in ChIP profiles a distinct peak at the 5′ end of genes can be observed [191]. In yeast, all three methylation states of H3K4 depend on the SET1 protein [208], which belongs to the COMPASS complex (Complex proteins Associated with Set1) [153, 202]. In mammalian cells, the situation is more complex, as a diversification and specialization of K4 methylating enzymes occurred. There are at least eight enzymes belonging to the MLL and SET1 families [222], which can methylate H3K4. Whereas the di- and trimethylated forms still mark active promoters [11, 16], K4me1 marks distant enhancers in mammals [81].

Methylation of H3K4 serves as an interaction surface for chromatin readers, which influence nearby chromatin and transcription. A large number of proteins directly binding to these marks have been described [253]. PHD fingers [130, 181], double tudor domains [87] and ZW type zinc fingers [80] are domains that recognize trimethylated H3K4. The basal transcription factor complex TFIID for example, directly binds to H3K4me3 via the PHD finger domain of its subunit TAF3 and this binding seems to

play a role for efficient gene expression [252]. Another example is the interaction of the ING (inhibitor of growth) proteins with H3K4me3. ING2 binds to H3K4me3 via its PHD finger domain on promoters of proliferation genes upon DNA damage and represses their expression [220]. In this way, genes specifically marked to be transcribed can be silenced upon a cellular stress situation.

H3K36: Trimethylation of H3K36 marks actively transcribed genes. The PKMT SET2 associates with elongating RNA Polymerase II and methylates K36 in the body of transcribed genes [119, 267]. Exonic regions were shown to be enriched in trimethylation of K36 compared to intronic regions [113]. The PWWP domains of BRPF1 [254] and DNMT3A [48] were shown to directly bind to this modification. H3K36me3 on gene bodies acts as a transcriptionally repressive mark, and thereby suppresses cryptic transcription initiation [29, 99, 103]. Moreover, overexpression of the H3K36 demethylases JHD1 or RPH1 in *S. cerevisiae* bypasses the requirement for the positive elongation factor gene BUR1 [105]. Although H3K36me3 is associated with actively transcribed genes, its molecular function seems to be repressive.

H3K79: In contrast to the methylation sites discussed above, H3K79 is not located on the unstructured histone tail, but on the surface of the nucleosome core. H3K79 methylation was originally discovered in *S. cerevisiae* where it plays a role in heterochromatin formation [248], telomeric silencing [163], DNA damage response and checkpoint control [68, 265]. In yeast, DOT1 methylates H3K79 [163], and in mammals its homologue DOT1L possesses this activity [154]. So far, DOT1 and its homologues are the only PKMTs that were found to methylate K79. A demethylase has not been discovered yet. In chromatin immunoprecipitations H3K79me3 was primarily found on transcriptionally active regions in the genome [229, 259], despite the phenotypic associations of this modification with heterochromatic functions. 53BP1 directly interacts with H3K79me3, which plays a role in the DNA damage response [91]. The tandem tudor domain of 53BP1 recognizes methylated H3K79 at sites of DNA double strand breaks. Although methylation of H3K79 is regarded a mark of active transcription because of its occurrence on actively transcribed genes, it is implicated in a very wide variety of functions including transcriptional repression [164].

Lysine methylations associated with transcriptional repression

H3K9: Methylation of H3K9 is a classical repressive chromatin mark, which can be found at centromers, telomeres and repressed stretches of chromatin [115]. H3K9

methylation is among the best studied histone modifications, as it is one of the central hubs for the genetic phenomenon of position effect variegation. Position effect variegation describes the observation in *Drosophila*, that euchromatic regions which are rearranged into the vicinity of heterochromatin acquire a variegated pattern of expression [160]. Fly mutants with either a positive or negative effect on this phenomenon were isolated. It turned out that many of these gene products were linked to methylated H3K9. Su(var)3-9 is one of the PKMTs that methylate H3K9 [196] and Su(var)2-5, which is also called HP1 binds to di- and trimethylated H3K9. Based on the properties of methyl-K9 associated proteins, a model for heterochromatin spreading was suggested. HP1 binds to K9me3 and recruits Su(var)3-9, which in turn methylates adjacent histones to provide a new binding platform for HP1 [45]. Several additional proteins with an affinity for methylated H3K9 have been described. The abovementioned HP1 protein, of which the three isoforms HP1 α, β and γ are present in higher eukaryots, contains a chromodomain, which interacts with methylated H3K9 [9, 122]. The chromodomains of CDYL and CDYL2 also specifically recognize di- and trimethylated HK9 [60]. Moreover, direct interaction of MPHOSPH8 [112], UHRF1 [268] and UHRF2 [189], as well as ATRX [49, 56] with this chromatin mark were demonstrated. Although H3K9me3 is generally regarded as a repressive modification, it seems to have additional functions. It was also found in the coding region of actively transcribed genes [246] which is consistent with a report showing HP1 and H3K9me3 staining in heat-shocked genes in *Drosophila* [188]. Whereas di- and trimethylation of K9 are generally found in areas with repressed chromatin, K9me1 was found to be enriched in more active promoters [11].

H3K27: Trimethylation of H3K27 is a repressive mark, which plays a role in development especially in the epigenetic model systems X chromosome inactivation [180, 190] and parental imprinting [218]. Polycomb group proteins play a crucial role for establishing this modification and its consequences on adjacent chromatin and transcriptional regulation. Polycomb is a *Drosophila* mutant with improper body segmentation, and the polycomb gene was suggested to be a negative regulator of homeotic genes required for proper body segmentation [126]. Genes, leading to a phenotype reminiscent of polycomb, are generally referred to as Polycomb group (PcG). A simplified model describes two main polycomb complexes: The Polycomb repressive complex 2 (PRC2) contains the SET domain methyltransferase enhancer of zeste (EZH2 in humans) and methylates H3K27 [28, 44, 121, 157]. Trimethylated H3K27 recruits the Polycomb repressive complex 1 (PRC1) [28, 44, 121], which ubiquitinates H2A and leads to chromatin compaction. However, the system is far more complex as many specialized sub-

complexes exist [158]. Additional mechanisms like long non-coding RNAs and DNA methylation act in concert with polycomb group proteins in gene silencing [225]. Whereas trimethylation of H3K27 has been studied extensively, the two other methylation states of H3K27 are less well understood. H3K27me2 seems to have a similar distribution as H3K27me3 [11]. In contrast, monomethylation of H3K27 is enriched on pericentric heterochromatin [187]. Moreover, depletion of H3K27me1 in the vicinity of transcribed genes [247] as well as enrichment of H3K27me1 at active promoters [11] have been described.

H4K20: Methylation of H4K20 is associated with very diverse processes, including replication, DNA damage repair and transcriptional repression [8]. In mammals, PR-SET7 is the only enzyme generating monomethylated H4K20 [266]. Di- and trimethylation are catalyzed by SUV4-20H1 and SUV4-20H2 [211]. PHF8 demethylates H4K20me1 [13], however, so far no demethylase for the higher methylation states of H4K20 has been described.

During the cell cycle progressive methylation of H4K20 can be observed [186]. H4-K20me1 has been linked to active as well as to repressive loci. Whereas an enrichment of H4K20me1 on actively transcribed genes was reported [247], other studies link H3K20me1 to transcriptional repression [100] and X chromosome inactivation [111]. H4K20me1 shows a dynamic behaviour during the cell cycle and is enriched in S phase [100, 197]. Loss of the H4K20 methyltransferase PR-SET7, and consequently loss of H4K20me1, has severe effects on cell cycle progression and genome stability (reviewed in [13]). H4K20me2 has not been studied in detail, but it is also believed to be a repressive mark [8]. H4K20me3 is generally associated with repressed chromatin and was found in constitutive heterochromatin [114] and on telomeres [14]. Loss of trimethylation either by chemical inhibition [155] or knock-out of the responsible methyltransferases (SUV4-20H1, SUV4-20H2) [211] confirmed that H3K20me3 plays a role in gene repression.

1.1.5 Histone variants

Histone variants are specialized histone isoforms that differ in the primary amino acid sequence from their canonical paralogues. Canonical histones can be replaced by histone variants to generate nucleosomes with modified properties. Whereas the main functions of the canonical histones are packaging of the genome and gene regulation, histone variants are associated with many different processes, like DNA damage repair, transcription initiation and termination as well as sex chromosome condensation

and sperm packaging. Similar to PTMs on histones and nucleosome remodeling, the use of histone variants contributes to the regulatory repertoire of chromatin. Histone variants are not incorporated randomly into chromatin, but each of them shows a distinctive pattern. This incorporation is tightly controlled by histone chaperones, which ensure that variants are only deposited in their proper places. Genes encoding canonical histones are found in repeat arrays and their transcription is tightly coupled to DNA replication [146]. In contrast, histone variants are found as singly copy genes and are mostly constitutively expressed. The majority of histone variants are described for histone H3 and histone H2A. Most histone variants, like the H3 variants CENP-A and H3.3, or the H2A variants H2A.Z and H2A.X are of nearly universal occurrence in all eukaryotes [235].

H3 variants

Mammals have two canonical H3 variants: histone H3.1 and H3.2. Histone H3.1 differs from H3.2 in only one amino acid – a cysteine instead of a serine in position 96. H3.1 and H3.2 are both synthesized and incorporated into chromatin in a DNA replication dependent manner. Deposition of canonical H3 is catalyzed by a protein complex consisting of CAF-1, ASF1 and NASP [234]. Most H3 variants show very high identity to canonical H3 (Figure 1.1.7) but still exhibit different properties and genome localization.
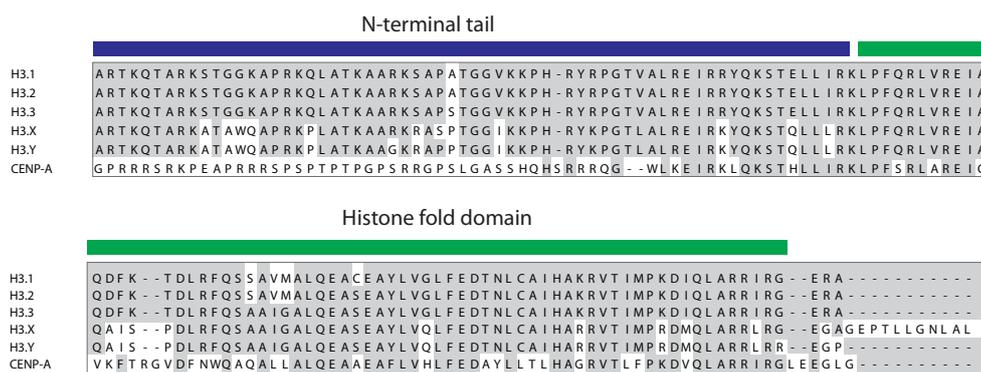

Figure 1.1.7: Amino acid sequence alignment of human H3 variants. All variants of histone H3, except for CENP-A, show very high identity with only very few amino acid exchanges.

H3.3: Two conserved differences distinguish histone H3.3 from the canonical histone H3. First, its expression is cell cycle independent and not coupled to DNA replication. Second, an amino acid substitution of residues 87-90 in the histone core region ('SAVM' in the canonical H3 and 'AAIG' in H3.3) is necessary and sufficient for selective deposition [2]. H3.3 is incorporated at active chromatin and a strong enrichment of H3.3 could be observed at actively transcribed rDNA arrays [2]. In line with its deposition at sites of active transcription, H3.3 is preferentially marked with activating histone modifications [150]. Nucleosomes containing H3.3 instead of canonical histone H3 were described to be less stable [97], which contributes to the formation of accessible chromatin structures at transcriptionally active loci. In addition to its enrichment on actively transcribed loci, H3.3 has been found on regulatory elements [98] and constitutive heterochromatin at telomeres [263]. The precise localization of H3.3 on chromatin demands a very specialized deposition machinery. Two separate chaperone complexes are described for H3.3. Deposition of H3.3 at actively transcribed regions is dependent on the histone chaperone HIRA [195]. ATRX and DAXX are essential for H3.3 deposition at telomeres and repression of the telomeric repeat containing RNA (TERRA) in a HIRA independent pathway [70]. DAXX directly interacts with H3.3, and the amino acid residues 80-94 of H3.3 but not H3.1 are necessary and sufficient for this specific interaction [127].

H3.X and H3.Y: Recently two novel histone H3 variants, H3.X and H3.Y, were discovered [262]. H3.Y expression is increased upon cellular stress and it seems to play a role in regulating cell growth and expression of cell cycle control genes [262].

CENP-A: The centromere specific H3 variant CENP-A [54] is essential for assembly of the kinetochor and for proper chromosome segregation [5]. It has 50-60 % identity to H3 in the histone fold domain, but no conservation in the N-terminal tail (Figure 1.1.7). CENP-A expression peaks in G_2 phase and it is incorporated into chromatin in telophase and early G_1 phase [224]. Purification of CENP-A associated histone chaperones identified HJURP (Holliday junction-recognizing protein) [52, 62] as an interaction partner. HJURP directly binds to the centromere targeting domain (CATD) of CENP-A and is necessary for centromeric deposition of CENP-A [62].

H2A variants

The H2A variants form a very heterogeneous group and are less conserved than H3 variants (Figure 1.1.8). Their genome-wide localization patterns and their very differ-

ent functional associations cover a wide array of chromatin templated processes.
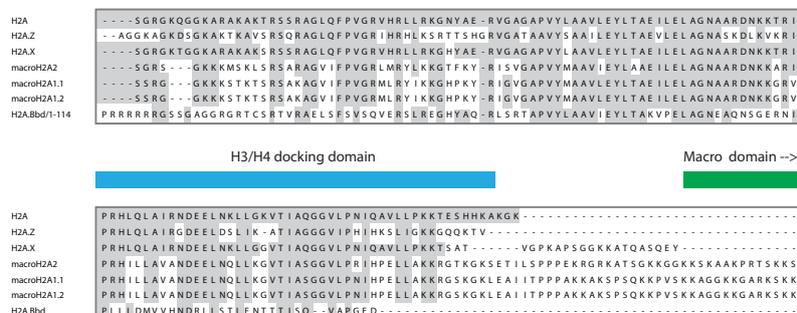
Figure 1.1.8: Amino acid sequence alignment of histone fold domains of human H2A variants. H2A variants show high conservation in the central part. H2A.BBd deviates most from canonical H2A (macro domains are omitted).

H2A.Z: H2A.Z is encoded by one gene in *S. cerevisiae* (Htz1) and two genes in vertebrates (H2A.Z.1 and H2A.Z.2). Despite only three amino acid difference between H2A.Z.1 and H2A.Z.2, H2A.Z.2 cannot rescue a H2A.Z.1 knock-out in mice [57], which points to non-redundant functions. H2A.Z containing nucleosomes are enriched adjacent to nucleosome free regions at transcription start sites, where they co-localize with H3.3 [98]. An *in vivo* study reported that H2A.Z/H3.3 nucleosomes are highly unstable [98]. However, no such instability could be detected in an *in vitro* study [238]. In *S. cerevisiae*, the SWR1 complex, which consists of 13 subunits including the ATPase SWR1P mediates the ATP dependent exchange of H2A for HZT1 [118]. The INO80 complex removes H2A.Z/H2B dimers and thus also controls H2A.Z localization [177]. In mammals, two complexes — the SRCAP complex and the TIP60 complex — are responsible for H2A.Z deposition (reviewed in [120]). The function of H2A.Z is controversially debated, as it seems to be involved in many, sometimes contradictory processes such as gene activation and silencing, nucleosome turnover, DNA repair, heterochromatin, boundary element and chromatin fiber formation [273].

H2A.X: H2A.X is mainly studied for its role in DNA double strand break repair. H2A.X contains a C-terminal Ser-Gln (Glu/Asp)-ϕ motif, where ϕ stands for a hydrophobic residue, in which the serine becomes phosphorylated upon DNA damage

[201]. Phosphorylated H2A.X is commonly referred to as γ -H2A.X. This phosphorylation is most likely accomplished by the ATM kinase [22]. γ -H2A.X foci around the site of a DNA double strand break form already one minute after induction of the break [201]. This signal plays an important role in the recruitment and assembly of the DNA damage repair machinery. The phosphorylated C-termini of γ -H2A.X serve as an interaction platform for chromatin readers. The DNA damage response proteins MDC1 (mediator of DNA damage checkpoint protein 1) [230] and NBS1 [110] have been shown to be directly recruited to DNA damage sites by binding to the phosphorylated C-terminus. Interestingly, H2A.X $^{-/-}$ or H2A.X point mutants defective for phosphorylation of Ser139 are viable and could perform the initial recruitment of DNA repair factors to sites of DNA double strand breaks [30]. Despite the fast kinetics of H2A.X phosphorylation, γ -H2A.X is not the initial recruiter of DNA repair factors. However, γ -H2A.X may be necessary to concentrate proteins in the vicinity of DNA lesions, as H2A.X $^{-/-}$ cells fail to form irradiation induced foci (IRIF).

MacroH2A: Among all H2A variants, macroH2A differs most from the canonical counterpart. In addition to the histone fold domain it contains a linker and a C-terminal macro domain [183]. The macro domain is about twice the size of the histone fold domain and protrudes out of the nucleosome. Two genes code for macroH2A in vertebrates, macroH2A.1 and macroH2A.2; macroH2A.1 can be alternatively spliced. MacroH2A.1 and macroH2A.2 are both enriched on the inactive X chromosome [31, 39] suggesting a function in gene silencing. On autosomes of human pluripotent cells, both macroH2A variants were found to occupy repressed key developmental genes. Moreover, macroH2A was necessary for the exact temporal activation of HOX gene clusters during neuronal development [24]. Several proteins were described to specifically interact with the macro domain, among them SPOP, HDAC1, HDAC2 and PARP1 [23]. The distinct localization pattern of macroH2A argues for a dedicated histone chaperone machinery for this variant. ATRX was recently suggested to serve not only as a histone chaperone for H3.3, but also for macroH2A [194].

H2A.Bbd: The histone variant H2A.Bbd is the least well understood histone variant. H2A.Bbd is excluded from the inactive X chromosome [32], hence its name "Barr body deficient". Among H2A variants, H2A.Bbd has the least conservation towards canonical H2A. It lacks the flexible C-terminus and has a unique N-terminal tail of six consecutive arginines (Figure 1.1.8). Furthermore, significant differences in the docking domain, which is responsible for the contacts to histone H3, contribute to structural alterations of H2A.Bbd containing nucleosomes. Only 118 base pairs of DNA are

wrapped around H2A.Bbd containing nucleosomes [10] as opposed to 146 in nucleosomes containing canonical H2A. These smaller nucleosomes are less stable *in vitro* [10] and *in vivo* [63]. The presence of H2A.Bbd not only affects the nucleosome stability, but the whole chromatin fiber is in a less compacted state when H2A.Bbd is present [272]. Although a biological function has not yet been attributed to H2A.Bbd, it is generally regarded to be associated with active gene expression.

1.2 Mass spectrometry-based quantitative proteomics

Mass spectrometry (MS) based proteomics developed over the last years from a technique applied by some specialists to an indispensable method for molecular cell biology [1]. The unbiased identification of proteins and protein modifications from complex mixtures greatly contributed to our current understanding of protein interactions, dynamics and post-translational modifications. Proteomics can be performed in an assumption free manner to identify all possible proteins – a so-called discovery approach. In contrast, targeted methods only analyze *a priori* defined subsets of the proteome and mostly aim at describing the behaviour of this fraction under multiple conditions. Current proteomic approaches allow the identification of several thousand proteins from complex organisms [46, 162], analysis of protein interactions [251] and mapping of post-translational modifications [35].

Mass spectrometry can be used to identify peptides as well as whole proteins. The analysis of intact proteins by MS, which is referred to as "top-down" mass spectrometry, is challenging and has severe technical limitations. An example for the complexity of top-down proteomics is the analysis of modification combinatorics on histone H4 [185]. In most proteomic studies, "bottom-up" proteomics (also called shotgun proteomics) approaches are applied in which proteins are digested into peptides prior to MS analysis. Peptides have better ionization efficiencies than proteins, produce less complex spectra and yield fragmentation spectra which are easier to interpret.

Figure 1.2.1 depicts the steps of a classical bottom-up proteomics workflow. Proteins derived from cells or tissue or a preceding biochemical experiment are digested into peptides. A fractionation step at the protein level, for example by one-dimensional gel-electrophoresis, or at the peptide level, for example by isoelectric focusing, can be included to reduce sample complexity. Peptides are further fractionated by nanoscale reverse phase chromatography and directly sprayed into the mass spectrometer via electrospray ionization. Contemporary MS instruments perform analysis of the intact masses of the peptides (MS1 scan, precursor mass) and in addition, selected peptides are isolated and fragmented and their mass is measured (MS2 scan). By transferring energy to the ions in the gas phase, they fragment in a characteristic manner at the peptide bonds [228]. Most commonly, collision induced dissociation (CID) is applied, in which the peptide collides with an inert gas. Alternative fragmentation techniques like higher energy collisional dissociation (HCD), electron transfer dissociation (ETD) and pulsed Q dissociation (PQD) are also frequently used. Ideally, the fragmentation generates a "ladder", from which the amino acid sequence can be directly derived. As most spectra contain only partial sequence information, statistical algorithms are
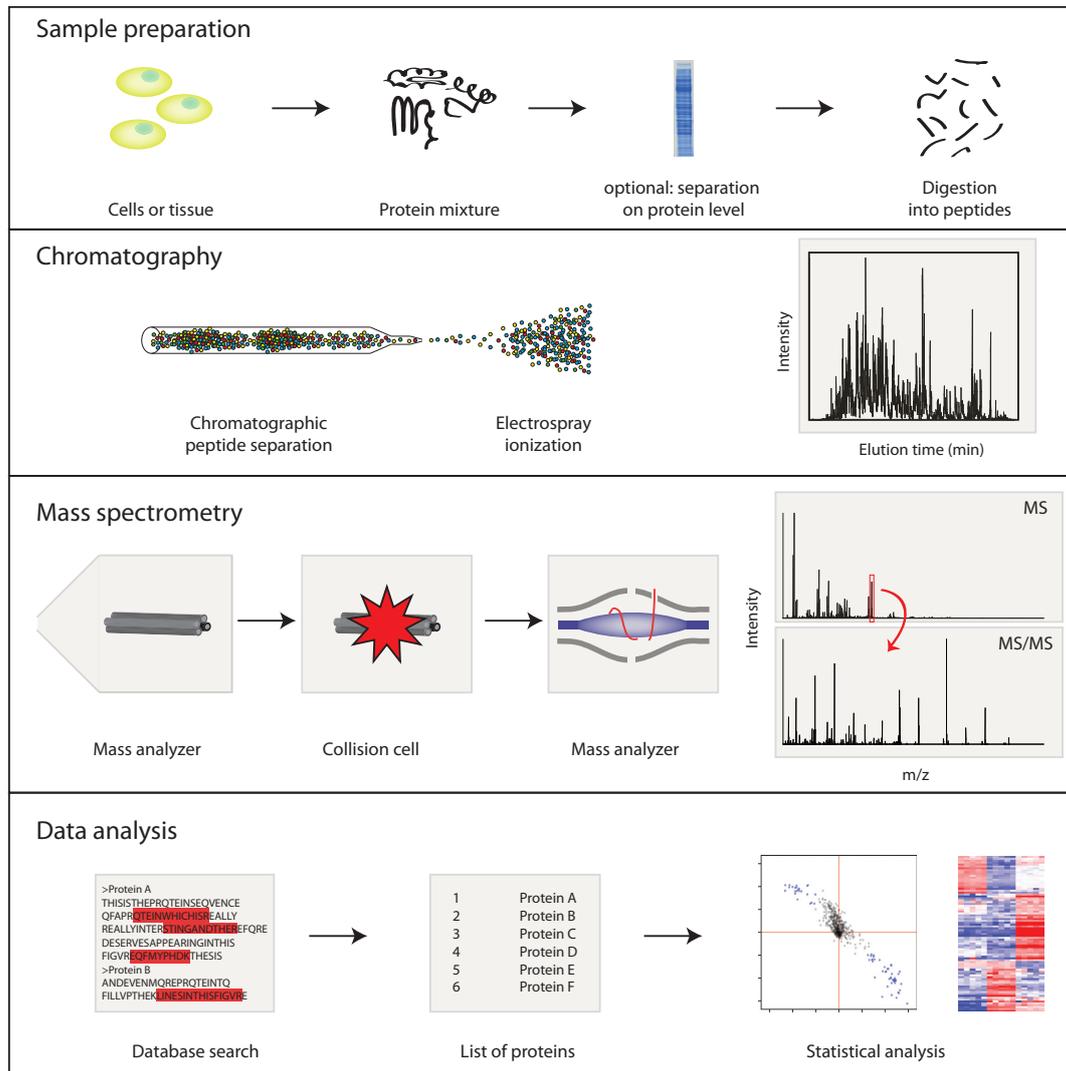
Figure 1.2.1: Bottom-up proteomics workflow Schematic depiction of the major steps in a proteomic experiment. During sample preparation proteins are digested into peptides. Nanoscale chromatography fractionates peptides and electrospray ionization transfers them into the mass spectrometer. In the mass spectrometer, masses of the intact and fragmented peptides are measured. The data analysis pipeline identifies peptides, infers protein identities and determines regulated proteins by statistical analysis (modified from [35]).

applied to determine the best match in a database search. A search engine (e.g. Mascot [184] or Andromeda [42]) performs a database search in which the observed mass of the intact peptide (precursor mass) and its fragment masses are matched with the *in silico* digested and fragmented peptides derived from a protein database. Identified

peptides are reassembled into proteins and statistical analysis is performed to identify proteins which are regulated significantly.

1.2.1 Contemporary mass spectrometry instrumentation

Analysis of peptides in the mass spectrometer is the central step in a proteomic experiment. A large variety of MS instruments is currently available in which different physical principles are exploited to manipulate and analyze ions in the gas phase. Different mass analyzers, fragmentation principles and mass detectors have been combined. The application range of an instrument depends on which parts are incorporated and how they are combined. A thorough understanding of the underlying principles is crucial to choose the instrument required for the desired application. In general, contemporary MS instruments for proteomics are capable of recording parent masses at high resolution and accuracy in combination with a high sequencing speed for fragment ions.

Mass analyzers

The mass analyzer is the core element of every mass spectrometer. Depending on the underlying physical principles, mass analyzers exhibit unique characteristics which make them suitable for different tasks. The following properties are used to describe the performance of mass analyzers: Mass precision describes the "repeatability", meaning the variation between several measurements for the same mass. The term mass accuracy describes the deviation of the measured to the theoretical mass [41]. Current instruments can achieve mass accuracy in the low parts per million (ppm) range. Resolution is a dimensionless number calculated by dividing the mass of an observed peak by its width. Resolution is important for proper quantification, separation of neighboring peaks, and also influences mass accuracy by the separation of isotope clusters. A high "dynamic range", which describes the ratio of the strongest signal to the weakest signal that can still be detected in a spectrum, is a prerequisite for sampling deeply into a complex peptide mixture.

Two of the most common mass analyzers, which were also used in this thesis, the linear ion trap and the Orbitrap analyzer, are described below. Time of flight (TOF) mass analyzers are also popular, however, as they were not used in this thesis, they will not be discussed.

Ion trap: The linear ion trap is a very versatile mass analyzer capable of mass selection, fragmentation and detection. The trap consists of four hyperbolic rods, each of
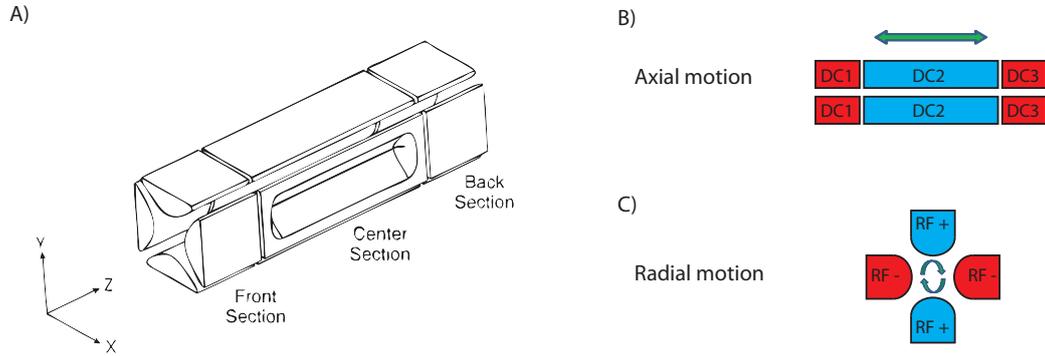
Figure 1.2.2: The ion trap mass analyzer A) Schematic view of a linear two dimensional ion trap. B) Application of DC voltage on the front and back sections forms a potential well, which traps ions in z direction. C)Ions stored in the ion trap follow a radial motion guided by the application of an RF voltage to the quadrupole rods (from [214]).

which is cut into three axial sections (Figure 1.2.2 A). Each section has a discrete DC level which generates a potential well and traps ions in axial direction (Figure 1.2.2 B). The rods are paired, and a radio frequency (RF) voltage is applied to the rod pairs. This leads to a potential well in radial direction confining the trajectories of the ions (Figure 1.2.2 C). To reduce ion motion and dispersion, a dampening gas (usually helium) is introduced into the trap. Ions in the ion trap collide with the helium gas leading to a loss of kinetic energy. The stability of the ions in the ion trap can be described by the Mathieu equations:

$$a = \frac{8zeU'}{m(x^2+y^2)\Omega^2}$$
$$q = \frac{4zeV'}{m(x^2+y^2)\Omega^2}$$

with

m = mass of a trapped ion

e = charge of a trapped ion

z = number of charges on the trapped ion

V′ = RF power (amplitude of RF oscillation)

U′ = DC offset

Ω′ = frequency of RF

x = distance from the center of the trap to the X rods

y = distance from the center of the trap to the Y rods

Combinations of a and q leading to stable ion trajectories can be seen in Figure 1.2.3. With some assumptions, these equations can be simplified. The geometric parameters x and y, as well as Ω' are fixed by the machine design. As the DC offset is never changed, it is set to 0 in the equation. This leads to a value of 0 for a, and leaves a one dimensional stability definition for ions which only depends on q. Ions are stable in the ion trap as long as their q value is below 0.908.
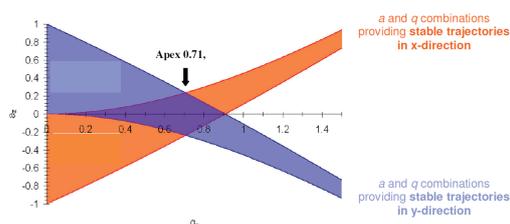

Figure 1.2.3: Stability of ions in the ion trap Depiction of a and q combinations which lead to stable ion trajectories in x and y direction. The overlapping area indicates a-q combinations under which ions are stable in the ion trap (from [206]).

Moreover, a smaller ion will always have a larger q than a larger ion. By modulating the RF amplitude and thus raising the q value, ions can be ejected from the ion trap in a size dependent manner. This is used for two processes. First of all, by selectively removing ions of a specific m/z, the ion trap can perform mass selection. Second, the ejection can be guided towards the slits on the rods behind which the multipliers are positioned (see figure 1.2.2). By successively increasing the m/z of ejected ions and detecting them with the multipliers, mass spectra can be recorded. Although ion traps can be used to obtain full scan spectra of the precursor masses, in proteomics they are often only employed for fragmentation. For fragmenting selected ion populations, a m/z range of interest is isolated. Afterwards, the q of these ions is reduced to the so-called activation q – a standard value would be 0.25. As fragmentation generates many ions that are smaller, many of those would be lost if their q would be above 0.908. The smallest mass fragment, which can still be observed after fragmentation, can be calculated as activation q divided by 0.908 times the precursor mass. At a standard activation q of 0.25, fragments with a mass about a quarter of the precursor mass cannot be retained in the ion trap, which is referred to as the " ⅓ mass cutoff". In summary, the ion trap is capable of storing, isolating, fragmenting and, in combination with a multiplier, detecting ions. It has a very high sensitivity and a high sequencing speed, but mass accuracy and mass resolution are relatively low compared to high resolution devices like TOF or Orbitrap analyzers. In hybrid instruments, ion traps are preferentially used for fragmentation as only few ions are necessary and fast cycle times can be achieved. However, due to the ⅓ mass cutoff for collision induced dissociation, fragment ions in the low mass range cannot be observed.

Orbitrap: The Orbitrap traps and measures ions in an electrostatic field. Its design is based on the Kingdon trap [107], and the actual Orbitrap was first described in 2000 by Alexander Makarov [142]. The Orbitrap consists of an inner spindle-like central electrode, surrounded by an outer barrel-like split electrode (Figure 1.2.4). Before entering the Orbitrap cell, ions are accumulated and stored in the C-trap. The C-trap is an RF-only quadrupole in the shape of the letter "C". From there, ions are injected as a compacted package into the Orbitrap cell off its plane of symmetry (red arrow in Figure 1.2.4).

Once in the Orbitrap, two forces are acting on the ions. First, a radial force generated by the radial field E_r attracts ions towards the central electrode. If the centrifugal force produced by the tangential velocity equals the attractive force towards the central electrode generated by the electrostatic field, the ions orbit in a circular trajectory around the central electrode. This oscillation in radial direction is highly dependent on the initial energy of the ions. Second, an axial field is generated in the Orbitrap, which is zero at the equator plane and increases with the distance from the center. Ions are attracted towards the equator plane, traverse it and upon entering the other half
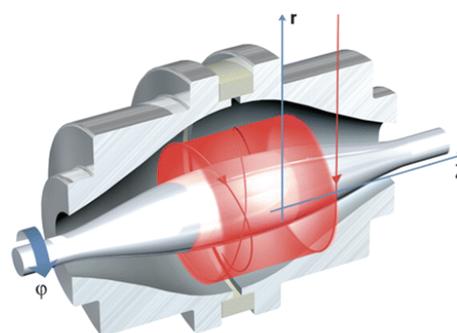
Figure 1.2.4: The Orbitrap mass analyzer Schematic cross section of an Orbitrap mass analyzer. Blue arrows indicate radial (r) and axial (z) directions. Red arrow indicates ion movement (from [215]).

of the Orbitrap cell, a force opposite to their movement direction pulls them back to the equator plane. This force increases with distance from the equator plane until the kinetic energy in axial direction is zero. Now ions are accelerated back to the equator plane. These forces generate an axial oscillation. Axial and radial movements sum up to a stable spiral-like trajectory around the central electrode. Importantly, the axial oscillation component is independent of the initial energy of the ion and depends only on the mass to charge ratio m/z. A Fourier transformation can convert the frequency readout generated from axial oscillation of all ions present in the Orbitrap into an m/z spectrum. The Orbitrap mass analyzer is capable of generating high resolution, high mass accuracy measurements in a short time frame compatible with chromatography coupled settings. Its performance is comparable with a Fourier transform ion cyclotron resonance (FTICR) cell, however no large superconducting magnets are needed and a

higher resolution in the high mass range can be achieved [215]. As the Orbitrap is not capable of performing fragmentation, it is normally coupled to ion selection and fragmentation devices like a linear ion trap, or a quadrupole and a dedicated collision cell.

Mass spectrometers containing the Orbitrap as a mass analyzer

The Orbitrap analyzer is exclusively incorporated into mass spectrometers manufactured by Thermo Fisher Scientific and currently five different instruments are equipped with the Orbitrap cell. LTQ Orbitrap, Orbitrap Velos and Orbitrap Elite are hybrid instruments, which use the Orbitrap for high accuracy and precision recording of precursor masses and HCD fragment masses, but include a line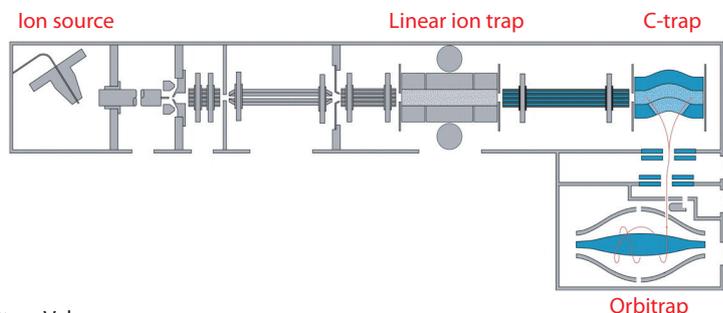ar ion trap capable of rapid peptide fragmentation. The benchtop instruments Exactive and Q Exactive contain the Orbitrap as sole mass analyzer, which is used for recording precursor as well as fragment ions. Measurements for this thesis were performed on the LTQ Orbitrap, Orbitrap Velos and Q Exactive and these instruments will be introduced in more detail (Figure 1.2.5).

LTQ Orbitrap: The LTQ Orbitrap was the first mass spectrometer that incorporated an Orbitrap mass analyzer [171] (Figure 1.2.5 A). This hybrid instrument consists of an Orbitrap cell for measuring precursor ion masses at high resolution and a linear ion trap for rapid acquisition of fragment spectra. In the beginning of a recording cycle, ions are guided through the ion optics and the linear ion trap and are accumulated in the C-trap. From there, a compacted package of ions is transferred into the Orbitrap cell and a prescan at low resolution is recorded to define the most abundant ions. TopN methods, in which the N most abundant ions (often five or ten) from the MS1 scan are chosen for fragmentation, are routinely used in data dependent acquisition. Isolation, fragmentation and measurement of these ions is performed in the ion trap, concurrent with acquisition of the high-resolution spectrum of the precursor masses in the Orbitrap. One cycle, including an MS1 scan in the Orbitrap at a resolution of 60,000 and five fragmentation events, takes around 2.5 seconds. The high mass accuracy of the Orbitrap can be even further increased by injecting ambient ions from laboratory air as internal recalibration standard [171]. An upgrade (called Orbitrap XL) contains a dedicated collision cell for HCD fragmentation [172].

Orbitrap Velos: The Orbitrap Velos was released after the LTQ Orbitrap [174], and has the same principle design (Figure 1.2.5 B). It is also a hybrid instrument consisting

A) LTQ Orbitrap

Ion source   Linear ion trap   C-trap

Orbitrap

B) Orbitrap Velos

Ion source   Double linear ion trap   C-trap   HCD cell

S-Lense

Orbitrap

C) Q Exactive

C-trap   HCD cell

S-Lense
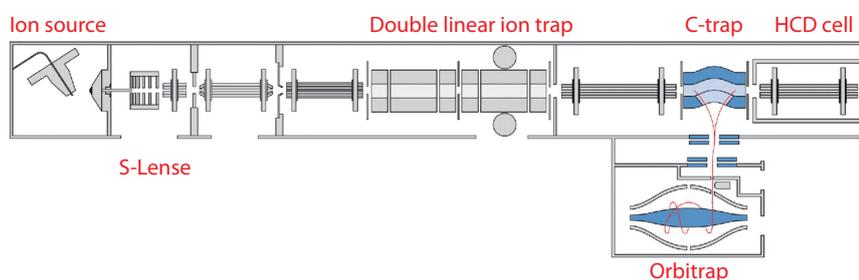
Orbitrap

Ion source
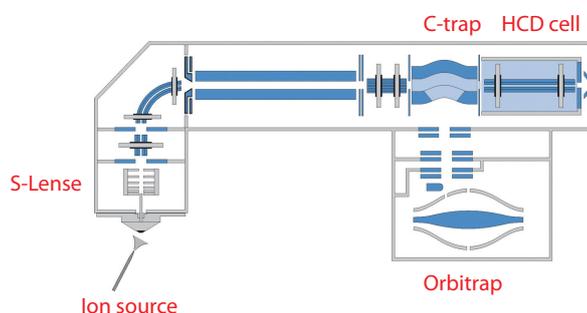
Figure 1.2.5: Three important members of the Orbitrap family A) The LTQ Orbitrap was the first hybrid instrument containing an Orbitrap cell [171]. B) The Orbitrap Velos is an improved hybrid instrument [174]. C) The benchtop instrument Q Exactive contains only an Orbitrap mass analyzer [152].

of a linear ion trap and an Orbitrap mass analyzer. The front part was significantly modified: an S-lens replaces the tube lens/skimmer and allows better transmission of ions into the instrument, thus increasing the sensitivity. The linear ion trap in the LTQ Orbitrap was replaced by a dual linear ion trap. The first ion trap is operated at a higher pressure of helium bath gas (5.0×10^{-3} Torr) which allows very efficient trapping, iso-

lation and fragmentation of ions. Ions are transferred into the second trap operated at lower pressure (3.5×10^{-4} Torr), in which mass spectra can be recorded at higher speed. HCD fragmentation could already be performed in the Orbitrap XL. However, a large number of ions had to be accumulated due to inefficient ion transfer. Because of improvements in design and electronics, five to ten times more ions per unit time enter the HCD cell in the Orbitrap Velos. This development has made HCD fragmentation suitable for standard proteomic experiments. The availability of rapid fragmentation and scanning in the ion trap, or efficient quadrupole-like fragmentation in the C-trap combined with high resolution mass analysis in the Orbitrap, allows two analysis strategies. In the high-low strategy, precursor masses are recorded at high resolution and high mass accuracy in the Orbitrap, with concomitant rapid analysis of fragment ions at low resolution and low mass accuracy in the linear ion trap. In contrast, the high-high strategy makes use of the improved HCD setup, and the fragment spectra are also recorded at high resolution and high mass accuracy in the Orbitrap. As all mass measurements are performed in the Orbitrap cell, full scans and fragmentation scans are recorded successively. The high resolution recording of fragment spectra allows deconvolution of multiply charged fragment ions and the higher mass accuracy allows smaller mass tolerances thereby increasing confidence in spectra matching. In addition, fragmentation in the HCD cell does not suffer from the ⅓ mass cutoff, and produces spectra that contain more information. In summary, the Orbitrap Velos has improved sensitivity and speed compared to the Orbitrap XL and enables efficient shotgun proteomics experiments with the high-high strategy.

Q Exactive: The Q Exactive is the latest member of the Orbitrap family and is based mainly on the Exactive. The Exactive is a benchtop instrument with only one mass analyzer, an Orbitrap, which can only perform precursor mass detection and all ion fragmentation [66]. An additional quadrupole in the Q Exactive enables isolation of selected ions to perform data dependent acquisition [152]. Proteomic measurements on the Q Exactive yield high-high data similar to HCD experiments with the Orbitrap Velos. Improved sensitivity and increased sequencing speed can be achieved with the Q Exactive due to a shorter ion path, the lack of a linear ion trap and further improvements on electronics and software. The introduction of a benchtop instrument, which is at least equally powerful for shotgun proteomics as the Orbitrap Velos, is a major step forward to make high quality mass spectrometry available for the larger biological community.

1.2.2 Quantitative proteomics

Proteomics has developed into a powerful technique for biology and biochemistry by providing the means to reliably identify many proteins in complex mixtures in a relatively small time frame. However, information about the presence or absence of a protein in a sample is in most cases not enough to draw valid biological conclusions. It is more important to obtain quantitative information on proteins, as often a change in protein abundance rather than their appearance or disappearance *per se* is responsible for a biological effect. Mass spectrometry can be used to obtain not only qualitative, but also this quantitative information. There are two principle types of quantification. Absolute quantification aims to determine the absolute amount of a protein in a solution or a cell system and yields concentrations or copy numbers per cell. In relative quantification, only the abundance difference between two samples is determined.

Two basic strategies can be applied to obtain quantitative information: label-free and stable isotope labeling approaches (Figure 1.2.6). Whereas label-free approaches simply prepare and measure the samples separately, labeling approaches introduce stable isotopes which generate a mass difference. Stable isotopes can be introduced at different stages during the experiment (Figure 1.2.6). Differently labeled samples can be distinguished in the mass spectrometer and this allows to combine proteins (or peptides) and analyze them together. Parallel sample processing steps and, to an even greater extend, separate measurements as in label-free approaches, introduce variability which reduces the precision of the quantification. A more accurate quantification allows more reliable identification of significantly changed protein hits from the observed protein population. The earlier samples can be combined, the more accurate the quantification will be. In principle, metabolic labeling produces the most accurate quantification, whereas label-free approaches are more prone to accumulate variability and demand more replicates and more sophisticated statistical analysis.

Relative quantification by stable isotope labeling

Isotopic labeling strategies are always based on introducing defined stable isotopes into a sample to make it distinguishable from another sample by mass spectrometry. It can be either achieved by chemical derivatization of an unlabeled sample or by exploiting metabolic pathways to incorporate heavy isotopes.

Chemical labeling: Chemical labeling strategies use a reactive group on a polypeptide to fuse it to an isotopic label. All samples that are to be compared are treated in
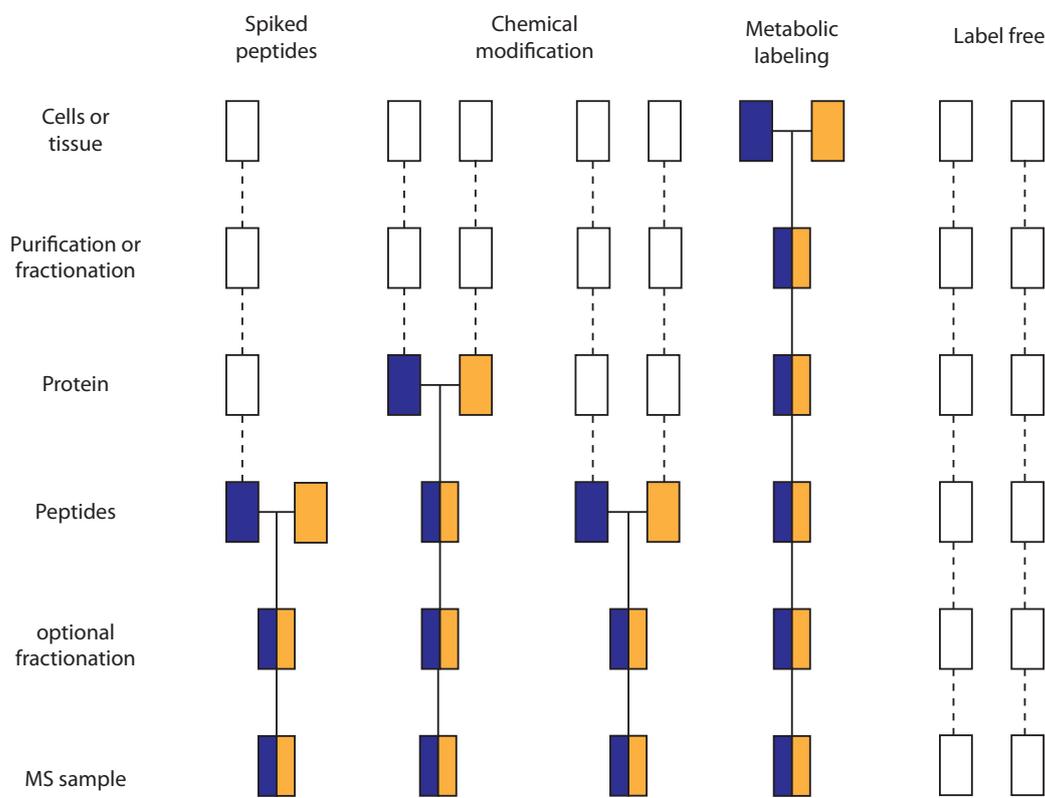
Figure 1.2.6: Labeling strategies and their impact on quantitative accuracy The scheme depicts typical stable isotope labeling and label-free workflows. Empty boxes represent samples without a label which cannot be distinguished in the mass spectrometer. Once samples are isotopically labeled (represented by colored boxes) they can be distinguished in the mass spectrometer and are pooled. The earlier the samples are pooled, the less variability is introduced during the sample workflow (modified from [176]).

the same manner, but using isotopically different reagents. For quantification at the MS1 level, reagents are used that introduce a mass difference between the peptides. The advantage of these methods is high quantification accuracy; however, samples increase in complexity by a factor of two for double labeling. The ICAT (isotope-coded affinity tag) reagent consists of a thiol specific reactive group, a linker which contains the isotope label and a biotin group for affinity enrichment [76]. Only cysteine containing peptides can be labeled and are subsequently enriched via the biotin moiety prior to MS analysis. Although this approach is very specific and reduces complexity, only a subset of peptides can be labeled and quantification of many proteins will rely on very few data points. Another method to introduce a mass shift is dimethyl labeling

[20]. Up to three different isotopomeres of formaldehyde react with alpha and epsilon amino groups to form dimethyl amines. This reaction adds two methyl groups to all lysine side chains and all free N-termini and achieves labeling of all peptides present. A different approach performs quantification on the MS2 level. TMT (tandem mass tag) [241] and iTRAQ (isobaric Tag for Relative and Absolute Quantitation) [203] use isobaric tags. Isobaric tags consist of a reporter group and a balancing group, which add up to the same mass for all tags, hence the name isobaric. Pooled samples which have been treated with isobaric reagents generate a single ion cluster in the MS1 space for every peptide. Upon fragmentation, the different reporter ions are released and their intensity is used for relative quantification. This approach can be easily multiplexed, without concomitant increase in the complexity in the MS1 space. However, quantification at the MS2 level can have some disadvantages. Standard collision induced fragmentation in the ion trap does typically not cover fragments in the low mass region. Instead, pulsed Q dissociation (PQD) or a triple quadrupole like fragmentation (HCD) must be used. Furthermore, every peptide quantification is based on a single observation in a fragmentation event, whereas in MS1 based methods a peptide is observed during consecutive full scans allowing several quantification events. Finally, co-eluting peptides, which are in the fragmentation window, also contribute their reporter ions which leads to ratio dampening [151].

In summary, chemical labeling provides a possibility to perform isotope labeling based quantification on material that was initially unlabeled . These chemical labeling methods have the disadvantage of additional processing steps that can introduce variability and artifacts.

Metabolic labeling: Metabolic labeling strategies already introduce the isotopic atoms through growth medium or food. This can be done in a global manner, e.g. by replacing all nitrogen atoms by heavy nitrogen [170]. Unfortunately, this approach produces broad isotope distributions which are complicated to analyze and it is therefore only used for specialized applications in plant and bacterial biology. A very defined incorporation can be achieved by replacing essential amino acids in the growth medium with their heavy counterpart, an approach termed SILAC (stable isotope labeling with amino acids in cell culture) [175]. Labeling all proteins with heavy arginine and lysine in combination with usage of the protease trypsin which cleaves C-terminal to these amino acids [173] for digestion ensures that every peptide contains at least one labeled amino acid (except the C-terminal peptide of the protein). Two isotope clusters can be observed for every peptide, forming a so-called SILAC pair. From the intensities

of the SILAC pair, a ratio can be directly assigned to the identified peptide. In principle, nearly every cell line can be SILAC-labeled, including cell lines that demand more sophisticated culturing like mouse and human embryonic stem cells [72, 199]. Moreover, whole organisms are also amenable to SILAC labeling. A lysine auxotroph *Saccharomyces cerevisiae* strain [74], *Drosophila melanogaster* [233], *Mus musculus* [116] and *Caenorhabditis elegans* [123] were successfully labeled.

By spiking in a heavy-labeled human cell line to human samples (e.g. tumour biopsies), the high-accuracy SILAC based quantification can be applied for human samples which are otherwise not accessible for metabolic labeling. As the internal standard (the heavy labeled cell line) is the same in all samples, the "ratio of ratios" allows a direct comparison of protein abundance between different samples. Combining several representative cell lines to a super-SILAC mix further enhances quantification accuracy [67].

Relative quantification by label-free approaches

Label-free proteomics aims at performing quantification without the introduction of stable isotopes. In general, these approaches have to cope with higher variability because sample preparation and measurement are performed separately. As a consequence, a more complex statistical analysis is required. Comparing the number of peptide spectra recorded for a protein in two samples is the most straightforward relative quantification. This spectral counting approach [134] correctly classifies highly regulated proteins. However, especially proteins with few sequenced peptides cannot be quantified accurately and this approach is generally prone to a high false negative rate. Better results can be obtained by an intensity-based label-free quantification as it is computed in the MaxQuant software platform [40, 136]. To overcome experimentally introduced variability, the algorithm contains several normalization steps.

Absolute quantification

To obtain absolute protein concentrations, a defined amount of standard, in most approaches a heavy-isotope-labeled reference, needs to be spiked into the sample. Labeled synthetic peptides can be used for this purpose in an approach often referred to as AQUA (for absolute quantification) [108]. To control for variability introduced during sample preparation (mainly missed cleavages and protein adsorption), heavy protein fragments or full length proteins can be spiked in before digestion [78, 269]. The abovementioned methods can provide very accurate quantification, but their high-throughput capability is severely limited as for every protein to be quantified a separate

standard has to be spiked in.

Although the intensity between two different peptides cannot be directly used to infer quantitative information, approaches were developed that can estimate absolute amounts without using isotope labeled spike in standards. The empirical abundance index (emPAI), for example, is computed as ten to the power of the number of observed peptides divided by the number of theoretical peptides minus one. Interestingly, the emPAI shows direct proportionality to the absolute protein amount [92].

A more refined method is to calculate a so-called iBAC (intensity-based absolute quantification) intensity [213]. In this method the intensities of all identified peptides of a protein are summed up, divided by the number of theoretically observed peptides and log transformed. To perform absolute quantification, a non-labeled standard of accurately quantified proteins is spiked into the sample before sample preparation. Using iBAC intensities for the standard proteins, the absolute protein amount of all identified proteins can be estimated using a linear regression [213].

1.2.3 Interaction proteomics

MS-based proteomics is a powerful tool for studying protein-protein interactions. Its major power is the unbiased identification and quantification of proteins, thereby allowing the rapid and reliable identification of interacting proteins without prior knowledge. Classical approaches used very clean purifications, for example by employing a TAP-tagging strategy [198] and defined all identified proteins as specific interactors. The TAP-tagging approach was successfully applied in large-scale interactome studies [64, 65, 117]. However, using highly-sensitive contemporary mass spectrometers, background binding proteins will be identified which give rise to false positive interactors. A superior alternative combines truly quantitative proteomics techniques with affinity enrichments. True interactors are separated from background binders via the ratio observed between specific affinity purification and control purification. Furthermore, as every protein is assigned a ratio, it is not necessary to perform rigorous purifications, but instead single step purifications and mild washing conditions can be applied [251]. This strategy opens up the field of interaction proteomics to any kind of protein interaction as long as a bait can be immobilized. High quality results could be achieved for protein-protein interactions [88], modification dependent peptide-protein interactions [79, 212, 252], sequence specific DNA-protein [25, 156] or RNA-protein interactions [26] as well as interactions with small molecules [219]. As discussed above, several semiquantitative approaches based on spectra-counting have been developed [143, 167, 209, 227], however, the focus here will be on SILAC and intensity-based label-

free approaches because they were used in this thesis.

SILAC-based interaction proteomics

SILAC currently offers the best and most reliable quantification method for interaction proteomics. Proteins are already metabolically labeled before the pull-down, thus labeling artifacts or incomplete labeling, which can happen for chemical labeling approaches, are circumvented. The experimental design of SILAC-based interaction screens in general follows the same standard principles: In the forward experiment, the specific pull-down is performed with the heavy and the control pull-down with the light labeled extracts. For the reverse experiment labels are swapped (Figure 1.2.7 A). Pull-downs are generally performed separately to avoid subunit exchange reactions [102]. Beads are pooled after washing and bound proteins are eluted together. Whether samples are fractionated or not depends on the complexity of the sample. For specific elutions single MS runs are usually sufficient.
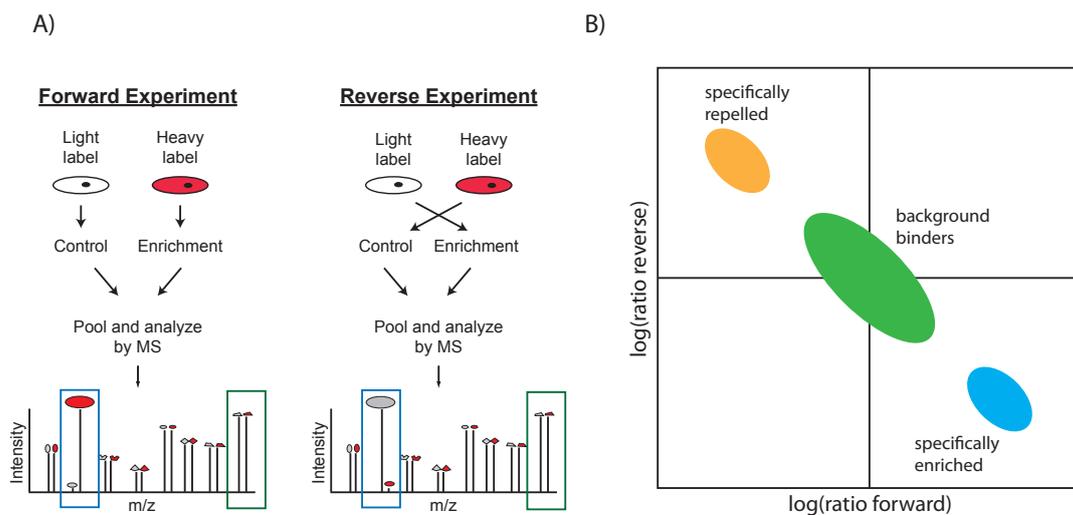
Figure 1.2.7: SILAC based interaction workflows A) Overview of a typical SILAC forward and reverse experiment. B) Ratio-ratio plot of a SILAC interaction experiment. Background cloud in green, proteins specifically binding (blue) can be found in the lower right quadrant, proteins specifically repelled (orange) can be found in the upper left quadrant.

Nonspecific binders show a ratio of around one in both experiments, whereas specific outliers have a high ratio in the forward and a low ratio in the reverse experiment. For visualization, logarithmized ratios from the forward and the reverse experiment are

plotted in a so-called ratio-ratio plot (Figure 1.2.7 B). Specific outliers can be found in the lower right quadrant. In some cases, enrichment on both baits makes biological sense (e.g. when comparing modified and unmodified peptides as baits) and outliers can be found in the lower right and the upper left quadrants. Ideally the background cloud forming around 0 is compressed and dense, and outliers are clearly offset. In almost all cases a visual inspection of the data is sufficient to pinpoint the specific hits, however, if necessary a statistical significance value can provide a robust p-value [25]. In most SILAC interaction studies, two enrichment experiments are compared: a specific pull-down and a control pull-down. For some questions, it is desirable to directly compare two samples to a control. Triple labeling approaches enable such an analysis of three different states. This allows, for example, the dissection of co-operative effects of recruiting modifications on histone tails. Enhanced binding of TFIID to H3K4me3 upon acetylation of K9 and K14 was discovered using a triple labeling strategy [252]. In another study, the composition of the interactome of APC (Adenomatous polyposis coli) and AXIN1, two important proteins in the Wnt pathway, were investigated in their native and stimulated state [86].

Label-free interaction proteomics

Label-free approaches were recently shown to yield comparable results to SILAC-based quantification for interaction studies [88]. Samples from specific pull-down and control are prepared and measured separately, which gives rise to a higher variability. This can be minimized by automation, either by using a robotic workstation [88] or by setting up interaction experiments on 96-well plates as applied in this thesis. Intensity-based label-free quantification was performed by the MaxQuant software platform by computing a label-free intensity at the protein level [136]. In contrast to a SILAC approach which is very intuitive to analyze, label-free approaches need a more complex statistical analysis to define outliers. A modified t-test statistic [245] that takes reproducibility and fold change into account has proven to provide a good separation of outliers from background binding proteins [88]. An additional parameter (termed s_0) is introduced, which puts more weight on the relative difference between the groups.

$$d(i) = \frac{\bar{x}_I(i) - \bar{x}_U(i)}{s(i) + s_0}$$

Although label-free approaches need more replicates and are more complex to analyze than SILAC experiments, they offer some advantages. First of all, any protein source can directly be used. This allows interaction experiments from cells, which are compli-

cated to label and from organisms which are otherwise not accessible for SILAC-based studies. Near unlimited multiplicity is another advantage of label-free approaches. Whereas in SILAC a maximum of three samples can be directly compared, label-free approaches allow the comparison of any number of baits.

Full-length protein-protein interactions

Protein-protein interactions are usually studied by enriching the protein of interest from a cell or tissue extract and analyzing the co-purified proteins. Pull-downs can be performed using antibodies against endogenous proteins, or by expression of tagged proteins and purifying them with an antibody against the tag. For the latter, good antibodies are available and furthermore a generic pull-down setup can be applied, making it very convenient for large scale interaction studies. Using FLAG-tagged protein over-expression, the protein interactions of 5,000 individually tagged *Drosophila* proteins were analyzed [75]. However, the addition of a tag to a protein can interfere with protein-protein interactions by occluding interaction surfaces or by preventing proper protein folding. Moreover, protein over-expression can lead to artifacts. For example, mislocalization of the bait protein into cellular compartments where it normally would not be present, can force unphysiological protein interactions.

Co-IPs using antibodies against endogenous proteins and isotype antibodies as control circumvent many of the abovementioned problems. Endogenous Co-IPs heavily rely on the quality of the antibodies. First of all, antibodies against the bait proteins need to be available. Second, they need to be highly specific as cross-reactions with other proteins would generate artifacts. Although Co-IPs of endogenous proteins are not easily streamlined, a large interaction screen in human cells was recently performed [143]. To overcome the problem of cross-reactivity, the QUICK approach (QUantitative Immuno precipitation Combined with a Knockdown) was developed [216]. By using a cell line in which the protein of interest is knocked down, the same antibody can be used for Co-IP and control. Proteins which cross-react with the antibody will be equally enriched in both purifications, and thereby not lead to false positive interactors.

BAC TransgeneOmics [192] is a powerful method to generate cell lines with tagged proteins at near endogenous expression levels. BACs (bacterial artificial chromosomes), encoding the gene of interest including introns and the gene-specific promoter, are modified by recombineering to include a GFP-tag and an antibiotic resistance marker. These modified BACs are transfected into cell lines where they stably integrate into the genome. As the whole genomic region including the endogenous promoter is used, the resulting tagged proteins are expressed at near endogenous level. Moreover, cell

cycle-dependent protein expression patterns as well as cell type-specific isoforms can be obtained. The GFP tag is a very versatile tool, as it can be used for immunofluorescence and ChIP assays [192]. It is also an excellent tag for protein-protein interaction studies [242]. The combination of BAC TransgenOmics and quantitative mass spectrometry, termed QUBIC (Quantitative BAC InteraCtomics), is a powerful approach in interaction proteomics [88, 89]. Due to the large interest in GFP as a purification tag, GFP nanotraps were developed recently [204]. These are engineered proteins based on a single chain antibody from Llama which show excellent binding affinities to GFP. Due to its much smaller size, nanotraps do not generate as many peptides as normal antibodies which would interfere with the subsequent MS analysis.

Modification-dependent protein-protein interactions

Many protein interactions in a cell are not constitutive, but only take place after a specific stimulus. One way to accomplish this in a cell is by making protein interactions dependent on post-translational modifications. Several protein domains binding to a partner protein in a modification-dependent manner have evolved. For example, SH2 domains specifically bind to phosphorylated tyrosines [179], bromo domains bind to acetylated lysines [47] and many binding domains for methylated lysines are described [237].
A Peptide pull-downs approach using modified and unmodified bait peptides to screen for modification-dependent protein-protein interactions is a robust method. Bait peptides are coupled to beads via a biotin moiety and incubated with protein extracts. Quantitative mass spectrometry (e.g. SILAC) is used to separate background binders from modification-dependent protein interactions [212]. The quantitative read-out is absolutely crucial, as a large number of proteins will bind unspecifically to the unstructured peptide bait [251]. This workflow was successfully applied to study phosphotyrosine binders [79, 217] and readers of lysine trimethylation [252].

2 Results

2.1 Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers

2.1.1 Project aim and summary

Histone modifications play a crucial role in the regulation of chromatin-associated processes and very often serve as binding platforms to recruit so-called chromatin readers. Knowing all the possible binding partners for these modifications is a first prerequisite for a detailed investigation of their functions. Mass spectrometry-based proteomics offers powerful and unbiased methods to identify novel chromatin readers by combining peptide pull-downs of modified and unmodified peptides with a SILAC-based quantitative read-out [212]. The direct binding of the general transcription factor TFIID to H3K4me3 has been discovered in such an approach [252].
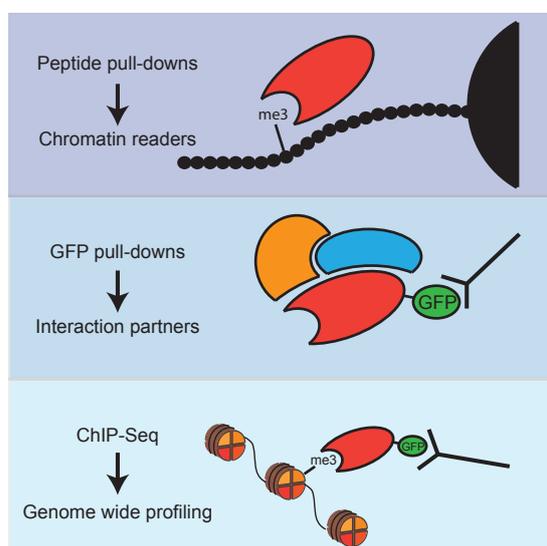
Figure 2.1.1: **Three techniques were combined to study trimethyl lysine readers** Peptide pull-downs were used to identify proteins associated with the respective chromatin marks. GFP pull-downs of selected hits from the peptide pull-downs were used to define chromatin reader complexes. ChIP-Seq of selected chromatin readers was used to verify *in vivo* the interaction with the chromatin mark and to define the genome wide binding pattern.

We used this workflow to screen for chromatin readers of the major trimethylation sites on histone H3 and H4. The screen included the activating H3K4me3 and H3K36me3 marks, as well as the repressive H3K9me3, H3K27me3 and H4K20me3 marks. To as-

sign proteins enriched in our peptide pull-down screens into complexes, we employed the BAC recombineering technology [192] to generate stable cell lines expressing GFP-tagged potential new readers under their endogenous promoter. We then performed GFP pull-downs to identify their interaction partners to define an interaction network. Combining this network with prior biochemical knowledge, we were able to develop hypotheses about direct binders. We demonstrated biochemically that the SAGA complex subunit SGF29 directly binds to H3K4me3 via its double tudor domain, and that the PWWP domain of NPAC is necessary for H3K36me3 binding. We further used our stable cell lines for ChIP-Seq profiling. By comparing the genome-wide binding of chromatin readers to the ChIP-Seq profiles of the actual chromatin marks, we could verify the interaction with the histone modification *in vivo*. In addition, we observed that some H3K4me3 readers were binding to all occurrences of this modification, whereas others only bound a subset. In summary, this was the first large scale screen for chromatin readers of the major trimethyl lysine marks. It combined proteomics, protein biochemistry and ChIP-Seq (Figure 2.1.1) and provided a detailed and unbiased view on five important epigenetic marks.

2.1.2 Contribution

This project was initiated and coordinated by Michiel Vermeulen, who also co-supervised the first part of this PhD thesis. When I joined the project, the initial peptide pull-down screen as well as some biochemical verification was already done. I established the protocol for GFP pull-downs of nuclear protein complexes and performed most protein-protein interaction studies. To increase our confidence in the data, I repeated some of the peptide pull-downs to confirm new chromatin readers. I confirmed the association of the novel chromatin reader NPAC with the H3K36me3 chromatin mark and showed furthermore that the PWWP domain is necessary for binding. Finally, I analyzed the proteomic data and prepared all proteomic figures and tables for the publication.

2.1.3 Publication

This project was published as a Resource article in 2010:

Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers

Michiel Vermeulen[*], H. Christian Eberl[*], Filomena Matarese[*], Hendrik Marks, Sergei Denissov, Falk Butter, Kenneth K. Lee, Jesper V. Olsen, Anthony A. Hyman, Henk G. Stunnenberg and Matthias Mann

[*] these authors contributed equally

Resource

Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers

Michiel Vermeulen,[1,6,7,*] H. Christian Eberl,[1,6] Filomena Matarese,[2,6] Hendrik Marks,[2] Sergei Denissov,[2] Falk Butter,[1] Kenneth K. Lee,[3] Jesper V. Olsen,[1,5] Anthony A. Hyman,[4] Henk G. Stunnenberg,[2,*] and Matthias Mann[1,*]
[1]Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, D-82152 Martinsried, Germany
[2]Department of Molecular Biology, Faculty of Science, Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen, Geert Grooteplein 26 Zuid, 6525 GA Nijmegen, The Netherlands
[3]Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA
[4]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
[5]Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen, Denmark
[6]These authors contributed equally to this work
[7]Present address: Department of Physiological Chemistry and Cancer Genomics Centre, University Medical Center Utrecht, Utrecht, The Netherlands
*Correspondence: m.vermeulen-3@umcutrecht.nl (M.V.), h.stunnenberg@ncmls.ru.nl (H.G.S.), mmann@biochem.mpg.de (M.M.)
DOI 10.1016/j.cell.2010.08.020

SUMMARY

Trimethyl-lysine (me3) modifications on histones are the most stable epigenetic marks and they control chromatin-mediated regulation of gene expression. Here, we determine proteins that bind these marks by high-accuracy, quantitative mass spectrometry. These chromatin "readers" are assigned to complexes by interaction proteomics of full-length BAC-GFP-tagged proteins. ChIP-Seq profiling identifies their genomic binding sites, revealing functional properties. Among the main findings, the human SAGA complex binds to H3K4me3 via a double Tudor-domain in the C terminus of Sgf29, and the PWWP domain is identified as a putative H3K36me3 binding motif. The ORC complex, including LRWD1, binds to the three most prominent transcriptional repressive lysine methylation sites. Our data reveal a highly adapted interplay between chromatin marks and their associated protein complexes. Reading specific trimethyl-lysine sites by specialized complexes appears to be a widespread mechanism to mediate gene expression.

INTRODUCTION

In the eukaryotic nucleus, DNA is wrapped around an octamer of histone proteins, which constitute the nucleosomes. Rather than merely serving as a means to store genetic material, nucleosomes play an active role in regulating processes such as transcription, DNA repair, and apoptosis. The N-terminal tails of the four core histones that protrude from the core structure of the nucleosome are subject to a variety of posttranslational modifications such as acetylation, methylation, and phosphorylation. One role of these modifications is the recruitment of regulatory proteins that in turn exert their function on chromatin (Jenuwein and Allis, 2001; Kouzarides, 2007).

The major lysine methylation sites on the N terminus of histone H3 and histone H4 with a clearly defined biological function are H3K4me3, H3K9me3, H3K27me3, H3K36me3, and H4K20me3, which are associated with different functional states of chromatin. H3K4me3 is almost exclusively found on promoter regions of actively transcribed genes while H3K36me3 is linked to transcription elongation. H3K9me3, H3K27me3, and H4K20me3 are generally found on silent heterochromatic regions of the genome. Part of the functional distinction between these methylation sites relates to the proteins interacting with them. A number of these "chromatin readers" for various histone methyl lysine sites have already been identified and characterized (Kouzarides, 2007; Shilatifard, 2006; Taverna et al., 2007), but this list is unlikely to be exhaustive. To obtain a comprehensive map of the histone methyl lysine interactome, unbiased screening methods are required.

Mass spectrometry (MS)-based proteomics is increasingly used in functional biological studies and has proved to be a powerful tool to characterize histone modifications (Garcia et al., 2007; Vermeulen and Selbach, 2009). For protein-protein interactions a quantitative format is desirable, as this enables to distinguish specific and background binders (Vermeulen et al., 2008). In particular, the technology of stable isotope labeling by amino acids in cell culture (SILAC) (Ong et al., 2002) can be used to expose peptide baits bearing a posttranslational modification to "heavy" SILAC-labeled cell extracts, whereas the unmodified peptide is exposed to "light" labeled cell extract. Binders specific to the modified form of the peptide appear in mass spectra with a significant ratio between heavy and light form of the protein. Using this approach, we discovered that TFIID binds to H3K4me3, thereby providing a link between
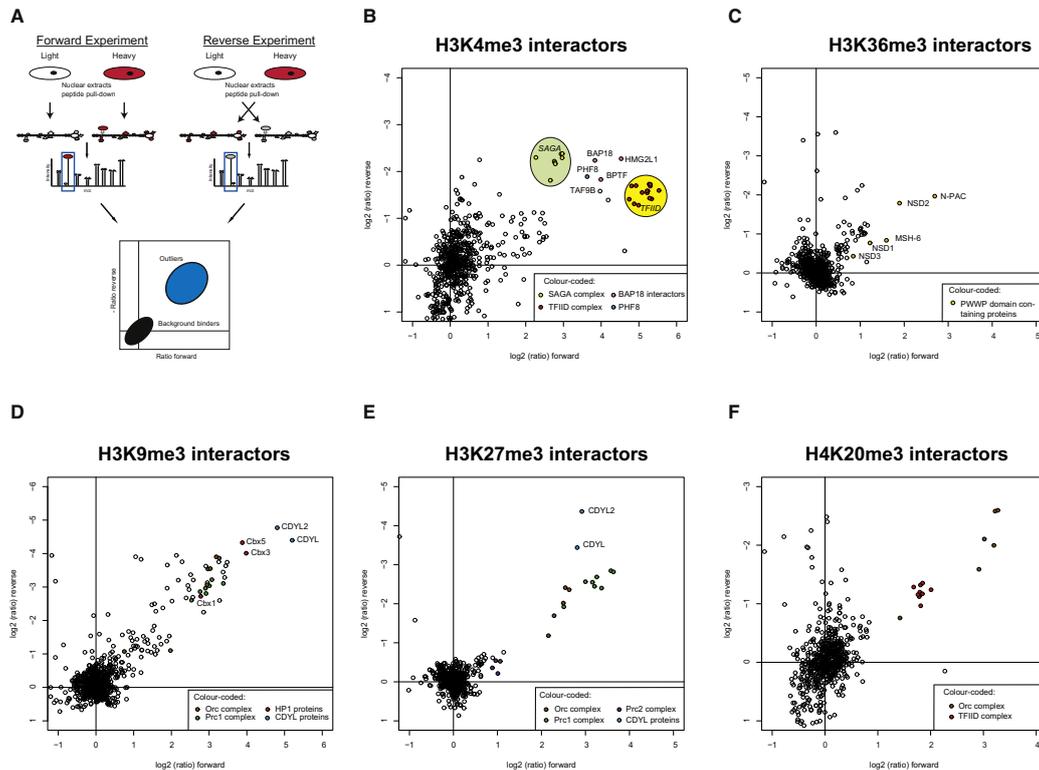
Figure 1. A Histone Peptide Pulldown Approach Using SILAC Technology
(A) Schematic representation of the experimental approach (M indicates trimethyl lysine).
(B) The H3K4me3 interactome. Proteins are plotted by their SILAC-ratios in the forward (x axis) and reverse (y axis) SILAC experiment. Specific interactors should lie close to the diagonal in the upper right quadrant. The two major transcriptional coactivator complexes that were found to interact with this mark (TFIID and SAGA) are encircled. TAF9b, which is localized between TFIID and SAGA in the figure, is a shared subunit between these two complexes.
(C) The H3K36me3 interactome. Proteins carrying a PWWP domain are colored yellow.
(D–F) The interactome of H3K9me3, H3K27me3 and H4K20me3, respectively. Note that the ORC complex, including LRWD1, binds to these three marks. See also Figure S1 and Table S1.

this modification and activation of transcription (Vermeulen et al., 2007).

Here, we refine this technology and perform an unbiased interaction screen for the known activating and repressive trimethyl histone marks on H3 and H4. We apply the BAC-GFP transgeneOmics technology (Poser et al., 2008) to characterize chromatin readers and their complexes. Chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) with the same BAC-GFP lines identifies the in vivo target genes, which are found to overlap with the histone marks they interact with. This integrative approach provides not only an interactome of the studied histone marks, including many previously uncharacterized factors, but also mechanistic insights into epigenetic regulation of gene expression.

RESULTS

A Large-Scale Methyl Lysine Interactome
To characterize the interactome of trimethyl-lysine chromatin marks, we developed an interaction screen based on a recently described technology (Vermeulen et al., 2007). In brief, nuclear extracts derived from HeLaS3 cells grown in "light" or "heavy" medium were incubated with immobilized biotinylated histone peptides (Figure 1A). After incubation, beads from both pull-downs were pooled, run on a one-dimensional PAGE gel, and subjected to in-gel trypsin digestion. The resulting peptide mixtures were measured by high-resolution on-line electrospray MS on a hybrid linear ion trap, Orbitrap (see Experimental Procedures). Computational analysis was done with the MaxQuant

algorithms (Cox and Mann, 2008), which enabled sub parts-per-million mass assignment and accurate quantitation even for very low abundance SILAC pairs. Eluates from methylated and non-methylated peptides each contained hundreds of proteins and are visually indistinguishable on 1D gels (Figure S1A available online). Nevertheless, the SILAC-ratios reliably retrieved specific binders even when they were hundred-fold less abundant than background binders (Figure S1B). We determined the interactome of the two activating marks H3K4me3 and H3K36me3 and three repressive marks, H3K9me3, H3K27me3, and H4K20me3 (Table S1; Figures 1B–1F). Each measurement identified between 600 and 1200 proteins at a confidence level of 99%. Of these, between 10 and 60 had highly significant ratios indicating specific binding to the respective marks.

In our previous study, we identified interactions of members of the TFIID complex with H3K4me3. Here, we performed the interaction screen in the "forward" and "reverse" format to obtain higher discrimination between specific baits and background. The forward experiment consists of incubating the modified peptide with heavy labeled cell lysate and the nonmodified peptide with light labeled cell lysate, whereas in the reverse format the labels are switched. These two experiments also constitute a biological replicate. With a minimum of two quantification events, every significant interactor is supported by at least four quantitative measurements. Plotting interaction data for H3K4me3 in a two-dimensional space and inverting the SILAC-ratios of the reverse experiment places the true interactors into the top right quadrant (Figure 1A). Nonlabeled contaminants, such as keratin and proteins derived from the medium will not change the ratio in the reverse experiment and are located in other quadrants. Furthermore, a number of other proteins, such as polypyrimidine tract-binding protein 2, were automatically filtered out because they show significant ratios only in one of the labeling experiments, and are color coded accordingly in Table S1. In some cases, interactions may be biophysically correct but they may not occur in vivo because of compartmentalization in the cell (for example, mitochondrial hsp60 binding to H3K9me3). We noticed that the entire TFIID protein complex clustered together in the two-dimensional plot, indicating very similar SILAC ratios in the forward and reverse experiments (Figure 1B). This prompted us to inspect the interaction plots for other protein complexes binding to specific chromatin marks.

Sgf29 Links the Human SAGA Complex to H3K4me3
The measured H3K4me3 interactome contained eight subunits of the human SAGA complex, which tightly clustered together in the two-dimensional plot (green circle in Figure 1B). Inspection of the sequences of all known SAGA subunits revealed a double Tudor domain in the C terminus of Sgf29 (Figure 2A). Double Tudor domains are known to have affinity for H3K4me3 (Huang et al., 2006). We therefore speculated that Sgf29 could be the subunit within the SAGA complex that directly binds to H3K4me3. To address this question, we used RNAi to knock down Sgf29 in HeLa cells (Figure 2B). The nuclear extracts derived from these cells as well as nuclear extracts derived from cells transfected with control oligonucleotides were used for peptide pulldowns. Western blotting shows that the SAGA subunit GCN5 only binds to H3K4me3 and not to H3K4me0

(Figure 2B). This binding is abolished upon knockdown of Sgf29, while GCN5 levels in these cells are similar to those in the cells treated with mock siRNA. These experiments also imply that, at least in mammalian cells, Sgf29 is responsible for the observed interaction between H3K4me3 and SAGA, and not CHD1, as has been suggested in yeast (Pray-Grant et al., 2005).

To biophysically characterize this interaction, we expressed Sgf29 as a recombinant protein in *E. coli* and used the induced bacterial lysates for histone peptide pulldowns. As shown in Figure 2C, Sgf29 binds to histone H3 peptides, with a clear preference for H3K4me3. This binding is specific as no interaction with other histone lysine methylation sites such as H3K9me3 or H3K36me3 was observed. Sgf29 binds to both H3K4me2 and H3K4me3 with a slight preference for H3K4me3 (Figure 2D). Based on sequence alignments between yeast, *Drosophila* and human Sgf29 we selected conserved and nonconserved residues for mutational analyses (Figure 2A). Results of nine pulldown experiments revealed that conserved residues in the second Tudor domain of Sgf29 are particularly important for H3K4me3 binding. As expected, mutating nonconserved residues did not affect the binding (Figure 2E). We used isothermal calorimetry experiments to measure the affinity of the interaction between Sgf29 and H3K4me3 (Figure 2F). The binding constant of 4 μM is comparable to that of other trimethyl-lysine marks to their readers and in particular to the interaction constant of the Tudor domain of JMJD2A, which is 10 μM (Huang et al., 2006). No affinity between Sgf29 and the unmethylated histone H3 peptide could be observed. Together, these results demonstrate that the human SAGA complex binds to H3K4me3 and that the double Tudor domain in its subunit Sgf29 is both necessary and sufficient to mediate this interaction.

Functional Insights into Chromatin Readers Using BAC transgeneOmics
Our screening of the H3K4me3 and H3K36me3 interactome, two lysine methylations associated with actively transcribed genes, revealed a large number of chromatin readers of unknown function. To gain insight into the molecular mechanism of their interaction with the lysine methylation sites, we tagged a selection of these proteins with GFP using the recently developed BAC transgeneOmics technology (Poser et al., 2008). In this strategy, a GFP-tagged fusion of the protein of interest is stably integratedpreserving the endogenous genomic context—in HeLa cells by recombineering (Zhang et al., 1998). Fusion proteins are therefore expressed at near endogenous levels, as demonstrated previously (Poser et al., 2008). Furthermore, we tested expression levels of several of the GFP-tagged BAC lines and found very similar expression levels to the endogenous proteins (Figures S2E–S2H).

Quantitative SILAC-based GFP pulldowns employing wild-type parental cells as control were optimized such that protein complexes can be identified and visualized in a single two hour MS analysis without the need to separate proteins on an SDS PAGE gel (Hubner et al., 2010). As a proof of principle we applied this workflow to the K4me3 binding protein Sgf29, which is known to assemble into either the SAGA or the ATAC complex (Nagy et al., 2010). Both SAGA and ATAC complex subunits copurified with GFP-Sgf29 demonstrating the applicability of single
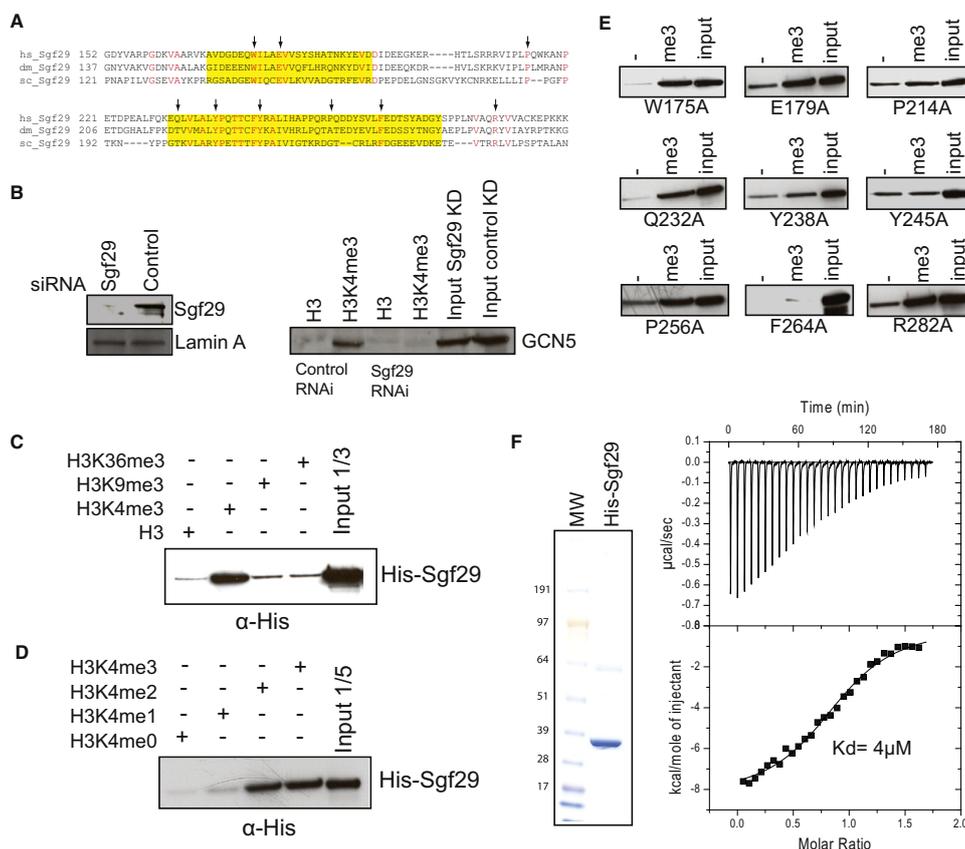
Cell

Figure 2. Sgf29 Links the SAGA Complex to H3K4me3

(A) Alignment of the C-terminal part of human, Drosophila, and yeast Sgf29. Tudor domains are indicated in yellow.

(B) siRNA experiments followed by peptide pulldowns show that Sgf29 links the SAGA complex to H3K4me3.

(C and D) Bacterial lysates expressing recombinant his-tagged Sgf29 were incubated with the indicated peptides. Following incubation and washes, the amount of bound Sgf29 protein was determined by western blotting using an anti-His antibody.

(E) Bacterial lysates expressing the indicated Sgf29 mutants were used for histone peptide pulldowns to determine their binding affinity for H3K4me3. The first lane represents peptides without the me3 modifications.

(F) Isothermal calorimetry experiment revealing the affinity of the full-length Sgf29 protein for H3K4me3.

step GFP affinity purification to identify protein-protein interactions for chromatin readers (Figure 3A; Table S2). We then applied this approach to the as-yet uncharacterized protein C17orf49, which we had found to interact with H3K4me3 (Figure 1B). C17orf49 is an 18 kDa protein that carries a SANT domain, which commonly occurs in chromatin associated proteins. Pulldown of the GFP fusion protein from stably transfected HeLa cells specifically copurified subunits of the human NuRF/BPTF complex (Figure 3B; Table S2). Strikingly, HMG2L1, another highly significant interactor of H3K4me3 (Figure 1B) is one of the most prominent interactors of

C17orf49. Thus, this experiment established C17orf49 and HMG2L1 as subunits of the human NuRF/BPTF complex. Their association with H3K4me3 is explained by their interaction with the H3K4me3 reader BPTF. We name the uncharacterized open reading frame C17orf49 as "BPTF associated protein of 18 kDa" (BAP18).

GATA zinc finger domain containing 1 (GATAD1) is another protein of unknown function that was identified as a H3K4me3 interactor. Using the GFP pulldown approach, we identified subunits of the Sin3b/HDAC complex, the H3K4me3-specific lysine demethylase Jarid1A/RBBP2, and the breast cancer
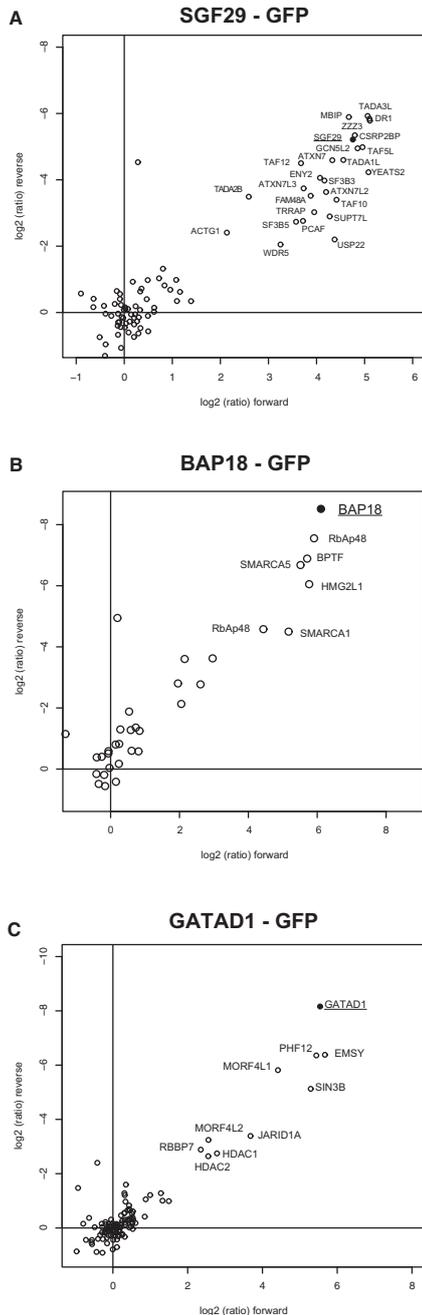
A SGF29 - GFP

B BAP18 - GFP

C GATAD1 - GFP

associated protein EMSY (Hughes-Davies et al., 2003) as interactors for GATAD1 (Figure 3C; Table S2). Because all of the subunits in this complex were identified as H3K4me3 readers with similar ratios, we hypothesized that they form an as-yet uncharacterized chromatin reading complex (Figures S2A–S2D). Jarid1a was recently reported to bind tightly to H3K4me3 with a Kd of 0.75 μM (Wang et al., 2009a) and therefore forms the direct link between the complex and the chromatin mark. Further evidence for our hypothesis comes from a subsequently published *Drosophila* Lid complex (Lee et al., 2009; Moshkin et al., 2009). Lid is the *Drosophila* homolog of the mammalian Jarid1 family of proteins, consisting of Jarid1a, Jarid1b and Jarid1c. The complex furthermore contains homologs of the Sin3 proteins, as well as an EMSY and GATAD1 homolog. In mammals, interactions between the Sin3/HDAC complex and Jarid1a have also been reported (van Oevelen et al., 2008). However, EMSY has not been tied to any of these proteins yet. EMSY is known to be a repressor of transcription (Hughes-Davies et al., 2003) but the mechanisms underlying this repressive activity are poorly understood. The identification of the above-described complex provides important clues as to how EMSY represses transcription. We hypothesize that gene repression involves histone deacetylation coupled with H3K4me3 demethylation.

Localizing the Chromatin Readers on the Genome

To further investigate the function of our proteins of interest in vivo, we performed ChIP-Seq profiling using an anti-GFP antibody on the BAC-GFP lines. Figure 4A shows a representative snapshot of the ChIP-Seq data. Profiling of GFP-tagged proteins interacting with H3K4me3 and H3K36me3 was performed on biological replicas and showed that the approach is highly reproducible (Pearson correlation >0.85; Figures S3F and S3G). In agreement with our peptide pulldown data, the identified H3K4me3 readers Sgf29, TRRAP, PHF8, GATAD1, and BAP18, are associated mainly with promoters (Figures S3A and S3B) and coincide with H3K4me3 marking (Figures 4B and 4C; Figure S3C). We also identified a small number of binding sites of H3K4me3 readers outside of annotated promoters (Figure S3A). As these are not associated with H3K4me3 (Figure S3B), the interactor proteins are apparently recruited to these loci by H3K4me3 independent mechanisms. Nevertheless, for each of these five proteins we observed a good genome-wide correlation with H3K4me3 (Pearson correlation BAP18: 0.71, GATAD1: 0.71, PHF8: 0.66, TRRAP: 0.66, SGF29: 0.55).

For Sgf29, TRRAP, and BAP18, it was expected that they would localize to promoters, as they are part of conserved complexes associated with active transcription– SAGA/ATAC, SAGA/NuA4, and BPTF/NuRF, respectively (Nagy et al., 2010; Wysocka et al., 2006). PHD finger protein 8 (PHF8) belongs to

Cell
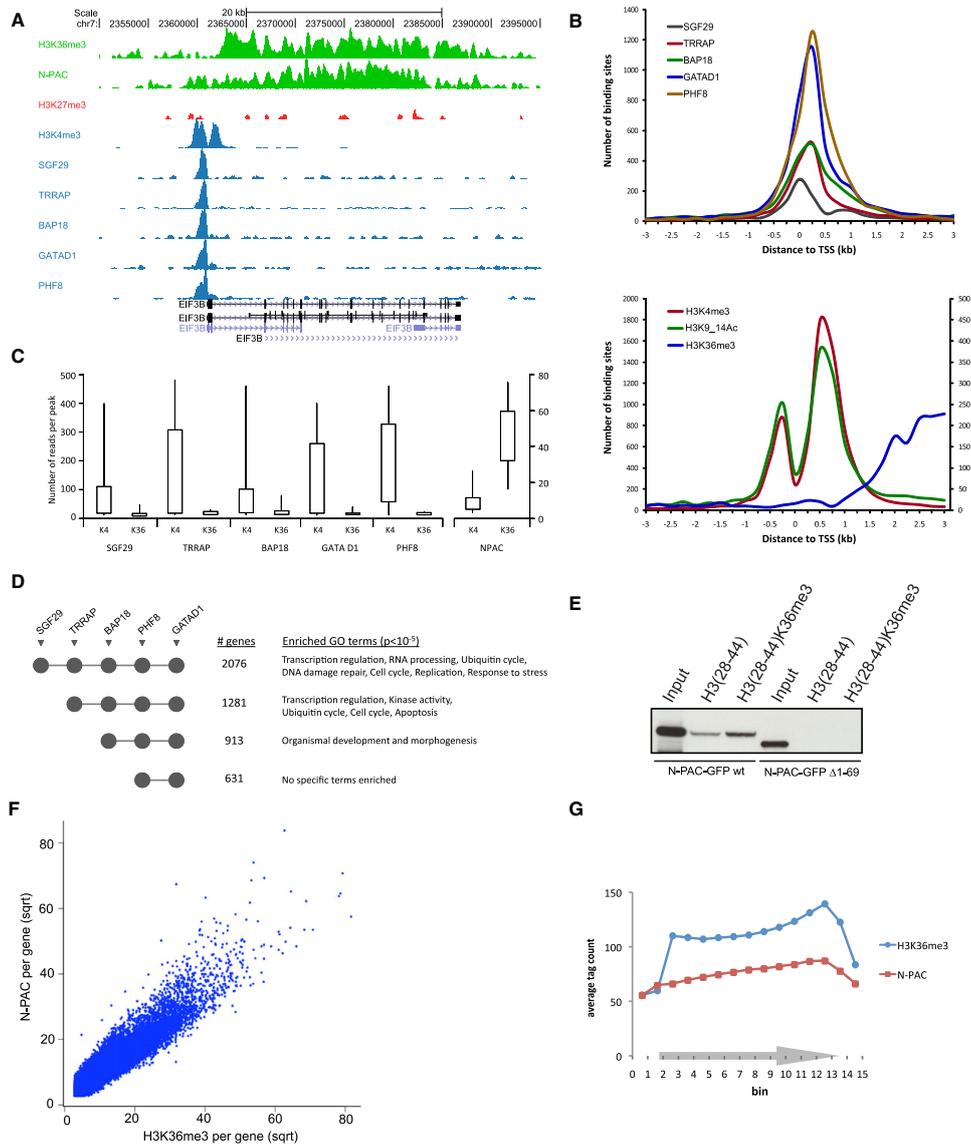
Figure 4. ChIP Sequencing of H3K4me3 and H3K36me3 Readers
(A) ChIP-Seq profiles of three histone modifications and the interactors across the Eif3B gene on human chromosome 7.
(B) Distance distribution of the binding sites for the H3K4me3 interactors and the three histone modifications relative to the closest transcription start site (TSS). x axis is in 1000 bp; on the y axis the number of binding sites is indicated. Values for H3K36me3 are plotted on a separate scale (right side).
(C) Number of reads for H3K4me3 and H3K36me3 (indicated with K4 and K36, respectively) within the binding sites for the H3K4me3 interacting proteins. The ends of the whiskers represent the 9th and 91st percentile, respectively. Values for SGF29, TRRAP, BAP18, PHF8, and GATAD1 are on the scale on the left side of the plot, while values for N-PAC are on a separate scale on the right.
(D) Promoters clustered by the binding sites for the H3K4me3 interacting proteins (Figure S3). Co-occurrence of binding sites is indicated with gray circles under

the JmJc domain-containing family of proteins that can remove methyl groups from arginine or lysine residues (Cloos et al., 2008). PHF8 can remove the repressive mark H3K9me2 (Horton et al., 2010), associating it with activation of transcription, which is in agreement with our ChIP-Seq analyses.

We found GATAD1 to interact with Jarid1a/EMSY/Sin3 (Figure 3C). Jarid1a is a JmJc domain-containing protein that demethylates H3K4me3 (Cloos et al., 2008). In addition, the GATAD1 purification enriched for components of the Sin3/HDAC transcriptional corepressor complex, including two histone deacetylases, HDAC1 and HDAC2. Despite the repressive enzymatic activities associated with GATAD1, our ChIP-Seq analysis reveals that this complex binds to promoters marked with H3K4me3. These data may be explained by invoking a mechanism of cyclical recruitment of "writers" and "erasers" to sites of active transcription (Wang et al., 2009c).

Interestingly, our ChIP-Seq analyses showed that many target genes can be occupied by each of the five H3K4me3 readers. Analysis of all identified target genes resulted in four discrete clusters (Figures S3D and S3E; Table S3). PHF8 and GATAD1 were the only factors found to be common to all clusters and therefore are likely to have a general role in transcription. The two largest clusters combined genes whose promoters were bound by Sgf29 and/or TRRAP, indicating that transcriptional regulation of these genes involves SAGA/NuA4-related complexes. Gene ontology (GO) annotation of the genes in these clusters revealed a number of highly enriched ($p < 10^{-5}$) functional terms that agree very well with the biological functions of these complexes (Figure 4D). For example, SAGA/ATAC and NuA4 complexes are crucial regulators of transcription, DNA repair, DNA replication, and the cell cycle (Squatrito et al., 2006). Distinct GCN5/PCAF-containing complexes function as coactivators and are involved in transcription factor and global histone acetylation (Nagy and Tora, 2007). SAGA was shown to regulate various stress-response genes (Huisinga and Pugh, 2004; Nagy et al., 2010), while TRRAP-containing complex NuA4 regulates apoptosis (Ikura et al., 2000; Tyteca et al., 2006). Thus, each functional category of the GO analysis corresponds to an established function of the SAGA and NuA4 complex, which independently validates the connection between the activating histone mark and its reader found in our experiments.

N-PAC, MSH-6, and NSD1 as well as NSD2 were identified as H3K36me3 interactors (Figure 1C; Table S2). Interestingly, these four proteins share a PWWP domain which is part of the Tudor domain "Royal Family" and includes the Tudor, chromo and MBT domains that can interact with methylated lysine residues. The PWWP domain of Set9 was recently identified as a reader for H4K20me1 (Wang et al., 2009b). Our peptide pulldown data

suggest that this domain is also capable of recognizing H3K36me3, which is associated with elongation of transcription and peaks in coding regions of genes (Shilatifard, 2006). Very recently the PWWP domain of Brpf1 was shown to bind specifically to H3K36me3 (Vezzoli et al., 2010). Indeed, deletion analyses revealed that the PWWP domain of N-PAC is necessary for H3K36me3 binding (Figure 4E). This PWWP domain mediated K36me3 binding is most likely direct, since purification of N-PAC-GFP from a BAC line did not reveal protein-protein interactions (data not shown). To investigate the genomic binding pattern of N-PAC, we generated the corresponding BAC-GFP line and performed ChIP-Seq analysis. Consistent with our peptide pulldown data, N-PAC binds to coding regions of active genes correlating with the presence of H3K36me3 (Figures 4C and 4F). N-PAC and H3K36me3 increase toward the 3′ end (Figures 4A and 4G). Together our data establish the PWWP domain as a putative binder of H3K36me3. In addition to a PWWP domain, N-PAC also contains an AT-hook that is often found in proteins that are associated with elongation of transcription and an enzymatic domain of unknown function. Our ChIP-Seq analysis revealed that both H3K36me3 and N-PAC are present almost exclusively over gene bodies (data not shown), and that the vast majority of H3K36me3 marked regions are also bound by N-PAC, indicating a broad or universal function of this protein in transcriptional elongation.

The Interactome of the Repressive Histone Methyl Marks

We next investigated the chromatin readers of H3K9me3, H3K27me3 and H4K20me3, histone methyl marks associated with gene repression (Figures 1D–1F). H3K9me3 yielded the richest set of interactors, including all three HP1 isoforms (CBX1, CBX3, and CBX5). The chromodomain-containing HP1 proteins are classical readers of H3K9me3 (Jenuwein and Allis, 2001) and our analysis confirms that they are restricted to this repressive modification. Two chromodomain proteins, CDYL and CDYL2, were identified as binders for both H3K9me3 and H3K27me3 but not H4K20me3. These proteins are members of a family of three chromodomain proteins, the third one being chromodomain Y protein, whose gene is located on the Y chromosome and whose expression is testis specific. Recently, direct binding of CDYL and CDYL2 to H3K9me3 and H3K27me3 has been reported (Fischle et al., 2008; Franz et al., 2009). As expected, Polycomb group proteins represent the major readers for H3K27me3, but many of these proteins were also identified as specific interactors for H3K9me3. Given the high degree of sequence identity surrounding H3K9 and H3K27 (TARKST and AARKSA for K9 and K27, respectively), it is not surprising to find Polycomb group proteins as interactors

the corresponding interactor names. Four major groups of promoters were identified, for which the number of genes within each group and highly enriched GO terms (p value $< 10^{-5}$) are listed.

(E) Full-length N-PAC-GFP and Δ1-69 N-PAC-GFP were transfected into HeLa Kyoto cells. Extracts from these cells were subsequently used for K36/K36me3 peptide pulldowns. Unlike the wild-type protein, Δ1-69 N-PAC-GFP, that lacks most of the PWWP domain, does not bind to H3K36me3.

(F) Dotplot showing the correlation between H3K36me3 and N-PAC ($R^2 = 0.86$). Every dot represents the number of N-PAC or H3K36me3 ChIP-Seq tags per gene.

(G) All genes containing H3K36me3 (>5 kb) were each divided in 15 bins followed by counting and averaging of the H3K36me3 and N-PAC ChIP-Seq tags within each bin.

See also Figure S3 and Tables S3 and S4.

Cell

for H3K9me3. Literature evidence also supports the interaction of Polycomb group proteins with H3K9me3, although their affinity for H3K27me3 is higher (Fischle et al., 2003b; Ringrose et al., 2004). Finally, we identified the origin recognition complex (ORC) as an interacting complex for all three repressive sites.

We purified complexes associated with the HP1 family members to ascertain if the H3K9me3 readers physically interact with them using BAC-GFP constructs (Figures 5A–5C). Among the specifically interacting proteins, known HP1 interactors were identified, such as chromatin assembly factors CHAF1A/ CHAF1B and ADNP (Lechner et al., 2005; Mandel et al., 2007). Two uncharacterized proteins, POGZ and Znf828, consistently interacted with high ratios with all HP1 family members. We confirmed the binding of POGZ to H3K9me3 by western blotting (Figure S1C). POGZ and Znf828 have an interesting domain structure and multiple zinc fingers, suggesting that these proteins may specifically bind DNA sequences. POGZ or POGO transposable element with a ZNF domain is a 1410 amino acid protein containing two domains that are also present in the centromeric protein B (CenPB). Next, we generated BAC-GFP constructs for these proteins. Pulldowns with POGZ and Znf828 reciprocally confirmed interaction with HP1 and, interestingly, with each other (Figures 5D and 5E). Additionally, POGZ interacted specifically with mitotic spindle checkpoint protein, Mad2l2. To substantiate this possible connection to a prominent cell cycle protein, we performed a GFP pulldown with a cell line of this protein, which clearly demonstrated reciprocal binding (Figure 5F). Thus, a combination of repressive mark interactors and full-length protein interactomes allows us to deconstruct the majority of protein interactions involved in the biology of the repressive marks.

We noticed that LRWD1 clusters together in the two-dimensional interaction plots with the ORC complex in the pulldowns of each of the repressive marks (Figures 1D–1F). LRWD1 has not been characterized but obtains its name from a leucine-rich repeat and a stretch of WD40 domains. To test if this protein is a subunit of the ORC complex, we generated the BAC-GFP cell line of Orc2L. Pulldown with this ORC subunit indeed demonstrated specific interaction with LRWD1 (Figure 5G). Furthermore, ChIP-Seq of the BAC LRWD1-GFP line revealed a strong enrichment on satellite repeats, correlating with high levels of H3K9me3 which is known to be enriched over satellites (Figure 5H) (Martens et al., 2005).

Triple SILAC Pulldowns Reveal Differential Fine-Tuning of Trimethyl Lysine Binding

The five trimethyl lysine marks that we screened for interactors are flanked by numerous residues that can also be subjected to posttranslational modifications. These modifications could, either agonistically or antagonistically, affect trimethyl lysine binding. To study such potential interplay between different posttranslational modifications (PTMs) occurring in close proximity on the histone H3 tail, we applied triple pulldown experiments involving a combination of methylation and other PTM marks, in this case acetylations or phosphorylations (Vermeulen et al., 2007). In this approach, cells are grown in three different SILAC media, each containing different stable isotopic versions of lysine and arginine. These extracts, which are distinguishable

by MS, are each incubated with a differently modified histone peptide (triple pulldown). Peptides appear as triplets in the MS spectra and a significant ratio between the first two peaks indicates specific binding to the H3K4me3 mark. The highest mass peak in the triplet originates from the eluate of the combinatorially modified peptide and its intensity compared with the eluate from the singly modified peptide (middle peak) indicates either agonistic or antagonistic binding or no effect. On genes that are actively being transcribed, H3K4me3 often co-occurs with acetylation of H3K9 and H3K14. A number of readers for H3K4me3 carry both a domain that recognizes H3K4me3 as well as one or multiple bromodomains, which bind to acetylated lysine residues. We therefore wondered whether these acetylations would function agonistically with H3K4me3 to bind H3K4me3 readers to the histone H3 tail. Consistent with our previous findings (Vermeulen et al., 2007), TFIID and BPTF bound more strongly to the H3K4me3 mark when it was flanked by acetylation on H3K9 and H3K14 acetylation (Figures S4A and S4B). In addition, we also observed—by quantitative proteomics and by western blotting—agonistic binding to the methylated and acetylated peptide for the SAGA complex (Sgf29 in Figures 6A and 6C). In contrast, recombinant Sgf29 does not display preferential H3K9,14Ac binding (Figure 6D), indicating that the observed effects in the triple pulldown are due to the agonistic binding effects of the Sgf29 double Tudor domain and the GCN5 bromodomain. Finally, we also observed agonistic binding of PHD finger protein 8 (PHF8) to H3K4me3 and H3K9,14 Ac (Figure 6B). PHF8 carries an H3K4me3-binding PHD finger (Horton et al., 2010), but it does not contain a bromodomain. Therefore, we hypothesize that this protein either carries an unidentified acetyl lysine binding motif, or interacts with an as-yet unidentified bromodomain-containing protein. These results indicate that agonistic H3K4me3 and H3K9,14Ac recognition occurs in several chromatin readers. The mechanisms are diverse; for example, a PHD finger domain can be combined with a bromodomain in one protein (BPTF), or in different subunits of the same complex (TAF3 PHD finger and TAF1 bromodomains in the TFIID complex). Moreover, a different recognition domain combination can be used (Tudor domain of Sgf29 with the bromodomain of GCN5 in the SAGA complex). Clearly, these chromatin readers have each evolved the ability to target combinatorially marked nucleosomes allowing regulation of specific subsets of genes.

To study potential antagonistic histone PTM crosstalk, we decided to focus on phosphorylations on the histone H3 tail. Phosphorylation of histone H3S10 results in the release of HP1 from chromatin during mitosis even though levels of H3K9me3 remain unchanged (Fischle et al., 2005). H3K27me3 is also flanked by a serine residue that can be phosphorylated (Winter et al., 2008). To investigate if these trimethylations co-occur with the respective adjacent phosphorylations, we analyzed our recent large-scale study of the proteome and the phosphoproteome of the cell cycle (Olsen et al., 2010). Indeed, we found the corresponding doubly modified peptides. Moreover cell cycle data indicates that they are specific for mitotic cells (Figures S4G–S4J). As shown in Figure 6E, H3S10 phosphorylation does not appear to drastically affect the binding of HP1 to H3K9me3. These results are in agreement with data reporting
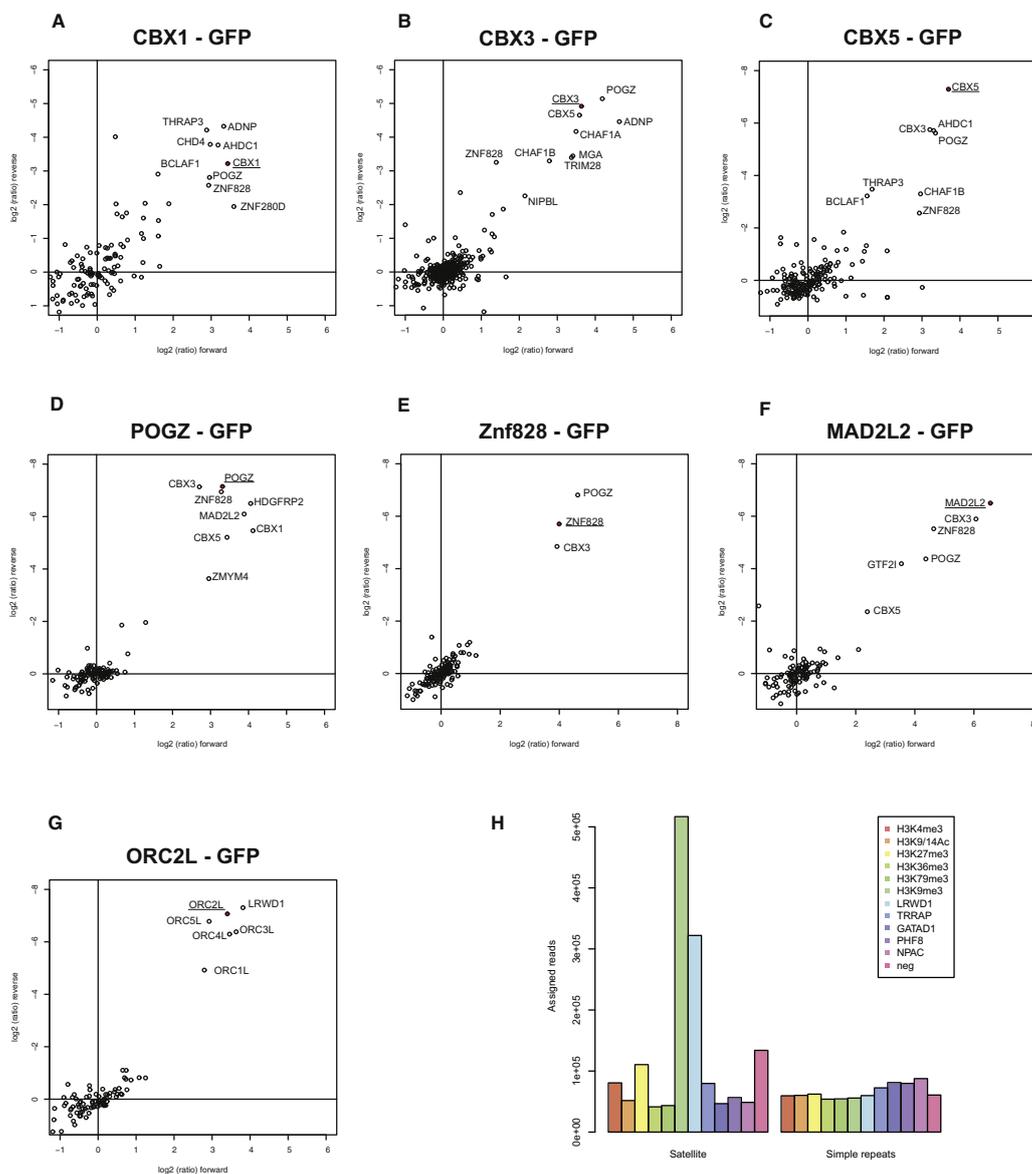
Figure 5. GFP Pulldowns for Readers of the Repressive Histone Marks

(A–G) GFP-fusions proteins expressed in SILAC-labeled HeLa cells were enriched on GFP-nanotrap beads. In each figure, the ratio of the identified proteins in the forward and reverse pulldown is plotted. Proteins interacting with the baits are indicated.

(H) The total number of ChIP-Seq reads present on either satellite repeats or simple repeats for the indicated proteins and histone marks is shown.

See also Table S2.

Figure 6. Triple SILAC Pulldowns Revealing Histone Modification Crosstalk

(A) Three-dimensional representation of the MS signal of an Sgf29 peptide identified in a triple pulldown SILAC experiment (the m/z scale is the x axis, the chromatographic retention time is the y axis, and the MS-signal is the z axis). Each group of signals represents the natural isotope pattern of the peptide. The relative intensities of the triplet peak of the Sgf29 peptide indicates the preference of binding to the modification states (unmethylated histone H3 peptide [left peak], H3K4me3 peptide [middle peak], and the double-modified H3K4me3/H3K9,14 Ac peptide [right peak]).

(B) Same as (A) for a PHF8 peptide identified in the same triple pulldown.

(C) Nuclear extracts derived from HeLa cells were incubated with the indicated histone peptides. The amount of Sgf29 protein bound to these peptides was determined by western blotting using an antibody against endogenous Sgf29.

(D) Bacterial lysates expressing recombinant His-tagged Sgf29 were incubated with the indicated peptides. Following incubation and washes, the amount of bound Sgf29 protein was determined by western blotting using an anti-His antibody. Note that Sgf29 does not bind to a peptide containing H3K9,14 acetylation and that the binding of Sgf29 to H3K4me3 is not affected by asymmetric dimethylation of H3R2.

(E–G) Three-dimensional representation of an HP1α (E), Orc5 (F), and CDYL (G) peptide identified in a triple pulldown SILAC exp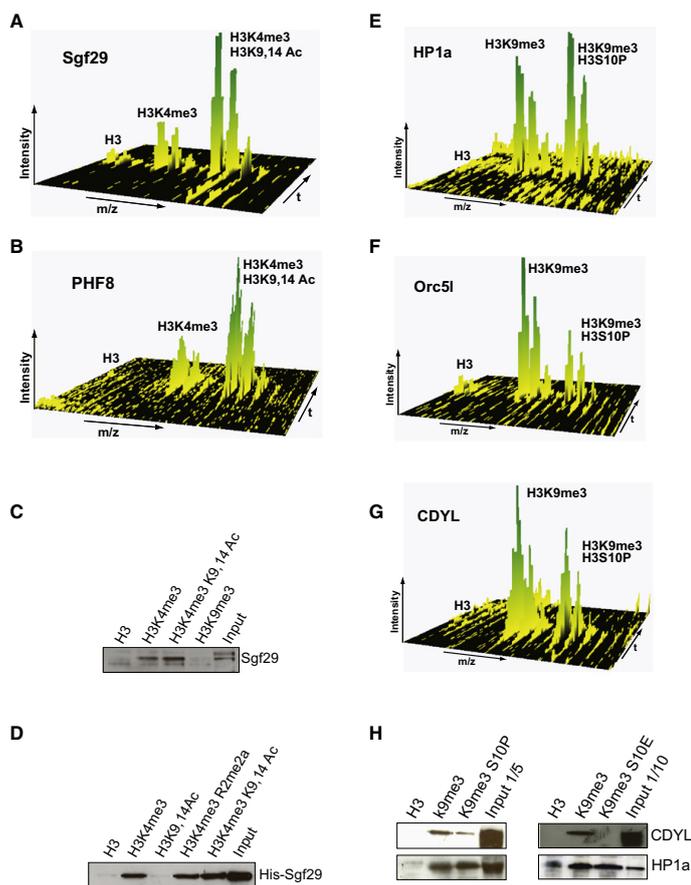eriment. The spectra show the MS-signal representing the relative binding of these peptides to the unmethylated histone H3 peptide (left peak), the H3K9me3 peptide (middle peak), and the double-modified H3K9me3/H3S10P peptide (right peak).

(H) Histone peptide pulldowns in HeLa nuclear extracts were performed with the indicated peptides. The amount of HP1α and CDYL binding to these peptides was determined by western blotting using an antibody against HP1α and CDYL. See also Figure S4.

stabilization of HP1 binding by H3S10 phosphorylation (Mateescu et al., 2004). Indicating that our assay can indeed reveal antagonistic effects, we observed that CDYL as well as the ORC complex subunits do show reduced H3K9me3 binding in combination with H3S10 phosphorylation (Figures 6F and 6G). These experiments were further confirmed by western blotting, also making use of a phosphomimetic peptide where H3S10 was mutated to glutamic acid (Figure 6H). Similarly, H3S28 phosphorylation destabilizes the binding of CDYL and ORC complex subunits to H3K27me3, whereas this phosphorylation only mildly affects the binding of Polycomb group proteins (Figures S4C–S4F). Taken together, these results suggest that phosphorylations on the N-terminal tails of histones selectively affect the binding of proteins to adjacent modified lysines residues. Such so-called phospho-methyl switches are quite common on core histones (Fischle et al., 2003a). We have also identified H3S57 and H3T80 as phosphorylation sites on histone H3 (for H3S57P and H3T80, Figures S4K and S4L), both of which

are adjacent to modified lysine residues. Thus, almost all of the modified lysine residues on histone H3 can be flanked by phosphorylated residues. An important function of these phosphorylation sites could be the differential regulation of protein binding to neighboring methylated or acetylated lysines in specific cellular situations and for specific genes.

DISCUSSION

Here, we have characterized the association of chromatin readers with histone trimethyl-lysine modifications by a combination of three technologies. The major findings from our integrated approach are visualized and summarized in Figure 7. High-accuracy, quantitative proteomics based on SILAC identified known and previously unknown binders to each of the chromatin marks. Plotting SILAC ratios from forward and reverse experiments grouped distinct protein clusters together, representing functional complexes. To investigate these complexes, we turned
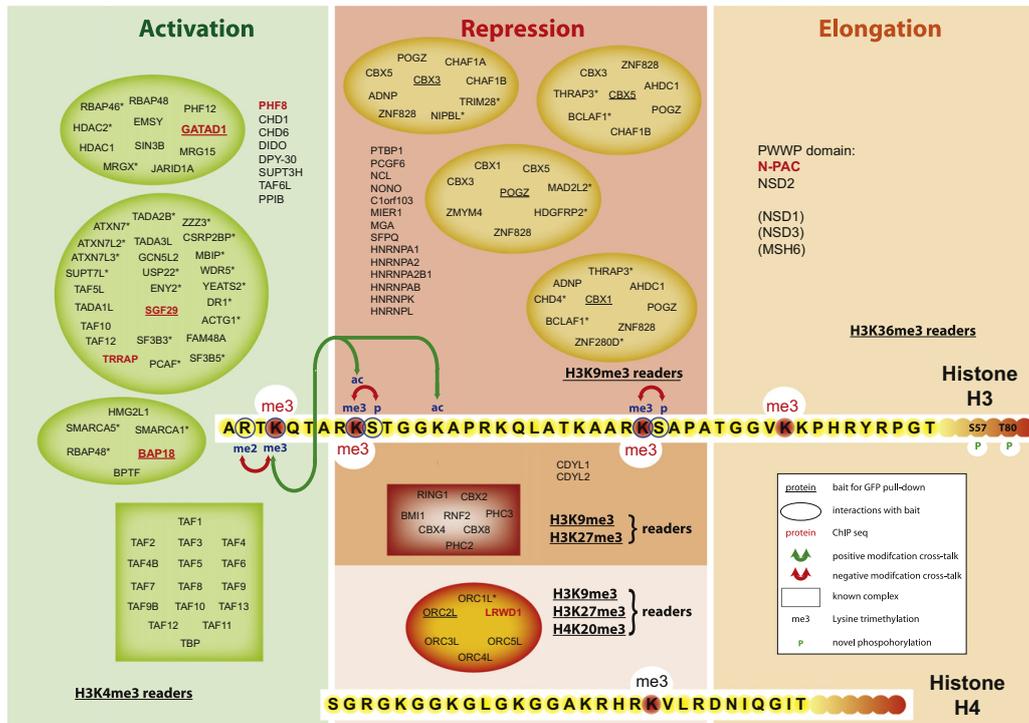
Figure 7. Visualization of the Histone Trimethyl-Lysine Interactome

Proteins interacting with the five trimethyl lysine marks are indicated. Encircled are proteins that were additionally identified in GFP pulldown experiments; baits in these pulldowns are underlined. Proteins in those circles marked with an asterisk were not identified as interactors in the peptide pulldowns. Proteins clustered in rectangles were identified in the peptide pulldowns and were previously shown to interact with each other (TFIID for H3K4me3 and PRC1 for H3K9me3 and H3K27me3). For proteins that are color coded red in vivo verification by ChIP-Seq is also provided. The arrows and associated labels indicate histone modification crosstalk investigated in this study. In the globular part of histone H3, two identified histone phosphorylations (H3S57P and H3T80), are indicated.

to the recently developed BAC-transgeneOmics technology (Poser et al., 2008), which allowed rapid generation of stable cell lines containing the entire gene of interest fused to GFP in its endogenous context. Therefore, this technology provides a generic "handle" for the members of chromatin reader complexes while maintaining endogenous control. We used these cell lines in a next round of SILAC-based quantitative interaction screens to establish physical interactions between the chromatin readers. Furthermore, the GFP-tag was utilized for chromatin immunoprecipitation followed by next generation DNA sequencing to localize the readers on the genome. The synergistic use of these three approaches enabled us to create data sets and reagents that provide a resource for researchers interested in epigenetic questions. While shown here for histone modifications, our approach can be extended to posttranslational modifications on other chromatin-associated proteins and to other cellular systems such as stem cells. Illustrating the usefulness of this resource, we were able to dissect several

mechanisms of chromatin reader associations with their chromatin marks starting from basic interaction data.

One such example is the human SAGA complex, all identified members of which clustered tightly in the two-dimensional interaction plot (Figure 1B). SAGA is a highly conserved complex, which plays key roles in the activation of transcription of RNA polymerase II target genes. However, the mechanisms of activation are not completely understood. In yeast, it has been suggested that CHD1 links the complex to H3K4me3 (Pray-Grant et al., 2005). However, this association is controversial as it has been reported that yeast CHD1 does not bind to H3K4me3 (Sims et al., 2005). While we identified human CHD1 as a specific binder to this mark, it did not co-cluster with the SAGA subunits in our H3K4me3 peptide pulldowns. Furthermore, we were not able to identify CHD1 as an interactor of the SAGA subunit Sgf29 in a GFP pulldown. Instead, starting with the observation that Sgf29, which we identified as a H3K4me3 interactor, has a double Tudor domain (Lee and Workman, 2007) and given

Cell

the fact that Tudor domains can bind methylated lysines (Huang et al., 2006), we established by biochemical and biophysical means that the double Tudor domain of Sgf29 forms the direct molecular link between SAGA and H3K4me3. This binding mode is likely conserved down to yeast, which has a homolog of Sgf29 that also contains a double Tudor domain (Figure 2A). Such conservation is not universal as it is not the case for association of TFIID with H3K4me3. This interaction is mediated by the PHD-finger domain of human TAF3, but yeast TAF3 lacks the PHD-finger domain (Vermeulen et al., 2007).

Bioinformatic analysis of the interactors of the activating H3K36me3 mark revealed that four of the most prominent specific interactors shared the same domain. This PWWP domain is part of the Tudor domain "Royal family" of domains (Maurer-Stroh et al., 2003) and therefore almost certainly mediates direct binding to H3K36me3. In agreement with this, deletion analysis revealed that the PWWP domain of N-PAC is essential for its interaction with H3K36me3 (Figure 4E).

In the interactome of the repressive marks we identified, in addition to expected heterochromatin associated proteins, several other proteins. Interaction studies with BAC GFP-fusion proteins uncovered many interactions with members of the HP1 family. This HP1 family and associated proteins represent a large portion of the H3K9me3 interactome and establish the HP1 proteins as interaction hubs in mediating repressive gene functions. Interestingly, several HP1 interactors contain zinc finger domains (such as POGZ and Znf828), which may serve to recruit HP1 to specific sites in the genome.

The origin recognition complex (ORC) has a key function in replication firing. It is known to localize to heterochromatic regions (Prasanth et al., 2004) and it interacted with all three repressive marks. LRWD1 grouped with the ORC complex members in the two-dimensional interaction plots. Pulldowns with an Orc2L BAC-GFP cell line demonstrated that LRWD1 is indeed an ORC complex subunit and ChIP-Seq experiments established that it co-enriches with H3K9me3 on satellite repeats. The WD40 repeat domain of LRWD1 may mediate the interaction of the ORC complex with the repressive marks as it was recently shown that the WD40 repeats of the Polycomb group protein EED directly binds to H3K27me3 (Margueron et al., 2009).

A triple-encoding variant of the SILAC peptide pulldown allowed us to directly address the question of agonistic and antagonistic binding to combinatorial histone modifications. These experiments recapitulated several known combinatorial interactions, such as the agonistic effects between H3K4me3 and nearby acetylations. The general conclusion from these experiments is that the trimethyl marks constitute the major docking sites for chromatin readers and that other nearby modifications fine-tune these primary interactions by augmenting or destabilizing specific interactions. For example, our data show that H3S10 phosphorylation destabilizes the ORC complex and CDYL binding to H3K9me3, whereas HP1 binding does not appear to be affected. Consistent with this paradigm, we have not been able to determine specific interactors with peptides bearing only the ancillary modifications. This is unlikely to be an artifact due to pulldowns with synthetic peptides because similar results are obtained when performing pulldowns

with entire immobilized nucleosomes carrying particular epigenetic marks (T. Bartke, M.V., M.M., and T. Kouzarides, unpublished data). In this context, mass spectrometry can also contribute by identifying and quantifying the combinatorially modified peptides in vivo, as shown for several examples here.

A striking finding that emerges from our integrative investigation into the nature of the relationship between histone marks and their readers is the degree of overlap between the known biological functions of the marks and the biological functions of their associated readers (Figure 7). Histone modifications are usually studied by techniques such as ChIP, ChIP-Seq, or immunofluorescence that associate them with particular genes or nuclear processes. The same holds true for transcription factors or other chromatin regulators. By its nature, our strategy combines investigation of chromatin marks and transcriptional regulators and is thereby uniquely suited as an integrative tool for the investigation of epigenetic regulation of gene expression.

EXPERIMENTAL PROCEDURES

Recombinant Protein Expression and ITC Calorimetry

Full-length Sgf29 constructs were expressed with an N-terminal His-tag and a maltose binding protein (MBP) domain using expression plasmid pETM44 (Novagen). For histone peptide pulldown experiments, crude induced bacterial lysates were used as described (Vermeulen et al., 2007). His-tag westerns were performed using a penta-His antibody (QIAGEN). For isothermal calorimetry (ITC) experiments the Sgf29 protein was enriched using Ni NTA beads after which the protein was further purified on a Superdex 200 column. ITC measurements were performed on a VP-ITC Microcal calorimeter (Microcal, Northampton, MA) at 25°C. During titration, 7 μl of H3K4me3 peptide (aa 1–17) at a concentration of 300 μM was injected into a solution of 25 μM Sgf29 protein.

GFP Pulldowns

Generation of the BACs with GFP-fusion constructs was done as described (Poser et al., 2008). Nuclear extracts from BAC-GFP-tagged or wild-type HeLa cells were SILAC labeled with heavy lysine (Isotec, Sigma). For CBX3, no BAC was available and SILAC-labeled HeLa cells were transfected with plasmid pBCHGN-CBX3 (Addgene). GFP nanotrap beads (Chromotek) were used to precipitate GFP-tagged proteins from these lysates. Approximately 500–1000 μg of nuclear extract was used per pulldown in a buffer containing 300 mM NaCl, 0.25% NP40, 0.5 mM DDT, 20 mM HEPES KOH (pH 7.9), and protease inhibitors. Following incubation and washes with the same buffer, beads from both pulldowns were combined, proteins were eluted with acidic glycine (0.1 M [pH 2.0]) and digested overnight with LysC (Wako Biochemicals, Japan) using the FASP protocol (Wisniewski et al., 2009) prior to LC/MS-MS analysis.

Mass Spectrometry of Proteins

Gel lanes representing each pulldown were cut into eight equally sized slices as described (Vermeulen et al., 2007). Peptide identification was performed on an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany) essentially as described (Olsen et al., 2004). Full-scan MS spectra were acquired with a resolution of 60,000 in the Orbitrap analyzer. For every full scan, the five most intense ions were fragmented in the linear ion trap. Raw data were processed and analyzed using the MaxQuant software (version 1.0.12.33) and searched with the Mascot search engine against a human IPI database 3.52 as described (Butter et al., 2010). Phosphopeptide enrichment of core histones and MS analysis of these were performed as described (Hurd et al., 2009).

Deposition of MS-Related Data

Mass spectrometric data for peptide pulldowns and GFP pulldowns, consisting of raw data files, unfiltered "proteingroups" tables, and identified

Cell

peptides, can be accessed at the TRANCHE repository (https://proteomecommons.org/) under the name "Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers."

Chromatin Immunoprecipitation and Deep Sequencing
ChIP experiments were performed using 3.3×10^6 cells per ChIP according to standard protocols (Denissov et al., 2007), with two minor modifications. Crosslinking of the cells was done on the culture plates for 20 min, while ChIP'ed DNA was purified by Qiaquick PCR purification Kit (QIAGEN cat. no. 28106). ChIP enrichment levels were analyzed by qPCR using specific primers (available upon request) for quality control. ChIP-Seq samples were prepared and analyzed according to the manufacturer (Illumina). Enriched regions were identified by FindPeaks (Fejes et al., 2008). Table S4 summarizes the ChIP-Seq output. For the repeat analysis of the H3K9me3 and LRWD1 ChIP-Seq profiles, mappings were performed by maq aligner (Li et al., 2008). For further information about the ChIP-Seq methods and data analysis see Extended Experimental Procedures. All ChIP-Seq data are present in the NCBI GEO SuperSeries GSE20303.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and five tables and can be found with this article online at doi:10.1016/j.cell.2010.08.020.

REFERENCES

Butter, F., Kappei, D., Buchholz, F., Vermeulen, M., and Mann, M. (2010). A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. EMBO Rep. *11*, 305–311.

Cloos, P.A., Christensen, J., Agger, K., and Helin, K. (2008). Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev. *22*, 1115–1140.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Denissov, S., van Driel, M., Voit, R., Hekkelman, M., Hulsen, T., Hernandez, N., Grummt, I., Wehrens, R., and Stunnenberg, H. (2007). Identification of novel functional TBP-binding sites and general factor repertoires. EMBO J. *26*, 944–954.

Fejes, A.P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M., and Jones, S.J. (2008). FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. Bioinformatics *24*, 1729–1730.

Fischle, W., Wang, Y., and Allis, C.D. (2003a). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475–479.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003b). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Fischle, W., Franz, H., Jacobs, S.A., Allis, C.D., and Khorasanizadeh, S. (2008). Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. J. Biol. Chem. *283*, 19626–19635.

Franz, H., Mosch, K., Soeroes, S., Urlaub, H., and Fischle, W. (2009). Multimerization and H3K9me3 binding are required for CDYL1b heterochromatin association. J. Biol. Chem. *284*, 35049–35059.

Garcia, B.A., Shabanowitz, J., and Hunt, D.F. (2007). Characterization of histones and their post-translational modifications by mass spectrometry. Curr. Opin. Chem. Biol. *11*, 66–73.

Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y., and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat. Struct. Mol. Biol. *17*, 38–43.

Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748–751.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J. Cell Biol. *189*, 739–754.

Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S.F., Milner, J., Brown, L.A., Hsu, F., Gilks, B., et al. (2003). EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell *115*, 523–535.

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol. Cell *13*, 573–585.

Hurd, P.J., Bannister, A.J., Halls, K., Dawson, M.A., Vermeulen, M., Olsen, J.V., Ismail, H., Somers, J., Mann, M., Owen-Hughes, T., et al. (2009). Phosphorylation of histone H3 Thr-45 is linked to apoptosis. J. Biol. Chem. *284*, 16575–16583.

Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell *102*, 463–473.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

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

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd. (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem. Biophys. Res. Commun. *331*, 929–937.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284–295.

Lee, N., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2009). The H3K4 Demethylase Lid Associates with and Inhibits Histone Deacetylase Rpd3. Mol. Cell. Biol. *29*, 1401–1410.

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.

Mandel, S., Rechavi, G., and Gozes, I. (2007). Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. Dev. Biol. *303*, 814–824.

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

Cell

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. *24*, 800–812.

Mateescu, B., England, P., Halgand, F., Yaniv, M., and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO Rep. *5*, 490–496.

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem. Sci. *28*, 69–74.

Moshkin, Y.M., Kan, T.W., Goodfellow, H., Bezstarosti, K., Maeda, R.K., Pilyugin, M., Karch, F., Bray, S.J., Demmers, J.A., and Verrijzer, C.P. (2009). Histone chaperones ASF1 and NAP1 differentially modulate removal of active histone marks by LID-RPD3 complexes during NOTCH silencing. Mol. Cell *35*, 782–793.

Nagy, Z., and Tora, L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. Oncogene *26*, 5341–5357.

Nagy, Z., Riss, A., Fujiyama, S., Krebs, A., Orpinell, M., Jansen, P., Cohen, A., Stunnenberg, H.G., Kato, S., and Tora, L. (2010). The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. Cell. Mol. Life Sci. *67*, 611–628.

Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol. Cell. Proteomics *3*, 608–614.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. *3*, ra3.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics *1*, 376–386.

Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Methods *5*, 409–415.

Prasanth, S.G., Prasanth, K.V., Siddiqui, K., Spector, D.L., and Stillman, B. (2004). Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. EMBO J. *23*, 2651–2663.

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434–438.

Ringrose, L., Ehret, H., and Paro, R. (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol. Cell *16*, 641–653.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu. Rev. Biochem. *75*, 243–269.

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

Squatrito, M., Gorrini, C., and Amati, B. (2006). Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends Cell Biol. *16*, 433–442.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Tyteca, S., Vandromme, M., Legube, G., Chevillard-Briet, M., and Trouche, D. (2006). Tip60 and p400 are both required for UV-induced apoptosis but play antagonistic roles in cell cycle progression. EMBO J. *25*, 1680–1689.

van Oevelen, C., Wang, J., Asp, P., Yan, Q., Kaelin, W.G., Jr., Kluger, Y., and Dynlacht, B.D. (2008). A role for mammalian Sin3 in permanent gene silencing. Mol. Cell *32*, 359–370.

Vermeulen, M., and Selbach, M. (2009). Quantitative proteomics: a tool to assess cell differentiation. Curr. Opin. Cell Biol. *21*, 761–766.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vermeulen, M., Hubner, N.C., and Mann, M. (2008). High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. Curr. Opin. Biotechnol. *19*, 331–337.

Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Gottgens, B., and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. Nat. Struct. Mol. Biol. *17*, 617–619.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009a). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847–851.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R., III, and Jia, S. (2009b). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Mol. Cell *33*, 428–437.

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

Winter, S., Fischle, W., and Seiser, C. (2008). Modulation of 14-3-3 interaction with phosphorylated histone H3 by combinatorial modification patterns. Cell Cycle *7*, 1336–1342.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.

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

Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. *20*, 123–128.

2.2 A map of general and specialized chromatin readers in mouse tissues generated by highly sensitive, label-free interaction proteomics

2.2.1 Project aim and summary

Quantitative interaction proteomics has been an invaluable tool for the discovery of chromatin readers. After the success of our trimethyl lysine interactome (2.1.1), we wanted to shift from cell lines to tissues to discover chromatin readers that are not present in standard *in vitro* systems. Our initial plan was to use the SILAC mouse [116] and perform peptide pull-downs from SILAC labeled mouse tissue extracts. However, due to a limited amount of available heavy mice, we decided to establish a label-free pipeline instead. In label-free approaches many peptide baits can be directly compared and we established a generic workflow that is applicable for any tissue from any organism. First, we demonstrated that we could obtain comparable results as with the established SILAC-based work-flow in our label-free approach. We then used this pipeline to screen for readers of the activating H3K4me3 and the repressive H3K9me3 mark from the four different mouse tissues brain, kidney, liver and testis. In these experiments, we obtained the currently most comprehensive list of chromatin readers for these marks from a proteomic screen. Interestingly, the majority of enriched proteins (direct binders and associated complex members) did not show differences between the tissues, arguing for general functions of these readers and reader complexes. Nevertheless, we detected several organ-specific chromatin readers, such as the brain-specific NuRD complex subunit CHD5, and several testis specific readers, like MBD3L or SSTY1 and SSTY2. In this project we significantly enlarged the list of chromatin readers obtained in our previous study [250]. In addition, we demonstrated the feasibility to move away from pure cell culture-based screens to tissue-based approaches (moving from *in vitro* to *in vivo*).

2.2.2 Contribution

The initial plan for this project was developed by me and Michiel Vermeulen. Almost all experiments were designed, carried out and analyzed by myself. The GFP pull-downs for ZMYND8 and ZNF687 were performed by Cornelia C. Spruijt according to a protocol that I developed previously [250]. The pull-down setup on 96-well plates was developed together with Christian D. Kelstrup.

2.2.3 Publication

The project was published as an Resource article in 2013:

A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics

H. Christian Eberl, Cornelia G. Spruijt, Christian D. Kelstrup, Michiel Vermeulen and Matthias Mann

Mol Cell. 2013 Jan 24;49(2):368-78

A Map of General and Specialized Chromatin Readers in Mouse Tissues Generated by Label-free Interaction Proteomics

H. Christian Eberl,[1] Cornelia G. Spruijt,[2] Christian D. Kelstrup,[3] Michiel Vermeulen,[2,*] and Matthias Mann[1,*]
[1]Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Am Klopferspitz 18,
D-82152 Martinsried, Germany
[2]Department of Molecular Cancer Research, University Medical Center Utrecht, Utrecht, The Netherlands
[3]Department for Proteomics, NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3b,
DK-2200 Copenhagen, Denmark
*Correspondence: m.vermeulen-3@umcutrecht.nl (M.V.), mmann@biochem.mpg.de (M.M.)
http://dx.doi.org/10.1016/j.molcel.2012.10.026

SUMMARY

Posttranslational modifications on core histones can serve as binding scaffolds for chromatin-associated proteins. Proteins that specifically bind to or "read" these modifications were previously identified in mass spectrometry-based proteomics screens based on stable isotope-labeling in cell lines. Here we describe a sensitive, label-free histone peptide pull-down technology with extracts of different mouse tissues. Applying this workflow to the classical activating and repressive epigenetic marks on histone H3, H3K4me3, and H3K9me3, we identified known and putative readers in extracts from brain, liver, kidney, and testis. A large class of proteins were specifically repelled by H3K4me3. Our screen reached near-saturation of direct interactors, most of which are ubiquitously expressed. In addition, it revealed a number of specialized readers in tissues such as testis. Apart from defining the chromatin interaction landscape in mouse tissues, our workflow can be used for peptides with different modifications and cell types of any organism.

INTRODUCTION

The genetic information of eukaryotes is stored in the nucleus by wrapping the DNA around octamers of histone proteins, forming the basic building blocks of chromatin, the nucleosomes (Luger et al., 1997). Besides compacting and storing DNA, nucleosomes play an active role in regulated processes such as transcription and DNA repair. Post-translational modifications (PTMs) of the N-terminal tails of the core histones often serve as docking sites for "chromatin readers," which can subsequently modify chromatin in *cis* or directly activate or repress transcription (Kouzarides, 2007). Prominent examples include the binding of HP1 proteins to H3K9me3 (K9me3) or the wide variety of H3K4me3 (K4me3) binding modules like, e.g., BPTF

(Li et al., 2006), ING proteins (Peña et al., 2006), SGF29 (Vermeulen et al., 2010), or PHF8 (Feng et al., 2010). A number of reader domains have evolved that recognize specific PTMs in a protein sequence. These domains form special binding pockets, which probe the surrounding amino acid sequence in addition to containing a very selective interaction surface discriminating the unmodified from the modified state of a specific amino acid (Taverna et al., 2007).

Histone modifications and their readers play important roles during cellular differentiation and development and in tumorigenesis (Berdasco and Esteller, 2010; Wang et al., 2009). They contribute to maintaining gene expression differences between tissues. Even at the bulk histone levels, differences in the modification pattern between tissues can be observed (Garcia et al., 2008). Clearly the repertoire of chromatin readers and associated proteins varies between cell types and developmental stages. A classical example is the PHD finger-containing protein RAG2, which is expressed in B cells during VDJ recombination. Its binding to K4me3 is crucial for the recombination event that these cells undergo during maturation (Matthews et al., 2007). Currently it is not known if RAG2 is an example for a larger group of specific chromatin readers or a specialized exception.

Mass spectrometry (MS)-based proteomics has played a crucial role in defining the global histone modification landscape in cells and in characterizing the subunit composition of chromatin-related protein complexes (reviewed in Eberl et al., 2011). A principal strength of MS-based methods is that they are hypothesis free, making them well suited to discovering new interactors (Vermeulen et al., 2008). The combination of histone peptide pull-downs from crude nuclear extracts with quantitative MS is a particularly powerful approach to identify novel chromatin readers. Pull-downs are performed with modified and unmodified peptides, and a quantitative filter distinguishes specific PTM readers from the vast amount of background binders that are typically present. We first applied this approach in HeLa cells that were metabolically labeled as heavy or light using SILAC (Ong et al., 2002) to identify TFIID as a reader for K4me3 (Vermeulen et al., 2007) and later characterized readers for five major lysine trimethylation sites on histone H3 and H4 (Vermeulen et al., 2010). Similar workflows

CrossMark

56

Molecular Cell

Tissue-Specific Chromatin Readers

identified proteins that specifically recognize combinations of histone modifications and DNA methylation (Bartke et al., 2010), and enabled the study of interactions with reconstituted modified nucleosomal arrays (Nikolov et al., 2011).

All of the abovementioned studies were performed in a single cancer cell line, which restricted the identifiable interactors to proteins and protein complexes expressed in that system. Because reader complexes could differ by cell type and tissue or developmental stage, we wished to remove this limitation and develop a label-free technology that would be applicable to any sample and organism. Investigation of the binding to the activating K4me3 and the repressive K9me3 mark across tissues resulted in a very high coverage of known reader complexes, most of which are ubiquitously expressed in all the tissues we screened. We also observe a large group of proteins that are repelled by the K4 trimethyl mark as well as tissue-specific subunits of chromatin reader complexes. Whereas the majority of chromatin reader complexes is conserved between tissues, some of the ubiquitously expressed chromatin reader complexes have evolved to contain tissue-specific subunits, which could enable regulation of tissue-specific target genes or fine-tune enzymatic activities. Some of these tissue-specific subunits of chromatin-reading complexes are DNA binding transcription factors which may serve to recruit reader complexes to tissue-specific target genes in the genome.

RESULTS

A Label-free Interaction Pipeline Allows Rapid Screening for Chromatin Readers

Our previous workflow required individual analysis of each pull-down including separation by 1D gel electrophoresis followed by LC-MS/MS analysis of eight fractions (Vermeulen et al., 2007, 2010). Here we placed Sepharose beads in wells with a coarsely meshed bottom, which are impenetrable for aqueous solutions under normal conditions but enable liquid removal by slow centrifugation. This allowed switching to a 96-well format, increasing throughput and reproducibility. Furthermore, we made use of the increased sequencing speed of a linear ion trap—Orbitrap mass spectrometer (Olsen et al., 2009)—as well as longer gradients, to reduce the measurement of pull-downs to single LC-MS/MS runs. Finally, we replaced isotope-based quantification by a sophisticated label-free quantification algorithm within the MaxQuant software suite (Luber et al., 2010).

To test this workflow, we performed SILAC-based and label-free peptide pull-downs in parallel for K4me3 readers from a mouse liver cell line (Table S1). The SILAC experiment was done in forward (i.e., incubating the modified peptide with the heavy and the unmodified peptide with the light extracts) and reverse (swapping of the labels). We found 46 proteins to be enriched and 23 proteins to be repelled by K4me3; these outliers encompassed many of the known K4me3 interactors (Figure 1A). Label-free pull-downs were performed in triplicate and analyzed by a modified t test (Tusher et al., 2001) (Figure 1B). The K4me3 mark enriched 49 proteins and specifically repelled 18. The large majority of the outliers were found in both experiments (blue in Figure 1C). Several proteins were only identified

or quantified in one of them (green in Figure 1C). In accordance with a previous comparison (Hubner et al., 2010), the larger dynamic range of the label-free experiment led to proteins only identified in this set of experiments (red in Figure 1C), whereas the higher quantitative accuracy of SILAC ensured statistical significance for borderline cases. For instance, the K4me3 interactor MORC3 or the K4me3-associated EMSY was significant in the SILAC experiment but close to threshold in the label-free experiment. The fact that some proteins are outliers in one experiment but not the other is expected based on the different statistical behavior of binders in label-free and SILAC analysis. Overall, we concluded that label-free quantification is a viable alternative to SILAC for discovering chromatin readers, especially if quantitative accuracy is further boosted by increasing the number of replicates.

Having established a label-free high-throughput histone peptide pull-down interaction screening platform, we decided to use it to screen for tissue-specific chromatin readers of the key activating and repressive histone modifications K4me3 and K9me3, respectively. Nuclear extracts were prepared from pooled mouse brain, liver, and kidney, and these were separately incubated with unmodified and K4me3- and K9me3-modified peptides (Figure 1D). Every pull-down was analyzed in quintuplicate to maximize statistical significance.

We tested significant binding between the three possible pairs of bait peptides for each organ (nine t test comparisons). Hierarchical clustering of all outliers generated in this way showed distinct groups (Figure 1E): enriched on K4me3 (115 proteins), enriched on K9me3 (64 proteins), and de-enriched on K4me3 (41 proteins) (Table S1).

Inspecting the group of proteins significantly binding to these chromatin marks, we found almost only proteins annotated to be nuclear and very few apparent interactors from unexpected cellular compartments. Of the 31 K4me3 binders found by both Vermeulen et al. (Vermeulen et al., 2010) and Nikolov et al. (Nikolov et al., 2011), our tissue-based screen included 28. For the repressive K9me3 mark, these studies had only 14 interactors in common, of which 11 are statistically significant in our data set. Thus our tissue-based screen appears to have reached very high coverage of previously established chromatin readers.

As an example of a tissue that cannot easily be mimicked in cell culture, we chose testis. This is a particularly interesting system to study chromatin readers, as sperm maturation and concomitant massive chromatin remodeling take place in this organ. Although nucleosomes are replaced to a large extent by protamines during sperm maturation, conventional histones, histone variants, and modifications such as K4me3 can still be detected in mature sperm cells in developmentally important loci (Hammoud et al., 2009). Because of the relatively low tissue mass, we performed pull-downs from total tissue extract. Although the different extraction procedure precludes a direct comparison to the pull-downs with the other organs, many of the same interactors were found, showing that chromatin readers can efficiently be retrieved even from total tissue extracts available in small amounts. In total we found 21 proteins associated with K4me3, 29 proteins associated with K9me3, and 19 proteins being repelled by K4me3 in testis (Figure 1F; Table S1).

58
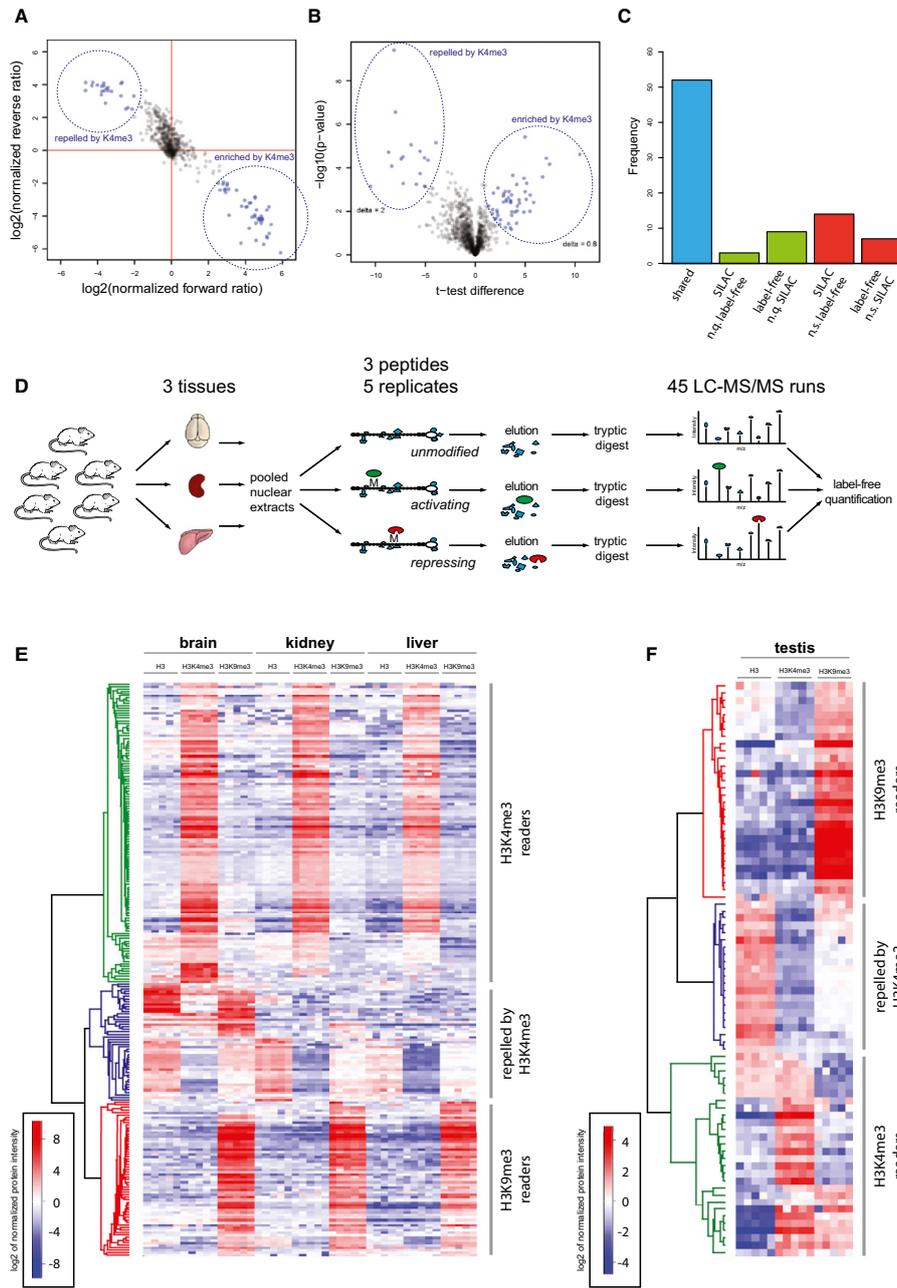
Molecular Cell
Tissue-Specific Chromatin Readers

General and Organ-Specific Chromatin-Associated Complexes

The large majority of reader proteins were found as specific binders in all three organs studied. Table 1 lists these proteins, grouped into known chromatin reader complexes where possible. We found ten such complexes for the K4me3 mark, and in most of these cases the entire set of established complex members were found as significant interactors. This indicates that our screen reached unprecedented coverage. Interestingly, the SET1 complex, which itself methylates H3K4, was one of the complexes bound to K4me3. In yeast, direct binding of SET1 complex member SPP1 to H3K4me3, which recruits yeast SET1, has been described (Shi et al., 2007); however, in mammals none such interaction has been described yet. We therefore tested the PHD finger of the complex member CXXC1 for binding to K4me3 and indeed observed a specific interaction with H3K4me3 (Figure 2A). Moreover, overexpressed CXXC1 devoid of the PHD finger still interacts with Set1 (Figure 2B). Furthermore, it shows a dominant-negative effect on Set1 binding to the H3K4me3 peptide (Figure 2C). Thus we conclude that CXXC1 recruits SET1 to H3K4me3.

The proteins associated with K9me3 encompass most of the known direct readers of this modification, including several that were only described very recently (Table 1). As expected among the specific binders to this repressive mark were many Polycomb group members as well as many HP1 interactors reported in a recent HP1 interactome study (Nozawa et al., 2010). It is noteworthy that both among the already known and the newly described K9me3-associated proteins were many with zinc finger motifs. These proteins could couple a DNA sequence specific readout of the repressive mark in a similar manner as already described for the HP1 interactor POGZ (Nozawa et al., 2010).

We tested several of the outliers of specific interest as well as some completely uncharacterized proteins by western blotting. In each of the cases, the western blot verified the result of our global analysis (Figure 2E).

Next, we inspected our quantitative data for tissue-specific chromatin readers and associated proteins. MS and western blotting found ZNF462 as a specific binder to K9me3 in brain and kidney but not in liver, where this protein appears not to be expressed (Figures 2D and 2E). ZNF462 is a zinc finger protein with a role in development (Massé et al., 2011), and its knockdown leads to mislocalization of HP1 alpha (Massé et al., 2010). In conjunction with the enrichment of ZNF462 on K9me3, this suggested that it is an HP1 alpha interactor. Indeed

ZNF462 is present in HP1 alpha immunoprecipitations from brain and kidney, but not from liver extracts (Figure 2D). Thus we conclude that ZNF462 is a tissue-specific and restricted HP1 interactor.

In brain extracts but none of the other extracts, CHD5 was enriched with the unmodified and K9me3-modified peptide as compared to K4me3. This was also confirmed by western blotting, which furthermore indicated absence of the protein in the input material in kidney and liver extracts (Figure 2E). To obtain insights into the function of CHD5, we performed interaction proteomics with the above-described platform but coupling an antibody against CHD5 to the beads. Members of the NuRD complex (MBD2/3, MTA1/2/3, GATAD2A/B, HDAC1/2, and RBBP7) were significantly enriched, except for CHD3 and CHD4 (Figure 2F). Together with a very recent report (Potts et al., 2011), this demonstrates that CHD5 is a member of a NuRD-like complex. The NuRD complex represses transcription by nucleosome remodeling and deacetylation (Tong et al., 1998; Xue et al., 1998). As its interaction with the H3 tail is mediated by the two PHD fingers of CHD3 or CHD4 (Mansfield et al., 2011), neither of which interacted with CHD5, we tested if CHD5 could take over this function. We expressed the PHD fingers of CHD5 and found that both bind to the unmodified peptide and are repelled by K4me3 (Figure 2G). The binding pattern of the CHD5 PHD fingers mirrors that of CHD4, whose two PHD fingers bivalently recognize both H3 tails on a single nucleosome (Musselman et al., 2012). We hypothesize that CHD5 takes the position of CHD3 or CHD4 in a neuronal NuRD complex and that it is responsible for binding to the H3 tail.

Several readers were exclusively found in testis, reflecting the unique chromatin-remodeling events in spermatogenesis. Among the known testis-specific readers and associated proteins, we detected MBD3L, a testis-specific NuRD subunit (Jiang et al., 2004) that clusters with other NuRD complex members in the typical repulsion pattern from K4me3. TRIM66 (TIF1 delta) is an HP1 interactor predominantly expressed in testis (Khetchoumian et al., 2004) and was enriched on the K9me3 modification. DNMT3A is a DNA methyltransferase preferentially expressed in cells undergoing de novo methylation such as testis, and was enriched on unmodified H3 as described before (Otani et al., 2009). In addition, the testis-specific proteins SSTY1 and SSTY2 were specifically enriched on K4me3. Both proteins are encoded in many copies on the Y chromosome of mice and are expressed during sperm development (Touré et al., 2004a). Deletions of these genes lead to severe sperm head defects and sterility (Touré et al., 2004b). Interestingly,

---

Figure 1. Label-free Quantification Is as Powerful as SILAC-Based Quantification

(A) Peptide pull-down H3K4me3 versus H3 unmodified SILAC forward and reverse; significant outliers are marked in blue.

(B) Same pull-down in label-free; outliers that show significance in modified t test-based analysis are marked in blue.

(C) Overlap of outliers between SILAC and parallel label-free experiment: blue, outliers that were identified and significant in both; green, outliers that were only identified in one experiment; red, outliers significant in one experiment but not in the other, n.q., not quantified; n.s., not significant.

(D) The workflow for screening chromatin readers from mouse tissue extracts is as follows: nuclear extract pools were prepared from mouse brain, liver, and kidney. Pull-downs were performed with each extract with three different peptides (H3 unmodified, K4me3 and K9me3 modified), resulting in a total of 45 samples. Samples were measured separately, and a label-free quantification algorithm was applied.

(E) Heat map of significant outliers from peptide pull-downs for H3K9me3 and H3K4me3 from brain, kidney, and liver nuclear extracts. Readers with the same pattern are clustered together and are indicated on the right (see also Table S1).

(F) Similar heat map as in (E) for testis. In contrast to (E), whole-cell extracts were used (see also Table S1).

Table 1. Chromatin Readers and Associated Proteins

Reader Group	Complex	Direct Binder	Complex Members
K4me3	TFIID	TAF3	TAF1, 2, 3, 4a, 4b, 5, 6, 7, 8, 9, 9b, 10, 11, 12, 13, TBP
	SAGA	SGF29	ATXN7, ATXN7L1, ATXN7L2, ATXN7L3, CHD1, FAM48A, USP22, TAF5L, TAF6L, SUPT3H, SUPT7L, TADA1L, SGF29
	SET1	CXXC1	ASH2L, SETD1A, SETD1B, CXXC1,
	NuA4 HAT	ING3	BRD8, DMAP1, EP400, EPC1, TIP60, ING3, MORF4L1, MORF4L2, RUVBL1, RUVBL2, YL1, YEATS4, MRGBP, TRRAP
	ATAC		TADA3L, CSRP2BP, GCN5L2, PCAF, SGF29, YEATS2, MBIP, TADA2L, ZZZ3
		JARID1A	EMSY, GATAD1, JARID1A, SIN3B, PHF12, MORF4L1,
	HBO1 (ING5 complex)	ING4/5	HBO1, ING4/5, PHF15, PHF16, PHF17, MEAF6, BRD1, BRPF3,
	SIN3A	ING2	ING2, SIN3A, SAP130, SAP30L, SUDS3, SAP180, ARID4A, BRMS1L
	MLL		DPY30, HCFC1, HCFC2, JMJD3, MLL2, MLL5, CHD8, RBBP5, MEN1
	NURF	BPTF	C17ORF49(BAP18), HMGB2L1, SMARCA1
	Not yet assigned to complexes	DIDO1, ING1, JHDM1D (KDM7), JHDM1B, JmJD2A, PHF8, MORC3, PHF13, PHF2, PHF23, SPIN1	BOD1L, BAF53B, EPC2, GTF2A1, H2AFV, JARID1B, JAZF1, PCYOX, MBTD1, SMARCA5, TADA2B, C11ORF84 homolog, SMC1A, SMC3, UBXD7,[c] EHMT1,[c] EHMT2,[c] BRWD1,[c] CRCP,[c] SSTY1,[c] SSTY2,[c] SLY,[c] SLX,[c] SLXL1,[c] KLHL36 [c]
K9me3		HP1 alpha	ADNP, AHDC1, FBXL11, ZNF828, POGZ, SENP7, RLF, NIPBL, PRR14, C1ORF103 homolog, ZNF462,[b] TRIM66,[c] CHAF1A[c]
		HP1 beta	
		HP1 gamma	
	ORC		LRWD1, ORC2
	Polycomb		SUZ12, RING1A, RING1B, EED, EZH1, EZH2, MGA, L3MBTL2, MAX, PCGF6, PHF1, CBX4
	Not yet assigned to complexes	CDYL, CDYL2, ATRX, MPHOSPH8, UHRF1, UHRF2	hypothetical protein LOC72123, ADNP2, PRDM10, HDGFRP2, HOMEZ, ZMYM2, ZMYM3, ZMYM4, ZMYM5, ZMYM6, SMCHD1, TRIM33, MIER1, MIER2, ZFP280C, ZFP280D, ZNF518B, PAP20, TRIM28, PPHLN1, NSD3, P91A, TRIM24, ZFP15, ZFP524, ZFP597, C19ORF68 homolog, FAM208A, SFRS2, SCAI, C19ORF68 homolog, UBR7,[c] PHF10,[c] KPNA3,[c] KPNA4[c]
repelled by K4me3	NuRD	CHD3, CHD4, CHD5[a]	RBAP48, RBAP46, HDAC1, HDAC2, MBD2, MBD3, MTA1, MTA2, MTA3, CHD3, CHD4, CHD5,[a] FOG2,[a] GATAD2A, GATAD2B, DOC1, MBD3L[c]
	NuRD associated	CHD4	ZNF687, ZMYND8, ZNF592, ZNF532
			RAI1, PHF14, TCF20
		BHC80 (PHF21A)	
	Not yet assigned to complexes	DNMT3A,[c] DNMT3B [c]	BCL7A, CFL1, DGKE, DHX30, FLYWCH1, PRMT5, PWWP2A, PPIG, KBTBD7, MYT1L,[a] PABP1, ZBTB43, ZNF428, GABRG1,[c] H1FX,[c] HAT1,[c] RPS10

Summary of all specific interaction partners for the investigated chromatin marks (for details, see Table S1). Proteins are grouped into complexes or interaction networks according to their description in literature.
[a]Only found in brain.
[b]Only found in brain and kidney.
[c]Only found in testis.

SPIN1, a known K4me3 reader (Wang et al., 2011), has 55% and 52% sequence identity toward SSTY1 and SSTY2, respectively. These proteins share the same domain, and the amino acids suggested to mediate the interaction with the modified lysine residue in SPIN1 (F141, Y170, and Y177) (Wang et al., 2011) are conserved. We therefore speculate that SSTY is a direct binder of K4me3 in testis. Additional testis-specific proteins that specifically bound to K4me3 include SLX, SLXL1, and SLY.

Complexes Specifically Repelled by K4 Trimethylation

Apart from readers for K4me3 and K9me3, our screen also identified a group of proteins that specifically showed reduced binding to the K4me3 modification (Table 1). Among these is the already-mentioned NuRD complex with its known subunits and BHC80, the first PHD finger-containing protein described to bind preferentially to unmodified H3K4 via its PHD finger (Lan et al., 2007). In proteomic data sets published so far, the
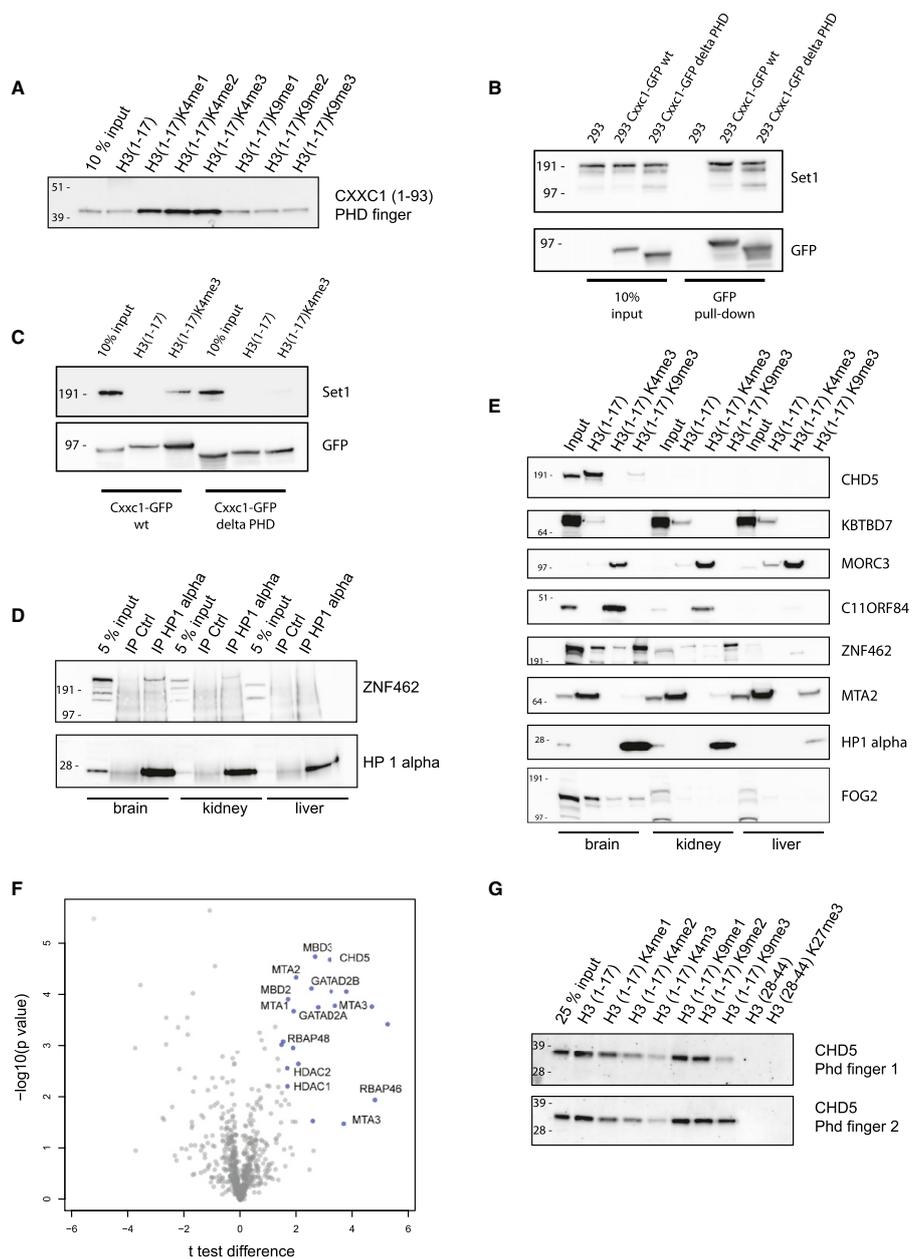
Figure 2. Verification of General and Tissue-Specific Chromatin Readers and Associated Proteins
(A) Peptide pull-down using purified CXXC1 PHD finger 1: specific binding of the SET1 complex subunit CXXC1 to H3K4me3.
(B) Overexpression of GFP-tagged mouse CXXC1 full-length and delta PHD in 293 cells. SET1 coprecipitates with both constructs.

(legend continued on next page)

focus has been on readers of modified amino acids, rather than proteins that are specifically repelled by a modification. We found 41 such proteins, all of which were repelled by K4me3, whereas no readers specifically repelled by K9me3 were apparent, in accordance with an absence of literature reports of proteins specifically recognizing unmodified H3K9. As all of these repelled proteins—with the exception of CHD5—showed nearly equal binding in all three tissues, they appear to perform general and non-tissue-specific functions.

To further elucidate these functions, we used cell line-based methods to assign them into complexes. Specifically, we employed the recently developed BAC technology (Poser et al., 2008) to perform SILAC-based GFP pull-downs of proteins expressed at endogenous levels (Hubner et al., 2010). We analyzed protein-protein interactions for three proteins not described in the context of reading unmodified histone H3 (Table S2). Of particular interest was a series of zinc finger proteins, including ZMYND8, a zinc finger protein that also contains a PWWP domain, a bromodomain, and a PHD type zinc finger. It interacts with CHD4, the NuRD complex member that is responsible for binding of the complex to unmodified and K9me3 (Musselman et al., 2009), thereby explaining the observed binding pattern (Figure 3A). The zinc finger proteins ZNF592, ZNF687, and ZNF532, which we also found to be enriched in our peptide pull-down, likewise specifically interacted with ZMYND8. Moreover, when pulling down ZNF687, we reciprocally enriched ZMYND8, as well as ZNF592 and ZNF532 (Figure 3B). CHD4 and further NuRD complex members specifically interacted with ZNF687 as well. The zinc finger proteins ZMYND8, ZNF592, and ZNF687 have been shown to form a subcomplex (Malovannaya et al., 2011), and our data now link them to the NuRD complex as auxiliary members. Given the large number of zinc fingers in these proteins, we hypothesize that some of them serve to recruit the NuRD complex to specific target genes in the genome.

Another protein associated with unmodified histone H3 was retinoic acid-induced protein 1 (RAI1), which is implicated in Smith-Magenis syndrome, a developmental disorder characterized by mental retardation and craniofacial and skeletal abnormalities (Slager et al., 2003). In the GFP pull-down we found PHF14, TCF20 (Kiaa0292), and HMG20A specifically associated with RAI1 (Figure 3C); these four proteins may form a chromatin-associated complex whose members possess several PHD fingers.

Chromatin Readers of the H3K4me1 Mark

To demonstrate extensibility of our pull-down methodology not only for specialized tissues (Figure 1F) but also for different baits, we performed pull-downs with brain and liver nuclear extracts for monomethylated H3K4 (Figure 4A, Table S3), a histone modification generally associated with enhancers

(Heintzman et al., 2007). We enriched for the known H3K4me1 readers CHD1 (Flanagan et al., 2005) and the TIP60 complex (Jeong et al., 2011) with its members EP400, EPC1, BRD8, YL1, and ING3. Interestingly, the H3K4me3 readers MORC3, Spindlin1, PHF2, and PHF23 were also significantly enriched compared to the unmodified peptide. In contrast, the large group of direct H3K4me3 interactors described above (Table 1) were not significantly enriched in the H3K4me1 pull-downs. Finally, we observed tissue-specific interactions, like the already observed FOG2 and CHD5, which are brain specific and repelled by H3K4me1, as well as ZHX2 and ZHX3, which are repelled by H3K4me1 in liver.

Deep Proteomic Quantification Supports Tissue-Binding Patterns of Chromatin Readers

Next we complemented our interaction studies by a deep proteomic profile of nuclear extracts across the tissues (biological triplicates; more than 5,000 proteins identified, see also Table S4). This demonstrated that organ-specific chromatin readers in our interaction screen also show organ-specific expression patterns. This is exemplified by the brain-specific CHD5 (Figure 4B). The testis-specific readers SSTY1 and SSTY2, as well as SLY or SLX, were not identified in brain, kidney, or liver. The HP1 interactor ZNF462, which was absent in the interaction screen in liver, also was not detected in the nuclear liver proteome. In line with the pull-down results, the large majority of chromatin readers observed in our screen showed approximately equal expression levels in all three tissue nuclear extracts (Figure 4C).

DISCUSSION

Here we have developed and demonstrated a high-resolution and high-accuracy workflow to detect interactions with modified peptides. It uses label-free quantification and is completely generic, as it can be used for any synthesizable peptide modification as well as any suitable protein extract. The technology is highly sensitive, streamlined, and scalable. The absence of any protein or peptide fractionation steps, with concomitant reduction in measurement time, enabled us to perform a relatively large number of replicates in different tissues, increasing statistical confidence. Compared to previous proteomics efforts on identifying chromatin readers, we obtained much improved coverage. This was evident, for instance, by the fact that subunits of chromatin reader complexes were in most cases completely recovered.

We applied our workflow to generate a reader map of interactors of the activating K4me3 and the repressive K9me3 chromatin mark from mouse tissue, which not only covers the large majority of known interactors, but also describes many associations for the first time. The increased depth and completeness of

(C) Peptide pull-down with HEK293 nuclear extracts overexpressing CXXC1-GFP WT and delta PHD. CXXC1 WT is enriched on the H3K4me3 peptide compared to the unmodified peptide. The delta PHD mutant only shows background binding. SET1 binding to H3K4me3 is seen in the CXXC1 WT extracts, but not when CXXC1 delta PHD is overexpressed, demonstrating that CXXC1 recruits SET1 to H3K4me3.
(D) HP1 alpha coIP is as follows: ZNF462—which is enriched on H3K9me3—is enriched from brain and kidney but not from liver extracts.
(E) Western blot verification of selected readers.
(F) CHD5-coIP from brain nuclear extracts, followed by label-free quantitative proteomics: CHD5 enriches members of the NuRD complex.
(G) Peptide pull-down using purified CHD5 PHD fingers reveals specific repulsion by H3K4me3.
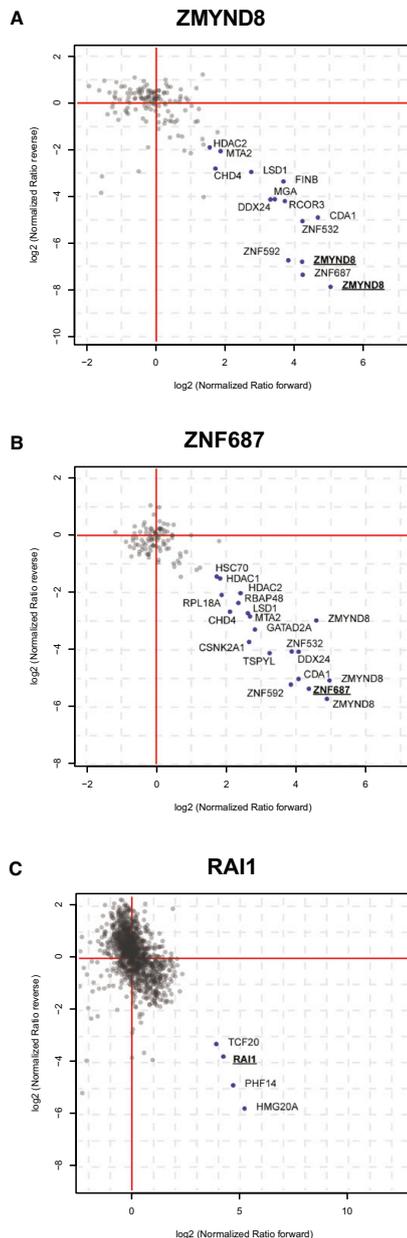
Tissue-Specific Chromatin Readers

Figure 3. Interaction Proteomics for Proteins Repelled by H3K4me3
SILAC GFP pull-downs from HeLa nuclear extracts for ZMYND8 (A), ZNF687 (B) and RAI1 (C); proteins are expressed at near-endogenous levels in HeLa

the measured interactome should make it a useful resource to the community. It also highlights the diversity and complexity of chromatin-associated proteins for these marks. This is especially apparent for the activating K4me3 mark, for which we recover 17 direct binders and most of their associated complex members as well as cofactors. These proteins represent a strikingly broad variety of different functions that they can perform on the surrounding chromatin, even including writing and erasing the K4me3 mark itself. Furthermore, some readers play a general role for gene expression, such as TFIID, whereas others are only important for expression of a specialized subset of genes. One important question that remains is how all these different chromatin readers are recruited to their specific target genes in the genome, since it is clear that different K4me3 reading complexes bind to distinct and only partially overlapping clusters of K4me3-marked genes in human cells (Vermeulen et al., 2010). Part of this specificity may be brought about by additional chromatin marks that serve to differentially enhance or reduce the binding of readers to genes. We have previously shown how such fine-tuning modifications including H3R2me2a and H3S10P can selectively enhance or repress the binding of readers to K4me3 and K9me3, respectively (Vermeulen et al., 2010). But beyond these auxiliary modifications, many of the chromatin-reading complexes described here most likely gain binding specificity for their target genes by DNA sequence-driven recruitment events.

The combination of DNA sequence-specific and histone modification-mediated recruitment of chromatin-associated complexes can best be seen on the repressive K9me3 mark, for which we describe new associated proteins. Among them, many harbor DNA binding modules like zinc finger domains. Furthermore, even a tissue-specific function can be connected to a general machinery by auxiliary factors like ZNF462 in brain and kidney, or TRIM66 in testis.

In addition to the interaction screen, we also used proteomics to correlate our results to organ-specific expression patterns. The large majority of chromatin readers showed similar expression patterns across the tissues. However, all tissue-specific binders also had tissue-specific expression patterns. This restricted expression suggests unique functions necessary in the respective tissue.

The combination of interaction and deep expression proteomics can also be used in an inverse approach: the tissue, cell type, or developmental stage-specific expression of a putative chromatin reader could guide subsequent targeted experiments to determine if this protein binds to a specific mark in those contexts.

In conclusion, advances in proteomics technology increasingly make it possible to move from in vitro cell culture to in vivo-derived tissues extracts. This allows surveying the binding of proteins expressed in diverse tissues, including ones not expressed in standard cell lines. In particular, it allows surveying the interactome in specialized tissues that cannot be easily mimicked in cell culture, such as testis. Scalable and accurate

cells. Interaction partners can be found on the right lower quadrant and are marked with their names.
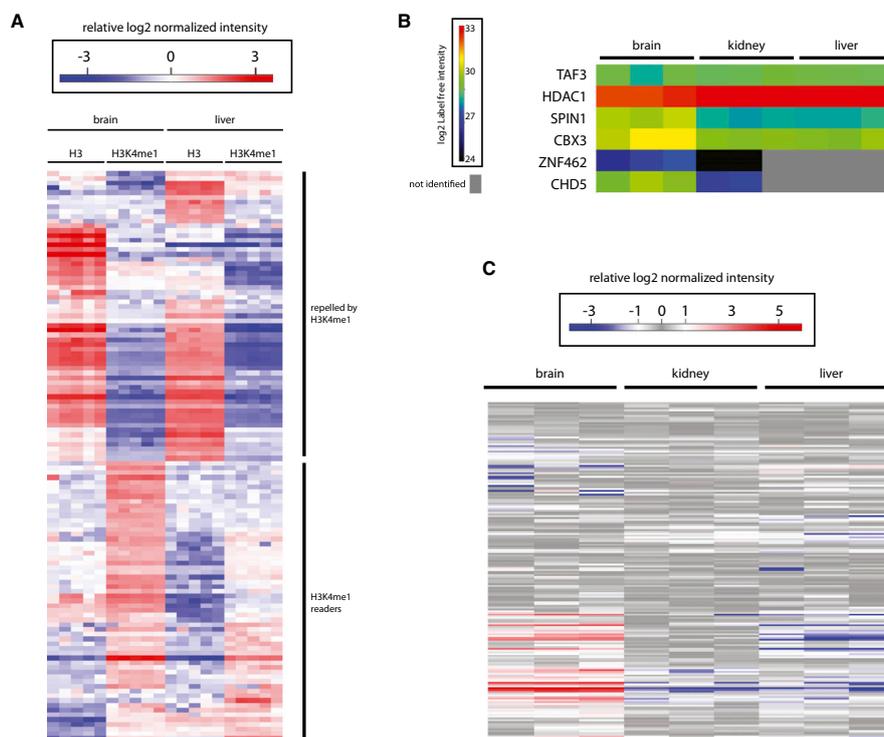
Figure 4. Extension of the Proteomic Screen
(A) Label-free interaction screen for readers of H3K4me1 from mouse brain and liver nuclear extracts.
(B) Protein expression profiles of selected chromatin readers. General chromatin readers show nearly equal expression levels over the analyzed tissues, whereas organ-specific chromatin readers show organ-specific expression profiles.
(C) Proteomic expression profiles of chromatin readers identified in this study.

mapping of the binders to single and combined chromatin marks should contribute to increased understanding of protein-chromatin interactions and their role in regulating tissue- and cell-type-specific gene expression programs and cell-fate decisions.

EXPERIMENTAL PROCEDURES

Extract Preparation

Nuclear extracts from cell lines were prepared as previously described (Vermeulen et al., 2007).

Nuclei from brain, liver and kidney were purified by homogenization followed by pelleting through a sucrose cushion, modified from Lavery and Schibler (1993). Nuclei were lysed in 2 volumes 420 mM NaCl, 20 mM HEPES (pH 7.9), 20% v/v glycerol, 2 mM MgCl$_2$, 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche), 0.5 mM DTT.

Testes were snap frozen in liquid nitrogen and grinded in a beadmill (2 × 3 min, 300 Hz), and 4 volumes lysis buffer (50 mM Tris [pH 8.0], 20 mM NaCl, 0.25% NP40, 1 mM MgCl$_2$, complete Protease inhibitor, 0.5 mM DTT) were added followed by sonication. Samples were incubated with Benzonase until no pellet was visible anymore and subsequently precleared at 15,000 g

for 10 min. Extracts were pooled for pull-downs to minimize variability introduced by extract preparation.

All mice were housed in the animal facility of the Max Planck Institute of Biochemistry, Munich. Animal experiments were approved by the Regierung of Oberbayern.

Peptide Pull-Downs

Peptide pull-downs were performed on 96-well plates modified from Vermeulen et al. (2010). In brief, histone peptides containing the N-terminal 17 amino acids of the histone H3 tail followed by two glycines and a biotinylated lysine were synthesized using the Fmoc strategy as described (Schulze and Mann, 2004). An excess of peptide was coupled to Sepharose streptavidin beads (GE Healthcare). Beads were transferred to 96-well multiscreen filter plates (Millipore, MSBVN1210). Nuclear extracts (400 μg total protein) in 200 μl incubation buffer (150 mM NaCl, 50 mM Tris [pH 8.0], 1% NP40, 0.5 mM DTT) were added and incubated for 3 hr at 4°C while gently shaking. Beads were washed three times (30 s, 60 g) with 200 μl wash buffer 1 (320 mM NaCl, 50 mM Tris [pH 8.0], 0.5% NP40), followed by five washes with wash buffer 2 (150 mM NaCl, 50 mM Tris [pH 8.0]) to minimize residual detergent. A volume of 25 μl 2 M urea, 1 mM DTT supplemented with 120 ng trypsin (Promega) was added and incubated for 30 min at room temperature and eluted, followed by two additional elution steps (50 μl 2 M urea, 5 mM

Molecular Cell

Tissue-Specific Chromatin Readers

iodoacetamide, 10 min incubation each). Proteins were digested overnight at room temperature.

LC-MS/MS Analysis

Samples were measured using the LTQ-Orbitrap Velos or Q Exactive proteomic pipeline. Raw mass spectrometric data were analyzed using the MaxQuant pipeline (Cox and Mann, 2008).

GFP Pull-Downs

GFP pull-downs were performed as described before (Vermeulen et al., 2010) followed by FASP and measurement as single run (ZMYND8 and ZNF687) or in gel digest and fractionation into eight slices (RAI1). All samples were measured on a LTQ Orbitrap Velos using 120 min segmented gradients.

Protein CoIPs

Chd5 antibody and rabbit control antibody or HP1 alpha antibody and goat control antibody were crosslinked to protein G sepharose (GE Healthcare) using dimethyl pimilidate. CHD5 coIPs were performed on 96-well multiscreen plates as described above using brain nuclear extracts (350 μg total protein). HP1 alpha coIPs were performed in tube (600 μg total protein) and eluted by boiling in loading buffer.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2012.10.026.

REFERENCES

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

Berdasco, M., and Esteller, M. (2010). Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev. Cell *19*, 698–711.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Eberl, H.C., Mann, M., and Vermeulen, M. (2011). Quantitative proteomics for epigenetics. ChemBioChem *12*, 224–234.

Feng, W., Yonezawa, M., Ye, J., Jenuwein, T., and Grummt, I. (2010). PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation. Nat. Struct. Mol. Biol. *17*, 445–450.

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

Garcia, B.A., Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2008). Tissue-specific expression and post-translational modification of histone H3 variants. J. Proteome Res. *7*, 4225–4236.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473–478.

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

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J. Cell Biol. *189*, 739–754.

Jeong, K.W., Kim, K., Situ, A.J., Ulmer, T.S., An, W., and Stallcup, M.R. (2011). Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. Nat. Struct. Mol. Biol. *18*, 1358–1365.

Jiang, C.L., Jin, S.G., and Pfeifer, G.P. (2004). MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex. J. Biol. Chem. *279*, 52456–52464.

Khetchoumian, K., Teletin, M., Mark, M., Lerouge, T., Cerviño, M., Oulad-Abdelghani, M., Chambon, P., and Losson, R. (2004). TIF1delta, a novel HP1-interacting member of the transcriptional intermediary factor 1 (TIF1) family expressed by elongating spermatids. J. Biol. Chem. *279*, 48329–48341.

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

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature *448*, 718–722.

Lavery, D.J., and Schibler, U. (1993). Circadian transcription of the cholesterol 7 alpha hydroxylase gene may involve the liver-enriched bZIP protein DBP. Genes Dev. *7*, 1871–1884.

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

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

Luber, C.A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M., and Mann, M. (2010). Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity *32*, 279–289.

Malovannaya, A., Lanz, R.B., Jung, S.Y., Bulynko, Y., Le, N.T., Chan, D.W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., et al. (2011). Analysis of the human endogenous coregulator complexome. Cell *145*, 787–799.

Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., and Mackay, J.P. (2011). Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. J. Biol. Chem. *286*, 11779–11791.

Massé, J., Laurent, A., Nicol, B., Guerrier, D., Pellerin, I., and Deschamps, S. (2010). Involvement of ZFPIP/Zfp462 in chromatin integrity and survival of P19 pluripotent cells. Exp. Cell Res. *316*, 1190–1201.

Massé, J., Piquet-Pellorce, C., Viet, J., Guerrier, D., Pellerin, I., and Deschamps, S. (2011). ZFPIP/Zfp462 is involved in P19 cell pluripotency and in their neuronal fate. Exp. Cell Res. *317*, 1922–1934.

Matthews, A.G., Kuo, A.J., Ramón-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature *450*, 1106–1110.

Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P., and Kutateladze, T.G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. Biochem. J. *423*, 179–187.

Musselman, C.A., Ramírez, J., Sims, J.K., Mansfield, R.E., Oliver, S.S., Denu, J.M., Mackay, J.P., Wade, P.A., Hagman, J., and Kutateladze, T.G. (2012). Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. Proc. Natl. Acad. Sci. USA *109*, 787–792.

Nikolov, M., Stützer, A., Mosch, K., Krasauskas, A., Soeroes, S., Stark, H., Urlaub, H., and Fischle, W. (2011). Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. Mol. Cell. Proteomics *10*. http://dx.doi.org/10.1074/mcp.M110.005371, M110, 005371.

Nozawa, R.S., Nagao, K., Masuda, H.T., Iwasaki, O., Hirota, T., Nozaki, N., Kimura, H., and Obuse, C. (2010). Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation. Nat. Cell Biol. *12*, 719–727.

Olsen, J.V., Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., et al. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol. Cell. Proteomics *8*, 2759–2769.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics *1*, 376–386.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. *10*, 1235–1241.

Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature *442*, 100–103.

Poser, I., Sarov, M., Hutchins, J.R., Hériché, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Methods *5*, 409–415.

Potts, R.C., Zhang, P., Wurster, A.L., Precht, P., Mughal, M.R., Wood, W.H., 3rd, Zhang, Y., Becker, K.G., Mattson, M.P., and Pazin, M.J. (2011). CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. PLoS ONE *6*, e24515. http://dx.doi.org/10.1371/journal.pone.0024515.

Schulze, W.X., and Mann, M. (2004). A novel proteomic screen for peptide-protein interactions. J. Biol. Chem. *279*, 10756–10764.

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

Slager, R.E., Newton, T.L., Vlangos, C.N., Finucane, B., and Elsea, S.H. (2003). Mutations in RAI1 associated with Smith-Magenis syndrome. Nat. Genet. *33*, 466–468.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

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

Touré, A., Grigoriev, V., Mahadevaiah, S.K., Rattigan, A., Ojarikre, O.A., and Burgoyne, P.S. (2004a). A protein encoded by a member of the multicopy Ssty gene family located on the long arm of the mouse Y chromosome is expressed during sperm development. Genomics *83*, 140–147.

Touré, A., Szot, M., Mahadevaiah, S.K., Rattigan, A., Ojarikre, O.A., and Burgoyne, P.S. (2004b). A new deletion of the mouse Y chromosome long arm associated with the loss of Ssty expression, abnormal sperm development and sterility. Genetics *166*, 901–912.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Vermeulen, M., Hubner, N.C., and Mann, M. (2008). High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. Curr. Opin. Biotechnol. *19*, 331–337.

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

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847–851.

Wang, W., Chen, Z., Mao, Z., Zhang, H., Ding, X., Chen, S., Zhang, X., Xu, R., and Zhu, B. (2011). Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes. EMBO Rep. *12*, 1160–1166.

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

2.3 H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization

2.3.1 Project aim and summary

Histone variants play an important role in chromatin biology by providing a means to alter the properties of nucleosomes and to generate distinct chromosome regions. One very important H2A variant is H2A.Z, which is implicated in transcriptional regulation, chromosome segregation and mitosis [51]. Two isoforms, encoded by two different genes, are known: H2A.Z.1 and H2A.Z.2. Here, we describe a novel splice variant of H2AZ.2, which we term H2A.Z.2.2. H2A.Z.2.2 is incorporated into chromatin in a replication-independent process by the H2A.Z-specific TIP60 and SRCAP chaperone complexes. Nucleosomes containing this variant exhibit major structural changes and are less stable than nucleosomes containing H2A or H2A.Z.2.1. These drastic effects on the nucleosomes can be attributed to the unique C-terminus of H2A.Z.2.2. This project adds a novel member with very special biochemical properties to the large list of H2A variants.

2.3.2 Contribution

This project was initiated and headed by the group of Sandra Hake at the Adolf Butenandt Institute (Ludwig Maximilian University, Munich). The main driving force behind the experiments was Clemens Boenisch. To answer the question of histone chaperones for this novel H2A variant, I performed protein pull-downs of H2A.Z.2.1 and H2A.Z.2.2 from the soluble nuclear fraction. In addition, we investigated the C-terminus of H2A.Z.2.2, which is strikingly different to the C-terminus of H2A.Z.2.1. As the C-terminus is described to be important for several protein-protein interactions, I used peptide pull-downs with both C-termini to screen for proteins binding there.

2.3.3 Publication

This project was published in 2012 as an article in Nucleic Acid Research:

H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization

Clemens Boenisch, Katrin Schneider, Sebastian Puenzeler, Sonja M. Wiedemann, Christina Bielmeier, Marco Bocola, H. Christian Eberl, Wolfgang Kuegel, Juergen Neumann, Elisabeth Kremmer, Heinrich Leonhardt, Matthias Mann, Jens Michaelis, Lothar Schermelleh and Sandra B. Hake

Nucleic Acids Res. 2012 Mar 29. (Epub ahead of print)

H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization

Clemens Bönisch[1], Katrin Schneider[2], Sebastian Pünzeler[1], Sonja M. Wiedemann[1], Christina Bielmeier[2], Marco Bocola[3], H. Christian Eberl[4], Wolfgang Kuegel[5], Jürgen Neumann[2], Elisabeth Kremmer[6], Heinrich Leonhardt[2,7], Matthias Mann[4,7], Jens Michaelis[4,7,8], Lothar Schermelleh[2,*] and Sandra B. Hake[1,7,*]

[1]Department of Molecular Biology, Adolf-Butenandt-Institute, Ludwig-Maximilians-University Munich, 80336 Munich, [2]Department of Biology, Biozentrum, Ludwig-Maximilians-University Munich, 82152 Planegg-Martinsried, [3]Department of Biochemistry II, University Regensburg, 93053 Regensburg, [4]Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, 82152 Martinsried, [5]Department of Chemistry, Ludwig-Maximilians-University Munich, [6]Institute of Molecular Immunology, Helmholtz Center Munich, German Research Center for Environmental Health, [7]Center for Integrated Protein Science Munich (CIPSM), 81377 Munich and [8]Department of Physics, Ulm University, 89081 Ulm, Germany

ABSTRACT

The histone variant H2A.Z has been implicated in many biological processes, such as gene regulation and genome stability. Here, we present the identification of H2A.Z.2.2 (Z.2.2), a novel alternatively spliced variant of histone H2A.Z and provide a comprehensive characterization of its expression and chromatin incorporation properties. Z.2.2 mRNA is found in all human cell lines and tissues with highest levels in brain. We show the proper splicing and *in vivo* existence of this variant protein in humans. Furthermore, we demonstrate the binding of Z.2.2 to H2A.Z-specific TIP60 and SRCAP chaperone complexes and its active replication-independent deposition into chromatin. Strikingly, various independent *in vivo* and *in vitro* analyses, such as biochemical fractionation, comparative FRAP studies of GFP-tagged H2A variants, size exclusion chromatography and single molecule FRET, in combination with *in silico* molecular dynamics simulations, consistently demonstrate that Z.2.2 causes major structural changes and significantly destabilizes nucleosomes. Analyses of deletion mutants and chimeric proteins pinpoint this property to its unique C-terminus. Our findings enrich the list of known human variants by an unusual protein belonging to the H2A.Z family that leads to the least stable nucleosome known to date.

INTRODUCTION

In the eukaryotic nucleus, DNA is packaged into chromatin. The fundamental unit of this structure is the nucleosome consisting of a histone octamer (two of each H2A, H2B, H3 and H4) that organizes ∼147 bp of DNA (1). In order to allow or prevent nuclear regulatory proteins access to the DNA, the chromatin structure has to be flexible and dynamic. Several mechanisms ensure controlled chromatin changes, one being the incorporation of specialized histone variants (2,3).

Variants of the histone H2A family are the most diverse in sequence and exhibit distinct functions (4,5), comprising DNA damage repair, transcriptional regulation, cell cycle control and chromatin condensation, though the exact mechanisms of action are not fully understood yet. Interestingly, the highest sequence variation among H2A variants is found in the C-terminus, suggesting that differences in structure and biological function might be

*To whom correspondence should be addressed. Tel: +49 89 2180 75435; Fax: +49 89 2180 75425; Email: sandra.hake@med.uni-muenchen.de
Correspondence may also be addressed to Lothar Schermelleh. Tel: +44 1865 613264; Fax: +49 89 2180 74236;
Email: lothar.schermelleh@bioch.ox.ac.uk
Present address:
Lothar Schermelleh, Department of Biochemistry, University of Oxford, South Park Road, Oxford OX1 3QU, UK.

primarily attributed to this domain (6–9). One of the best investigated and highly conserved but also functionally enigmatic histone variant is H2A.Z. This variant is essential in most eukaryotes and possesses unique functions (10,11). H2A.Z is involved in transcriptional regulation, chromosome segregation and mitosis, acting in an organism- and differentiation-dependent manner (12,13). Furthermore, H2A.Z has been implicated in regulating epigenetic memory (14) and in inhibiting read-through antisense transcription (15). In higher eukaryotes, H2A.Z might play a role in heterochromatin organization (16), genome stability and chromosome segregation (17). Despite many efforts to elucidate the exact biological functions of H2A.Z, its roles have been and remain controversial (18). Furthermore, deregulation of H2A.Z expression or localization seems to be connected to the development of several neoplasias (19–23). Interestingly, in vertebrates two non-allelic genes coding for two highly similar H2A.Z proteins, H2A.Z.1 and H2A.Z.2, exist (24) (previously named H2A.Z-1 and H2A.Z-2, prefixes were changed due to a new histone variant nomenclature; Talbert P.B., manuscript in preparation). They have a common origin in early chordate evolution, are both acetylated on the same N-terminal lysines (25–27) and might be ubiquitinated on either one of the two C-terminal lysines (28).

Here, we report the identification and structural characterization of H2A.Z.2.2 (Z.2.2), an unusual alternative splice form of H2A.Z. We show that Z.2.2 mRNA is expressed to different degrees in all human cell lines and tissues examined, with highest levels found in brain. Cell biological and biochemical analyses consistently reveal the presence of two distinct Z.2.2 populations within the cell. The majority of Z.2.2 is freely dispersed in the nucleus, whereas only a minority is stably incorporated into chromatin, most likely through the H2A.Z-specific p400/NuA4/TIP60 (TIP60) and SRCAP chaperone complexes. *In vivo* and *in vitro* analyses, in agreement with molecular dynamic (MD) simulations, demonstrate that due to its unique docking domain Z.2.2 chromatin incorporation leads to severely unstable nucleosomes. Our data provide compelling evidence that a novel H2A.Z variant exists in humans that plays a distinct and novel role in chromatin structure regulation.

MATERIALS AND METHODS

See Supplementary Materials and Methods section for detailed protocols.

Cell culture, transfection, FACS and cloning

Cell lines were grown in DMEM medium (PAA) supplemented with 10% FCS (Sigma) and 1% penicillin/streptomycin at 37°C and 5% CO_2. Cells were transfected using FuGene HD (Roche Applied Science) according to the manufacturer's instructions. For details on cell selection, FACS and cloning of expression plasmids see Supplementary Materials and Methods section.

RNA expression analysis

RNA isolation and cDNA generation were performed as previously described (29). Data were analyzed with the advanced relative quantification tool of the Lightcycler 480 (Roche) software including normalization to HPRT1 and HMBS levels. Statistical evaluation was done using *t*-test (two-tailed distribution, heteroscedastic). Total RNA from different human tissues was commercially acquired from: Applied Biosystems: normal lung, breast and tumor breast, lung and ovary; Biochain: tumor lung, breast, thyroid and bone, normal testis, cerebellum, cerebral cortex, hippocampus, thalamus and total fetal brain; amsbio: frontal lobe.

Histone extraction, RP–HPLC purification, sucrose gradient, cellular fractionation and salt stability experiments

Acid extraction of histones was done as previously described (30). Histones were separated by RP–HPLC as previously described (29). Fractions were dried under vacuum and stored at −20°C.

Details on MNase digest and sucrose gradient fractionation can be found in Supplementary Materials and Methods section.

Fractionation and salt stability experiments were carried out as described previously (31–33) with minor changes. For details on these methods see Supplementary Materials and Methods section.

Antibodies

For the generation of a Z.2.2-specific antibody (αZ.2.2), a peptide spanning the last C-terminal amino acids GGEKRRCS of Z.2.2 was synthesized (Peptide Specialty Laboratories GmbH) and coupled to BSA and OVA, respectively. Development of Z.2.2-specific monoclonal antibodies in rats was done as previously described (29). The αZ.2.2 clone 1H11-11 of rat IgG1 subclass was applied in this study. Rabbit αZ.2.2 antibody (rabbit 2, bleed 3) was generated by the Pineda-Antikörper-Service company using the identical peptide epitope followed by affinity purification. Following other primary antibodies were used: αGAPDH (sc-25778, Santa Cruz), αGFP (Roche Applied Science), αH2A (ab 13923, abcam), αH3 (ab1791, abcam) and αH2A.Z (C-terminus: ab4174, abcam; N-terminus: ab18263, abcam). Following secondary antibodies and detection kits were used in immunoblots: GFP-Z.2.2 and GFP-Bbd histones (αGFP) and endogenous Z.2.2 (αZ.2.2) were detected using HRP-conjugated secondary antibodies (Amersham) with ECL advance (Amersham), all other proteins were detected using ECL (Amersham). Detection of recombinant proteins to evaluate histone stoichiometry of *in vitro* assembled nucleosomes was carried out using IRDye-labeled secondary antibodies (LI-COR).

Fluorescence microscopy of cells and chromosomes

Preparation of cells and chromosome spreads for fluorescence microscopy was done as previously reported (34). Wide-field fluorescence imaging was performed on

a PersonalDV microscope system (Applied Precision) equipped with a 60×/1.42 PlanApo oil objective (Olympus), CoolSNAP ES2 interline CCD camera (Photometrics), Xenon illumination and appropriate filtersets. Iterative 3D deconvolution of image *z*-stacks was performed with the SoftWoRx 3.7 imaging software package (Applied Precision).

FRAP and exponential fitting

For details see Supplementary Materials and Methods section.

Stable isotope labeling with amino acids in cell culture (SILAC) and mass spectrometric identification of H2A.Z-specific chaperone complexes

HeLa cells expressing GFP-Z.2.1 or GFP-Z.2.2 were SILAC labeled and nuclear extracts were prepared as described before (35,36). High-resolution LC MS/MS analysis was performed on an Orbitrap platform: details on the experimental procedure are found in Supplementary Materials and Methods section. Mass spectrometric (MS) operation and raw data analysis (37) are described in Supplementary Materials and Methods section. A complete list of all proteins identified is found in Supplementary Table S1.

Immunofluorescence microscopy of cell cycle-dependent GFP-Z.2.1 and GFP-Z.2.2 chromatin incorporation

Details on the experimental labeling (38) and microscopy procedures are found in Supplementary Materials and Methods section.

Expression of recombinant human histone proteins in *Escherichia coli*, *in vitro* octamer and nucleosome reconstitution

Histones were expressed, purified and assembled into octamers as described (39) and mononucleosomes were assembled on DNA containing the 601-positioning sequence (40) according to (39,41). For details on *in vitro* octamer and nucleosome reconstitution, see Supplementary Materials and Methods section.

Single molecule Förster resonance energy transfer

Single molecule Förster resonance energy transfer (smFRET) single molecule burst analysis followed by the removal of multi-molecular events (42–45) are described in detail in the Supplementary Materials and Methods section.

Molecular modeling and MD simulations

The molecular modeling suite YASARA-structure version 9.10.29 was employed, utilizing the AMBER03 force field (46) for the protein and the general amber force field (GAFF) (47) throughout this study. The partial charges were computed using the AM1/BCC procedure (48) as implemented in YASARA structure (49). The starting point for molecular modeling was the crystal structure of a nucleosome core particle containing the histone variant H2A.Z (PDB 1F66) (50). Missing side chain atoms were added (Glu E 634). The missing N-terminal and C-terminal

residues were not modeled, although they might interact with the neighboring DNA, e.g. in the case of missing C-terminal residues in H2A.Z (119–128; GKKGQQKTV). All structures were solvated in a water box with 0.9% NaCl and neutralized (51). The structures were initially minimized using steepest descent and simulating annealing procedures. All deletions and mutations were introduced sequentially using YASARA structure. MD simulations were carried out at 300 K over 2.5 ns in an NPT ensemble using PME. All simulations were performed four times using various starting geometries. The 2.5 ns MD trajectories were sampled every 25 ps, resulting in 100 simulation frames per run, which were evaluated after an equilibration phase of 500 ps to derive statistical averages and properties of the corresponding variant. Finally, the interaction energy of H2A and H3 was calculated from a simulation of the solvated octamer and the isolated (H3–H4)$_2$ tetramer or the isolated respective H2A.Z–H2B dimer. The interaction energy is calculated as energy difference of the solvated octamer minus the solvated (H3–H4)$_2$ tetramer and H2A.Z–H2B dimer.

RESULTS

Alternative splicing of H2A.Z.2 occurs *in vivo*

Two non-allelic intron-containing genes with divergent promoter sequences that code for H2A.Z variants exist in vertebrates (24,27). In humans, the H2A.Z.2 (H2AFV) primary transcript is predicted to be alternatively spliced thereby generating five different gene products (Supplementary Figure S1A). Using PCR and confirmed by sequencing we detected not only H2A.Z.2.1 (Z.2.1) but also H2A.Z.2.2 (Z.2.2) mRNA, though none of the other splice variants in human cells (Supplementary Figure S1B) showing that the H2A.Z.2 primary transcript is indeed alternatively spliced *in vivo*. Interestingly, database searches found Z.2.2 mRNA to be predicted in chimpanzee (*Pan troglodytes*) and Northern white-cheeked gibbon (*Nomascus leucogenys*) as well. In addition, the coding sequence of the unique exon 6 was present downstream of the *H2AFV* locus of several other primate genomes, such as gorilla (*Gorilla gorilla gorilla*), macaque (*Macaca mulatta*), orangutan (*Pongo abelii*) and white-tufted-ear marmoset (*Callithrix jacchus*) (data not shown). In all of these primates, with the exception of marmoset, the resulting protein sequence, if translated, is 100% identical to the unique human Z.2.2 peptide. Further searches revealed that the genomes of horse, and to a certain extent also rabbit and panda bear, contain sequences downstream of their *H2AFV* loci that could, if translated, lead to proteins with some similarities to human Z.2.2, although they are much more divergent and even longer (rabbit, panda bear). Due to these differences, it is highly likely that those species do not express a Z.2.2 protein homolog. Surprisingly, we could not detect Z.2.2-specific sequences in mouse, rat or other eukaryotic genomes, suggesting that Z.2.2 might be primate specific.

Next, we wanted to determine to what degree all three H2A.Z mRNAs are expressed in different human cell lines and tissues and performed quantitative PCR (qPCR).

Z.2.2 mRNA was present to different degrees in all human cell lines and tissues tested, though less abundant than Z.1 and Z.2.1 mRNAs that are expressed in similar amounts (Supplementary Figure S1C and D). Z.2.2 constituted between 5% and 15% of total Z.2 transcripts in all cell lines and tissues, with the exception of brain, where it was statistically significant upregulated ($p = 1.7 \times 10^{-4}$; Figure 1A). In some regions of this particular organ Z.2.2 accounted for up to 50% of all Z.2 transcripts pointing toward an exciting brain-specific function of this novel variant.

Encouraged by our findings we next investigated whether the endogenous protein is present *in vivo*. The distinctive feature of Z.2.2 is its C-terminus that is 14 amino acids shorter and contains six amino acids differences compared to Z.2.1 (Figure 1B). Due to this shortened C-terminal sequence, ubiquitination sites at positions K120 and K121 (28) and part of the H3/H4 docking domain (50) are lost in Z.2.2. We generated antibodies against Z.2.2's unique C-terminal amino acids (αZ.2.2) in rats and rabbits and confirmed their specificity in immunoblots (IB) with recombinant Z.2.1 and Z.2.2

proteins (Supplementary Figure S1E and data not shown). We extracted histones from several human and mouse cell lines, purified them by reversed phase–high performance liquid chromatography (RP–HPLC) and analyzed obtained fractions by IB (Figure 1C). Using αZ.2.2 (polyclonal rabbit), we observed a signal of the calculated weight of Z.2.2 that elutes shortly before Z.1- and Z.2.1-containing fractions in all human samples. Similar results were obtained with a monoclonal αZ.2.2 rat antibody (data not shown). In agreement with the finding that Z.2.2-specific exon 6 sequences are mainly restricted to primate genomes, we could detect Z.2.2 protein in human but not in mouse cells (Figure 1C). In summary, our data show that Z.2.2 protein indeed exists *in vivo*, albeit at a low expression level.

GFP-Z.2.2 is partially incorporated into chromatin

Having demonstrated the existence of this novel variant *in vivo*, we next sought to clarify whether Z.2.2 constitutes a bona fide histone by being part of the chromatin structure. Due to high background of all our αZ.2.2 antibodies in IB with cell extracts (data not shown), we generated
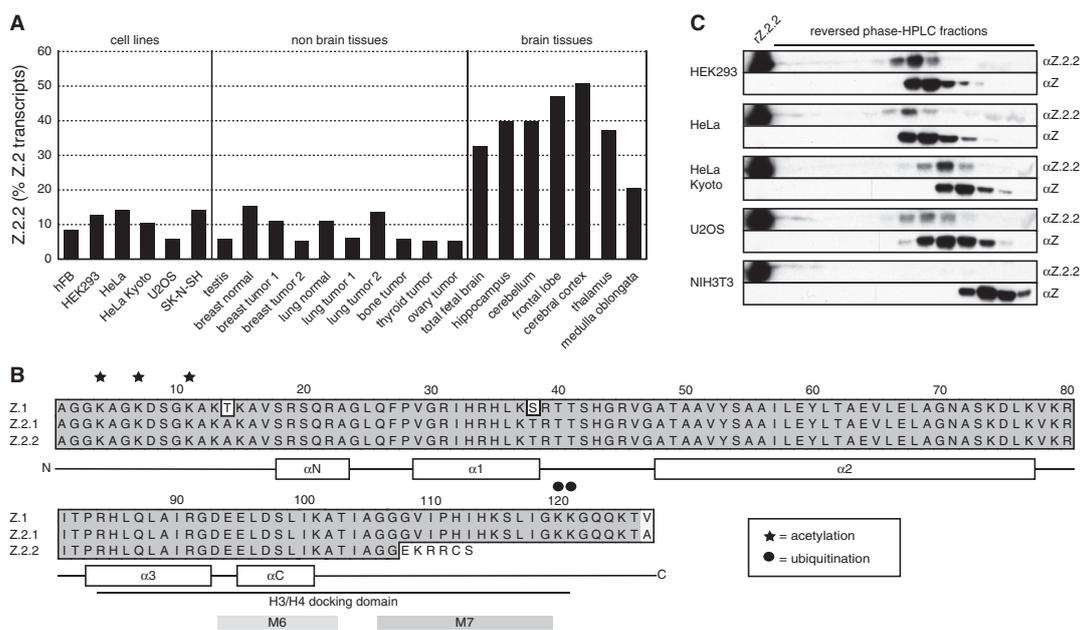
Figure 1. Identification of Z.2.2. (**A**) qPCR with cDNA from different human cell lines and tissues using primers specific for Z.2.1 and Z.2.2. Data were normalized to HPRT1 and HMBS expression levels. Controls generated without reverse transcriptase (no RT) were used to assess amplification threshold. Shown are the levels of Z.2.2 mRNA as percentages of total Z.2 transcripts (Z.2.1 + Z.2.2). For an evaluation of absolute expression levels see Supplementary Figure S1C and D. (**B**) Amino acid alignment of human Z.1, Z.2.1 and Z.2.2 proteins using ClustalW Alignment (MacVector 10.0.2). Identical amino acids are highlighted in dark gray, similar amino acids in light gray and changes are set apart on white background. Known acetylation sites are depicted with stars and ubiquitination sites with circles. A schematic representation of the secondary structure of Z.1 and Z.2.1 is shown below the alignment, including depiction of the H3/H4 docking domain (50). M6 and M7 boxes indicate regions important for H2A.Z-specific biological functions in *D. melanogaster* (60). (**C**) IB analyses of RP–HPLC purified fractions from different human (HEK293, HeLa, HeLa Kyoto and U2OS) and mouse (NIH3T3) cell lines using a polyclonal rabbit αZ.2.2 and αH2A.Z (αZ, C-terminal) antibodies. Recombinant Z.2.2 protein (rZ.2.2) was loaded in the first lane as positive control for αZ.2.2 antibody. Similar results were obtained when using a monoclonal rat αZ.2.2 antibody (data not shown).

HeLa Kyoto cell lines stably expressing GFP-tagged H2A variants (HK-GFP cells) for subsequent analyses. Expression levels of GFP-tagged histone variants were determined by FACS (Supplementary Figure S2A) and by comparing expression levels of GFP-tagged variants with endogenous H2A.Z proteins in IB (Supplementary Figure S2B). GFP-Z.1 and -Z.2.1 were expressed in similar amounts as the endogenous H2A.Z protein, and GFP-Z.2.2 expression levels were considerably lower than those of other GFP-tagged H2A variants, with the exception of GFP-H2A.Bbd (Barr body deficient; Bbd). These data show that all GFP-H2A variants were not expressed in abnormal amounts in cell clones used for further analyses.

In fluorescence microscopy, GFP-Z.2.2 exhibited a sole but rather diffuse nuclear distribution similar to GFP-Bbd, suggesting that both variants might have similar properties (Figure 2A). Additionally, GFP-Z.2.2 was detected in condensed mitotic chromosomes, with a faint residual staining in the surrounding area (Figure 2B), suggesting that it is incorporated into chromatin, although to a lesser extent than other GFP-H2A variants. To discriminate between a potential non-specific DNA binding and nucleosomal incorporation of Z.2.2 we purified mononucleosomes by sucrose gradient centrifugation. GFP-Z.2.2 was detected by IB in fractions containing mononucleosomes (Figure 2C), suggesting that Z.2.2 is indeed a nucleosomal constituent.

To analyze the extent of Z.2.2 chromatin incorporation in more detail, we isolated soluble (*sol*) and chromatin (*chr*) fractions from HK-GFP cells. IB analyses revealed, as expected, that similar to GFP-Bbd, GFP-Z.2.2 is predominantly nuclear soluble, with only minor amounts present in chromatin (Figure 3A). Based on fractionation and fluorescence imaging results, we hypothesized that this novel variant behaves in a different manner as compared to other H2A variants with regard to chromatin exchange mobility *in vivo*. To test this prediction, we performed fluorescence recovery after photobleaching (FRAP) experiments with HK-GFP cells. Using spinning disk confocal microscopy we monitored the kinetic behavior of H2A variants with variable intervals over 2 min (short-term) up to several hours (long-term) after bleaching a $5 \mu m \times 5 \mu m$ square nuclear region (Figure 3B and Supplementary Figure S3). As expected, GFP alone showed the highest mobility. In contrast, GFP-H2A, -Z.1 and -Z.2.1 showed a slow recovery, which is in agreement with a previous report (52). GFP-Bbd has been described to exhibit low nucleosomal stability and a fast FRAP kinetic (53), which we also observed in our experiment. Interestingly, GFP-Z.2.2 showed an even faster recovery than GFP-Bbd, with ∼80% of initial fluorescence reached after 1 min. Careful assessment and bi-exponential fitting of FRAP data allowed us to also calculate ratios of fractions with fast, intermediate and slow recovery and their respective half-time of recovery ($t_{1/2}$) as an indication of exchange rate thereby revealing quantitative differences between Z.2.2 and other H2A variants (Figure 3D, Supplementary Figure S3C and E). For Z.2.2 as well as for Bbd, we identified a fast fraction of unbound or very transiently
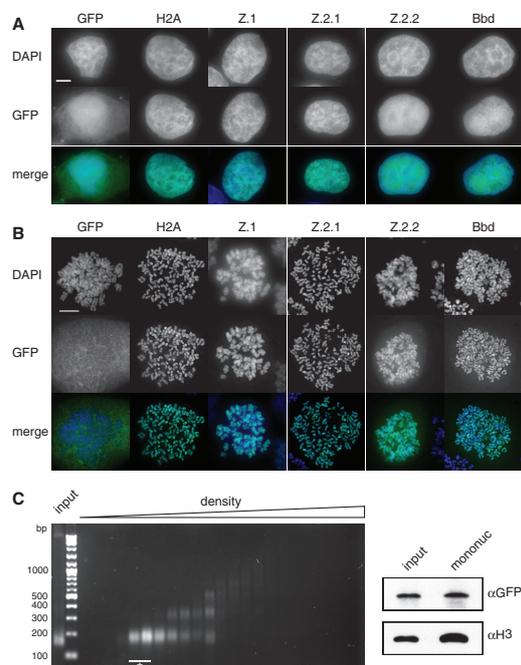
Figure 2. Z.2.2 localizes to the nucleus and is partially incorporated into chromatin. (**A**) Fluorescence imaging of stably transfected HeLa Kyoto cells shows nuclear localization of all GFP-H2A variants (middle). DNA was counterstained with DAPI (top). Overlay of both channels in color is shown at the bottom (Merge; GFP: green, DAPI: blue). Scale bar = 5 μm. (**B**) Deconvolved images of metaphase spreads of HeLa Kyoto cells stably expressing GFP-H2A variants (middle). Merged images in color are shown below (GFP: green; DAPI: blue). Scale bar = 10 μm. (**C**) Chromatin from HeLa Kyoto cells stably expressing GFP-Z.2.2 was digested with MNase followed by a purification of mononucleosomes using sucrose gradient centrifugation. Isolated DNA from subsequent sucrose gradient fractions was analyzed by agarose gel electrophoresis (left). Fractions containing pure mononucleosomes (marked with asterisk) were combined and analyzed by IB (right) using αGFP antibody for the presence of GFP-Z.2.2 (top), and αH3 (bottom).

interacting molecules (78%, $t_{1/2} \sim 1.1$ s and 52%, $t_{1/2} \sim 2.5$ s, respectively; for comparison GFP $t_{1/2} \sim 0.4$ s) and a substantially slower fraction with a $t_{1/2}$ in the range of 7–9 min. In contrast, GFP-H2A, -Z.1 and -Z.2.1 showed no fast mobile fraction but intermediate slow fractions with $t_{1/2}$ in the range of 8–17 min and a second even slower class exchanging with a $t_{1/2}$ of a few hours. For comparison, we measured the linker histone H1.0 (54–57) and the histone binding protein HP1α (58,59), both DNA-associated proteins, and found that HP1α shows an overall much faster recovery than all H2A variants. In contrast to Z.2.2 and Bbd, no unbound fraction of H1.0 was detected. More importantly, with regards to the bound Z.2.2 and Bbd fractions overall H1.0 showed a faster recovery, arguing against an unspecific DNA-association of Z.2.2 and Bbd. In agreement with cell biological and biochemical analyses,

these data clearly demonstrate that a large fraction of the splice variant Z.2.2 is very rapidly exchanged or chromatin unbound, and a minor population is incorporated into chromatin.

Z.2.2's unique docking domain, but not its shortened length, weakens chromatin association

The functional importance of specific C-terminal domains of H2A.Z has previously been demonstrated by nucleo-somal structure analyses (7,50) and in rescue experiments in flies (60). Since the C-terminus of Z.2.2 is shorter and has a distinct sequence when compared to Z.1 and Z.2.1, it is not clear which of these features determines Z.2.2's unusual chromatin-association.

Therefore, we generated deletion and domain-swap constructs (Supplementary Figure S3D) for FRAP experiments (short-term: Figure 3C and long-term: Supplementary Figure S3B). Surprisingly, C-terminal deletions of GFP-H2A (H2A^{111}) and GFP-Z.2.1 (Z.2.1^{113}) to mimic the shortened length of Z.2.2 did not affect their original mobility in short-term and only modestly in long-term FRAP. Hence, the mere shortening of the C-terminus is not sufficient to weaken stable chromatin association.

To investigate whether the unique six C-terminal amino acids of Z.2.2 are sufficient to generate highly mobile proteins, we created a further C-terminally truncated GFP-H2A construct (H2A^{105}) and added the Z.2.2 specific C-terminal six amino acids (H2A^{105}+CZ.2.2). Although both mutant constructs are slightly more mobile than H2A^{111}, their indistinguishable recovery kinetics demonstrate that the unique six C-terminal amino acids of Z.2.2 alone are not sufficient to cause its extreme mobility *in vivo*.

To explore whether the complete Z.2.2 docking domain is able to induce high-protein mobility, we transferred the respective domain of either Z.2.1 (amino acids 91–127) or Z.2.2 (amino acids 91–113) onto a C-terminally truncated H2A (H2A^{88}+CZ.2.1 and H2A^{88}+CZ.2.2, respectively). Interestingly, only the docking domain of Z.2.2, but not the one of Z.2.1, confers high mobility. In conclusion, the six unique C-terminal amino acids of Z.2.2 prevent chromatin-association of a large proportion of this protein, but only when present in the context of the preceding H2A.Z-specific docking domain sequence.

Z.2.2 interacts with H2A.Z-specific TIP60 and SRCAP chaperone complexes and is deposited into chromatin outside of S-phase

Our so far obtained data strongly imply that at least a minor amount of the cellular Z.2.2 protein is incorporated into nucleosomes. Since previous studies have shown that evolutionary conserved Swr1-related ATP-dependent chromatin remodelers specifically exchange canonical H2A–H2B with H2A.Z–H2B dimers within nucleosomes (10,61), we wondered if such complexes are also able to actively deposit Z.2.2 into chromatin. HK cells and HK cells stably expressing GFP-Z.2.1 or -Z.2.2 were SILAC labeled, soluble nuclear proteins isolated, GFP-tagged
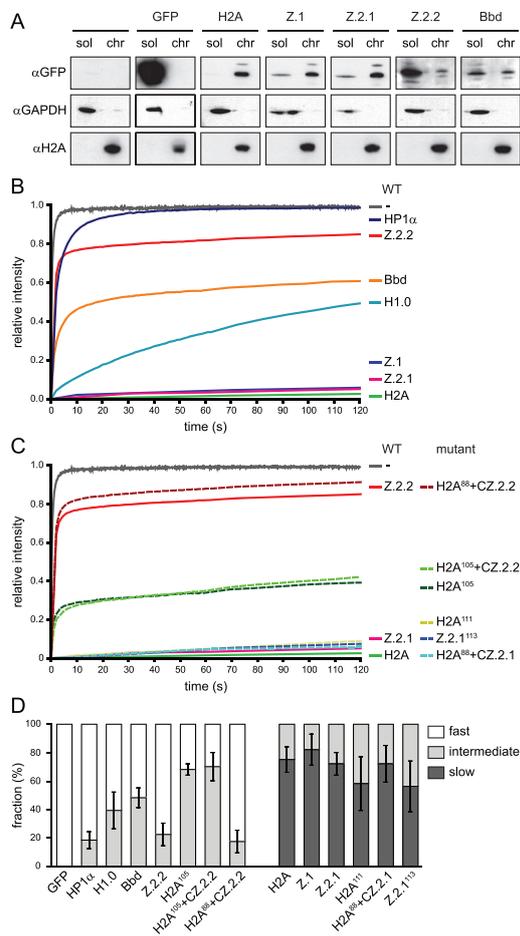
Figure 3. The majority of Z.2.2 protein is nuclear soluble and highly mobile in a sequence-dependent manner. (**A**) HK-GFP cells were subjected to biochemical fractionation. Fractions *sol* and *chr* of identical cell equivalents were probed in IB with αGFP (top), αH2A (middle) and αGAPDH (bottom). (**B**) FRAP quantification curves of average GFP signal relative to fluorescence intensity prior to bleaching are depicted for GFP, GFP-tagged wild-type H2A variants, linker histone H1.0 and heterochromatin protein 1α (HP1α). Mean curves of 10–29 cells are shown for each construct. Error bars are omitted for clarity. (**C**) FRAP quantification curves similar to (**B**) are depicted for GFP, GFP-tagged wild-type H2A, Z.2.1, Z.2.2 and mutant constructs. (**D**) Quantitative evaluation of FRAP curves. Plot shows calculated mobility fraction sizes of different wild-type and mutant H2A variant constructs, as well as H1.0 and HP1α, based on bi-exponential fitting of FRAP data. Error bars indicate SD (see Supplementary Figure S3 for long-term FRAP and for numerical values).

Z.2.1 and Z.2.2-associated proteins precipitated using GFP nanotrap beads and identified by quantitative mass spectrometry (Figure 4 and Supplementary Table S1 for a complete list of all identified proteins). Whereas the majority of proteins are background binders clustering

C

protein name	ratio H/L Z.2.1	ratio L/H Z.2.2 (ratio H/L)	complex
EP400	25,43	81,32 (0,012)	TIP60
EPC1	25,18	49,73 (0,020)	
TIP60	12,86	21,30 (0,047)	
BRD8	12,75	18,42 (0,054)	
ING3	12,34	30,86 (0,032)	
PAF400	11,10	28,32 (0,035)	
MRGBP	4,95	3,34 (0,299)	
MEAF6	4,68	5,16 (0,194)	
MRG15	*2,79*	*2,05 (0,488)*	
DMAP1	22,85	65,40 (0,015)	TIP60 / SRCAP
TIP49B	22,05	57,00 (0,018)	
TIP49A	21,42	55,18 (0,018)	
GAS41	15,24	54,06 (0,018)	
YL1	14,24	31,44 (0,032)	
BAF53A	10,10	20,98 (0,048)	
ARP6	27,52	71,25 (0,014)	SRCAP
SRCAP	25,51	59,24 (0,017)	
ZNHIT1	NA	19,93 (0,050)	
GFP	23,82	108,88 (0,009)	

Figure 4. Z.2.2 associates with H2A.Z-specific SRCAP and TIP60 chaperone complexes. GFP-pull-downs for H2A.Z-specific chaperone complexes are shown. HK cells stably expressing GFP-Z.2.1 (**A**) and GFP-Z.2.2 (**B**) were SILAC-labeled and subjected to single-step affinity purifications of soluble nuclear proteins in a 'forward' (GFP-Z.2.1) or 'reverse' (GFP-Z.2.2) pull-down using GFP nanotrap beads. In
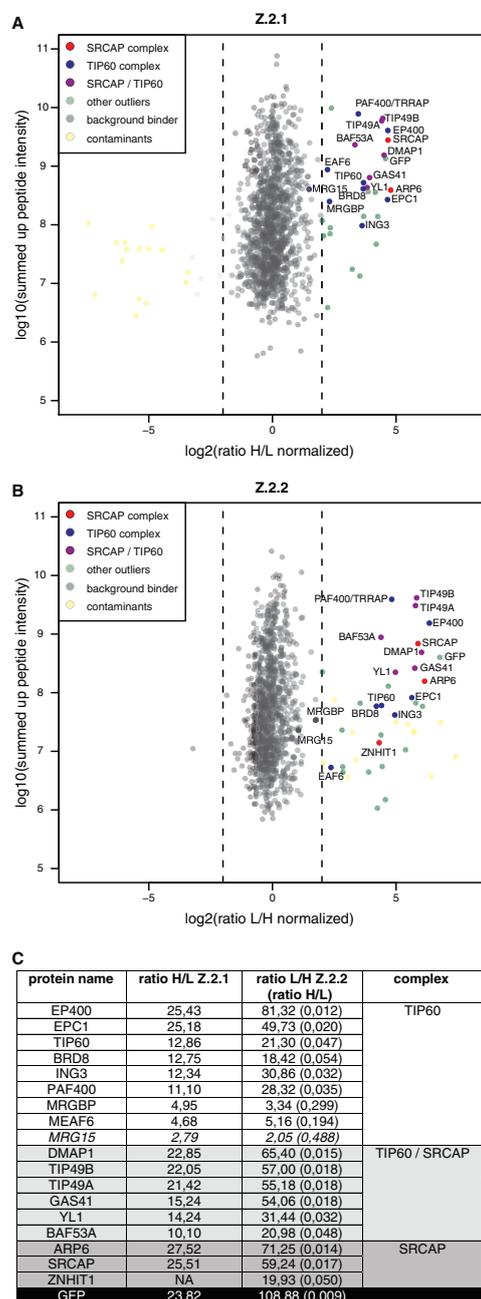
around 0, specific interactors can be found on the right side having a high ratio H/L or ratio L/H for Z.2.1 and Z.2.2, respectively. In accordance with previous studies (62–65), we found GFP-Z.2.1 to be part of two major complexes, the SRCAP and the p400/NuA4/TIP60 (TIP60) complexes (Figure 4A), as we were able to detect all of their thus far identified members, with the exception of actin, as significant outliers. Interestingly, GFP-Z.2.2 also associated with both SRCAP and TIP60 complexes (Figure 4B), showing an almost identical binding composition as GFP-Z.2.1 (Figure 4C). These results strongly imply that Z.2.2 is, similar to other H2A.Z variants, actively deposited into chromatin through specific chaperone complexes.

Based on these results, we predicted that Z.2.1 and Z.2.2 should be incorporated into chromatin in a highly similar spatial manner. Since both SRCAP and TIP60 chaperone complexes are evolutionary conserved between different species, we tested mouse C127 cells that do not express endogenous Z.2.2 for their ability to deposit GFP-Z.2.2. Hereby we should be able to distinguish whether SRCAP and TIP60 complexes are sufficient for deposition, or if other potential primate-specific factors are needed. GFP-Z.2.1 and -Z.2.2 were transiently expressed in C127 cells, S-phase stages highlighted by EdU-incorporation and co-localization patterns visualized by fluorescence microscopy (Figure 5). GFP-Z.2.1 and -Z.2.2 showed an almost identical chromatin localization and deposition pattern, suggesting that Z.2.2 is, like Z.2.1, deposited through SRCAP and TIP60 complexes. In accordance with a recent study, we observed an enrichment of both H2A.Z variants in facultative heterochromatin regions in interphase nuclei (66). Surprisingly, although H2A.Z is expressed in all cell cycle phases (67), and GFP-Z.2.1 and -Z.2.2 expression is driven by a constitutive active promoter, chromatin deposition of both proteins is underrepresented at replication foci. This result underlines our findings that Z.2.2 interacts with all members of both TIP60 and SRCAP complexes and is actively and not passively deposited, as would have been the case during S-phase when nucleosomes are highly exchanged.

Structural changes in Z.2.2's C-terminus prevent histone octamer folding and enhance DNA breathing on structurally destabilized nucleosomes

Our findings thus far imply that Z.2.2 is incorporated into nucleosomes and most likely targeted by TIP60 and

Figure 4. Continued
each panel the ratio of the identified proteins after MS is plotted. Proteins known to interact with H2A.Z are indicated in the following way: members of the SRCAP complex in red, members of the TIP60 complex in blue and shared subunits in purple. Potential novel H2A.Z-interacting proteins are shown as green dots ('other outliers') and are distinguished from background binders (gray dots) and contaminants (yellow dots). See also Supplementary Table S1 for a list of all identified proteins. (**C**) List of the SRCAP and TIP60 complex members and their normalized binding intensity to Z.2.1 or Z.2.2. Note that for comparison reasons the obtained H/L ratios of GFP-Z.2.2 binders (numbers in brackets) were calculated in the corresponding L/H ratios. See also Supplementary Table S1 for a list of all identified proteins and their normalized H/L ratios.
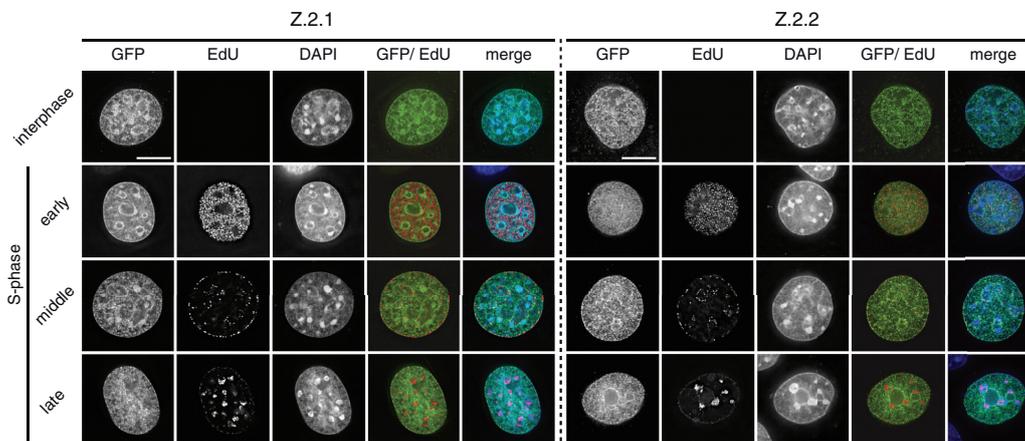
Figure 5. Z.2.1 and Z.2.2 are actively deposited into chromatin and are under-represented at replication foci. C127 cells transiently expressing GFP-Z.2.1 (left) and -Z.2.2 (right) were pulse labeled with EdU to visualize replication foci and to identify S-phase stages. DNA was counterstained with DAPI and analyzed by wide-field deconvolution microscopy. To remove the unbound fraction in GFP-Z.2.2 expressing cells, an *in situ* extraction was performed prior to fixation. Cells in early, middle and late S-phases were distinguished due to their characteristic differential EdU replication labeling patterns of eu- and heterochromatic regions. Merged images in color are shown alongside (GFP: green; EdU: red; DAPI: blue). Scale bar = 5 µm.
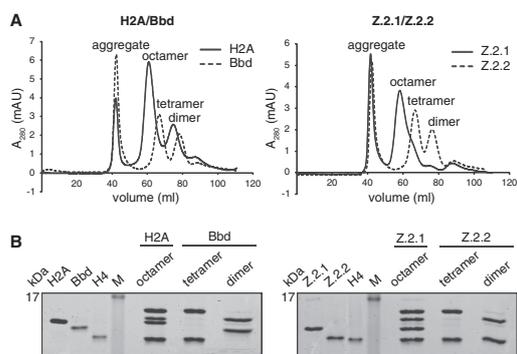

Figure 6. Z.2.2 does not constitute stable histone octamers with H2B, H3 and H4 *in vitro*. (**A**) Size exclusion chromatography of refolding reactions using recombinant human H3, H4 and H2B proteins together with either H2A (solid line) or Bbd (dashed line) (left overlay) or with either Z.2.1 (solid line) or Z.2.2 (dashed line) (right overlay). Peaks corresponding to aggregates, histone octamers, tetramers or dimers are labeled respectively. (**B**) Fractions corresponding to H2A-containing octamers, Bbd-containing tetramers and dimers (left) or Z.2.1-containing octamers and Z.2.2-containing tetramers and dimers (right) were analyzed by 18% SDS–PAGE and stained with Coomassie brilliant blue.

SRCAP complexes. Then why does a large fraction of the cellular Z.2.2 protein pool shows a high mobility and is freely dispersed in the nucleus? One plausible possibility is that Z.2.2 severely destabilizes nucleosomes due to its divergent C-terminal docking domain and is hence rapidly exchanged. To test this hypothesis, we used an *in vitro* reconstitution system. Recombinant human H2A

variants together with H3, H2B and H4 (Supplementary Figure S4A) were refolded by dialysis, and formed complexes purified by size exclusion chromatography. As expected, both H2A and Z.2.1 containing samples readily formed histone octamers (Figure 6A, solid lines). Bbd served as a negative control, because it has been demonstrated to not form octamers under these conditions (41), a result we also observed (Figure 6A left, dotted line). Interestingly, in accordance with our FRAP data, Z.2.2 behaved like Bbd in that it only formed Z.2.2–H2B dimers, but did not complex together with (H3–H4)$_2$ tetramers to generate octamers (Figure 6A right, dotted line), which was further confirmed by SDS–PAGE analyses of the separate fractions (Figure 6B). Thus, like for Bbd the incorporation of Z.2.2 destabilizes the interface between Z.2.2–H2B dimers and (H3–H4)$_2$ tetramers in a C-terminal sequence dependent manner (Supplementary Figure S4B and C). In conclusion, the Z.2.2 docking domain is sufficient to prevent octamer formation.

Although no Z.2.2 containing histone octamers could be generated *in vitro*, our results using GFP-Z.2.2 strongly suggest that Z.2.2 can be part of nucleosomes. To test this *in vitro* and to evaluate the effect of Z.2.2 on nucleosome stability, we reconstituted mononucleosomes by mixing Z.2.2–H2B dimers, (H3–H4)$_2$ tetramers and DNA containing a 'Widom 601' DNA positioning sequence in a 2:1:1 ratio. As controls, we reconstituted H2A or Z.2.1 containing nucleosomes by mixing octamers and DNA in a 1:1 ratio. As expected, analysis of all nucleosomes by native PAGE showed a single band before and after heat shift (Figure 7A), indicating a unique position on the 'Widom 601' DNA template. Purification of nucleosomes from a native gel and analysis of the protein content by
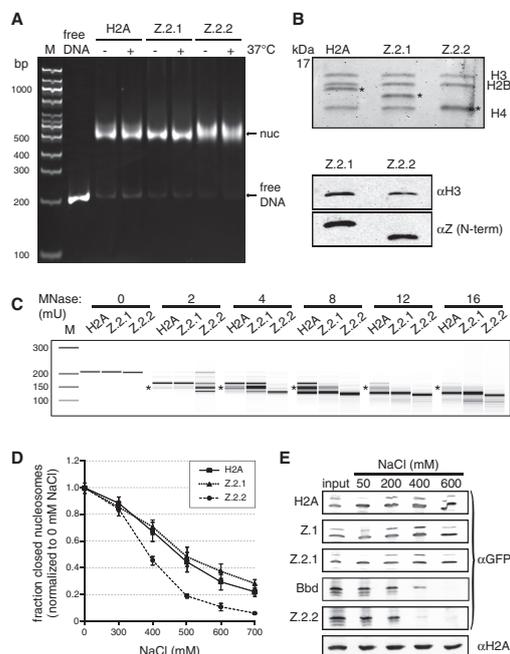
Figure 7. Z.2.2-containing nucleosomes are less resistant to MNase digestion and increased ionic strength. (**A**) H2A, Z.2.1 or Z.2.2 containing nucleosomes were assembled on DNA by salt gradient deposition, incubated at 4°C or 37°C to evaluate DNA positioning and separated by a native 5% PAGE gel. (**B**) Agarose-gel-electro-eluted material from (A) was analyzed by 18% SDS–PAGE and Coomassie stained to evaluate stoichiometry of histones after nucleosome assembly (top). Stars indicate H2A variants that were used for assembly. Further evaluation of histone stoichiometry after nucleosome assembly was done by IB using a LI-COR instrument (bottom). Assembled nucleosomes containing Z.2.1 or Z.2.2 were immunoblotted and the amount of histones was visualized using an αH3 antibody (top) and an N-terminal αZ antibody (recognizes all H2A.Z variants, bottom). (**C**) Mononucleosomes containing either H2A, Z.2.1 or Z.2.2 were digested with increasing concentrations of MNase and extracted DNA was separated using Bioanalyzer. Stars indicate DNA length of 146 bp. For detailed electropherogram analyses of fragment lengths in each sample see Supplementary Figure S5. (**D**) Mononucleosomes containing either H2A, Z.2.1 or Z.2.2 histones together with double dye labeled DNA were incubated with increasing amounts of salt. smFRET measurement values of each salt concentration were normalized to 0 mM NaCl. Error bars represent SEM of six measurements. (**E**) Chromatin from HK-GFP cells was isolated and incubated with increasing amounts of salt. Chromatin-bound histones were precipitated and detected by IB using αGFP antibody. Staining with αH2A was used as loading control.

SDS–PAGE (Coomassie staining and immunoblot) showed that Z.2.2 was indeed incorporated into nucleosomes (Figure 7B). All nucleosomes were further evaluated for their resistance to MNase cleavage as an indicator of stably organized nucleosomes and to determine nucleosomal DNA length (Figure 7C and Supplementary Figure S5). We observed fragments corresponding to protected nucleosomal DNA with the length of

146 bp for all variant nucleosomes tested. The appearance of smaller, subnucleosomal fragments indicates that DNA breathing occurred (68). Interestingly, DNA of Z.2.2 nucleosomes is less protected, since subnucleosomal fragments were obtained at lower MNase concentrations than with H2A or Z.2.1 nucleosomes. Additionally, at higher MNase concentrations a stable DNA fragment of about 120 bp was most abundant for Z.2.2 nucleosomes (Supplementary Figure S5), indicating that this might be the preferred DNA length wrapped around this octamer. These data suggest that increased DNA breathing occurs in Z.2.2 nucleosomes, which as a result might be less stable. To quantify nucleosome stability *in vitro* we measured salt-dependent changes in nucleosome structure using smFRET (69). In line with the results presented above, Z.2.2 containing recombinant nucleosomes lost their compact structure at lower salt concentrations than Z.2.1 or H2A-containing ones (Figure 7D). To investigate whether the observed Z.2.2-dependent nucleosome destabilization is true in the context of chromatin, we isolated chromatin from HK cells expressing GFP-H2A variants and incubated it with buffer containing increasing amounts of salt. Histones that remained stable chromatin components were precipitated and detected by IB (Figure 7E). As observed with FRET techniques, Z.2.2-containing nucleosomes disintegrated between 200 and 400 mM NaCl, and were therefore even less stable than Bbd-containing ones. In summary, incorporation of Z.2.2 leads to a severely reduced nucleosome stability due to C-terminal sequence dependent changes in its docking domain and subsequent loss of its interaction with histone H3.

Our FRAP data suggest that the Z.2.2 C-terminal amino acids might have a direct influence on the nucleosomal structure by affecting interactions with DNA and/or adjacent histones. Based on the existing structural data (50), we performed MD simulations of nucleosomes containing Z.1 (Supplementary Figure S7) or Z.2.2. In addition, we also included the deletion mutant Z.2.1[113], which did not show any change in short-term FRAP (Figure 3C), but some increase in mobility in long-term FRAP (Supplementary Figure S3B) in our assay. These *in silico* models revealed that changes in the C-terminus of H2A.Z strongly affect its protein structure (Figure 8A). Strikingly, different statistical descriptors over the MD-trajectory like distance and mobility (B-factor) show in contrast to Z.1 and Z.2.1[113] unique properties for the Z.2.2 tail. Only Z.2.2 leads to a substantial structural change in the C-terminus resulting in an increased distance to histone H3, which in turn makes a hydrogen bond interaction between peptide backbone NH of Cys112 in Z.2.2 and the oxygen in the Gln55 side chain in H3 impossible (Figure 8B). Additionally, an increase in the B-factor for Z.2.2 indicates a substantially enhanced mobility of Z.2.2's C-terminus (Figure 8C). We also calculated the Z.2.2–H3 interaction energy and observed a switch from negative to positive values in the case of Z.2.2 suggesting that this histone variant destabilizes the nucleosome (Figure 8D). In summary, these data suggest that the C-terminal sequence of Z.2.2 leads to a more dynamic structure that in turn loses binding to histone
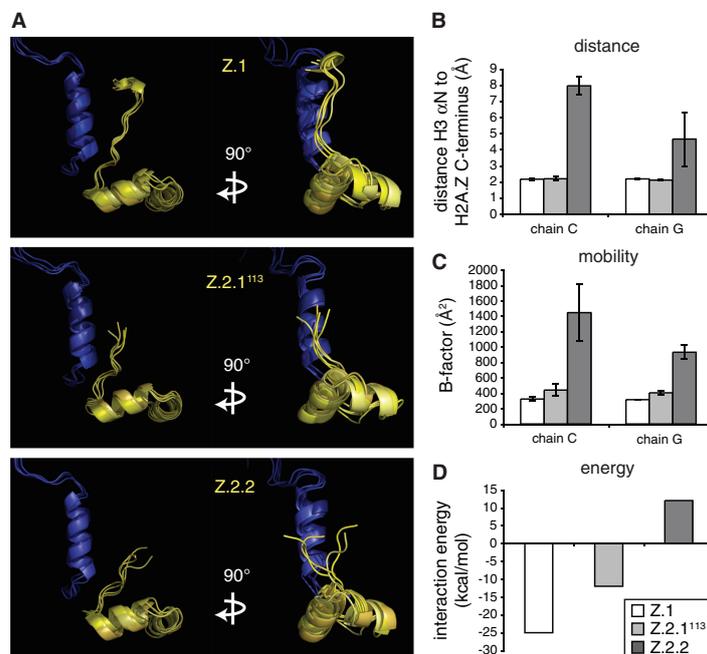
Figure 8. Unique Z.2.2 C-terminal amino acids cause significant changes in protein and nucleosome structure. (**A**) *In silico* models of Z.1, Z.2.1[113] and Z.2.2 C-terminal C-chains (yellow; from amino acids 84 to C-terminus, including the complete docking domain) together with the E-chain of histone H3 (blue; amino acids 33–60, including αN-helix). Side (left) and frontal views (right) of four MD simulations are shown respectively. See Supplementary Figure S7 for complete *in silico* model of H2A.Z-containing nucleosome. (**B**) Simulated distances between peptide backbone NH of amino acids 112 in H2A.Z (His or Cys, respectively) variants and the oxygen in the Gln55 sidechain in H3 based on *in silico* nucleosome models containing either Z.1 (white), Z.2.1[113] (light gray) or Z.2.2 (dark gray) proteins. Error bars represent SD of four independent simulations. (**C**) Simulated mobility measuring B-factor values between amino acids 108 and 113 in respective H2A.Z variant C-termini. Error bars represent SD of four independent simulations. (**D**) Simulated interaction energy between tetramer versus respective H2A.Z variant-containing dimers.

H3 and destabilizes the nucleosomal structure, providing a reasonable explanation for the observed *in vivo* and *in vitro* data.

DISCUSSION

In this work, we have identified a previously unknown histone H2A.Z variant and provide a comprehensive characterization of its nucleosomal properties. This alternatively spliced variant, Z.2.2, is present to different degrees in all human cell lines and tissues investigated, with a significant enrichment in brain. Z.2.2 contains a shortened and in six amino acids divergent C-terminus compared to Z.1 and Z.2.1 that is necessary, but only not sufficient, to weaken chromatin association. Only in the context of the unique Z.2.2 docking domain does the C-terminal sequence negatively affect nucleosome stability *in vitro* and *in vivo*. To our knowledge, Z.2.2 has the strongest destabilizing effect on nucleosomal structure compared to other histone H2A variants reported to date.

Only one other histone variant, macroH2A, has been shown thus far to be alternatively spliced (70). Here, like our observation with H2A.Z, two independent genes *mH2A1* and *mH2A2* exist in mammals, with only mH2A1 being alternatively spliced resulting in functional different proteins (71). In our study, we demonstrate that the human H2A.Z.2 (H2AFV) primary transcript is alternatively spliced generating Z.2.1 and Z.2.2 mRNAs and proteins. These observations suggest that Z.2.2 is tightly regulated in a tissue-specific manner through alternative splicing and/or RNA stability. Our findings now raise the intriguing possibility that alternative splicing of histone variants might not be rare but more common than previously thought. If true, it will be of interest to reevaluate other intron-containing histone variant genes with regard to their possible alternative transcripts and protein products.

Bioinformatic genome analyses revealed the existence of Z.2.2-specific sequences only in humans, old and new world primates and to some extend in other mammals, with the exclusion of mouse, rat and even lower eukaryotes. It remains to be seen, whether Z.2.2's evolution is indeed limited to primates only. Primate-specific gene products have been often identified in human brain and reproductive tissues (72), supporting the notion that

their RNAs and proteins might be essential to adaptive changes leading to human development and further speculates that primate-specific genes might be important in reproductive function and disease. Since we have found Z.2.2 transcripts to be strongly enriched in brain samples of higher brain function in comparison to other tissues and cell types, it will be of great interest to determine in future studies, if this novel variant might play an important functional role in this particular organ. These observations also raise the interesting question of how alternative splicing and/or differential stability of H2AFV transcripts are tissue specifically regulated.

Another intriguing feature of Z.2.2 is its influence on nucleosome stability. Although Z.2.2 localizes exclusively to the nucleus, only a minor proportion is stably incorporated into chromatin. The only other exception in humans known thus far is Bbd, which has previously been demonstrated to destabilize the nucleosome structure (41,53,73). Bbd, similar to Z.2.2, is a shorter H2A variant with an unusual C-terminus and a considerable different primary histone fold sequence that might explain its ability to destabilize nucleosomes. In agreement, a recent study demonstrated that the incomplete C-terminal docking domain of Bbd results in structural alterations in nucleosomes and that those are in turn associated with an inability of the chromatin remodeler RSC to both remodel and mobilize nucleosomes (8). Z.2.2, on the other hand, is identical to Z.2.1, except that its C-terminus is 14 amino acids shorter and in six amino acids altered. How can this small change in Z.2.2's primary sequence lead to such drastic effects on chromatin association?

We show that Z.2.2 can be part of a bona fide nucleosome and that it interacts with the H2A.Z-specific TIP60 and SRCAP chaperone complexes. These complexes have been shown to catalyze the exchange of H2A–H2B dimers with H2A.Z–H2B dimers in nucleosomes and our finding therefore suggests that both complexes are also involved in an active chromatin incorporation of Z.2.2. Supporting this idea is the observation that both Z.2.1 and Z.2.2 are incorporated into chromatin in a replication-independent manner, even in mouse cells that do not express endogenous Z.2.2. Both H2A.Z variants are not primarily deposited at replication foci, not even in middle S-phase when facultative heterochromatin is replicated, where the majority of the H2A.Z protein pool is found in interphase cells (66). Our findings are in agreement with a model proposed by Hardy and Robert, in which H2A.Z variants are randomly deposited into chromatin by specific chaperone complexes in a replication-independent manner coupled to a subsequent targeted H2A.Z depletion (74). As a consequence, an enrichment of H2A.Z at non-transcribed genes and heterochromatin regions over several cell generations can be observed (74). It might be possible that INO80 facilitates this eviction function, as it has been shown to exchange nucleosomal H2A.Z–H2B dimers with free H2A–H2B dimers (75). It will be of interest in future studies to determine whether Z.2.2 exchange is subjected to a similar mechanism. Taken together, our findings strongly imply that Z.2.2 is actively deposited into chromatin through the interaction with evolutionary conserved chaperone complexes. Nevertheless, a large fraction of Z.2.2 protein is not chromatin bound and we have mapped the region crucial for high FRAP mobility to its docking domain. In addition to Z.2.2's unique C-terminal amino acids this region spans the highly conserved acidic patch responsible for deposition (76), the M6 region that is functionally essential in fly H2A.Z (60) and required for the interaction with the SWR1 complex in yeast (77). Strikingly, *in silico* simulation of Z.2.2 predicted dynamic structural changes that in turn weaken interaction with histone H3 and destabilize the nucleosome structure. Such a gross structural alteration explains why Z.2.2 is not able to form stable octamers *in vitro* and leads to enhanced DNA breathing in a nucleosomal context. Hence, Z.2.2 incorporation into chromatin disrupts nucleosomes more easily and supports a model in which Z.2.2 is more rapidly exchanged than Z.2.1.

What functional outcome might Z.2.2 cause when incorporated into chromatin? And how is the variant composition of Z.2.2 containing nucleosomes? It has been shown that a special class of nucleosomes containing both H2A.Z and H3.3 variants exists in humans (78). These nucleosomes are enriched at promoters, enhancers and insulator region and promote the accessibility of transcription factors to these DNA regions (78), most likely due to their extreme sensitivity to disruption (79). Since these studies nicely demonstrate that differential nucleosome stabilities due to the incorporation of different histone variants influence transcriptional regulation, it is tempting to speculate that Z.2.2 might also affect chromatin-related processes. Future experiments will shed light on Z.2.2 function(s), especially with regard to its increased expression in human brain areas, and explain why and where nucleosomal destabilization is needed. This is of particular interest, since Bbd that also leads to nucleosomal destabilization is almost exclusively present in testis (80–82) in contrast to the apparently ubiquitously expressed Z.2.2, possibly pointing toward distinct roles of both destabilizing H2A variants in different tissues. Our data suggest that additional interesting, yet unidentified, histone variants may exist and await their discovery.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table S1, Supplementary Figures S1–S7 and Supplementary Materials and Methods.

ACKNOWLEDGEMENTS

manuscript. The authors thank especially the Hake lab and members of the Adolf-Butenandt Institute for constructive discussions.

REFERENCES

1. van Holde,K.E. (1988) *Chromatin*. Springer, New York.
2. Bonisch,C., Nieratschker,S.M., Orfanos,N.K. and Hake,S.B. (2008) Chromatin proteomics and epigenetic regulatory circuits. *Exp. Rev. Proteomics*, **5**, 105–119.
3. Bernstein,E. and Hake,S.B. (2006) The nucleosome: a little variation goes a long way. *Biochem. Cell Biol.*, **84**, 505–517.
4. Boulard,M., Bouvet,P., Kundu,T.K. and Dimitrov,S. (2007) Histone variant nucleosomes: structure, function and implication in disease. *Subcell. Biochem.*, **41**, 71–89.
5. Redon,C., Pilch,D., Rogakou,E., Sedelnikova,O., Newrock,K. and Bonner,W. (2002) Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.*, **12**, 162–169.
6. Ausio,J. and Abbott,D.W. (2002) The many tales of a tail: carboxyl-terminal tail heterogeneity specializes histone H2A variants for defined chromatin function. *Biochemistry*, **41**, 5945–5949.
7. Wang,A.Y., Aristizabal,M.J., Ryan,C., Krogan,N.J. and Kobor,M.S. (2011) Key functional regions in the histone variant H2A.Z C-terminal docking domain. *Mol. Cell. Biol.*, **31**, 3871–3884.
8. Shukla,M.S., Syed,S.H., Goutte-Gattat,D., Richard,J.L., Montel,F., Hamiche,A., Travers,A., Faivre-Moskalenko,C., Bednar,J., Hayes,J.J. *et al*. (2011) The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling. *Nucleic Acids Res.*, **39**, 2559–2570.
9. Vogler,C., Huber,C., Waldmann,T., Ettig,R., Braun,L., Izzo,A., Daujat,S., Chassignet,I., Lopez-Contreras,A.J., Fernandez-Capetillo,O. *et al*. (2010) Histone H2A C-terminus regulates chromatin dynamics, remodeling, and histone H1 binding. *PLoS Genet.*, **6**, e1001234.
10. Billon,P. and Cote,J. (2011) Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. *Biochim Biophys Acta*, **1819**, 290–302.
11. Marques,M., Laflamme,L., Gervais,A.L. and Gaudreau,L. (2010) Reconciling the positive and negative roles of histone H2A.Z in gene transcription. *Epigenetics*, **5**, 267–272.
12. Draker,R. and Cheung,P. (2009) Transcriptional and epigenetic functions of histone variant H2A.Z. *Biochem. Cell Biol.*, **87**, 19–25.
13. Svotelis,A., Gevry,N. and Gaudreau,L. (2009) Regulation of gene expression and cellular proliferation by histone H2A.Z. *Biochem. Cell Biol.*, **87**, 179–188.
14. Brickner,D.G., Cajigas,I., Fondufe-Mittendorf,Y., Ahmed,S., Lee,P.C., Widom,J. and Brickner,J.H. (2007) H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol.*, **5**, e81.
15. Zofall,M., Fischer,T., Zhang,K., Zhou,M., Cui,B., Veenstra,T.D. and Grewal,S.I. (2009) Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature*, **461**, 419–422.
16. Rangasamy,D., Berven,L., Ridgway,P. and Tremethick,D.J. (2003) Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J.*, **22**, 1599–1607.
17. Rangasamy,D., Greaves,I. and Tremethick,D.J. (2004) RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. *Nat. Struct. Mol. Biol.*, **11**, 650–655.
18. Zlatanova,J. and Thakar,A. (2008) H2A.Z: view from the top. *Structure*, **16**, 166–179.
19. Dryhurst,D., McMullen,B., Fazli,L., Rennie,P.S. and Ausio,J. (2012) Histone H2A.Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated transcription and is upregulated in a model of prostate cancer progression. *Cancer Lett.*, **315**, 38–47.
20. Valdes-Mora,F., Song,J.Z., Statham,A.L., Strbenac,D., Robinson,M.D., Nair,S.S., Patterson,K.I., Tremethick,D.J., Stirzaker,C. and Clark,S.J. (2012) Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Res.*, **22**, 307–321.
21. Conerly,M.L., Teves,S.S., Diolaiti,D., Ulrich,M., Eisenman,R.N. and Henikoff,S. (2010) Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Res.*, **20**, 1383–1390.
22. Svotelis,A., Gevry,N., Grondin,G. and Gaudreau,L. (2010) H2A.Z overexpression promotes cellular proliferation of breast cancer cells. *Cell Cycle*, **9**, 364–370.
23. Hua,S., Kallen,C.B., Dhar,R., Baquero,M.T., Mason,C.E., Russell,B.A., Shah,P.K., Liu,J., Khramtsov,A., Tretiakova,M.S. *et al*. (2008) Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression. *Mol. Syst. Biol.*, **4**, 188.
24. Eirin-Lopez,J.M., Gonzalez-Romero,R., Dryhurst,D., Ishibashi,T. and Ausio,J. (2009) The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. *BMC Evol. Biol.*, **9**, 31.
25. Beck,H.C., Nielsen,E.C., Matthiesen,R., Jensen,L.H., Sehested,M., Finn,P., Grauslund,M., Hansen,A.M. and Jensen,O.N. (2006) Quantitative proteomic analysis of post-translational modifications of human histones. *Mol. Cell Proteomics*, **5**, 1314–1325.
26. Bonenfant,D., Coulot,M., Towbin,H., Schindler,P. and van Oostrum,J. (2006) Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Mol. Cell Proteomics*, **5**, 541–552.
27. Dryhurst,D., Ishibashi,T., Rose,K.L., Eirin-Lopez,J.M., McDonald,D., Silva-Moreno,B., Veldhoen,N., Helbing,C.C., Hendzel,M.J., Shabanowitz,J. *et al*. (2009) Characterization of the histone H2A.Z-1 and H2A.Z-2 isoforms in vertebrates. *BMC Biol.*, **7**, 86.
28. Sarcinella,E., Zuzarte,P.C., Lau,P.N., Draker,R. and Cheung,P. (2007) Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. *Mol. Cell. Biol.*, **27**, 6457–6468.
29. Wiedemann,S.M., Mildner,S.N., Bonisch,C., Israel,L., Maiser,A., Matheisl,S., Straub,T., Merkl,R., Leonhardt,H., Kremmer,E. *et al*. (2010) Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J. Cell. Biol.*, **190**, 777–791.
30. Shechter,D., Dormann,H.L., Allis,C.D. and Hake,S.B. (2007) Extraction, purification and analysis of histones. *Nat. Protoc.*, **2**, 1445–1457.
31. Wysocka,J., Reilly,P.T. and Herr,W. (2001) Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol. Cell. Biol.*, **21**, 3820–3829.
32. Wessel,D. and Flugge,U.I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.*, **138**, 141–143.
33. Zhang,H., Roberts,D.N. and Cairns,B.R. (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, **123**, 219–231.

34. Hake,S.B., Garcia,B.A., Kauer,M., Baker,S.P., Shabanowitz,J., Hunt,D.F. and Allis,C.D. (2005) Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. *Proc. Natl Acad. Sci. USA*, **102**, 6344–6349.

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

36. Vermeulen,M., Mulder,K.W., Denissov,S., Pijnappel,W.W., van Schaik,F.M., Varier,R.A., Baltissen,M.P., Stunnenberg,H.G., Mann,M. and Timmers,H.T. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, **131**, 58–69.

37. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

38. Salic,A. and Mitchison,T.J. (2008) A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl Acad. Sci. USA*, **105**, 2415–2420.

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

40. Thastrom,A., Lowary,P.T., Widlund,H.R., Cao,H., Kubista,M. and Widom,J. (1999) Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.*, **288**, 213–229.

41. Bao,Y., Konesky,K., Park,Y.J., Rosu,S., Dyer,P.N., Rangasamy,D., Tremethick,D.J., Laybourn,P.J. and Luger,K. (2004) Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J.*, **23**, 3314–3324.

42. Gansen,A., Valeri,A., Hauger,F., Felekyan,S., Kalinin,S., Toth,K., Langowski,J. and Seidel,C.A. (2009) Nucleosome disassembly intermediates characterized by single-molecule FRET. *Proc. Natl Acad. Sci. USA*, **106**, 15308–15313.

43. Muller,B.K., Zaychikov,E., Brauchle,C. and Lamb,D.C. (2005) Pulsed interleaved excitation. *Biophys. J.*, **89**, 3508–3522.

44. Nir,E., Michalet,X., Hamadani,K.M., Laurence,T.A., Neuhauser,D., Kovchegov,Y. and Weiss,S. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B*, **110**, 22103–22124.

45. Lee,N.K., Kapanidis,A.N., Wang,Y., Michalet,X., Mukhopadhyay,J., Ebright,R.H. and Weiss,S. (2005) Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys. J.*, **88**, 2939–2953.

46. Duan,Y., Wu,C., Chowdhury,S., Lee,M.C., Xiong,G., Zhang,W., Yang,R., Cieplak,P., Luo,R., Lee,T. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.

47. Wang,J., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.

48. Jakalian,A., Jack,D.B. and Bayly,C.I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, **23**, 1623–1641.

49. Krieger,E., Darden,T., Nabuurs,S.B., Finkelstein,A. and Vriend,G. (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*, **57**, 678–683.

50. Suto,R.K., Clarkson,M.J., Tremethick,D.J. and Luger,K. (2000) Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.*, **7**, 1121–1124.

51. Krieger,E., Nielsen,J.E., Spronk,C.A. and Vriend,G. (2006) Fast empirical pKa prediction by Ewald summation. *J. Mol. Graph Model.*, **25**, 481–486.

52. Higashi,T., Matsunaga,S., Isobe,K., Morimoto,A., Shimada,T., Kataoka,S., Watanabe,W., Uchiyama,S., Itoh,K. and Fukui,K. (2007) Histone H2A mobility is regulated by its tails and acetylation of core histone tails. *Biochem. Biophys. Res. Commun.*, **357**, 627–632.

53. Gautier,T., Abbott,D.W., Molla,A., Verdel,A., Ausio,J. and Dimitrov,S. (2004) Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep.*, **5**, 715–720.

54. Lever,M.A., Th'ng,J.P., Sun,X. and Hendzel,M.J. (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, **408**, 873–876.

55. Th'ng,J.P., Sung,R., Ye,M. and Hendzel,M.J. (2005) H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. *J. Biol. Chem.*, **280**, 27809–27814.

56. Misteli,T., Gunjan,A., Hock,R., Bustin,M. and Brown,D.T. (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature*, **408**, 877–881.

57. Stasevich,T.J., Mueller,F., Brown,D.T. and McNally,J.G. (2010) Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. *EMBO J.*, **29**, 1225–1234.

58. Schmiedeberg,L., Weisshart,K., Diekmann,S., Meyer Zu Hoerste,G. and Hemmerich,P. (2004) High- and low-mobility populations of HP1 in heterochromatin of mammalian cells. *Mol. Biol. Cell.*, **15**, 2819–2833.

59. Cheutin,T., McNairn,A.J., Jenuwein,T., Gilbert,D.M., Singh,P.B. and Misteli,T. (2003) Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science*, **299**, 721–725.

60. Clarkson,M.J., Wells,J.R., Gibson,F., Saint,R. and Tremethick,D.J. (1999) Regions of variant histone His2AvD required for Drosophila development. *Nature*, **399**, 694–697.

61. Lu,P.Y., Levesque,N. and Kobor,M.S. (2009) NuA4 and SWR1-C: two chromatin-modifying complexes with overlapping functions and components. *Biochem. Cell Biol.*, **87**, 799–815.

62. Cai,Y., Jin,J., Florens,L., Swanson,S.K., Kusch,T., Li,B., Workman,J.L., Washburn,M.P., Conaway,R.C. and Conaway,J.W. (2005) The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. *J. Biol. Chem.*, **280**, 13665–13670.

63. Ruhl,D.D., Jin,J., Cai,Y., Swanson,S., Florens,L., Washburn,M.P., Conaway,R.C., Conaway,J.W. and Chrivia,J.C. (2006) Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. *Biochemistry*, **45**, 5671–5677.

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

65. Auger,A., Galarneau,L., Altaf,M., Nourani,A., Doyon,Y., Utley,R.T., Cronier,D., Allard,S. and Cote,J. (2008) Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. *Mol. Cell. Biol.*, **28**, 2257–2270.

66. Hardy,S., Jacques,P.E., Gevry,N., Forest,A., Fortin,M.E., Laflamme,L., Gaudreau,L. and Robert,F. (2009) The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. *PLoS Genet.*, **5**, e1000687.

67. Wu,R.S., Tsai,S. and Bonner,W.M. (1982) Patterns of histone variant synthesis can distinguish G0 from G1 cells. *Cell*, **31**, 367–374.

68. Ramaswamy,A., Bahar,I. and Ioshikhes,I. (2005) Structural dynamics of nucleosome core particle: comparison with nucleosomes containing histone variants. *Proteins*, **58**, 683–696.

69. Ha,T., Enderle,T., Ogletree,D.F., Chemla,D.S., Selvin,P.R. and Weiss,S. (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA*, **93**, 6264–6268.

70. Rasmussen,T.P., Huang,T., Mastrangelo,M.A., Loring,J., Panning,B. and Jaenisch,R. (1999) Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. *Nucleic Acids Res.*, **27**, 3685–3689.

71. Kustatscher,G., Hothorn,M., Pugieux,C., Scheffzek,K. and Ladurner,A.G. (2005) Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.*, **12**, 624–625.

72. Tay,S.K., Blythe,J. and Lipovich,L. (2009) Global discovery of primate-specific genes in the human genome. *Proc. Natl Acad. Sci. USA*, **106**, 12019–12024.

73. Angelov,D., Verdel,A., An,W., Bondarenko,V., Hans,F., Doyen,C.M., Studitsky,V.M., Hamiche,A., Roeder,R.G., Bouvet,P. *et al.* (2004) SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays. *EMBO J.*, **23**, 3815–3824.

74. Hardy,S. and Robert,F. (2010) Random deposition of histone variants: a cellular mistake or a novel regulatory mechanism? *Epigenetics*, **5**, 368–372.

75. Papamichos-Chronakis,M., Watanabe,S., Rando,O.J. and Peterson,C.L. (2011) Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell*, **144**, 200–213.

76. Jensen,K., Santisteban,M.S., Urekar,C. and Smith,M.M. (2011) Histone H2A.Z acid patch residues required for deposition and function. *Mol. Genet. Genomics.*, **285**, 287–296.

77. Wu,W.H., Alami,S., Luk,E., Wu,C.H., Sen,S., Mizuguchi,G., Wei,D. and Wu,C. (2005) Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nat. Struct. Mol. Biol.*, **12**, 1064–1071.

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

79. Jin,C. and Felsenfeld,G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.*, **21**, 1519–1529.

80. Chadwick,B.P. and Willard,H.F. (2001) A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J. Cell. Biol.*, **152**, 375–384.

81. Ishibashi,T., Li,A., Eirin-Lopez,J.M., Zhao,M., Missiaen,K., Abbott,D.W., Meistrich,M., Hendzel,M.J. and Ausio,J. (2010) H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. *Nucleic Acids Res.*, **38**, 1780–1789.

82. Soboleva,T.A., Nekrasov,M., Pahwa,A., Williams,R., Huttley,G.A. and Tremethick,D.J. (2011) A unique H2A histone variant occupies the transcriptional start site of active genes. *Nat. Struct. Mol. Biol.*, **19**, 25–30.

2.4 Quantitative proteomics for epigenetics

2.4.1 Overview

Mass spectrometry-based proteomics has matured into a powerful method to answer questions in several biological areas. In the chromatin field proteomics, has made important contributions by mapping post-translational modifications on histones and defining protein complexes. The broad interest of the chromatin community in quantitative proteomics prompted us to write a proteomics review dedicated to chromatin biology. We introduce the technical knowledge which is necessary to design and judge proteomic experiments. The review then focuses on quantitative approaches and highlights the importance to perform a proper quantification to generate high quality data. Selected examples are discussed to illustrate the power of proteomics.

2.4.2 Contribution

I drafted the whole review except for the interaction proteomics section, which was drafted by Michiel Vermeulen. The draft was edited by all authors.
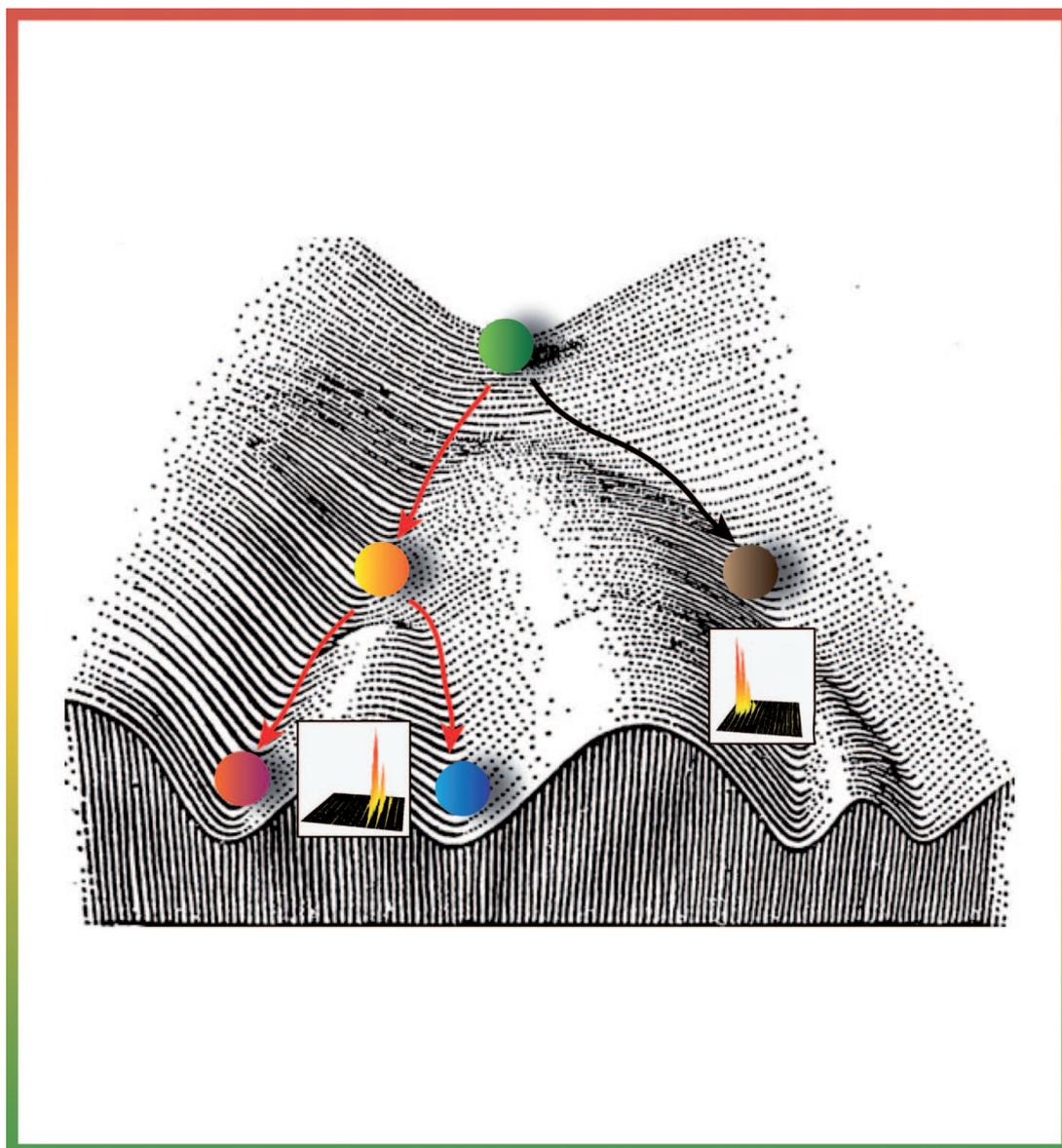
2.4.3 Publication

This review article was published in 2011:

Quantitative proteomics for epigenetics

H. Christian Eberl, Matthias Mann and Michiel Vermeulen

Chembiochem 2011, 12, 224-234

CHEMBIOCHEM

Quantitative Proteomics for Epigenetics

H. Christian Eberl,[a] Matthias Mann,*[a] and Michiel Vermeulen*[b]

84

REVIEWS

Mass spectrometry has made many contributions to the chromatin field through the mapping of histone modifications and the identification of protein complexes involved in gene regulation. MS-based proteomics has now evolved from the identification of single protein spots in gels to the identification and quantification of thousands of proteins in complex mixtures. Quantitative approaches also allow comparative and time-resolved analysis of post-translational modifications. An important emerging field is the unbiased interaction analysis of proteins with other proteins, defined protein modifications, specific DNA and RNA sequences, and small molecules. Quantitative proteomics can also accurately monitor whole proteome changes in response to perturbation of the gene expression machinery. We provide an up-to-date review of modern quantitative proteomic technology and its applications in the field of epigenetics.

Introduction

Epigenetics, a term first coined by Waddington in the 1950s, is defined as the study of inheritable changes in phenotype and gene expression not resulting from changes in the underlying DNA sequence.[1] Epigenetic regulation of gene expression plays a key role during cellular differentiation or de-differentiation, lineage commitment, and development.[2] Moreover, many enzymes that are involved in establishing the epigenetic landscapes are deregulated in cancer or are mutated in inheritable neurological diseases.[3] An example of the latter case can be found in PHF8, the PHD finger of which specifically recognizes the histone-3 lysine-4 trimethylation (H3K4me3) mark. When mutated, this gene is associated with cleft lip/palate and X-linked mental retardation.[4] Because of the importance of epigenetics in maintaining cell identity and initiating cell fate changes, great efforts to dissect the underlying principles and mechanisms have been made over recent years. The major determinants for epigenetic control of gene expression are methylation of cytosines in CpG dinucleotides[5] and post-translational modifications of core histones.[6] Chromatin, previously thought to be a static entity, is now appreciated to be a highly dynamic structure. It is regulated by protein complexes that act as "writers", which add histone modifications, and "erasers", which remove them. Additionally, chromatin "readers" specifically recognize certain modifications and exert their function at the site of recruitment.[7] Moreover, ATP-dependent chromatin remodeling enzymes use ATP hydrolysis to change the positions of nucleosomes on DNA and thereby influence gene expression.[8] Given patterns of histone modifications on genes or larger chromosomal domains are thought to establish the accessibility states and to modulate the transcriptional activity of genes. The specific combination of diverse post-translational modifications (PTMs) of the core histones has been termed the "histone code",[9] but the extent to which this code dictates gene expression remains controversial.[10]

Mass spectrometry has greatly contributed to epigenetic research through the mapping of PTMs on core histones[11] and the identification of subunits of purified chromatin-associated protein complexes. Technological advances in MS-based proteomics now provide the tools for the identification of thousands of proteins in complex mixtures.[12] Until recently, however, most proteomic data were obtained from low-resolution instruments and were purely qualitative in nature. This approach results in the identification of many proteins, but in the absence of quantitative information it is not easy to differentiate between background or contaminating proteins and proteins genuinely associated with the process under investigation. Fortunately, modern high-resolution instruments, combined with sophisticated data analysis, have greatly improved the reliability of data. In addition, methods that allow researchers to quantify the relative abundances of proteins in two or more samples have been developed. These developments hold great promise for answering epigenetic questions.

Here we argue that modern proteomics experiments demand high-accuracy data and quantitative read-out to separate functional candidates from background. Although only a few groups so far use truly quantitative approaches, we believe that these technologies can now be broadly applied. We review the use of MS-based quantitative proteomics for the epigenetics and chromatin fields and describe its application in the study of protein modifications, interactions, and expression dynamics.

State-of-the-Art Quantitative Proteomics

Instrumentation and workflow

MS-based proteomics is a relatively young but rapidly advancing field that, thanks to constant technological improvements, is becoming a powerful complement to genomic approaches.[13]

The primary function of the mass spectrometer in proteomics is the identification of peptides and proteins. Although it is possible to analyze intact proteins by MS, which is referred to as top-down mass spectrometry, this approach has severe technical limitations. In epigenetic research it is only used for special purposes, such as the analysis of intact histones.[14] In "bottom-up" proteomics, proteins are digested into peptides and these are analyzed by the mass spectrometer much more efficiently (Figure 1). Depending on the complexity of the

[a] H. C. Eberl, Prof. M. Mann
Department of Proteomics and Signal Transduction
Max Planck Institute for Biochemistry
Am Klopferspitz 18, 82152 Martinsried (Germany)
Fax: (+ 49) 89-8578-2219
E-mail: mmann@biochem.mpg.de

[b] Dr. M. Vermeulen
Department of Physiological Chemistry and Cancer Genomics Centre
University Medical Center Utrecht
Universiteitsweg 100, 3584 CG Utrecht (The Netherlands)
Fax: (+ 31) 88-75-68101
E-mail: m.vermeulen-3@umcutrecht.nl

M. Mann, M. Vermeulen and H. C. Eberl

sample to be analyzed, a subfractionation on the peptide or protein level can be added to the workflow. After this optional fractionation, peptides are usually separated by on-line reversed-phase nanoscale high-performance liquid chromatography (nano-HPLC). At the end of the packed column the effluent is directly electrosprayed by means of an applied voltage. Solvent droplets containing the peptides rapidly evaporate, and protonated peptides are injected into the mass spectrometer. To identify a peptide, the mass of the intact peptide (precursor mass) is determined from the MS spectrum, followed by isolation and fragmentation of the peptide. Peptide fragmenta-

H. Christian Eberl received his MSc in biochemistry from the Technical University Munich in 2008. Since 2008 he has been pursuing his PhD at the Max Planck Institute of Biochemistry in Martinsried in the research group of Prof. Matthias Mann, applying quantitative mass spectrometry to study protein–protein interactions related to chromatin.

Matthias Mann studied physics and mathematics at the University of Göttingen and received his doctorate in chemical engineering from Yale University in 1988. He contributed to the development of electrospray mass spectrometry (Nobel Prize in Chemistry in 2002 for M.M.'s supervisor John B. Fenn). After stations at the European Molecular Biology Laboratory (EMBL) as group leader and the University of Southern Denmark (SDU) as director of the Center for Experimental BioInformatics (CEBI), in 2005 M.M. became a director at the Max Planck Institute of Biochemistry in Martinsried. He is also affiliated with the NNF Center for Protein Research (University of Copenhagen).

Michiel Vermeulen performed his doctoral work in the laboratory of Henk Stunnenberg at the University of Nijmegen in the Netherlands and received his PhD in 2006. He then joined the lab. of Matthias Mann in Munich, Germany, where, amongst other things, he pioneered the application of quantitative proteomics technology to study interactions between nuclear proteins and modified histone tails. In March 2009 he was appointed as assistant professor at the University Medical Center Utrecht in the Netherlands, where he continues to use quantitative proteomics technology to study chromatin structure and function.

tion is achieved by energy transfer to the ionized peptides in the gas phase and a fragmentation spectrum of the peptide (MS/MS or tandem mass spectrum) is recorded. The standard fragmentation method is collision-induced dissociation (CID), in which the peptide collides with an inert gas at low pressure. Alternatively, electron capture dissociation (ECD)[15] or electron transfer dissociation (ETD)[16] activate peptides by very fast processes that minimize energy randomization and can thereby preserve labile PTMs.

CID causes peptide fragmentation at amide bonds, resulting in a series of fragments that differ in their masses by single amino acids (for an introduction see ref. [17]). In an ideal fragmentation spectrum one would therefore be able to "read" the sequence of a peptide in the spectrum (de novo sequencing). Most fragmentation spectra only contain partial sequence information, however, and so statistical algorithms are used to find best matches in an amino acid database (Figure 2A). The more accurately the parent mass is measured, the smaller the applied search window, thus reducing the number of possible matches. In addition, the more complete the fragmentation spectrum, the more confident the identification.

The complex peptide mixtures that are typical in "shotgun proteomics" experiments contain many thousands of peptides. Modern mass spectrometers therefore provide high-quality data in combination with high MS/MS sequencing speed. Other key parameters are the mass spectrometric resolution (a dimensionless number calculated by dividing the width of a peak by its mass) and the "dynamic range" (the ratio of the strongest signal to the weakest signal that can still be detected in a spectrum). Today most mass spectrometers are so-called hybrid instruments, usually either as a combination of a quadruple mass filter and a time-of-flight analyzer or as a combination of a linear ion trap and an Orbitrap analyzer. Both types of hybrid instruments offer sequencing speeds of several MS/MS spectra per second. Orbitrap analyzers are based on frequency detection and offer routine resolution of more than 50 000 with matching mass accuracy. The dynamic ranges in single spectra are in the range of 1000 to 10 000 for both types of instrument. A new linear ion trap Orbitrap instrument (LTQ-Orbitrap Velos) allows cycles of one MS followed by 20 MS/MS events in only 2.5 s. It is also routinely capable of recording MS/MS spectra at high resolution either by CID or by "higher-energy collisional dissociation" (HCD) methods.[18]

Quantification

The second function of the mass spectrometer is to provide quantitative information relating to relative or absolute peptide abundance. This is not straightforward, though, because MS in itself is not quantitative, mainly due to the different ionization efficiencies of different peptides. During the last decade, several methods that add a quantitative dimension to mass spectrometric measurements have been developed.[19] These methods can be divided into two groups: label-free and stable isotope labeling approaches.

The most straightforward label-free quantification approaches are spectral counting[20] and exponentially modified
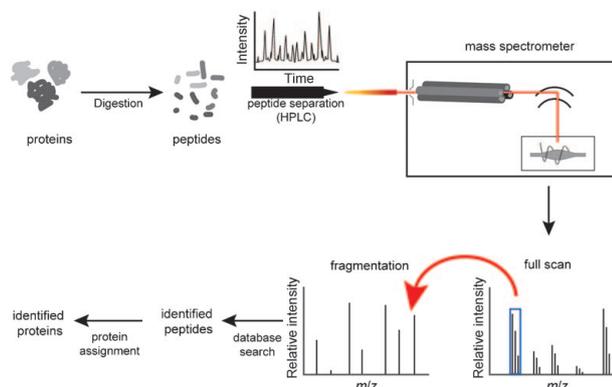
REVIEWS

Figure 1. Shotgun proteomics workflow. Proteins are digested to peptides, separated by nanoHPLC, and electro-sprayed on-line into the mass spectrometer (a linear ion trap—Orbitrap is depicted). Precursor masses recorded in full scans (MS scans) and selected and isolated peptides are fragmented and recorded again (MS/MS spectrum). Peptides are identified by database search and assigned to proteins.
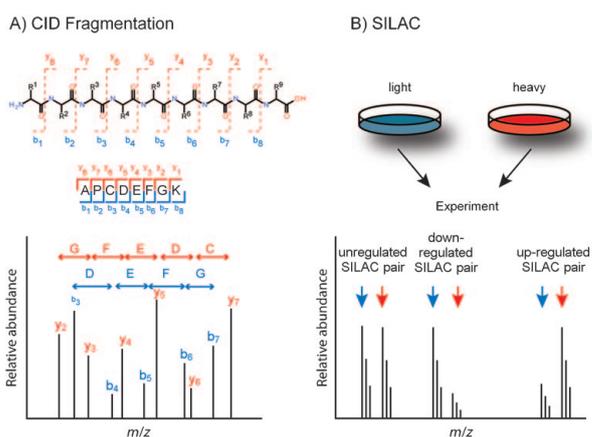
In high-resolution data, peptide peaks are clearly separated from each other and from chemical background. The peptide peak areas can then be integrated and compared over different LC MS/MS runs. Because the ionization efficiencies of the same peptides remain the same, label-free approaches are potentially quite accurate. A general disadvantage, however, is that samples have to be processed and measured separately and that the accuracy is therefore compromised by experimental variability between these runs. Sophisticated label-free algorithms attempt to correct for these errors; examples include Super-hirn[22] and MaxQuant.[23]

Stable isotope labeling approaches make peptides from two experimental states distinguishable by generating either "light" (normal) or "heavy" versions. A peptide common to two experiments is therefore present as a light and heavy pair in the same mass spectrum and the ratio of their signals is the ratio of the relative protein abundances under the two sets of experimental conditions. This quantification is therefore much more accurate than label-free quantification methods. Stable isotope labels can be introduced by different means: either by chemical derivatization (chemical labeling) or through cellular metabolism (metabolic labeling).

In chemical labeling, the reactive groups present in peptides (usually thiol or amine groups) are used to couple with an isotope-containing tag.[19b] Currently, the TMT[24] and iTRAQ methods are popular forms of chemical labeling. These create labeled peptides with identical masses, and the relative quantification is performed in the MS/MS spectra, in which each tag generates specific fragment ions. TMT and iTRAQ allow multiplexing, and up to eight sets of conditions can in principle be compared in the same experiment. Other examples of chemical labeling include stable-isotope dimethyl labeling[25] and propionic anhydride derivatization.[26] Chemical labeling is applicable to all sample types. A disadvantage of the method is the inevitable presence of chemical side products, which can interfere with the analysis of rare PTMs.

Figure 2. Mass spectrometry read-out. A) Standard CID fragmentation: upon energy transfer peptides break statistically at their amide bonds and give rise to characteristic fragmentation patterns. B) SILAC principle: on mixing of samples from two SILAC states, SILAC pairs can be observed for every peptide. These can be equal or can show up- or down-regulation.

protein abundance index (emPAI).[21] They rely on the fact that abundant proteins in a sample will usually generate more MS/MS fragmentation spectra than low-abundance proteins. An advantage of these methods is their applicability to any qualitative dataset, including low-resolution data. Spectral counting and emPAI can estimate the absolute amounts of proteins in a sample, as well as the relative amounts under two or more different sets of conditions. However, a significant drawback is that the correlation between the number of times that peptides are sequenced and protein amount is only approximate. In particular, proteins with low peptide counts show high quantification variability.

Metabolic labeling strategies introduce heavy isotopes through the growth medium or food. The entire proteome is labeled, meaning that samples can be mixed early during the experiment, minimizing experimental errors. This can be done in a global manner by replacing all nitrogen atoms in the medium by ^{15}N,[27] although this technique is generally restricted to specialized applications, such as bacterial and plant spe-

cies, because it leads to broad isotope distributions in the heavy forms. More targeted incorporation is achieved by replacing essential amino acids with their heavy counterparts, an approach known as SILAC (stable isotope labeling by amino acids in cell culture).[28] Arginine and lysine labeling are most commonly used. The protease trypsin cleaves at the C terminal of arginine and lysine residues, ensuring that every tryptic peptide contains at least one labeled amino acid (except the C-terminal peptide of the protein). For every peptide, two isotope clusters can be observed—one from each sample, a so-called SILAC pair (Figure 2 B). The mass difference between the SILAC pairs is exactly the mass difference between the light and heavy amino acids. Generally Arg10 and Lys8 are used, generating shifts of 10.0083 Da and 8.0142 Da, respectively. A ratio can be directly assigned for every identified peptide, and this ratio indicates whether a protein is up- or down-regulated, or unchanged, over two experiments. A large variety of different established cell lines, including mouse and human ES cells, have already been labeled.[29]

As well as cell lines, whole organisms can also be metabolically labeled.[30] Thus far, *S. cerevisiae* (lysine auxotroph strains growing in minimal medium with heavy lysine),[31] *Drosophila melanogaster* (feeding on heavy yeast),[32] and mice (by means of a special lysine-free diet supplemented with heavy lysine)[33] have been labeled and used for functional in vivo proteomics experiments.

The methods described above allow relative quantification between two or more samples. To obtain absolute concentrations, a defined amount of a heavy-isotope-labeled reference standard needs to be spiked into the sample. Labeled synthetic peptides can be used for this purpose (sometimes called AQUA for absolute quantification).[34] Spiking in partial or full-length proteins also controls for variability introduced by the digestion step.[35] This technique can also be applied for the absolute quantification of modification sites.[36]

Mapping and Relative Quantification of Post-translational Modifications

PTMs play an important role in chromatin research: histones, the basic building blocks of chromatin, are subjected to a large variety of PTMs and so are many other chromatin-associated proteins. The most reliable and unbiased method to identify and quantify PTMs is MS.

Technical considerations

Three different fragmentation techniques are generally used for analyzing PTMs: 1) collision-induced dissociation (CID) in a linear ion trap, 2) CID in quadrupole-TOFs or LTQ-Orbitraps (where it is called HCD), or 3) electron transfer dissociation (ETD), which has been coupled to diverse instrument types. Their different properties are summarized in Table 1. Fragments have most commonly been measured in ion traps, allowing fast sequencing with high sensitivity, albeit with low mass accuracy. Recent technological advances now also allow fast and sensitive measurement of fragments in the Orbitrap analyzer. This "high–high" strategy (high mass accuracy and resolution on precursor and product ion mass) allows deisotoping of product ions and results in higher-confidence assignment, especially of modified and of large peptides.

PTMs of interest are usually added "in silico" to candidate peptide sequences in the amino acid sequence database during the search. Only PTMs that are included as options in the search can be identified. Furthermore, to keep the "search space" manageable, the number of modifications that can be considered needs to be limited. A more unbiased way to detect PTMs is the "blind database search"[37] or "modificomb"[38] approach. Firstly a basic database search without consideration of modifications is carried out. The idea is then to match the unidentified peptides to already identified peptides by placing the mass difference between them on each amino acid. If a match is then obtained, the mass of the modification and its localization has been determined. This method holds great promise, in particular for the identification of novel PTMs that are not commonly considered.

One potential caveat in MS-based PTM identifications is the fact that many modifications can be in vitro artifacts rather than caused by in vivo enzymatic activity. A relevant example is the characteristic di-glycine tag on lysine used for mapping of ubiquitination sites. Iodoacetamide, which is commonly used for reduction of cysteines during MS workflows, generates an artifact with exactly the same elemental composition as the di-glycine tag.[39]

Quantification of PTMs

Two strategies for quantification of PTMs are available: either labeling of the peptides themselves by a stable isotope method as described above, or introduction of the label into the modification. The first approach can be used for any modification and is mainly applied to achieve differential quantification of a modification of interest in two different samples. The latter approach has so far only been described for methylation.

Table 1. Overview of fragmentation techniques.					
Fragmentation technique	Ion series	Mass cut-off	Labile modifications	Fragmentation efficiency	Fragmentation of long peptides
CID	y and b	1/3[b]	neutral loss	good	moderate
HCD/triple quad fragmentation	mostly y[a]	no	mostly preserved	good	moderate
ETD	c and z	no	preserved	lower[c]	good
[a] Low mass b ions are also routinely detected. [b] Because of the operating principle of ion trap instruments, fragment masses below m/z of 1/3 of the precursor mass cannot be stabilized in the ion trap and are lost;[c] often a substantial fraction of the precursor ion population remains after ETD.					

In "heavy methyl SILAC",[40] methionine containing a deuterated methyl group is added to the culture medium and incorporated into *S*-adenosyl methionine. This compound is the sole methyl group donor in cells, meaning that all cellular methyl groups can be labeled. This allows monitoring of methylation dynamics and stability in pulse experiments. In this way, Zee et al. demonstrated that methylations associated with active genes turn over more rapidly than those associated with repressed genes.[41]

PTM analysis after protein enrichment

The identification of PTMs on proteins in principle requires observation of all the peptides that together cover the sequence of the entire protein. Furthermore, PTMs are usually of low abundance and therefore not commonly sequenced in complex peptide mixtures. PTM analysis therefore usually requires an enrichment step of the protein or proteins of interest to reduce the dynamic range and to facilitate the sequencing of all proteolytic peptides, including those carrying PTMs. Most modification analysis in the epigenetics field has been carried out with core histones, because these basic and highly abundant proteins can easily be purified to near homogeneity.[42] Methylation pattern changes upon G9a knock-down have been studied by a chemical-labeling strategy, which revealed not only a decrease in H3K9 methylation but also a concomitant increase in K14 acetylation.[43] A methyl/acetyl switch was found in a SILAC study that used a knock-down of the Prc2 component Suz12 in mouse ES cells. This knock-down led to reduction of K27 di- and trimethylation but also to an increase in K27 acetylation.[44] A SILAC approach was likewise applied to compare PTMs on histones H3 and H4 between breast cancer cell lines and normal epithelial breast cells, which produced evidence for cancer-specific methylation patterns.[45] The ETD fragmentation technique has been applied to large histone fragments, enabling quantification of 74 unique H4 modification combinations in differentiating human ES cells by a label-free approach.[46]

Most studies directed towards single proteins have so far focused on histones. However, recent reports suggest that histone-associated proteins can also be heavily modified.[47] HP1 seems to carry histone-code-like patterns,[48] and very intensively studied complexes such as TFIID and SAGA bear a vast range of PTMs.[49] Although no quantitative experiments have yet been done on nonhistone chromatin-associated proteins, it is likely that the role of post-transcriptional regulation by dynamic PTMs will be very important in this case as well.

PTM analysis after modification-directed enrichment

When focusing on a single type of PTM, an enrichment strategy for this modification is applied to the entire sample. These large-scale studies provide a wealth of data for more targeted follow-up research. Lysine-acetylated peptides, for example, can be enriched by appropriate antibodies.[50] Choudhary et al. applied this strategy to identify 3600 lysine acetylation sites in human cells—most of them novel—as well as quantified acety-

lation changes upon treatment with the deacetylase inhibitors suberoylanilide hydroxamic acid and MS-275.[51] These studies revealed a hitherto unimagined diversity of cellular processes that are regulated by this modification.

Another PTM that has frequently been the target of large-scale proteomics studies is phosphorylation. Major techniques available for enrichment for phosphorylated peptides are:[52] 1) enrichment of peptides containing phosphorylated tyrosines by antibodies specific for this modification, 2) immobilized metal affinity chromatography (IMAC) utilizing coordination of phosphopeptides by metal ions, 3) titanium-dioxide-based enrichment, typically with a specificity enhancing reagent, and 4) enrichment in the flow through and early fractions in strong cation exchange chromatography (SCX), because of the presence of additional charge on phosphate at acidic pH, which leads to reduced retention times of phosphorylated peptides. ATM/ATR-signaling-mediated phosphorylation dynamics during the DNA damage response have been studied by an antibody enrichment strategy.[53] Our group recently applied a titanium dioxide enrichment strategy to identify and quantify more than 10 000 phosphorylation sites throughout the HeLa cell cycle and developed an algorithm to determine the occupancies or stoichiometries for thousands of these sites.[54] We suggest that mining of the data from large-scale, high-accuracy screens could provide useful leads for studies in the chromatin field.

Interaction Proteomics

For more than a decade, MS has been successfully employed to determine protein–protein interactions. Most commonly, the "bait" protein of interest and its associated binding partners have been purified to near homogeneity either by conventional column chromatography or by tandem affinity tagging approaches.[55] Purified proteins are separated by one-dimensional SDS PAGE, bands are excised and "in gel" digested, and proteins are identified by MS. Although this methodology has been very successful even in a high-throughput mode,[56] the rapidly increasing sensitivities of modern mass spectrometers render this approach prone to false positive determinations of interaction partners. Purification to homogeneity, even when using tandem affinity purifications, can be difficult to achieve and is usually the preserve of specialized protein biochemistry laboratories.

Quantitative proteomics technology can be used to solve these problems. When affinity purification is performed with a quantitative abundance readout—by SILAC , for example—enriched proteins can easily be distinguished from background proteins by their quantitative ratios (Figure 3). This approach elegantly sidesteps the problem of background proteins and allows purifications to be performed in a single affinity purification step at low stringency. This in turn potentially retains low-affinity but functionally relevant interactions that would otherwise be lost.[57]

There are different approaches to the identification of interaction partners of full-length proteins: endogenous protein enrichment with antibodies and tag-based purification of trans-
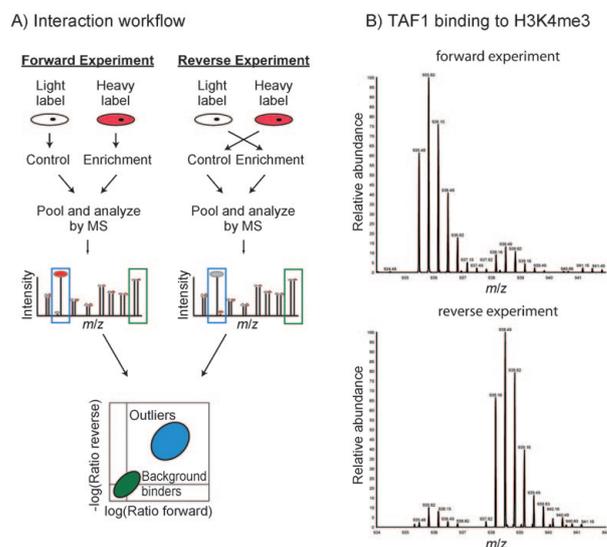
M. Mann, M. Vermeulen and H. C. Eberl

Figure 3. Workflow for interaction determination by SILAC. A) The experiment is done in "forward" and "reverse" modes; the pull-down for forward experiments is performed with heavy lysate and the control bait is incubated with light lysate and vice versa in the reverse experiment. Eluates are mixed after the enrichment step and processed, and measured. SILAC ratios from the two experiments are plotted against each other. Background and unspecific binders are clustered around 1:1 (0 on a logarithmic scale) and outliers (true binders) can be found in the top right quadrant. B) Pull-down with the peptide containing H3K4me3 and the unmethylated control peptide results in significant and reciprocal SILAC ratios for peptides from the TFIID complex member TAF1.

tween the basal transcription factor TFIID and H3K4me3.[62] This work was extended to all major tri-methylation marks and combined with ChIP-Seq and BAC-GFP pull-downs to define the lysine trimethyl-interactome,[63] indicating that this approach holds great promise for the deciphering of the histone PTM interactome. Another interesting application of quantitative proteomics is the estimation of dissociation constants between proteins of interest and their interaction partners in lysates.[64]

Protein interactions with methylated DNA or specific interactors of single-nucleotide polymorphisms can be studied by use of immobilized oligonucleotides (Figure 4 D).[65] Two recent studies found a protein binding to a single nucleotide polymorphism at the IGF2 locus, thereby identifying the long-sought repressor responsible for a "muscle" versus "fat" phenotype in domestic pigs.[66] Given the increasing importance of RNA in the field of epigenetics,[67] the application of quantitative RNA pull-downs[68] will likely be of great value for investigation of protein interactions that are mediated by noncoding RNAs in the mammalian nucleus.

Analyses of interactions between small molecules and proteins have received much attention in signal transduction and drug discovery, but not yet in the chromatin field. Small molecules are immobilized through flexible linkers on affinity matrices and used for binding partner enrichment.[69] This approach was applied to the characterization of kinase inhibitor

genes. For enrichment of proteins of interest with specific antibodies, isotype antibodies are mostly used as controls. Because of cross-reactivity there is a high risk of co-purifying false positives, however. This can be circumvented by performing the same immunoprecipitation in two SILAC-labeled cell populations but knocking down the protein of interest in one of them, an approach termed quantitative immunoprecipitation combined with knock-down, or QUICK[58] (Figure 4 A). An alternative is to apply protein tagging and purification with a tag-specific antibody, such as Flag or Myc; see for example ref. [59] (Figure 4 B). Another option is to apply BAC transgeneOmics technology to tag proteins of interest with GFP, which is also an excellent purification tag. A particular attraction of this "QUBIC" approach is that the bait is expressed at near endogenous expression levels.[60]

In the context of epigenetics, interaction mapping followed by MS "read-out" has successfully been used with peptide baits that only differ by a single functional group, such as a particular histone PTM, to identify "readers" of that mark.[61] Without a quantitative filter, however, these peptide pull-downs are very challenging and it is difficult to pinpoint specific interactors in the presence of a large excess of nonspecific binders. The quantitative proteomic approaches discussed above can again be used to overcome these problems (Figure 4 C). As an example, we have performed pull-downs with methylated histone peptides, which revealed an interaction be-

Figure 4. Quantitative interaction experiments. A) Quantitative immunoprecipitation combined with a knock-down (QUICK) uses an antibody against the endogenous protein, together with a pull-down from a knock-down of the target protein as the control (note: antibody cross-reactivity will not lead to false positives because only proteins associating with the knocked-down protein can produce SILAC ratios). B) Tagged proteins are often used for pull-down experiments, but note that the tag can interfere with some interactions, in which case the tag should be placed at the other terminus of the bait protein. C) Peptide pull-downs allow screening for PTM-dependent interactions. D) DNA pull-downs immobilize short DNA stretches on beads to screen for sequence-specific interactions (note: the same principle applies to RNA–protein interactions).

REVIEWS

specificity.[70] Promiscuous inhibitors have also been employed as effective tools for enrichment of low-abundance kinases.[71] Although a tremendous amount of work has been performed to characterize histone deacetylase (HDAC) inhibitors, their specificity is still not fully known. Recent reports suggest that HDAC inhibitors are far more promiscuous than expected.[51] In a pioneering study, Taunton et al. applied an immobilization strategy based on immobilized trapoxin and trichostatin to enrich for HDACs and thereby identified histone deacetylase 1.[72] Clearly, chemical proteomics approaches hold great promise for deeper characterization of HDAC inhibitor specificity.

Analyzing Proteome and Modification Changes and Dynamics

Differential gene expression between two different cellular states is commonly determined by mRNA profiling, with the aid of microarrays and recently also by massive parallel sequencing technology. Although these methods achieve impressive coverage of global mRNA expression, they are based on the fundamental assumption that mRNA levels are an accurate proxy for protein abundance. This assumption is not always true, however: post-transcriptional regulation of protein expression such as by miRNAs or by targeted protein degradation, for example, escapes detection in mRNA profiling experiments. In principle, MS-based proteomics technology covers all these effects, and so one of the primary goals of the field is to sequence proteomes as comprehensively as possibly. Current technology allows the quantitative analysis of the complete yeast proteome[73] and the identification of more than 7000 proteins in HeLa cells.[74] We predict that quantitative proteomics will soon be able to analyze the majority of cellular proteins.

In an alternative approach, a selected number of proteins can be analyzed by so-called "selected ion monitoring" (MRM). The MRM method requires triple quadrupole mass spectrometers and monitors predetermined transitions between precursor and fragments for each peptide of interest.[75] For every protein of interest, a peptide is chosen. Its precursor mass, together with specific fragment masses, provides a unique identifier in the mass spectrometric data. The first part of the triple quadrupole instrument is used for isolation of the precursor mass, the second part serves as a collision chamber, and the third is set to transmit only a predefined fragment ion. In contrast with the "discovery approach" discussed above, this "targeted approach" monitors only peptides defined beforehand. Quantification in MRM is achieved either through the MRM signals alone (a label-free approach) or, more accurately, by comparing the MRM signal with that of a spiked-in and labeled synthetic peptide.

The discrepancy between mRNA and protein abundance was clearly apparent in a recent study that monitored proteome changes upon ISWI knock-down in *Drosophila*.[76] The most striking example was Acf-1, a direct interaction partner of ISWI. This protein was depleted in the proteome data—presumably as a result of destabilization caused by the absence of

its interaction partner—but not affected at the mRNA level. Another study determined differences in nuclear proteomes between undifferentiated ES cells and fully differentiated cells.[77] Interestingly, the BAF chromatin remodeling complex was found to be significantly more highly expressed in ES cells. Four-factor reprogramming was significantly improved by co-transfecting the BAF complex members Brg1 and BAF155.

Whereas static differences between two cell populations can be analyzed by classical quantitative proteomics, induced differences can be analyzed by pulsed SILAC strategies.[78] A stimulus that causes proteome alterations is administered together with a switch in the SILAC amino acids. From this moment on, new protein synthesis can be monitored by the appearance of new SILAC pairs. This strategy is especially useful for studying transcriptional regulation—by miRNAs, for example—or protein turnover rates. Selbach et al. used this approach to study the effects of over-expression of five different miRNAs in HeLa cells.[79] Every miRNA mildly affected the expression of hundreds of genes and the data refined our knowledge of miRNA target characteristics. A concurrent study using a classical SILAC approach came to similar conclusions.[80] Furthermore, pulsed SILAC experiments have been applied to distinguish between "old" and newly synthesized histones. This enabled investigation of the progressive methylation of H4K20 and acetylation of H4K16[81] and the time-dependent establishment of histone modifications after histone incorporation into chromatin.[82] Another study used the pulsed SILAC strategy in combination with pull-down experiments for histones H3.1 and H3.3 in an elegant determination of the degree of H3–H4 tetramer splitting.[83] By timing the expression of tagged H3.1 or H3.3 with cell cycle arrest and SILAC label switch this study demonstrated that significant amounts of H3.3–H4 tetramers do indeed split, whereas H3.1–H4 tetramers stay intact. Here, the label pulse was applied to distinguish between interaction partners that were already present before the pulse and those that only interacted after the pulse.

Outlook

Although MS-based proteomics has made a major impact in the field of chromatin research and epigenetics, this contribution has thus far mainly been restricted to the identification of protein spots in gels and to the mapping of histone modifications. We anticipate that the recent dramatic developments in quantitative proteomics technology that we have outlined in this review will be of great value in deciphering the mechanisms underlying epigenetic regulation of gene expression. Two areas will be of particular importance: in vivo quantitative proteomics and locus-specific interaction proteomics.

Whereas current studies mainly use cell culture systems, labeling of model organisms will become more common practice in the future. The advantage of analyzing biological processes in vivo comes at the cost of more demanding experiments. This implies more elaborate sample preparation schemes and also requires whole-animal labeling. As described above, several of the commonly used model organisms have already successfully been labeled, and efforts to extend this list

CHEMBIOCHEM

are being made. An interesting recent trend is to use SILAC simply as an internal standard for cell culture and especially for in vivo work. This approach decouples SILAC labeling from the biological experiments, which has many practical advantages. These strategies can also be applied to study tissue samples and cell-type-specific processes that can only be investigated in vivo.[84]

Studying histone modifications or protein interactions in bulk chromatin yields valuable insights. These studies, however, only provide a global picture, whereas histone modifications are not distributed equally over the genome. They are found in very specific locations: H3K4me3, for instance, is found almost exclusively on transcription start sites.[85] Also, for many complexes, subunits and modifications may vary depending on their localization. Determination of chromatin interactions and modifications in a locus-specific manner would therefore be a very attractive goal. Chromatin immunoprecipitation (ChIP) determines the localization of a protein or histone modification of interest in the genome. The proteomic equivalent of performing these experiments would be a ChIP-like enrichment followed by the unbiased detection and analysis of the protein complement and the modification status. One elegant approach in which chromatin is isolated and sheared and an enrichment step is applied is called mChIP.[86,87] mChIP was successfully applied to study of H2A and its variant Htz1p interactions as well as several nonhistone chromatin proteins.[86] The even more elaborate PICH approach[88] in principle allows DNA-sequence-specific enrichment: cross-linked chromatin is sheared and the desired locus is isolated by use of complementary DNA probes. An unrelated sequence is used as control and associated proteins are identified by MS. This technique successfully identified telomere-associated proteins in HeLa cells. Telomeres are present 92 times in a normal diploid human cell. It would be desirable to apply PICH-like techniques in combination with quantitative proteomics to single loci. Because of sensitivity and dynamic range limits, however, this has not yet been accomplished. The major hurdles are not only the required sensitivity, but also relate to the challenge of distinguishing background binders from true interactors. Even more challenging, but also very rewarding, will be time course or stimulation experiments, in which only a few proteins display stimulus- or time-dependent changes.

In summary, the power of proteomics lies in the unbiased detection of proteins and their modifications. In combination with precise and reliable quantification this provides a powerful toolbox for functional approaches and systematic analyses of proteomes, sub-proteomes, and protein complexes. MS-based quantitative proteomics is now at a stage where it can be used not only for descriptive experiments but also to provide read-outs in directed functional experiments. Ongoing technological advances and new protocols should enable epigenetics researchers to answer many relevant questions in the chromatin field by MS—and also to generate numerous new questions.

[1] a) C. H. Waddington, *Endeavour* **1942**, *1*, 18; b) S. L. Berger, T. Kouzarides, R. Shiekhattar, A. Shilatifard, *Genes Dev.* **2009**, *23*, 781.

[2] W. Reik, *Nature* **2007**, *447*, 425.

[3] a) P. Chi, C. D. Allis, G. G. Wang, *Nat. Rev. Cancer* **2010**, *10*, 457; b) S. Gopalakrishnan, B. O. Van Emburgh, K. D. Robertson, *Mutat. Res.* **2008**, *647*, 30.

[4] J. Qiu, G. Shi, Y. Jia, J. Li, M. Wu, S. Dong, J. Wong, *Cell. Res.* **2010**, *20*, 908.

[5] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89.

[6] T. Kouzarides, *Cell* **2007**, *128*, 693.

[7] a) S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025; b) J. S. Lee, E. Smith, A. Shilatifard, *Cell* **2010**, *142*, 682.

[8] L. Ho, G. R. Crabtree, *Nature* **2010**, *463*, 474.

[9] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41.

[10] O. J. Rando, H. Y. Chang, *Annu. Rev. Biochem.* **2009**, *78*, 245.

[11] B. A. Garcia, J. Shabanowitz, D. F. Hunt, *Curr. Opin. Chem. Biol.* **2007**, *11*, 66.

[12] a) R. Aebersold, M. Mann, *Nature* **2003**, *422*, 198; b) L. Liao, D. B. McClatchy, J. R. Yates, *Neuron* **2009**, *63*, 12.

[13] J. Cox, M. Mann, *Cell* **2007**, *130*, 395.

[14] a) N. Siuti, N. L. Kelleher, *Nat. Methods* **2007**, *4*, 817; b) S. M. Eliuk, D. Maltby, B. Panning, A. L. Burlingame, *Mol. Cell. Proteomics* **2010**, *9*, 824.

[15] R. A. Zubarev, N. L. Kelleher, F. W. McLafferty, *J. Am. Chem. Soc.* **1998**, *120*, 3265.

[16] J. E. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528.

[17] H. Steen, M. Mann, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699.

[18] J. V. Olsen, J. C. Schwartz, J. Griep-Raming, M. L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E. R. Wouters, M. Senko, A. Makarov, M. Mann, S. Horning, *Mol. Cell. Proteomics* **2009**, *8*, 2759.

[19] a) S. E. Ong, M. Mann, *Nat. Chem. Biol.* **2005**, *1*, 252; b) M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster, *Anal. Bioanal. Chem.* **2007**, *389*, 1017; c) M. Vermeulen, M. Selbach, *Curr. Opin. Cell Biol.* **2009**, *21*, 761.

[20] H. Liu, R. G. Sadygov, J. R. Yates III, *Anal. Chem.* **2004**, *76*, 4193.

[21] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, *Mol. Cell. Proteomics* **2005**, *4*, 1265.

[22] L. N. Mueller, O. Rinner, A. Schmidt, S. Letarte, B. Bodenmiller, M. Y. Brusniak, O. Vitek, R. Aebersold, M. Muller, *Proteomics* **2007**, *7*, 3470.

[23] C. A. Luber, J. Cox, H. Lauterbach, B. Fancke, M. Selbach, J. Tschopp, S. Akira, M. Wiegand, H. Hochrein, M. O'Keeffe, M. Mann, *Immunity* **2010**, *32*, 279.

[24] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. Mohammed, C. Hamon, *Anal. Chem.* **2003**, *75*, 1895.

[25] J. L. Hsu, S. Y. Huang, N. H. Chow, S. H. Chen, *Anal. Chem.* **2003**, *75*, 6843.

[26] J. E. Syka, J. A. Marto, D. L. Bai, S. Horning, M. W. Senko, J. C. Schwartz, B. Ueberheide, B. Garcia, S. Busby, T. Muratore, J. Shabanowitz, D. F. Hunt, *J. Proteome Res.* **2004**, *3*, 621.

[27] Y. Oda, K. Huang, F. R. Cross, D. Cowburn, B. T. Chait, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6591.

[28] S. E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, M. Mann, *Mol. Cell. Proteomics* **2002**, *1*, 376.

[29] a) J. Graumann, N. C. Hubner, J. B. Kim, K. Ko, M. Moser, C. Kumar, J. Cox, H. Scholer, M. Mann, *Mol. Cell. Proteomics* **2008**, *7*, 672; b) T. A. Prokhorova, K. T. Rigbolt, P. T. Johansen, J. Henningsen, I. Kratchmarova, M. Kassem, B. Blagoev, *Mol. Cell. Proteomics* **2009**, *8*, 959; c) R. Tian, S. Wang, F. Elisma, L. Li, H. Zhou, L. Wang, D. Figeys, *Mol. Cell. Proteomics* **2010**; DOI: 10.1074/mcp.M110.000679.

[30] A. J. Heck, J. Krijgsveld, *Expert Rev. Proteomics* **2004**, *1*, 317.

[31] A. Gruhler, J. V. Olsen, S. Mohammed, P. Mortensen, N. J. Faergeman, M. Mann, O. N. Jensen, *Mol. Cell. Proteomics* **2005**, *4*, 310.

[32] M. D. Sury, J. X. Chen, M. Selbach, *Mol. Cell. Proteomics* **2010**.

[33] M. Kruger, M. Moser, S. Ussar, I. Thievessen, C. A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fassler, M. Mann, *Cell* **2008**, *134*, 353.

[34] D. S. Kirkpatrick, S. A. Gerber, S. P. Gygi, *Methods* **2005**, *35*, 265.

[35] a) V. Brun, A. Dupuis, A. Adrait, M. Marcellin, D. Thomas, M. Court, F. Vandenesch, J. Garin, *Mol. Cell. Proteomics* **2007**, *6*, 2139; b) S. Hanke, H. Besir, D. Oesterhelt, M. Mann, *J. Proteome Res.* **2008**, *7*, 1118.

[36] S. Singh, M. Springer, J. Steen, M. W. Kirschner, H. Steen, *J. Proteome Res.* **2009**, *8*, 2201.

[37] D. Tsur, S. Tanner, E. Zandi, V. Bafna, P. A. Pevzner, *Nat. Biotechnol.* **2005**, *23*, 1562.

[38] M. M. Savitski, M. L. Nielsen, R. A. Zubarev, *Mol. Cell. Proteomics* **2006**, *5*, 935.

[39] M. L. Nielsen, M. Vermeulen, T. Bonaldi, J. Cox, L. Moroder, M. Mann, *Nat. Methods* **2008**, *5*, 459.

[40] S. E. Ong, G. Mittler, M. Mann, *Nat. Methods* **2004**, *1*, 119.

[41] B. M. Zee, R. S. Levin, B. Xu, G. LeRoy, N. S. Wingreen, B. A. Garcia, *J. Biol. Chem.* **2010**, *285*, 3341.

[42] D. Shechter, H. L. Dormann, C. D. Allis, S. B. Hake, *Nat. Protoc.* **2007**, *2*, 1445.

[43] M. D. Plazas-Mayorca, J. S. Bloom, U. Zeissler, G. Leroy, N. L. Young, P. A. Dimaggio, L. Krugylak, R. Schneider, B. A. Garcia, *Mol. Biosyst.* **2010**; DOI: 10.1039/C003307C.

[44] H. R. Jung, D. Pasini, K. Helin, O. N. Jensen, *Mol. Cell. Proteomics* **2010**, *9*, 838.

[45] A. Cuomo, S. Moretti, S. Minucci, T. Bonaldi, *Amino Acids* **2010**; DOI: 10.1007/s00726-010-0668-2.

[46] D. Phanstiel, J. Brumbaugh, W. T. Berggren, K. Conard, X. Feng, M. E. Levenstein, G. C. McAlister, J. A. Thomson, J. J. Coon, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4093.

[47] J. Huang, S. L. Berger, *Curr. Opin. Genet. Dev.* **2008**, *18*, 152.

[48] G. LeRoy, J. T. Weston, B. M. Zee, N. L. Young, M. D. Plazas-Mayorca, B. A. Garcia, *Mol. Cell. Proteomics* **2009**, *8*, 2432.

[49] N. Mischerikow, G. Spedale, A. F. Altelaar, H. T. Timmers, W. W. Pijnappel, A. J. Heck, *J. Proteome Res.* **2009**, *8*, 5020.

[50] a) S. C. Kim, R. Sprung, Y. Chen, Y. Xu, H. Ball, J. Pei, T. Cheng, Y. Kho, H. Xiao, L. Xiao, N. V. Grishin, M. White, X. J. Yang, Y. Zhao, *Mol. Cell* **2006**, *23*, 607; b) Q. Wang, Y. Zhang, C. Yang, H. Xiong, Y. Lin, J. Yao, H. Li, L. Xie, W. Zhao, Y. Yao, Z. B. Ning, R. Zeng, Y. Xiong, K. L. Guan, S. Zhao, G. P. Zhao, *Science* **2010**, *327*, 1004; c) S. Zhao, W. Xu, W. Jiang, W. Yu, Y. Lin, T. Zhang, J. Yao, L. Zhou, Y. Zeng, H. Li, Y. Li, J. Shi, W. An, S. M. Hancock, F. He, L. Qin, J. Chin, P. Yang, X. Chen, Q. Lei, Y. Xiong, K. L. Guan, *Science* **2010**, *327*, 1000.

[51] C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, M. Mann, *Science* **2010**, *325*, 834.

[52] B. Macek, M. Mann, J. V. Olsen, *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 199.

[53] S. Matsuoka, B. A. Ballif, A. Smogorzewska, E. R. McDonald III, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi, S. J. Elledge, *Science* **2007**, *316*, 1160.

[54] J. V. Olsen, M. Vermeulen, A. Santamaria, C. Kumar, M. L. Miller, L. J. Jensen, F. Gnad, J. Cox, T. S. Jensen, E. A. Nigg, S. Brunak, M. Mann, *Sci. Signal* **2010**, *3*, ra3.

[55] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, *Nat. Biotechnol.* **1999**, *17*, 1030.

[56] a) A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, C. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, *Nature* **2002**, *415*, 141; b) A. C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, G. Superti-Furga, *Nature* **2006**, *440*, 631; c) N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St. Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, J. F. Greenblatt, *Nature* **2006**, *440*, 637.

[57] M. Vermeulen, N. C. Hubner, M. Mann, *Curr. Opin. Biotechnol.* **2008**, *19*, 331.

[58] M. Selbach, M. Mann, *Nat. Methods* **2006**, *3*, 981.

[59] a) A. B. Mak, Z. Ni, J. A. Hewel, G. I. Chen, G. Zhong, K. Karamboulas, K. Blakely, S. Smiley, E. Marcon, D. Roudeva, J. Li, J. B. Olsen, C. Wan, T. Punna, R. Isserlin, S. Chetyrkin, A. C. Gingras, A. Emili, J. Greenblatt, J. Moffat, *Mol. Cell. Proteomics* **2010**, *9*, 811; b) M. E. Sowa, E. J. Bennett, S. P. Gygi, J. W. Harper, *Cell* **2009**, *138*, 389.

[60] a) I. Poser, M. Sarov, J. R. Hutchins, J. K. Heriche, Y. Toyoda, A. Pozniakovsky, D. Weigl, A. Nitzsche, B. Hegemann, A. W. Bird, L. Pelletier, R. Kittler, S. Hua, R. Naumann, M. Augsburg, M. M. Sykora, H. Hofemeister, Y. Zhang, K. Nasmyth, K. P. White, S. Dietzel, K. Mechtler, R. Durbin, A. F. Stewart, J. M. Peters, F. Buchholz, A. A. Hyman, *Nat. Methods* **2008**, *5*, 409; b) N. C. Hubner, A. W. Bird, J. Cox, B. Splettstoesser, P. Bandilla, I. Poser, A. Hyman, M. Mann, *J. Cell. Biol.* **2010**, *189*, 739.

[61] a) M. Stucki, J. A. Clapperton, D. Mohammad, M. B. Yaffe, S. J. Smerdon, S. P. Jackson, *Cell* **2005**, *123*, 1213; b) J. Wysocka, T. Swigut, T. A. Milne, Y. Dou, X. Zhang, A. L. Burlingame, R. G. Roeder, A. H. Brivanlou, C. D. Allis, *Cell* **2005**, *121*, 859; c) D. W. Chan, Y. Wang, M. Wu, J. Wong, J. Qin, Y. Zhao, *Proteomics* **2009**, *9*, 2343.

[62] M. Vermeulen, K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann, H. T. Timmers, *Cell* **2007**, *131*, 58.

[63] M. Vermeulen, H. C. Eberl, F. Matarese, H. Marks, S. Denissov, F. Butter, K. K. Lee, J. V. Olsen, A. A. Hyman, H. G. Stunnenberg, M. Mann, *Cell* **2010**, *142*, 967.

[64] K. Sharma, C. Weber, M. Bairlein, Z. Greff, G. Keri, J. Cox, J. V. Olsen, H. Daub, *Nat. Methods* **2009**, *6*, 741.

[65] a) G. Mittler, F. Butter, M. Mann, *Genome Res.* **2009**, *19*, 284; b) C. G. Spruijt, S. J. Bartels, A. B. Brinkman, J. V. Tjeertes, I. Poser, H. G. Stunnenberg, M. Vermeulen, *Mol. Biosyst.* **2010**, *6*, 1700.

[66] a) E. Markljung, L. Jiang, J. D. Jaffe, T. S. Mikkelsen, O. Wallerman, M. Larhammar, X. Zhang, L. Wang, V. Saenz-Vash, A. Gnirke, A. M. Lindroth, R. Barres, J. Yan, S. Stromberg, S. De, F. Ponten, E. S. Lander, S. A. Carr, J. R. Zierath, K. Kullander, C. Wadelius, K. Lindblad-Toh, G. Andersson, G. Hjalm, L. Andersson, *PLoS Biol.* **2009**, *7*, e1000256; b) F. Butter, D. Kappei, F. Buchholz, M. Vermeulen, M. Mann, *EMBO Rep.* **2010**, *11*, 305.

[67] a) A. Kanhere, K. Viiri, C. C. Araujo, J. Rasaiyaah, R. D. Bouwman, W. A. Whyte, C. F. Pereira, E. Brookes, K. Walker, G. W. Bell, A. Pombo, A. G. Fisher, R. A. Young, R. G. Jenner, *Mol. Cell* **2010**, *38*, 675; b) M. C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, H. Y. Chang, *Science* **2010**, *329*, 689.

[68] F. Butter, M. Scheibe, M. Morl, M. Mann, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10626.

[69] U. Rix, G. Superti-Furga, *Nat. Chem. Biol.* **2009**, *5*, 616.

[70] a) K. Godl, O. J. Gruss, J. Eickhoff, J. Wissing, S. Blencke, M. Weber, H. Degen, D. Brehmer, L. Orfi, Z. Horvath, G. Keri, S. Muller, M. Cotten, A. Ullrich, H. Daub, *Cancer Res.* **2005**, *65*, 6919; b) M. Bantscheff, D. Eberhard, Y. Abraham, S. Bastuck, M. Boesche, S. Hobson, T. Mathieson, M. Perrin, M. Raida, C. Rau, V. Reader, G. Sweetman, A. Bauer, T. Bouwmeester, C. Hopf, U. Kruse, G. Neubauer, N. Ramsden, J. Rick, B. Kuster, G. Drewes, *Nat. Biotechnol.* **2007**, *25*, 1035.

[71] H. Daub, J. V. Olsen, M. Bairlein, F. Gnad, F. S. Oppermann, R. Korner, Z. Greff, G. Keri, O. Stemmann, M. Mann, *Mol. Cell* **2008**, *31*, 438.

[72] J. Taunton, C. A. Hassig, S. L. Schreiber, *Science* **1996**, *272*, 408.

93

CHEMBIOCHEM

M. Mann, M. Vermeulen and H. C. Eberl

[73] L. M. de Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Frohlich, T. C. Walther, M. Mann, *Nature* **2008**, *455*, 1251.

[74] J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, *Nat. Methods* **2009**, *6*, 359.

[75] a) P. Picotti, B. Bodenmiller, L. N. Mueller, B. Domon, R. Aebersold, *Cell* **2009**, *138*, 795; b) A. Wolf-Yadlin, S. Hautaniemi, D. A. Lauffenburger, F. M. White, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 5860.

[76] T. Bonaldi, T. Straub, J. Cox, C. Kumar, P. B. Becker, M. Mann, *Mol. Cell* **2008**, *31*, 762.

[77] N. Singhal, J. Graumann, G. Wu, M. J. Arauzo-Bravo, D. W. Han, B. Greber, L. Gentile, M. Mann, H. R. Scholer, *Cell* **2010**, *141*, 943.

[78] B. Schwanhäusser, M. Gossen, G. Dittmar, M. Selbach, *Proteomics* **2009**, *9*, 205.

[79] M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, N. Rajewsky, *Nature* **2008**, *455*, 58.

[80] D. Baek, J. Villen, C. Shin, F. D. Camargo, S. P. Gygi, D. P. Bartel, *Nature* **2008**, *455*, 64.

[81] J. J. Pesavento, H. Yang, N. L. Kelleher, C. A. Mizzen, *Mol. Cell. Biol.* **2008**, *28*, 468.

[82] A. N. Scharf, T. K. Barth, A. Imhof, *Nucleic Acids Res.* **2009**, *37*, 5032.

[83] M. Xu, C. Long, X. Chen, C. Huang, S. Chen, B. Zhu, *Science* **2010**, *328*, 94.

[84] T. Geiger, J. Cox, P. Ostasiewicz, J. R. Wisniewski, M. Mann, *Nat. Methods* **2010**, *7*, 383.

[85] a) B. E. Bernstein, M. Kamal, K. Lindblad-Toh, S. Bekiranov, D. K. Bailey, D. J. Huebert, S. McMahon, E. K. Karlsson, E. J. Kulbokas, III, T. R. Gingeras, S. L. Schreiber, E. S. Lander, *Cell* **2005**, *120*, 169; b) H. Santos-Rosa, R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, T. Kouzarides, *Nature* **2002**, *419*, 407.

[86] J. P. Lambert, L. Mitchell, A. Rudner, K. Baetz, D. Figeys, *Mol.Cell Proteomics* **2009**, *8*, 870.

[87] M. Abu-Farha, J. P. Lambert, A. S. Al-Madhoun, F. Elisma, I. S. Skerjanc, D. Figeys, *Mol. Cell. Proteomics* **2008**, *7*, 560.

[88] J. Dejardin, R. E. Kingston, *Cell* **2009**, *136*, 175.

94

3 Discussion

3.1 Summary of projects

In this thesis, quantitative proteomic approaches were applied to study protein-protein interactions related to chromatin. The two main projects focused on chromatin readers, which specifically bind to modified histone tails.

In a SILAC approach based on an already established workflow [252], we screened for binders of the major lysine trimethylation sites on histones H3 and H4. Our approach was highly sensitive and specific, as we retrieved many of the known interactors for the investigated chromatin marks. In addition, several proteins could be linked to histone modifications for the first time. We could demonstrate that the SAGA complex binds to H3K4me3 via the double tudor domain of SGF29. The PWWP domain was very frequently observed among H3K36me3-associated proteins and we suggested that it is a H3K36me3 binding module. We showed that the interaction of the H3K36me3 reader NPAC with the histone tail is dependent on its PWWP domain. Recognition of this chromatin mark by the PWWP domain was independently shown for BRPF1 [254] and DNMT3A [48]. Moreover, we performed genome-wide profiling of histone trimethylation sites and selected readers. ChIP-Seq profiles of H3K4me3 and its readers showed a significant overlap, thus providing *in vivo* evidence for the interaction observed by the *in vitro* pull-downs. The profiles of five H3K4me3-associated proteins (BAP18, GATAD1, PHF8, TRRAP and SGF29) were analyzed in detail. Whereas some promoters were occupied by all investigated chromatin readers, some promoters only recruited a subset of readers. PHF8 and GATAD1 were found on almost all promoters, suggesting that they play a general role in transcription. In contrast, SGF29 and TRRAP bound promoters were enriched, for instance, for genes involved in DNA damage repair and DNA replication, linking them to the transcription of genes related to specific biological pathways. Considering the large amount of proteins which specifically bind to H3K4me3, it is obvious that further determinants – DNA sequences, transcription factors or additional histone modifications – define the binding pattern on the genome. In summary, this project provided the first large-scale and systematic interaction study of chromatin marks and reported a wealth of new interactors for further investigation.

A label-free interaction pipeline to screen for chromatin readers from mouse tissue extracts was established to overcome the limitations of cell lines, because many proteins were not retrieved in the SILAC-based screen from HeLa cells [125], although they are biochemically known to directly bind. We decided to investigate binders to the well-studied activating H3K4me3 and the repressive H3K9me3 marks from different mouse tissue extracts. In comparison to the previous project we significantly increased the number of identified proteins associated with the respective chromatin marks. This is partly due to improvements in the sensitivity and sequencing speed of the mass spectrometers, but also to an improved pull-down protocol. The data encompasses almost all known chromatin readers. Although we screened from very different tissues (brain, liver and kidney) the vast majority of chromatin-associated proteins was the same. One example for an organ-specific chromatin reader is the brain specific NuRD complex subunit CHD5, which directly interacts with the histone tail via its two PHD fingers. Another example is the zinc finger protein ZNF462, which is recruited to H3K9me3 via binding to HP1 α and was enriched from kidney and brain but not from liver nuclear extracts. Additional experiments performed with mouse testis extracts provided a surprisingly large number of chromatin associated proteins which were not found in the other tissues. Whereas brain, liver and kidney are already differentiated and chromatin associated processes are maintaining the status quo, large chromatin rearrangements are occurring in testis, due to processes such as meiosis, chromatin compaction, parental imprinting and substitution of canonical histones. This screen shows the feasibility to switch to more complicated systems and describes chromatin readers which could not be observed from standard cancer cell lines.

In addition to marking chromatin by histone modifications, it is also very effectivly indexed by the incorporation of histone variants [77]. Whereas histone modifications are established by so-called chromatin writers, the analogous function for histone variants – the selective incorporation of a histone variant – is performed by histone chaperones. Pulling-down histones from the soluble nuclear fraction in combination with quantitative mass spectrometry is the method of choice to identify histone chaperones. We found that the novel H2A.Z splice variant H2A.Z.2.2 associates with the SRCAP and TIP60 chaperoning complexes, apparently utilizing the same principle chaperoning machineries as H2A.Z.2.1.

One of the long term goals of our research efforts was to introduce quantitative proteomics to the epigenetics and chromatin community. The attention that our SILAC-based chromatin reader screen received [125] already made many biologists aware of the possibilities that contemporary proteomics offers to study protein-protein interactions. To provide further incentives, we wrote a review which explains the various

quantitative approaches and highlights selected proteomic studies.

3.2 Proteomic approaches to investigate chromatin readers

Several biochemical approaches to screen for proteins specifically binding to histone modifications are currently used. The most straightforward method is to combine peptide pull-downs with quantitative proteomics [212] which was used in both projects described in this thesis, as well as subsequently by other groups [33, 166]. The experimental workflow is very robust and can be performed quickly without complicated bait preparations. Peptide based approaches are restricted to modifications on unfolded protein stretches, as for example the histone tails. Combinatorial effects of two modifications can only be studied if these modifications are in close vicinity, as e.g. phosphorylation of serine 10 and trimethylation of lysine 9 on the H3 N-terminal tail [61, 250].

Often, one histone modification can attract many possible readers and only the combination of several signals, including DNA methylation or other histone modifications, produces the necessary specificity. A good example for this is BPTF, which binds to one histone tail which is trimethylated at lysine 4 with its PHD finger and to another histone tail which is acetylated at lysine 16 via its bromodomain [205]. These multivalent interactions cannot be observed in simple peptide pull-downs. Nucleosomes with a specific modification pattern can be generated *in vitro* by ligating a modified peptide to purified tailless nucleosomes by expressed protein ligation [159]. This allows the generation of more elaborate baits like mononucleosomes [12] or nuclesosome arrays [166] which can contain a combination of histone modifications. Moreover, the DNA used to assemble these nucleosomes can also play a role for protein binding, and methylated or unmethylated DNA can be used. Although approaches assembling nucleosomes are very powerful, more chromatin readers could be retrieved in a direct comparison from peptide pull-downs [166]. Especially when studying modifications on the histone tails, far away from the nuclesosome core, interactions between the chromatin reader and the nuclesosome are of secondary importance. Almost none of the chromatin readers, for which solid evidence exists in the literature, were only found in nucleosome pull-downs.

In the approaches discussed above, it is not possible to distinguish the proteins that directly interact with the bait from co-purifying members of protein complexes. In our projects, we adessed this question by carefully examining the domain structure of enriched proteins. As already many histone modification binding domains are described

[145, 237], 'educated guesses' can generate a reduced list of putative direct binders which can be tested. In addition, by either using published protein-protein interactions or performing protein pull-down experiments, proteins can be grouped in complexes. At least one protein mediating the interaction to the modified histone tail has to be present in each complex. Although this approach is very effective in finding chromatin readers with already known binding domains, it is very difficult to identify new domains which were not described yet. A direct method crosslinks the direct binders to the bait peptide [131]. A photo-reactive cross-linker is placed in close proximity to the modified residue on the bait peptide. After incubation with extracts, the cross-linker was activated by a laser pulse and the chromatin reader was covalently attached to the peptide. Indirect binders could be removed by applying very stringent washing conditions. This approach was combined with a SILAC read-out and only direct binders were enriched, among them the novel H3K4me3 reader MORC3.

In summary, peptide pull-downs are the easiest and most straight-forward approach to study chromatin readers. They yield the best coverage of chromatin readers, however, extensive follow-up experiments to determine direct readers and to investigate the influence of additional chromatin readers are necessary.

3.3 Performance of label-free quantification for peptide pull-downs

One major part of this thesis was the establishment of a label-free interaction workflow for peptide pull-downs. The results were compared to SILAC, which is the gold standard for quantification in interaction proteomics [251]. The label-free approach performed equally well in peptide pull-downs. The same has been observed before when studying full length protein interactions [88].

As label-free approaches do not require SILAC labeling, the question arises whether label-free approaches will replace SILAC for interaction proteomics. However, the SILAC workflow has many intrinsic advantages. SILAC-based quantification is the most robust quantification and does not depend on the long-term stability of the mass spectrometric setup, as quantification is performed in the same measurement and not over consecutive ones. Moreover, the analysis is more straightforward and intuitive and needs less complicated statistics. Finally, due to its high precision, only two replicates are necessary (forward and reverse experiment), and automation of the pull-downs is not strictly necessary. Taken together, SILAC is the quantification method of choice for any medium sized interaction project. In addition, studies expecting only

small ratio changes need the highest quantification accuracy possible and thus also benefit from the SILAC approach.

Label-free approaches are a powerful addition to the proteomic interaction tool box. Any project, independent of bait or size of the screen, in which the proteins cannot be labeled has to use label-free approaches. In projects in which both label-free and SILAC approaches can be used, label-free approaches can be beneficial under certain circumstances. A major advantage of label-free approaches is that an unlimited number of baits can be compared, whereas SILAC is generally restricted to a maximum of three. Moreover, the more samples are compared, the better the statistics become. Hence, for large scale studies with many different baits, label-free statistics will perform better. Another advantage is that cells do not need to be labeled and experiments can be performed faster.

In conclusion, label-free approaches complement SILAC-based approaches. The choice which one to use depends on the biological system and the size of the study.

3.4 Chromatin readers of H3K4me3

Trimethylation of H3K4 is probably the most extensively studied chromatin mark. Despite tremendious research efforts, only a limited number of readers has been described before. Using a SILAC-based quantitative proteomics approach, we increased the number of H3K4me3 readers. We demonstrated that the activating SAGA complex binds to H3K4me3 via the double tudor domain of SGF29. Moreover, we describe a novel complex consisting of GATAD1, JARID1A, EMSY, PHF12, SIN3B and HDACs, which is also recruited to H3K4me3. Interestingly, in contrast to the SAGA complex or TFIID [252], this novel complex is likely to have a repressive effect on gene expression. SIN3B is a transcriptional co-repressor [6], as well as EMSY [90]. JARID1A binds directly to H3K4me3 [256] but is also capable of removing this methyl mark [37, 93, 109].

As several known readers were missing in the chromatin reader screen [125], we performed a label-free interaction screen from mouse tissue extracts to cover as many chromatin readers as possible. In total we linked 100 proteins to the H3K4me3 mark, of which 17 were already shown to be direct binders and we reached near comprehensiveness with this screen. One well-known H3K4me3 reader we did not observe is RAG2, which is expressed in B cells and necessary for V(D)J recombination [135, 149, 193]. As this protein is expressed in only very specific cell types, one would not expect to detect it from whole tissue extracts. The majority of chromatin readers did not show any tissue preference in our experiments, which argues for generic, cell type independent

functions of H3K4me3 and its readers. However, as the distribution of the H3K4me3 mark varies between different cell types, the actual establishment of this mark very likely follows a cell type-specific pattern. Furthermore, it is noteworthy that there are many complexes, which can bind directly to H3K4me3. Their sheer number and diversity emphasizes the complexity of protein interactions at H3K4me3 marked promoters. Whereas many of the readers perform functions that can be related to active gene expression like chromatin remodeling, lysine acetylation or removal of repressive histone marks, others perform repressive functions. As it is unlikely that all H3K4me3 binders compete at the same time for binding, a model in which binding to the H3K4me3 mark is orchestrated by further mechanism seems more appropriate. A cyclic binding behavior on active genes has already be shown for histone acetyltransferases and deacetylases [258] and this principle could also be generally applicable to the H3K4me3 mark. As was already observed before, some readers only bind a subset of possible H3K4me3 marks, as for example the SAGA complex [250] or Spindlin 1 [257]. Genome-wide profiling of all H3K4me3 binders will be necessary to separate them into general and gene subset-specific H3K4me3 chromatin readers.

3.5 Chromatin readers of H3K9me3

Although the repressive H3K9me3 mark is more widespread on the genome than the activating H3K4me3 mark, fewer chromatin readers were found for this mark. The major interaction hubs on this modification are the HP1 proteins (CBX1, CBX3 and CBX5), which interact directly with H3K9me3 via their chromo domain [94, 122, 165]. HP1 proteins are engaged in many protein-protein interactions [169]. For example, the HP1 chromoshadow domain can bind to proteins containing a PxVxL motif [239]. Among the known interactors of HP1 proteins is a large number of zinc finger domain containing proteins. As zinc finger domains very often recognize DNA sequence motifs, these HP1 interactors could provide a link from the genomic information of the underlying DNA sequence to the epigenetic information of the methylation state.

In addition to HP1, our screen for H3K9me3 binders yielded several other proteins which are described to directly bind to H3K9me3. CDYL and CDYL2 also both contain a chromo domain which binds directly to trimethylated H3K9 [60]. MPHOSPH8 [112], ATRX [49], and UHRF2 [101] are also known H3K9me3 readers. In contrast to HP1, these proteins have not been described to possess such large interaction networks.

The majority of readers of the H3K9me3 mark can be found equally in brain, liver and kidney suggesting general functions for gene repression. Two auxiliary proteins, both

binding to HP1, are found in a tissue-specific manner. ZFP462, which plays a role in pluripotency and development [147, 148] was enriched with the H3K9me3 peptide in brain and kidney; however it could not be detected in liver. We further showed that ZFP462 is recruited to H3K9me3 via binding to HP1 α. Another example for a tissue-specific HP1 interactor is TRIM66, which is only expressed in testis. TRIM66 interacts with HP1 via its PxVxL motif and forms discrete foci at the centromeric chromocenter [104].

3.6 Follow-up based on newly developed technologies

This thesis focused on chromatin readers for lysine trimethylations. Our data significantly contributes to the knowledge of chromatin readers, especially for trimethylation of lysines 4 and 9 on the H3 tail. However, with these findings and technologies in hand, certain follow-up experiments might be worth performing.

The projects focused on lysine trimethylation, as this modification was already well studied and known to serve as an interaction hub. Using the high-throughput interaction platform described here, broader screens including other modifications and also more combinations of these modifications could easily be analyzed. Especially readers for lysine monomethylation or the recently described lysine crotonylation could yield interesting new insights into chromatin biology. By analyzing peptides bearing different combinations of well studied and additional modifications, it should be possible to assign complexes. All members of a stable complex should follow the same trend and cluster together.

Peptide pull-downs in combination with quantitative mass spectrometry offer a powerful method to discover interactions of unstructured protein parts with specific readers. The concept of binding to an unstructured protein sequence (either modified or unmodified) has been widely studied in the field of chromatin biology. In signal transduction, the SH2 (Src homology 2) domain, which binds to sequences containing a phosphorylated tyrosine [179, 207], has been studied in detail [79]. Another example for binding to an unstructured peptide is the TPR (tetratricopeptide repeat) domain of HOP1, which recognizes the C-terminal 'EEVD' peptide sequence on HSP90 and HSP70 [210]. Using sophisticated prediction tools [132], disordered regions, which often contain linear motifs, can be discovered and used for peptide interaction screens. Many chromatin-associated proteins contain partially unstructured regions [53], which could be modified in a similar manner to histone tails. Moreover, many reader domains are present on chromatin associated proteins, which do not appear to target

histone modifications. It is thus not unlikely that those domains target modified linear motifs on other chromatin associated complexes. The high throughput interaction pipeline described here could offer the possibility to screen many unstructured protein sequences for interacting proteins in an unbiased manner.

Apart from further technical possibilities offered by the developments described above, the biological insights gained from our peptide pull-downs raise several interesting questions: The major question is why are there so many readers for the chromatin marks and what other determinants guide their binding behavior? We already used ChIP-Seq to dissect on which active promoters SGF29, PHF8, GATAD1, BAP18 and TRRAP are binding. Performing a similar analysis for the increased list of readers from tissue extracts could foster the understanding about reading chromatin marks.

In addition, subunits on many chromatin associated complexes can be exchanged, leading to a diverse array of subcomplexes. One example is the neuronal-specific NuRD complex in which CHD3 or CHD4 is replaced by CHD5. The NuRD complex has even more diversity, as also other positions can be occupied by different proteins, e.g. the deacetylase position can be occupied by HDAC1 or HDAC2. By performing pull-down experiments with all subunits, interaction networks could be obtained to describe the subcomplexes. Using this information in conjunction with genome-wide profiling should shed light on the composition of these subcomplexes and their function.

References

[1] R. Aebersold and M. Mann. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, Mar 2003.

[2] K. Ahmad and S. Henikoff. The histone variant h3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell*, 9(6):1191–1200, Jun 2002.

[3] Alberts, Johnson, Lewis., Raff., Roberts., and Walter. *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.

[4] C. D. Allis, S. L. Berger, J. Cote, S. Dent, T. Jenuwien, T. Kouzarides, L. Pillus, D. Reinberg, Y. Shi, R. Shiekhattar, A. Shilatifard, J. Workman, and Y. Zhang. New nomenclature for chromatin-modifying enzymes. *Cell*, 131(4):633–636, Nov 2007.

[5] R. C. Allshire and G. H. Karpen. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet*, 9(12):923–937, Dec 2008.

[6] D. E. Ayer, Q. A. Lawrence, and R. N. Eisenman. Mad-max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor sin3. *Cell*, 80(5):767–776, Mar 1995.

[7] L. A. Baker, C. D. Allis, and G. G. Wang. Phd fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat Res*, 647(1-2):3–12, Dec 2008.

[8] L. Balakrishnan and B. Milavetz. Decoding the histone h4 lysine 20 methylation mark. *Crit Rev Biochem Mol Biol*, 45(5):440–452, Oct 2010.

[9] A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. Selective recognition of methylated lysine 9 on histone h3 by the hp1 chromo domain. *Nature*, 410(6824):120–124, Mar 2001.

[10] Y. Bao, K. Konesky, Y.-J. Park, S. Rosu, P. N. Dyer, D. Rangasamy, D. J. Tremethick, P. J. Laybourn, and K. Luger. Nucleosomes containing the histone variant h2a.bbd organize only 118 base pairs of dna. *EMBO J*, 23(16):3314–3324, Aug 2004.

[11] A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837, May 2007.

[12] T. Bartke, M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, and T. Kouzarides. Nucleosome-interacting proteins regulated by dna and histone methylation. *Cell*, 143(3):470–484, Oct 2010.

[13] D. B. Beck, H. Oda, S. S. Shen, and D. Reinberg. Pr-set7 and h4k20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev*, 26(4):325–337, Feb 2012.

[14] R. Benetti, S. Gonzalo, I. Jaco, G. Schotta, P. Klatt, T. Jenuwein, and M. A. Blasco. Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *J Cell Biol*, 178(6):925–936, Sep 2007.

[15] S. L. Berger, T. Kouzarides, R. Shiekhattar, and A. Shilatifard. An operational definition of epigenetics. *Genes Dev*, 23(7):781–783, Apr 2009.

[16] B. E. Bernstein, M. Kamal, K. Lindblad-Toh, S. Bekiranov, D. K. Bailey, D. J. Huebert, S. McMahon, E. K. Karlsson, E. J. Kulbokas, 3rd, T. R. Gingeras, S. L. Schreiber, and E. S. Lander. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 120(2):169–181, Jan 2005.

[17] E. Bernstein and C. D. Allis. Rna meets chromatin. *Genes Dev*, 19(14):1635–1655, Jul 2005.

[18] S. R. Bhaumik, E. Smith, and A. Shilatifard. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol*, 14(11):1008–1016, Nov 2007.

[19] A. Bird. Perceptions of epigenetics. *Nature*, 447(7143):396–398, May 2007.

[20] P. J. Boersema, T. T. Aye, T. A. B. van Veen, A. J. R. Heck, and S. Mohammed. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics*, 8(22):4624–4632, Nov 2008.

[21] C. J. Brown, A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe, R. Tonlorenzi, and H. F. Willard. A gene from the region of the human x inactivation centre is expressed exclusively from the inactive x chromosome. *Nature*, 349(6304):38–44, Jan 1991.

[22] S. Burma, B. P. Chen, M. Murphy, A. Kurimasa, and D. J. Chen. Atm phosphorylates histone h2ax in response to dna double-strand breaks. *J Biol Chem*, 276(45):42462–42467, Nov 2001.

[23] M. Buschbeck and L. Di Croce. Approaching the molecular and physiological function of macroh2a variants. *Epigenetics*, 5(2):118–123, Feb 2010.

[24] M. Buschbeck, I. Uribesalgo, I. Wibowo, P. Rué, D. Martin, A. Gutierrez, L. Morey, R. Guigó, H. López-Schier, and L. Di Croce. The histone variant macroh2a is an epigenetic regulator of key developmental genes. *Nat Struct Mol Biol*, 16(10):1074–1079, Oct 2009.

[25] F. Butter, D. Kappei, F. Buchholz, M. Vermeulen, and M. Mann. A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. *EMBO Rep*, 11(4):305–311, Apr 2010.

[26] F. Butter, M. Scheibe, M. Mörl, and M. Mann. Unbiased rna-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A*, 106(26):10626–10631, Jun 2009.

[27] P. Byvoet, G. R. Shepherd, J. M. Hardin, and B. J. Noland. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch Biochem Biophys*, 148(2):558–567, Feb 1972.

[28] R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. Role of histone h3 lysine 27 methylation in polycomb-group silencing. *Science*, 298(5595):1039–1043, Nov 2002.

[29] M. J. Carrozza, B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W.-J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. Histone h3 methylation by set2 directs deacetylation of coding regions by rpd3s to suppress spurious intragenic transcription. *Cell*, 123(4):581–592, Nov 2005.

[30] A. Celeste, O. Fernandez-Capetillo, M. J. Kruhlak, D. R. Pilch, D. W. Staudt, A. Lee, R. F. Bonner, W. M. Bonner, and A. Nussenzweig. Histone h2ax phosphorylation is dispensable for the initial recognition of dna breaks. *Nat Cell Biol*, 5(7):675–679, Jul 2003.

[31] B. P. Chadwick and H. F. Willard. Histone h2a variants and the inactive x chromosome: identification of a second macroh2a variant. *Hum Mol Genet*, 10(10):1101–1113, May 2001.

[32] B. P. Chadwick and H. F. Willard. A novel chromatin protein, distantly related to histone h2a, is largely excluded from the inactive x chromosome. *J Cell Biol*, 152(2):375–384, Jan 2001.

[33] D. W. Chan, Y. Wang, M. Wu, J. Wong, J. Qin, and Y. Zhao. Unbiased proteomic screen for binding proteins to modified lysines on histone h3. *Proteomics*, 9(9):2343–2354, May 2009.

[34] P. Chi, C. D. Allis, and G. G. Wang. Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer*, 10(7):457–469, Jul 2010.

[35] C. Choudhary and M. Mann. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol*, 11(6):427–439, Jun 2010.

[36] J. Chow and E. Heard. X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol*, 21(3):359–366, Jun 2009.

[37] J. Christensen, K. Agger, P. A. C. Cloos, D. Pasini, S. Rose, L. Sennels, J. Rappsilber, K. H. Hansen, A. E. Salcini, and K. Helin. Rbp2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell*, 128(6):1063–1076, Mar 2007.

[38] P. A. C. Cloos, J. Christensen, K. Agger, A. Maiolica, J. Rappsilber, T. Antal, K. H. Hansen, and K. Helin. The putative oncogene gasc1 demethylates tri- and dimethylated lysine 9 on histone h3. *Nature*, 442(7100):307–311, Jul 2006.

[39] C. Costanzi and J. R. Pehrson. Histone macroh2a1 is concentrated in the inactive x chromosome of female mammals. *Nature*, 393(6685):599–601, Jun 1998.

[40] J. Cox and M. Mann. Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 26(12):1367–1372, Dec 2008.

[41] J. Cox and M. Mann. Computational principles of determining and improving mass precision and accuracy for proteome measurements in an orbitrap. *J Am Soc Mass Spectrom*, 20(8):1477–1485, Aug 2009.

[42] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann. Andromeda: a peptide search engine integrated into the maxquant environment. *J Proteome Res*, 10(4):1794–1805, Apr 2011.

[43] P. Cubas, C. Vincent, and E. Coen. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, 401(6749):157–161, Sep 1999.

[44] B. Czermin, R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta. Drosophila enhancer of zeste/esc complexes have a histone h3 methyltransferase activity that marks chromosomal polycomb sites. *Cell*, 111(2):185–196, Oct 2002.

[45] J. R. Danzer and L. L. Wallrath. Mechanisms of hp1-mediated gene silencing in drosophila. *Development*, 131(15):3571–3580, Aug 2004.

[46] L. M. F. de Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Fröhlich, T. C. Walther, and M. Mann. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455(7217):1251–1254, Oct 2008.

[47] C. Dhalluin, J. E. Carlson, L. Zeng, C. He, A. K. Aggarwal, and M. M. Zhou. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, 399(6735):491–496, Jun 1999.

[48] A. Dhayalan, A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin, and A. Jeltsch. The dnmt3a pwwp domain reads histone 3 lysine 36 trimethylation and guides dna methylation. *J Biol Chem*, 285(34):26114–26120, Aug 2010.

[49] A. Dhayalan, R. Tamas, I. Bock, A. Tattermusch, E. Dimitrova, S. Kudithipudi, S. Ragozin, and A. Jeltsch. The atrx-add domain binds to h3 tail peptides and reads the combined methylation state of k4 and k9. *Hum Mol Genet*, 20(11):2195–2203, Jun 2011.

[50] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder, and T. J. Richmond. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306(5701):1571–1573, Nov 2004.

[51] R. Draker and P. Cheung. Transcriptional and epigenetic functions of histone variant h2a.z. *Biochem Cell Biol*, 87(1):19–25, Feb 2009.

[52] E. M. Dunleavy, D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, Y. Nakamura, Y. Daigo, Y. Nakatani, and G. Almouzni-Pettinotti. Hjurp is a cell-cycle-

dependent maintenance and deposition factor of cenp-a at centromeres. *Cell*, 137(3):485–497, May 2009.

[53] H. J. Dyson and P. E. Wright. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6(3):197–208, Mar 2005.

[54] W. C. Earnshaw and N. Rothfield. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*, 91(3-4):313–321, 1985.

[55] J. Ernst and M. Kellis. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*, 28(8):817–825, Aug 2010.

[56] S. Eustermann, J.-C. Yang, M. J. Law, R. Amos, L. M. Chapman, C. Jelinska, D. Garrick, D. Clynes, R. J. Gibbons, D. Rhodes, D. R. Higgs, and D. Neuhaus. Combinatorial readout of histone h3 modifications specifies localization of atrx to heterochromatin. *Nat Struct Mol Biol*, 18(7):777–782, Jul 2011.

[57] R. Faast, V. Thonglairoam, T. C. Schulz, J. Beall, J. R. Wells, H. Taylor, K. Matthaei, P. D. Rathjen, D. J. Tremethick, and I. Lyons. Histone variant h2a.z is required for early mammalian development. *Curr Biol*, 11(15):1183–1187, Aug 2001.

[58] A. P. Feinberg. Phenotypic plasticity and the epigenetics of human disease. *Nature*, 447(7143):433–440, May 2007.

[59] G. J. Filion, J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. van Steensel. Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell*, 143(2):212–224, Oct 2010.

[60] W. Fischle, H. Franz, S. A. Jacobs, C. D. Allis, and S. Khorasanizadeh. Specificity of the chromodomain y chromosome family of chromodomains for lysine-methylated ark(s/t) motifs. *J Biol Chem*, 283(28):19626–19635, Jul 2008.

[61] W. Fischle, B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki, and C. D. Allis. Regulation of hp1-chromatin binding by histone h3 methylation and phosphorylation. *Nature*, 438(7071):1116–1122, Dec 2005.

[62] D. R. Foltz, L. E. T. Jansen, A. O. Bailey, J. R. Yates, 3rd, E. A. Bassett, S. Wood, B. E. Black, and D. W. Cleveland. Centromere-specific assembly of cenp-a nucleosomes is mediated by hjurp. *Cell*, 137(3):472–484, May 2009.

[63] T. Gautier, D. W. Abbott, A. Molla, A. Verdel, J. Ausio, and S. Dimitrov. Histone variant h2abbd confers lower stability to the nucleosome. *EMBO Rep*, 5(7):715–720, Jul 2004.

[64] A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084):631–636, Mar 2006.

[65] A.-C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.-A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–147, Jan 2002.

[66] T. Geiger, J. Cox, and M. Mann. Proteomics on an orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol Cell Proteomics*, 9(10):2252–2261, Oct 2010.

[67] T. Geiger, J. Cox, P. Ostasiewicz, J. R. Wisniewski, and M. Mann. Super-silac mix for quantitative proteomics of human tumor tissue. *Nat Methods*, 7(5):383–385, May 2010.

[68] M. Giannattasio, F. Lazzaro, P. Plevani, and M. Muzi-Falconi. The dna damage checkpoint response requires histone h2b ubiquitination by rad6-bre1 and h3 methylation by dot1. *J Biol Chem*, 280(11):9879–9886, Mar 2005.

[69] A. D. Goldberg, C. D. Allis, and E. Bernstein. Epigenetics: a landscape takes shape. *Cell*, 128(4):635–638, Feb 2007.

[70] A. D. Goldberg, L. A. Banaszynski, K.-M. Noh, P. W. Lewis, S. J. Elsaesser, S. Stadler, S. Dewell, M. Law, X. Guo, X. Li, D. Wen, A. Chapgier, R. C. DeKelver, J. C. Miller, Y.-L. Lee, E. A. Boydston, M. C. Holmes, P. D. Gregory, J. M. Greally, S. Rafii, C. Yang, P. J. Scambler, D. Garrick, R. J. Gibbons, D. R. Higgs, I. M. Cristea, F. D. Urnov, D. Zheng, and C. D. Allis. Distinct factors control histone variant h3.3 localization at specific genomic regions. *Cell*, 140(5):678–691, Mar 2010.

[71] M. G. Goll and T. H. Bestor. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, 74:481–514, 2005.

[72] J. Graumann, N. C. Hubner, J. B. Kim, K. Ko, M. Moser, C. Kumar, J. Cox, H. Schöler, and M. Mann. Stable isotope labeling by amino acids in cell culture (silac) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol Cell Proteomics*, 7(4):672–683, Apr 2008.

[73] S. A. Grigoryev and C. L. Woodcock. Chromatin organization - the 30nm fiber. *Exp Cell Res*, Feb 2012.

[74] A. Gruhler, J. V. Olsen, S. Mohammed, P. Mortensen, N. J. Faergeman, M. Mann, and O. N. Jensen. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics*, 4(3):310–327, Mar 2005.

[75] K. G. Guruharsha, J.-F. Rual, B. Zhai, J. Mintseris, P. Vaidya, N. Vaidya, C. Beekman, C. Wong, D. Y. Rhee, O. Cenaj, E. McKillip, S. Shah, M. Stapleton, K. H. Wan, C. Yu, B. Parsa, J. W. Carlson, X. Chen, B. Kapadia, K. VijayRaghavan, S. P. Gygi, S. E. Celniker, R. A. Obar, and S. Artavanis-Tsakonas. A protein complex network of drosophila melanogaster. *Cell*, 147(3):690–703, Oct 2011.

[76] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, 17(10):994–999, Oct 1999.

[77] S. B. Hake and C. D. Allis. Histone h3 variants and their potential role in indexing mammalian genomes: the "h3 barcode hypothesis". *Proc Natl Acad Sci U S A*, 103(17):6428–6435, Apr 2006.

[78] S. Hanke, H. Besir, D. Oesterhelt, and M. Mann. Absolute silac for accurate quantitation of proteins in complex mixtures down to the attomole level. *J Proteome Res*, 7(3):1118–1130, Mar 2008.

[79] S. Hanke and M. Mann. The phosphotyrosine interactome of the insulin receptor family and its substrates irs-1 and irs-2. *Mol Cell Proteomics*, 8(3):519–534, Mar 2009.

[80] F. He, T. Umehara, K. Saito, T. Harada, S. Watanabe, T. Yabuki, T. Kigawa, M. Takahashi, K. Kuwasako, K. Tsuda, T. Matsuda, M. Aoki, E. Seki, N. Kobayashi, P. Güntert, S. Yokoyama, and Y. Muto. Structural insight into the zinc finger cw domain as a histone modification reader. *Structure*, 18(9):1127–1139, Sep 2010.

[81] N. D. Heintzman, G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243):108–112, May 2009.

[82] E. Heitz. Das heterochromatin der moose. *Jahrb Wiss Bot*, 69:762–818, 1928.

[83] A. Hellman and A. Chess. Gene body-specific methylation on the active x chromosome. *Science*, 315(5815):1141–1143, Feb 2007.

[84] S. Henikoff and A. Shilatifard. Histone modification: cause or cog? *Trends Genet*, 27(10):389–396, Oct 2011.

[85] N. Herranz, N. Dave, A. Millanes-Romero, L. Morey, V. M. Díaz, V. Lórenz-Fonfría, R. Gutierrez-Gallego, C. Jerónimo, L. Di Croce, A. García de Herreros, and S. Peiró. Lysyl oxidase-like 2 deaminates lysine 4 in histone h3. *Mol Cell*, Apr 2012.

[86] M. Hilger and M. Mann. Triple silac to determine stimulus specific interactions in the wnt pathway. *J Proteome Res*, 11(2):982–994, Feb 2012.

[87] Y. Huang, J. Fang, M. T. Bedford, Y. Zhang, and R.-M. Xu. Recognition of histone h3 lysine-4 methylation by the double tudor domain of jmjd2a. *Science*, 312(5774):748–751, May 2006.

[88] N. C. Hubner, A. W. Bird, J. Cox, B. Splettstoesser, P. Bandilla, I. Poser, A. Hyman, and M. Mann. Quantitative proteomics combined with bac transgeneomics reveals in vivo protein interactions. *J Cell Biol*, 189(4):739–754, May 2010.

[89] N. C. Hubner and M. Mann. Extracting gene function from protein-protein interactions using quantitative bac interactomics (qubic). *Methods*, 53(4):453–459, Apr 2011.

[90] L. Hughes-Davies, D. Huntsman, M. Ruas, F. Fuks, J. Bye, S.-F. Chin, J. Milner, L. A. Brown, F. Hsu, B. Gilks, T. Nielsen, M. Schulzer, S. Chia, J. Ragaz, A. Cahn, L. Linger, H. Ozdag, E. Cattaneo, E. S. Jordanova, E. Schuuring, D. S. Yu, A. Venkitaraman, B. Ponder, A. Doherty, S. Aparicio, D. Bentley, C. Theillet, C. P. Ponting, C. Caldas, and T. Kouzarides. Emsy links the brca2 pathway to sporadic breast and ovarian cancer. *Cell*, 115(5):523–535, Nov 2003.

[91] Y. Huyen, O. Zgheib, R. A. Ditullio, Jr, V. G. Gorgoulis, P. Zacharatos, T. J. Petty, E. A. Sheston, H. S. Mellert, E. S. Stavridi, and T. D. Halazonetis. Methylated lysine 79 of histone h3 targets 53bp1 to dna double-strand breaks. *Nature*, 432(7015):406–411, Nov 2004.

[92] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann. Exponentially modified protein abundance index (empai) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*, 4(9):1265–1272, Sep 2005.

[93] S. Iwase, F. Lan, P. Bayliss, L. de la Torre-Ubieta, M. Huarte, H. H. Qi, J. R. Whetstine, A. Bonni, T. M. Roberts, and Y. Shi. The x-linked mental retardation gene smcx/jarid1c defines a family of histone h3 lysine 4 demethylases. *Cell*, 128(6):1077–1088, Mar 2007.

[94] S. A. Jacobs, S. D. Taverna, Y. Zhang, S. D. Briggs, J. Li, J. C. Eissenberg, C. D. Allis, and S. Khorasanizadeh. Specificity of the hp1 chromo domain for the methylated n-terminus of histone h3. *EMBO J*, 20(18):5232–5241, Sep 2001.

[95] R. H. Jacobson, A. G. Ladurner, D. S. King, and R. Tjian. Structure and function of a human tafii250 double bromodomain module. *Science*, 288(5470):1422–1425, May 2000.

[96] T. Jenuwein and C. D. Allis. Translating the histone code. *Science*, 293(5532):1074–1080, Aug 2001.

[97] C. Jin and G. Felsenfeld. Nucleosome stability mediated by histone variants h3.3 and h2a.z. *Genes Dev*, 21(12):1519–1529, Jun 2007.

[98] C. Jin, C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, and G. Felsenfeld. H3.3/h2a.z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet*, 41(8):941–945, Aug 2009.

[99] A. A. Joshi and K. Struhl. Eaf3 chromodomain interaction with methylated h3-k36 links histone deacetylation to pol ii elongation. *Mol Cell*, 20(6):971–978, Dec 2005.

[100] D. Karachentsev, K. Sarma, D. Reinberg, and R. Steward. Pr-set7-dependent methylation of histone h4 lys 20 functions in repression of gene expression and is essential for mitosis. *Genes Dev*, 19(4):431–435, Feb 2005.

[101] P. Karagianni, L. Amazit, J. Qin, and J. Wong. Icbp90, a novel methyl k9 h3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol Cell Biol*, 28(2):705–717, Jan 2008.

[102] C. A. Keetch, E. H. C. Bromley, M. G. McCammon, N. Wang, J. Christodoulou, and C. V. Robinson. L55p transthyretin accelerates subunit exchange and leads to rapid formation of hybrid tetramers. *J Biol Chem*, 280(50):41667–41674, Dec 2005.

[103] M.-C. Keogh, S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Chin, T. Punna, N. J. Thompson, C. Boone, A. Emili, J. S. Weissman, T. R. Hughes, B. D. Strahl, M. Grunstein, J. F. Greenblatt, S. Buratowski, and N. J. Krogan. Cotranscriptional set2 methylation of histone h3 lysine 36 recruits a repressive rpd3 complex. *Cell*, 123(4):593–605, Nov 2005.

[104] K. Khetchoumian, M. Teletin, M. Mark, T. Lerouge, M. Cerviño, M. Oulad-Abdelghani, P. Chambon, and R. Losson. Tif1delta, a novel hp1-interacting member of the transcriptional intermediary factor 1 (tif1) family expressed by elongating spermatids. *J Biol Chem*, 279(46):48329–48341, Nov 2004.

[105] T. Kim and S. Buratowski. Two saccharomyces cerevisiae jmjc domain proteins demethylate histone h3 lys36 in transcribed regions to promote elongation. *J Biol Chem*, 282(29):20827–20835, Jul 2007.

[106] T. H. Kim, L. O. Barrera, M. Zheng, C. Qu, M. A. Singer, T. A. Richmond, Y. Wu, R. D. Green, and B. Ren. A high-resolution map of active promoters in the human genome. *Nature*, 436(7052):876–880, Aug 2005.

[107] K. Kingdon. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys Rev*, 21:408–418, 1923.

[108] D. S. Kirkpatrick, S. A. Gerber, and S. P. Gygi. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods*, 35(3):265–273, Mar 2005.

[109] R. J. Klose, Q. Yan, Z. Tothova, K. Yamane, H. Erdjument-Bromage, P. Tempst, D. G. Gilliland, Y. Zhang, and W. G. Kaelin, Jr. The retinoblastoma binding protein rbp2 is an h3k4 demethylase. *Cell*, 128(5):889–900, Mar 2007.

[110] J. Kobayashi, H. Tauchi, B. Chen, S. Burma, S. Bruma, S. Tashiro, S. Matsuura, K. Tanimoto, D. J. Chen, and K. Komatsu. Histone h2ax participates the dna damage-induced atm activation through interaction with nbs1. *Biochem Biophys Res Commun*, 380(4):752–757, Mar 2009.

[111] A. Kohlmaier, F. Savarese, M. Lachner, J. Martens, T. Jenuwein, and A. Wutz. A chromosomal memory triggered by xist regulates histone methylation in x inactivation. *PLoS Biol*, 2(7):E171, Jul 2004.

[112] K. Kokura, L. Sun, M. T. Bedford, and J. Fang. Methyl-h3k9-binding protein mpp8 mediates e-cadherin gene silencing and promotes tumour cell motility and invasion. *EMBO J*, 29(21):3673–3687, Nov 2010.

[113] P. Kolasinska-Zwierz, T. Down, I. Latorre, T. Liu, X. S. Liu, and J. Ahringer. Differential chromatin marking of introns and expressed exons by h3k36me3. *Nat Genet*, 41(3):376–381, Mar 2009.

[114] N. Kourmouli, P. Jeppesen, S. Mahadevhaiah, P. Burgoyne, R. Wu, D. M. Gilbert, S. Bongiorni, G. Prantera, L. Fanti, S. Pimpinelli, W. Shi, R. Fundele, and P. B. Singh. Heterochromatin and tri-methylated lysine 20 of histone h4 in animals. *J Cell Sci*, 117(Pt 12):2491–2501, May 2004.

[115] T. Kouzarides. Chromatin modifications and their function. *Cell*, 128(4):693–705, Feb 2007.

[116] M. Krüger, M. Moser, S. Ussar, I. Thievessen, C. A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fässler, and M. Mann. Silac mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell*, 134(2):353–364, Jul 2008.

[117] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrín-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P.

Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Y. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, and J. F. Greenblatt. Global landscape of protein complexes in the yeast saccharomyces cerevisiae. *Nature*, 440(7084):637–643, Mar 2006.

[118] N. J. Krogan, M.-C. Keogh, N. Datta, C. Sawa, O. W. Ryan, H. Ding, R. A. Haw, J. Pootoolal, A. Tong, V. Canadien, D. P. Richards, X. Wu, A. Emili, T. R. Hughes, S. Buratowski, and J. F. Greenblatt. A snf2 family atpase complex required for recruitment of the histone h2a variant htz1. *Mol Cell*, 12(6):1565–1576, Dec 2003.

[119] N. J. Krogan, M. Kim, A. Tong, A. Golshani, G. Cagney, V. Canadien, D. P. Richards, B. K. Beattie, A. Emili, C. Boone, A. Shilatifard, S. Buratowski, and J. Greenblatt. Methylation of histone h3 by set2 in saccharomyces cerevisiae is linked to transcriptional elongation by rna polymerase ii. *Mol Cell Biol*, 23(12):4207–4218, Jun 2003.

[120] T. Kusch and J. L. Workman. Histone variants and complexes involved in their exchange. *Subcell Biochem*, 41:91–109, 2007.

[121] A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein. *Genes Dev*, 16(22):2893–2905, Nov 2002.

[122] M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. Methylation of histone h3 lysine 9 creates a binding site for hp1 proteins. *Nature*, 410(6824):116–120, Mar 2001.

[123] M. Larance, A. P. Bailly, E. Pourkarimi, R. T. Hay, G. Buchanan, S. Coulthurst, D. P. Xirodimas, A. Gartner, and A. I. Lamond. Stable-isotope labeling with amino acids in nematodes. *Nat Methods*, 8(10):849–851, 2011.

[124] C.-K. Lee, Y. Shibata, B. Rao, B. D. Strahl, and J. D. Lieb. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet*, 36(8):900–905, Aug 2004.

[125] D. Levy and O. Gozani. Decoding chromatin goes high tech. *Cell*, 142(6):844–846, Sep 2010.

[126] E. B. Lewis. A gene complex controlling segmentation in drosophila. *Nature*, 276(5688):565–570, Dec 1978.

[127] P. W. Lewis, S. J. Elsaesser, K.-M. Noh, S. C. Stadler, and C. D. Allis. Daxx is an h3.3-specific histone chaperone and cooperates with atrx in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A*, 107(32):14075–14080, Aug 2010.

[128] B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–719, Feb 2007.

[129] G. Li and D. Reinberg. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev*, 21(2):175–186, Apr 2011.

[130] H. Li, S. Ilin, W. Wang, E. M. Duncan, J. Wysocka, C. D. Allis, and D. J. Patel. Molecular basis for site-specific read-out of histone h3k4me3 by the bptf phd finger of nurf. *Nature*, 442(7098):91–95, Jul 2006.

[131] X. Li, E. A. Foley, K. R. Molloy, Y. Li, B. T. Chait, and T. M. Kapoor. Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *J Am Chem Soc*, 134(4):1982–1985, Feb 2012.

[132] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell. Protein disorder prediction: implications for structural proteomics. *Structure*, 11(11):1453–1459, Nov 2003.

[133] X. Ling, T. A. Harkness, M. C. Schultz, G. Fisher-Adams, and M. Grunstein. Yeast histone h3 and h4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation. *Genes Dev*, 10(6):686–699, Mar 1996.

[134] H. Liu, R. G. Sadygov, and J. R. Yates, 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*, 76(14):4193–4201, Jul 2004.

[135] Y. Liu, R. Subrahmanyam, T. Chakraborty, R. Sen, and S. Desiderio. A plant homeodomain in rag-2 that binds hypermethylated lysine 4 of histone h3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*, 27(4):561–571, Oct 2007.

[136] C. A. Luber, J. Cox, H. Lauterbach, B. Fancke, M. Selbach, J. Tschopp, S. Akira, M. Wiegand, H. Hochrein, M. O'Keeffe, and M. Mann. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity*, 32(2):279–289, Feb 2010.

[137] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648):251–260, Sep 1997.

[138] K. Luger and T. J. Richmond. Dna binding within the nucleosome core. *Curr Opin Struct Biol*, 8(1):33–40, Feb 1998.

[139] M. F. Lyon. Gene action in the x-chromosome of the mouse (mus musculus l.). *Nature*, 190:372–373, Apr 1961.

[140] N. Macdonald, J. P. I. Welburn, M. E. M. Noble, A. Nguyen, M. B. Yaffe, D. Clynes, J. G. Moggs, G. Orphanides, S. Thomson, J. W. Edmunds, A. L. Clayton, J. A. Endicott, and L. C. Mahadevan. Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. *Mol Cell*, 20(2):199–211, Oct 2005.

[141] K. Maeshima, S. Hihara, and M. Eltsov. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol*, 22(3):291–297, Jun 2010.

[142] Makarov. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*, 72(6):1156–1162, Mar 2000.

[143] A. Malovannaya, R. B. Lanz, S. Y. Jung, Y. Bulynko, N. T. Le, D. W. Chan, C. Ding, Y. Shi, N. Yucer, G. Krenciute, B.-J. Kim, C. Li, R. Chen, W. Li, Y. Wang, B. W. O'Malley, and J. Qin. Analysis of the human endogenous coregulator complexome. *Cell*, 145(5):787–799, May 2011.

[144] R. Margueron, P. Trojer, and D. Reinberg. The key to development: interpreting the histone code? *Curr Opin Genet Dev*, 15(2):163–176, Apr 2005.

[145] R. Marmorstein and S. L. Berger. Structure and function of bromodomains in chromatin-regulating complexes. *Gene*, 272(1-2):1–9, Jul 2001.

[146] W. F. Marzluff, P. Gongidi, K. R. Woods, J. Jin, and L. J. Maltais. The human and mouse replication-dependent histone genes. *Genomics*, 80(5):487–498, Nov 2002.

[147] J. Massé, A. Laurent, B. Nicol, D. Guerrier, I. Pellerin, and S. Deschamps. Involvement of zfpip/zfp462 in chromatin integrity and survival of p19 pluripotent cells. *Exp Cell Res*, 316(7):1190–1201, Apr 2010.

[148] J. Massé, C. Piquet-Pellorce, J. Viet, D. Guerrier, I. Pellerin, and S. Deschamps. Zfpip/zfp462 is involved in p19 cell pluripotency and in their neuronal fate. *Exp Cell Res*, 317(13):1922–1934, Aug 2011.

[149] A. G. W. Matthews, A. J. Kuo, S. Ramón-Maiques, S. Han, K. S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D. N. Ciccone, K. L. Walter, P. J. Utz, Y. Shi, T. G. Kutateladze, W. Yang, O. Gozani, and M. A. Oettinger. Rag2 phd finger couples histone h3 lysine 4 trimethylation with v(d)j recombination. *Nature*, 450(7172):1106–1110, Dec 2007.

[150] E. McKittrick, P. R. Gafken, K. Ahmad, and S. Henikoff. Histone h3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A*, 101(6):1525–1530, Feb 2004.

[151] P. Mertins, N. D. Udeshi, K. R. Clauser, D. R. Mani, J. Patel, S.-E. Ong, J. D. Jaffe, and S. A. Carr. itraq labeling is superior to mtraq for quantitative global proteomics and phosphoproteomics. *Mol Cell Proteomics*, Dec 2011.

[152] A. Michalski, E. Damoc, J.-P. Hauschild, O. Lange, A. Wieghaus, A. Makarov, N. Nagaraj, J. Cox, M. Mann, and S. Horning. Mass spectrometry-based proteomics using q exactive, a high-performance benchtop quadrupole orbitrap mass spectrometer. *Mol Cell Proteomics*, 10(9):M111.011015, Sep 2011.

[153] T. Miller, N. J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J. F. Greenblatt, and A. Shilatifard. Compass: a complex of proteins associated with a trithorax-related set domain protein. *Proc Natl Acad Sci U S A*, 98(23):12902–12907, Nov 2001.

[154] J. Min, Q. Feng, Z. Li, Y. Zhang, and R.-M. Xu. Structure of the catalytic domain of human dot1l, a non-set domain nucleosomal histone methyltransferase. *Cell*, 112(5):711–723, Mar 2003.

[155] T. B. Miranda, C. C. Cortez, C. B. Yoo, G. Liang, M. Abe, T. K. Kelly, V. E. Marquez, and P. A. Jones. Dznep is a global histone methylation inhibitor that reactivates developmental genes not silenced by dna methylation. *Mol Cancer Ther*, 8(6):1579–1588, Jun 2009.

[156] G. Mittler, F. Butter, and M. Mann. A silac-based dna protein interaction screen that identifies candidate binding proteins to functional dna elements. *Genome Res*, 19(2):284–293, Feb 2009.

[157] J. Müller, C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, M. B. O'Connor, R. E. Kingston, and J. A. Simon. Histone methyltransferase activity of a drosophila polycomb group repressor complex. *Cell*, 111(2):197–208, Oct 2002.

[158] J. Müller and P. Verrijzer. Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev*, 19(2):150–158, Apr 2009.

[159] T. W. Muir, D. Sondhi, and P. A. Cole. Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A*, 95(12):6705–6710, Jun 1998.

[160] Muller. Types of visible variations induced by x-rays in drosophila. *J. Genet.*, 22:299–334, 1930.

[161] C. A. Musselman, J. Ramírez, J. K. Sims, R. E. Mansfield, S. S. Oliver, J. M. Denu, J. P. Mackay, P. A. Wade, J. Hagman, and T. G. Kutateladze. Bivalent recognition of nucleosomes by the tandem phd fingers of the chd4 atpase is required for chd4-mediated repression. *Proc Natl Acad Sci U S A*, 109(3):787–792, Jan 2012.

[162] N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, and M. Mann. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*, 7:548, 2011.

[163] H. H. Ng, Q. Feng, H. Wang, H. Erdjument-Bromage, P. Tempst, Y. Zhang, and K. Struhl. Lysine methylation within the globular domain of histone h3 by dot1 is important for telomeric silencing and sir protein association. *Genes Dev*, 16(12):1518–1527, Jun 2002.

[164] A. T. Nguyen and Y. Zhang. The diverse functions of dot1 and h3k79 methylation. *Genes Dev*, 25(13):1345–1358, Jul 2011.

[165] P. R. Nielsen, D. Nietlispach, H. R. Mott, J. Callaghan, A. Bannister, T. Kouzarides, A. G. Murzin, N. V. Murzina, and E. D. Laue. Structure of the hp1 chromodomain bound to histone h3 methylated at lysine 9. *Nature*, 416(6876):103–107, Mar 2002.

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

[167] K. Ning, D. Fermin, and A. I. Nesvizhskii. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with rna-seq gene expression data. *J Proteome Res*, 11(4):2261–2271, Apr 2012.

[168] C. Nislow, E. Ray, and L. Pillus. Set1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell*, 8(12):2421–2436, Dec 1997.

[169] R.-S. Nozawa, K. Nagao, H.-T. Masuda, O. Iwasaki, T. Hirota, N. Nozaki, H. Kimura, and C. Obuse. Human pogz modulates dissociation of hp1alpha from mitotic chromosome arms through aurora b activation. *Nat Cell Biol*, 12(7):719–727, Jul 2010.

[170] Y. Oda, K. Huang, F. R. Cross, D. Cowburn, and B. T. Chait. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A*, 96(12):6591–6596, Jun 1999.

[171] J. V. Olsen, L. M. F. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, and M. Mann. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a c-trap. *Mol Cell Proteomics*, 4(12):2010–2021, Dec 2005.

[172] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, and M. Mann. Higher-energy c-trap dissociation for peptide modification analysis. *Nat Methods*, 4(9):709–712, Sep 2007.

[173] J. V. Olsen, S.-E. Ong, and M. Mann. Trypsin cleaves exclusively c-terminal to arginine and lysine residues. *Mol Cell Proteomics*, 3(6):608–614, Jun 2004.

[174] J. V. Olsen, J. C. Schwartz, J. Griep-Raming, M. L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E. R. Wouters, M. Senko, A. Makarov, M. Mann, and S. Horning. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics*, 8(12):2759–2769, Dec 2009.

[175] S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 1(5):376–386, May 2002.

[176] S.-E. Ong and M. Mann. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol*, 1(5):252–262, Oct 2005.

[177] M. Papamichos-Chronakis, S. Watanabe, O. J. Rando, and C. L. Peterson. Global regulation of h2a.z localization by the ino80 chromatin-remodeling enzyme is essential for genome integrity. *Cell*, 144(2):200–213, Jan 2011.

[178] E. Passarge. Emil heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *Am J Hum Genet*, 31(2):106–115, Mar 1979.

[179] T. Pawson. Specificity in signal transduction: from phosphotyrosine-sh2 domain interactions to complex cellular systems. *Cell*, 116(2):191–203, Jan 2004.

[180] B. Payer and J. T. Lee. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet*, 42:733–772, 2008.

[181] P. V. Peña, F. Davrazou, X. Shi, K. L. Walter, V. V. Verkhusha, O. Gozani, R. Zhao, and T. G. Kutateladze. Molecular mechanism of histone h3k4me3 recognition by plant homeodomain of ing2. *Nature*, 442(7098):100–103, Jul 2006.

[182] M. T. Pedersen and K. Helin. Histone demethylases in development and disease. *Trends Cell Biol*, 20(11):662–671, Nov 2010.

[183] J. R. Pehrson and V. A. Fried. Macroh2a, a core histone containing a large nonhistone region. *Science*, 257(5075):1398–1400, Sep 1992.

[184] D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, Dec 1999.

[185] J. J. Pesavento, C. R. Bullock, R. D. LeDuc, C. A. Mizzen, and N. L. Kelleher. Combinatorial modification of human histone h4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J Biol Chem*, 283(22):14927–14937, May 2008.

[186] J. J. Pesavento, H. Yang, N. L. Kelleher, and C. A. Mizzen. Certain and progressive methylation of histone h4 at lysine 20 during the cell cycle. *Mol Cell Biol*, 28(1):468–486, Jan 2008.

[187] A. H. F. M. Peters, S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. H. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J. H. A. Martens, and T. Jenuwein. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell*, 12(6):1577–1589, Dec 2003.

[188] L. Piacentini, L. Fanti, M. Berloco, B. Perrini, and S. Pimpinelli. Heterochromatin protein 1 (hp1) is associated with induced gene expression in drosophila euchromatin. *J Cell Biol*, 161(4):707–714, May 2003.

[189] G. Pichler, P. Wolf, C. S. Schmidt, D. Meilinger, K. Schneider, C. Frauer, K. Fellinger, A. Rottach, and H. Leonhardt. Cooperative dna and histone binding by uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem*, 112(9):2585–2593, Sep 2011.

[190] K. Plath, J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning, and Y. Zhang. Role of histone h3 lysine 27 methylation in x inactivation. *Science*, 300(5616):131–135, Apr 2003.

[191] D. K. Pokholok, C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, T. I. Lee, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4):517–527, Aug 2005.

[192] I. Poser, M. Sarov, J. R. A. Hutchins, J.-K. Hériché, Y. Toyoda, A. Pozniakovsky, D. Weigl, A. Nitzsche, B. Hegemann, A. W. Bird, L. Pelletier, R. Kittler, S. Hua, R. Naumann, M. Augsburg, M. M. Sykora, H. Hofemeister, Y. Zhang, K. Nasmyth, K. P. White, S. Dietzel, K. Mechtler, R. Durbin, A. F. Stewart, J.-M. Peters, F. Buchholz, and A. A. Hyman. Bac transgeneomics: a high-throughput method for exploration of protein function in mammals. *Nat Methods*, 5(5):409–415, May 2008.

[193] S. Ramón-Maiques, A. J. Kuo, D. Carney, A. G. W. Matthews, M. A. Oettinger, O. Gozani, and W. Yang. The plant homeodomain finger of rag2 recognizes histone h3 methylated at both lysine-4 and arginine-2. *Proc Natl Acad Sci U S A*, 104(48):18993–18998, Nov 2007.

[194] K. Ratnakumar, L. F. Duarte, G. LeRoy, D. Hasson, D. Smeets, C. Vardabasso, C. Bönisch, T. Zeng, B. Xiang, D. Y. Zhang, H. Li, X. Wang, S. B. Hake, L. Schermelleh, B. A. Garcia, and E. Bernstein. Atrx-mediated chromatin association of histone variant macroh2a1 regulates ?-globin expression. *Genes Dev*, 26(5):433–438, Mar 2012.

[195] D. Ray-Gallet, J.-P. Quivy, C. Scamps, E. M.-D. Martini, M. Lipinski, and G. Almouzni. Hira is critical for a nucleosome assembly pathway independent of dna synthesis. *Mol Cell*, 9(5):1091–1100, May 2002.

[196] S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, and T. Jenuwein. Regulation of chromatin structure by site-specific histone h3 methyltransferases. *Nature*, 406(6796):593–599, Aug 2000.

[197] J. C. Rice, K. Nishioka, K. Sarma, R. Steward, D. Reinberg, and C. D. Allis. Mitotic-specific methylation of histone h4 lys 20 follows increased pr-set7 expression and its localization to mitotic chromosomes. *Genes Dev*, 16(17):2225–2230, Sep 2002.

[198] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Séraphin. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, 17(10):1030–1032, Oct 1999.

[199] K. T. G. Rigbolt, T. A. Prokhorova, V. Akimov, J. Henningsen, P. T. Johansen, I. Kratchmarova, M. Kassem, M. Mann, J. V. Olsen, and B. Blagoev. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal*, 4(164):rs3, 2011.

[200] P. J. J. Robinson, L. Fairall, V. A. T. Huynh, and D. Rhodes. Em measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A*, 103(17):6506–6511, Apr 2006.

[201] E. P. Rogakou, D. R. Pilch, A. H. Orr, V. S. Ivanova, and W. M. Bonner. Dna double-stranded breaks induce histone h2ax phosphorylation on serine 139. *J Biol Chem*, 273(10):5858–5868, Mar 1998.

[202] A. Roguev, D. Schaft, A. Shevchenko, W. W. Pijnappel, M. Wilm, R. Aasland, and A. F. Stewart. The saccharomyces cerevisiae set1 complex includes an ash2 homologue and methylates histone 3 lysine 4. *EMBO J*, 20(24):7137–7148, Dec 2001.

[203] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin. Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 3(12):1154–1169, Dec 2004.

[204] U. Rothbauer, K. Zolghadr, S. Muyldermans, A. Schepers, M. C. Cardoso, and H. Leonhardt. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics*, 7(2):282–289, Feb 2008.

[205] A. J. Ruthenburg, H. Li, T. A. Milne, S. Dewell, R. K. McGinty, M. Yuen, B. Ueberheide, Y. Dou, T. W. Muir, D. J. Patel, and C. D. Allis. Recognition of a mononucleosomal histone modification pattern by bptf via multivalent interactions. *Cell*, 145(5):692–706, May 2011.

[206] J. Rykl. *LTQ Orbitrap Biotech operations Training course manual*. Thermo Fisher Scientific, November 2008.

[207] I. Sadowski, J. C. Stone, and T. Pawson. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of fujinami sarcoma virus p130gag-fps. *Mol Cell Biol*, 6(12):4396–4408, Dec 1986.

[208] H. Santos-Rosa, R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. T. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. Active genes are trimethylated at k4 of histone h3. *Nature*, 419(6905):407–411, Sep 2002.

[209] M. E. Sardiu, Y. Cai, J. Jin, S. K. Swanson, R. C. Conaway, J. W. Conaway, L. Florens, and M. P. Washburn. Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc Natl Acad Sci U S A*, 105(5):1454–1459, Feb 2008.

[210] C. Scheufler, A. Brinker, G. Bourenkov, S. Pegoraro, L. Moroder, H. Bartunik, F. U. Hartl, and I. Moarefi. Structure of tpr domain-peptide complexes: critical elements in the assembly of the hsp70-hsp90 multichaperone machine. *Cell*, 101(2):199–210, Apr 2000.

[211] G. Schotta, R. Sengupta, S. Kubicek, S. Malin, M. Kauer, E. Callén, A. Celeste, M. Pagani, S. Opravil, I. A. De La Rosa-Velazquez, A. Espejo, M. T. Bedford, A. Nussenzweig, M. Busslinger, and T. Jenuwein. A chromatin-wide transition

to h4k20 monomethylation impairs genome integrity and programmed dna re-arrangements in the mouse. *Genes Dev*, 22(15):2048–2061, Aug 2008.

[212] W. X. Schulze and M. Mann. A novel proteomic screen for peptide-protein inter-actions. *J Biol Chem*, 279(11):10756–10764, Mar 2004.

[213] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, May 2011.

[214] J. C. Schwartz, M. W. Senko, and J. E. P. Syka. A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom*, 13(6):659–669, Jun 2002.

[215] M. Scigelova, M. Hornshaw, A. Giannakopulos, and A. Makarov. Fourier trans-form mass spectrometry. *Mol Cell Proteomics*, 10(7):M111.009431, Jul 2011.

[216] M. Selbach and M. Mann. Protein interaction screening by quantitative immuno-precipitation combined with knockdown (quick). *Nat Methods*, 3(12):981–983, Dec 2006.

[217] M. Selbach, F. E. Paul, S. Brandt, P. Guye, O. Daumke, S. Backert, C. Dehio, and M. Mann. Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe*, 5(4):397–403, Apr 2009.

[218] K. Sha. A mechanistic view of genomic imprinting. *Annu Rev Genomics Hum Genet*, 9:197–216, 2008.

[219] K. Sharma, C. Weber, M. Bairlein, Z. Greff, G. Kéri, J. Cox, J. V. Olsen, and H. Daub. Proteomics strategy for quantitative protein interaction profiling in cell extracts. *Nat Methods*, 6(10):741–744, Oct 2009.

[220] X. Shi, T. Hong, K. L. Walter, M. Ewalt, E. Michishita, T. Hung, D. Carney, P. Peña, F. Lan, M. R. Kaadige, N. Lacoste, C. Cayrou, F. Davrazou, A. Saha, B. R. Cairns, D. E. Ayer, T. G. Kutateladze, Y. Shi, J. Côté, K. F. Chua, and O. Gozani. Ing2 phd domain links histone h3 lysine 4 methylation to active gene repression. *Nature*, 442(7098):96–99, Jul 2006.

[221] Y. Shi, F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi. Histone demethylation mediated by the nuclear amine oxidase homolog lsd1. *Cell*, 119(7):941–953, Dec 2004.

[222] A. Shilatifard. Molecular implementation and physiological roles for histone h3 lysine 4 (h3k4) methylation. *Curr Opin Cell Biol*, 20(3):341–348, Jun 2008.

[223] M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson. Histone h4-k16 acetylation controls chromatin structure and protein interactions. *Science*, 311(5762):844–847, Feb 2006.

[224] M. C. C. Silva and L. E. T. Jansen. At the right place at the right time: novel cenp-a binding proteins shed light on centromere assembly. *Chromosoma*, 118(5):567–574, Oct 2009.

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

[226] B. C. Smith and J. M. Denu. Chemical mechanisms of histone lysine and arginine modifications. *Biochim Biophys Acta*, 1789(1):45–57, Jan 2009.

[227] M. E. Sowa, E. J. Bennett, S. P. Gygi, and J. W. Harper. Defining the human deubiquitinating enzyme interaction landscape. *Cell*, 138(2):389–403, Jul 2009.

[228] H. Steen and M. Mann. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol*, 5(9):699–711, Sep 2004.

[229] D. J. Steger, M. I. Lefterova, L. Ying, A. J. Stonestrom, M. Schupp, D. Zhuo, A. L. Vakoc, J.-E. Kim, J. Chen, M. A. Lazar, G. A. Blobel, and C. R. Vakoc. Dot1l/kmt4 recruitment and h3k79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol*, 28(8):2825–2839, Apr 2008.

[230] G. S. Stewart, B. Wang, C. R. Bignell, A. M. R. Taylor, and S. J. Elledge. Mdc1 is a mediator of the mammalian dna damage checkpoint. *Nature*, 421(6926):961–966, Feb 2003.

[231] B. D. Strahl and C. D. Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–45, Jan 2000.

[232] B. D. Strahl, R. Ohba, R. G. Cook, and C. D. Allis. Methylation of histone h3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in tetrahymena. *Proc Natl Acad Sci U S A*, 96(26):14967–14972, Dec 1999.

[233] M. D. Sury, J.-X. Chen, and M. Selbach. The silac fly allows for accurate protein quantification in vivo. *Mol Cell Proteomics*, 9(10):2173–2183, Oct 2010.

[234] H. Tagami, D. Ray-Gallet, G. Almouzni, and Y. Nakatani. Histone h3.1 and h3.3 complexes mediate nucleosome assembly pathways dependent or independent of dna synthesis. *Cell*, 116(1):51–61, Jan 2004.

[235] P. B. Talbert and S. Henikoff. Histone variants–ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol*, 11(4):264–275, Apr 2010.

[236] M. Tan, H. Luo, S. Lee, F. Jin, J. S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, Q. Zhu, J. Wysocka, Y. Ye, S. Khochbin, B. Ren, and Y. Zhao. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, 146(6):1016–1028, Sep 2011.

[237] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol*, 14(11):1025–1040, Nov 2007.

[238] A. Thakar, P. Gupta, T. Ishibashi, R. Finn, B. Silva-Moreno, S. Uchiyama, K. Fukui, M. Tomschik, J. Ausio, and J. Zlatanova. H2a.z and h3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. *Biochemistry*, 48(46):10852–10857, Nov 2009.

[239] A. Thiru, D. Nietlispach, H. R. Mott, M. Okuwaki, D. Lyon, P. R. Nielsen, M. Hirshberg, A. Verreault, N. V. Murzina, and E. D. Laue. Structural basis of hp1/pxvxl motif peptide interactions and hp1 localisation to heterochromatin. *EMBO J*, 23(3):489–499, Feb 2004.

[240] G. Thomas, H. W. Lange, and K. Hempel. [relative stability of lysine-bound methyl groups in arginie-rich histones and their subfrations in ehrlich ascites tumor cells in vitro]. *Hoppe Seylers Z Physiol Chem*, 353(9):1423–1428, Sep 1972.

[241] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. A. Mohammed, and C. Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Anal Chem*, 75(8):1895–1904, Apr 2003.

[242] L. Trinkle-Mulcahy, S. Boulon, Y. W. Lam, R. Urcia, F.-M. Boisvert, F. Vandermoere, N. A. Morrice, S. Swift, U. Rothbauer, H. Leonhardt, and A. Lamond. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J Cell Biol*, 183(2):223–239, Oct 2008.

[243] P. Trojer, G. Li, R. J. Sims, 3rd, A. Vaquero, N. Kalakonda, P. Boccuni, D. Lee, H. Erdjument-Bromage, P. Tempst, S. D. Nimer, Y.-H. Wang, and D. Reinberg. L3mbtl1, a histone-methylation-dependent chromatin lock. *Cell*, 129(5):915–928, Jun 2007.

[244] Y.-i. Tsukada, J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst, and Y. Zhang. Histone demethylation by a family of jmjc domain-containing proteins. *Nature*, 439(7078):811–816, Feb 2006.

[245] V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*, 98(9):5116–5121, Apr 2001.

[246] C. R. Vakoc, S. A. Mandat, B. A. Olenchock, and G. A. Blobel. Histone h3 lysine 9 methylation and hp1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell*, 19(3):381–391, Aug 2005.

[247] C. R. Vakoc, M. M. Sachdeva, H. Wang, and G. A. Blobel. Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol Cell Biol*, 26(24):9185–9195, Dec 2006.

[248] F. van Leeuwen, P. R. Gafken, and D. E. Gottschling. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell*, 109(6):745–756, Jun 2002.

[249] B. van Steensel, J. Delrow, and S. Henikoff. Chromatin profiling using targeted dna adenine methyltransferase. *Nat Genet*, 27(3):304–308, Mar 2001.

[250] M. Vermeulen, H. C. Eberl, F. Matarese, H. Marks, S. Denissov, F. Butter, K. K. Lee, J. V. Olsen, A. A. Hyman, H. G. Stunnenberg, and M. Mann. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, 142(6):967–980, Sep 2010.

[251] M. Vermeulen, N. C. Hubner, and M. Mann. High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol*, 19(4):331–337, Aug 2008.

[252] M. Vermeulen, K. W. Mulder, S. Denissov, W. W. M. P. Pijnappel, F. M. A. van Schaik, R. A. Varier, M. P. A. Baltissen, H. G. Stunnenberg, M. Mann, and H. T. M. Timmers. Selective anchoring of tfiid to nucleosomes by trimethylation of histone h3 lysine 4. *Cell*, 131(1):58–69, Oct 2007.

[253] M. Vermeulen and H. T. M. Timmers. Grasping trimethylation of histone h3 at lysine 4. *Epigenomics*, 2(3):395–406, Jun 2010.

[254] A. Vezzoli, N. Bonadies, M. D. Allen, S. M. V. Freund, C. M. Santiveri, B. T. Kvinlaug, B. J. P. Huntly, B. Göttgens, and M. Bycroft. Molecular basis of histone h3k36me3 recognition by the pwwp domain of brpf1. *Nat Struct Mol Biol*, 17(5):617–619, May 2010.

[255] Waddington. The epigenotype. *Endeavour*, 1:18–20, 1942.

[256] G. G. Wang, J. Song, Z. Wang, H. L. Dormann, F. Casadio, H. Li, J.-L. Luo, D. J. Patel, and C. D. Allis. Haematopoietic malignancies caused by dysregulation of a chromatin-binding phd finger. *Nature*, 459(7248):847–851, Jun 2009.

[257] W. Wang, Z. Chen, Z. Mao, H. Zhang, X. Ding, S. Chen, X. Zhang, R. Xu, and B. Zhu. Nucleolar protein spindlin1 recognizes h3k4 methylation and stimulates the expression of rrna genes. *EMBO Rep*, 12(11):1160–1166, Nov 2011.

[258] Z. Wang, C. Zang, K. Cui, D. E. Schones, A. Barski, W. Peng, and K. Zhao. Genome-wide mapping of hats and hdacs reveals distinct functions in active and inactive genes. *Cell*, 138(5):1019–1031, Sep 2009.

[259] Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*, 40(7):897–903, Jul 2008.

[260] J. R. Whetstine, A. Nottke, F. Lan, M. Huarte, S. Smolikov, Z. Chen, E. Spooner, E. Li, G. Zhang, M. Colaiacovo, and Y. Shi. Reversal of histone lysine trimethylation by the jmjd2 family of histone demethylases. *Cell*, 125(3):467–481, May 2006.

[261] J. Widom and A. Klug. Structure of the 300a chromatin filament: X-ray diffraction from oriented samples. *Cell*, 43(1):207–213, Nov 1985.

[262] S. M. Wiedemann, S. N. Mildner, C. Bönisch, L. Israel, A. Maiser, S. Matheisl, T. Straub, R. Merkl, H. Leonhardt, E. Kremmer, L. Schermelleh, and S. B. Hake. Identification and characterization of two novel primate-specific histone h3 variants, h3.x and h3.y. *J Cell Biol*, 190(5):777–791, Sep 2010.

[263] L. H. Wong, H. Ren, E. Williams, J. McGhie, S. Ahn, M. Sim, A. Tam, E. Earle, M. A. Anderson, J. Mann, and K. H. A. Choo. Histone h3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. *Genome Res*, 19(3):404–414, Mar 2009.

[264] A. Wutz. Xist function: bridging chromatin and stem cells. *Trends Genet*, 23(9):457–464, Sep 2007.

[265] R. Wysocki, A. Javaheri, S. Allard, F. Sha, J. Côté, and S. J. Kron. Role of dot1-dependent histone h3 methylation in g1 and s phase dna damage checkpoint functions of rad9. *Mol Cell Biol*, 25(19):8430–8443, Oct 2005.

[266] B. Xiao, C. Jing, G. Kelly, P. A. Walker, F. W. Muskett, T. A. Frenkiel, S. R. Martin, K. Sarma, D. Reinberg, S. J. Gamblin, and J. R. Wilson. Specificity and mechanism of the histone methyltransferase pr-set7. *Genes Dev*, 19(12):1444–1454, Jun 2005.

[267] T. Xiao, H. Hall, K. O. Kizer, Y. Shibata, M. C. Hall, C. H. Borchers, and B. D. Strahl. Phosphorylation of rna polymerase ii ctd regulates h3 methylation in yeast. *Genes Dev*, 17(5):654–663, Mar 2003.

[268] S. Xie, J. Jakoncic, and C. Qian. Uhrf1 double tudor domain and the adjacent phd finger act together to recognize k9me3-containing histone h3 tail. *J Mol Biol*, 415(2):318–328, Jan 2012.

[269] M. Zeiler, W. L. Straube, E. Lundberg, M. Uhlen, and M. Mann. A protein epitope signature tag (prest) library allows silac-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol Cell Proteomics*, 11(3):O111.009613, Mar 2012.

[270] X. Zhang, H. Wen, and X. Shi. Lysine methylation: beyond histones. *Acta Biochim Biophys Sin (Shanghai)*, 44(1):14–27, Jan 2012.

[271] Y. Zhang and D. Reinberg. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*, 15(18):2343–2360, Sep 2001.

[272] J. Zhou, J. Y. Fan, D. Rangasamy, and D. J. Tremethick. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat Struct Mol Biol*, 14(11):1070–1076, Nov 2007.

[273] J. Zlatanova and A. Thakar. H2a.z: view from the top. *Structure*, 16(2):166–179, Feb 2008.

Acknowledgements

First of all, I want to express my deep gratitude towards Prof. Matthias Mann for offering me the unique opportunity to work in his lab. Thanks a lot for all the support, the scientific freedom, the possibility to visit many conferences and the unique working atmosphere.

I would also like to thank Michiel Vermeulen for supervision and support. Working with you was both, scientifically exciting and personally a lot of fun Your enthusiasm and passion for science were a great source of motivation for me. The office was not the same anymore when you and the "jackpot dance" left for Utrecht.

I would like to thank my collaborators for inspiring and succesful projects, and a lot of scientific fun:
Sandra Hake and Clemens Boenisch for a very nice collaboration on histone chaperones.
Filomena Matarese, Hendrik Marks, Sergey Denissov and Henk Stunnenberg for the amazing work that let to our Cell paper.
Michiel Vermeulen and Nelleke Spruit for contributing to the organ specific chromatin reader project.
Christian Kelstrup for scientific discussions and help with the 96-well plate based peptide pull-down setup.
Anne Frohn and Gunter Meister for the partially painful but finally succesful Ago interaction project.

I would also like to thank my TAC committee, consisting of Matthias Mann, Michiel Vermeulen, Marc Schmidt-Supprian and Jürg Müller for support and discussions.

Thanks a lot to all my collegues in the department who generated a very pleasent working atmosphere.

Special thanks goes to the members of the *evil office* Falk, Tar and Dirk for many dis-

cussions, laughter and science nerd talk.

Also thanks to former office members Maxi, Michael and Annette, who were not as evil as the current team, but nevertheless great company.

Thanks to all staff members of the department who helped me in various ways: Korbi for making the instruments run. Bianca for help in cell culture. Steven for help with mouse work.

Special thanks goes also to Marco, Charo, Felix, Jürgen and Markus for scientific discussions.

Not forgetting former lab members, who helped me in various ways: Chris Kelstrup, Nina, Maxi, Boris, Johannes.

I would also like to thank my parents and my family for always supporting me, accepting that I sometimes forgot everything around me and always being there when I needed them.

Finally, thanks a lot to Anne. Your love and support made the last years despite all the stress from work a great time.

Curriculum Vitae

Personal details

Name   Hans Christian Eberl
Date of birth   August 25th 1982
Place of birth   Eggenfelden, Germany
Nationality   German

Publications

A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics
Eberl HC, Spruijt CG, Kelstrup CD, Vermeulen M, Mann M
Mol Cell. 2013 Jan 24;49(2):368-78.

RNA-dependent and -independent Argonaute2 protein interaction networks in mammalian cells
Frohn A, **Eberl HC**, Stöhr J, Glasmacher E, Rüdel S, Heissmeyer V, Mann M, Meister G
Mol Cell Proteomics. 2012 Nov;11(11):1442-56.

H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome destabilization.
Bönisch C, Schneider K, Pünzeler S, Wiedemann SM, Bielmeier C, Bocola M, **Eberl HC**, Kuegel W, Neumann J, Kremmer E, Leonhardt H, Mann M, Michaelis J, Schermelleh L and Hake SB.
Nucleic Acids Res. 2012 Jul;40(13):5951-64.

Quantitative proteomics for epigenetics.

Eberl HC, Mann M and Vermeulen M.
Chembiochem. 2011 Jan 24;12(2):224-34.

Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers.
Vermeulen M[*], **Eberl HC**[*], Matarese F[*], Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG and Mann M.
Cell. 2010 Sep 17;142(6):967-80.
[*] these authors contributed equally

Hsp90 is regulated by a switch point in the C-terminal domain.
Retzlaff M, Stahl M, **Eberl HC**, Lagleder S, Beck J, Kessler H and Buchner J.
EMBO Rep. 2009 Oct;10(10):1147-53.

Investigation of protein-tyrosine phosphatase 1B function by quantitative proteomics.
Mertins P, **Eberl HC**, Renkawitz J, Olsen JV, Tremblay ML, Mann M, Ullrich A and Daub H.
Mol Cell Proteomics. 2008 Sep;7(9):1763-77.

Conference attendence and presentations

Oral presentations

2011 **10th HUPO World Congress**, Geneva
Label free quantitative proteomics for the discovery of novel modification dependent histone binding factors

2010 **4th HEROIC meeting**, Paris
Proteomic analysis of POGZ, a domesticated transposable element

2009 **3rd HEROIC meeting**, Vienna
Proteomic analysis of Cbx7 in mouse embryonic stem cells

Poster presentations

2012 **Keystone meeting Proteomes and Interactomes**, Stockholm
A chromatin reader map derived from mouse tissues

2011 **EMBO Conference Series Chromatin and Epigenetics**, Heidelberg
Quantitative proteomics for the discovery of novel histone binding factors

2010 **Keystone meeting Molecular Basis for Chromatin Structure and Regulation**, Taos, New Mexico, USA
Quantitative proteomics for the discovery of novel histone binding factors
awarded with a Keystone Travel Award