PhD degree in Molecular Medicine

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Faculty of Medicine

Settore disciplinare: BIO/10

Establishment and optimization of the ChroP approach, combining ChIP and MS-based proteomics, for the characterization of the chromatome at distinct functional domains

Monica Soldi

IFOM-IEO Campus, Milan

Matricola n. R08407

Supervisor: Dr. Tiziana Bonaldi
IFOM-IEO Campus, Milan

Added co-Supervisor: Dr. Gioacchino Natoli
IFOM-IEO Campus, Milan

Anno accademico 2011-2012

To my family: Andrea and Paolo!

TABLE OF CONTENT

FIGURES and TABLE INDEX

LIST OF ABBREVIATIONS

ChIP	Chromatin immunoprecipitation
N-ChIP	Native chromatin immunoprecipitation
X-ChIP	Crosslinking chromatin immunoprecipitation
MS	Mass spectrometry
hPTM	Histone post-translational modification
H3K9me3	Histone H3 Lysine 9 tri-methylation
H3K4me3	Histone H3 Lysine 4 tri-methylation
ChIP-Seq	ChIP-Sequencing
SILAC	Stable isotope labeling of amino acids in cell culture
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
MALDI-TOF	Matrix assisted laser desorption ionization-time of flight
LC	Liquid chromatography
LTQ-FT-ICR	Linear ion trap-fourier transform ion cyclotron resonance
XIC	Extracted ion chromatogram
MS/MS	Tandem mass spectrometry
AUC	Area under the curve
DDR	DNA damage response
DSB	Double strand break
FDR	False discovery rate
WB	Western blot
ESI	Electrospray ionization
CID	Collision induced dissociation
me1	Mono-methylation
me2	Di-methylation
me3	Tri-methylation
ac	Acetylation

di-Ac	Di-acetylation
tri-Ac	Tri-acetylation
tetra-Ac	Tetra-acetylation
γ-H2A.X	Histone H2A.X lysine 139 phosphorylation
H2A.X Tyr142p	Histone H2A.X tyrosine 142 phosphorylation
ACN	Acetonitrile
APS	Ammonium persulfate
TEMED	Tetramethylethylenediamine
LPS	Lypopolysaccharides
PEN/STEP	Penicilin/Streptomycin
LDS	Lithium dodecyl sulfate
DMSO	Dimethyl sulfoxide

1. ABSTRACT

In eukaryotes the genetic information is stored in chromatin, a highly structured nucleoprotein complex that mediates the coordinated regulation of gene expression. The basic unit of chromatin is the nucleosome, which consists of 147bp of DNA wound around a histone octamer core containing two copies of the histones H2A, H2B, H3 and H4. Changes in chromatin structure, which do not involve the nucleotide sequence, can translate into transient or heritable adjustments in gene expression. Various mechanisms modulate chromatin states: among them a pivotal role is played by covalent histone post-translational modifications (hPTMs), for which the repertoires of combinations and positions are extremely varied. In addition to hPTM patterns, chromatin is characterized by the local enrichment of a distinct set of histone variants; binding proteins, including various ATP-dependent chromatin remodelling complexes; DNA methylation and differential nucleosome density. Together, these components contribute to the establishment of specific “chromatin landscapes”, defining the functional state of the genome in that territory.

Chromatin immunoprecipitation (ChIP) and Mass Spectrometry (MS) are complementary strategies to investigate the epigenetic components of chromatin. ChIP followed by deep sequencing (ChIP-Seq) allows genome-wide profiling of hPTMs and binders at individual genes and regulatory regions, up to a resolution of inividual nucleosomes. However, ChIP does not inform about the protein portion of chromatin, knowledge instead offered by MS-based proteomics. At the level of individual histones, MS enables to detect virtually all hPTMs in an unambiguous and unbiased fashion and to reveal interplays between them. Yet, MS analysis on bulk chromatin preparations limits the inspection of PTMs to a global view, with no information about their patterning in distinct functional regions. Nowadays, a global investigation of synergies between histone PTMs,

variants, and chromatin-associated proteins in a locus-specific manner remains a very attractive unachieved goal.

During the course of my PhD, I contributed in this direction developing and optimizing a global, quantitative proteomic strategy, named ChroP (*Chro**matin** P**roteomics*), for the analysis of the protein component of distinct chromatin regions, enriched by modified and preparative version of ChIP. I developed two ChroP protocols, which differ in the step of chromatin IP: the Native-ChIP (N-ChIP) was used to dissect histone PTM patterns whereas the Crosslinking ChIP (X-ChIP) was used in combination with SILAC-based interactomics to characterize proteins interacting with the domains of interest. I used the antibodies against tri-methylated Lysine 9 and Lysine 4 on histone H3 (H3K9me3 and H3K4me3) to enrich functionally distinct and non-overlapping chromatin regions from HeLa nuclei. High-resolution MS of the fractionated nucleosomes enabled a dissection of the domain-specific composition in terms of histone modifications, variants and non-histonic proteins, which we refer to as the *modificome* and *interactome*.

First of all, I observed the expected combinatorial enrichment of silent modifications in H3K9me3, and of active ones in H3K4me3. The accordance of my results with previous studies allows to validate the robustness of the approach and, with this confidence, I could investigate novel PTMs. Remarkably, ChroP exhibited a unique and peculiar strength in revealing PTMs associations not only at the intra-molecular level within H3, but also across the different core histones, within the same nucleosome.

The SILAC-based investigation of co-associated proteins revealed a number of histone variants and multi-protein complexes differentially enriched in the two functional territories. Some of them confirmed several previously described interactions, thereby validating our method. In addition, I identified numerous novel interactors, suggesting potential novel roles and regulating chromatin pathways for these proteins. A representative case was the variant H2A.X with the WICH chromatin-remodeling complex, accumulating in the H3K9me3 regions. I thus propose a *higher local density* model for

H2A.X in heterochromatin and provide evidence that this accumulation, together with the recruitment of WICH, represents an additional level of modulation of the DNA damage response (DDR) in this chromatin compartment.

The ChroP approach is relatively easy to setup, given the limited changes made to the conventional N- and X- ChIP protocols. Hence, ChroP emerges as a potential useful tool to dissect chromatin composition and understand how all the distinct protein components can act in a concerted manner to enforce a specific chromatin status.

2. INTRODUCTION

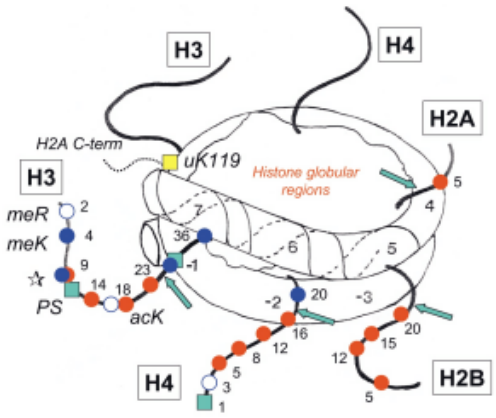
2.1 Chromatin, epigenetics and histone post-translational modifications

Chromatin is a highly ordered nucleoprotein complex that mediates both the DNA compaction into the eukaryotic nucleus and the regulation of gene expression. At the structural level, the basic unit of chromatin is the nucleosome, consisting of 147bp of DNA wound around an octamer core containing one histone H3-H4 tetramer and two histone H2A-H2B dimers (1, 2). Functionally, chromatin is organized into two distinct regions: euchromatin is less condensed and generally permissive for transcription, whereas heterochromatin is highly condensed and transcriptionally silent. Heterochromatin is classified as being either constitutive or facultative. In constitutive heterochromatin, the DNA remains condensed throughout the cell cycle. In facultative heterochromatin however the DNA can loose its condensed form and become transcriptionally active in response to distinct signals (3-5).

Changes in chromatin structure that do not involve the nucleotide sequence can translate into heritable adjustments of gene expression and thus constitute an “epigenetic memory” system of the cell (6-10). Epigenetic inheritance can be explained through a step-wise model proposing that “epigenator, initiator and maintainer” factors operate sequentially and synergistically to enforce and maintain specific functional states of the genome (11). The “epigenator”, a signal emanating from the external environment, is translated by an “initiator” into a specific chromatin/DNA functional state, which is sustained by a number of different “maintainer” factors. These include the methylation of cytosine in CpG islands (12, 13), covalent post-translational modifications of histones (hPTMs) and, in light of more recent studies, the activities of non-coding RNAs (ncRNA) (14, 15).

Among the epigenetic maintainers listed, histone PTMs are largely recognized as key regulators of chromatin structure and function. hPTMs include acetylation, ubiquitination and sumoylation of Lysines; different methylation degrees of Arginines and Lysines; phosphorylation of Serines, Threonines and Tyrosines; ADP-ribosylation of Arginines, Glutamic and Aspartic acids; deimination (or citrullination) of Arginine; Proline isomerization (Figure 1) (16-18), and in addition some less-characterized modifications.

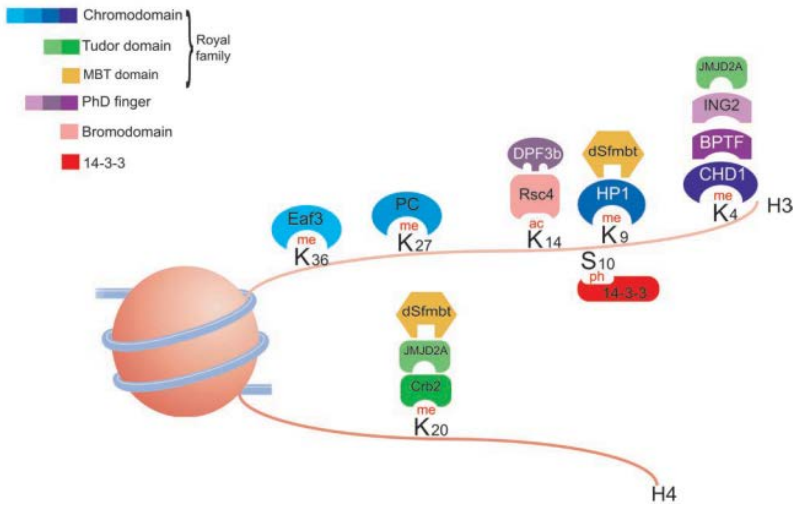
Figure 1. Histone post-translational modifications. The nucleosome core particle, with the N-terminal tail of core histone and the annotation of sites of post-translational modification. Numbers along the DNA indicate each complete helical turn on either side of the dyad axis. Sites marked by green arrows are susceptible to cutting by trypsin in intact nucleosomes. Most important modifications are Acetyl Lysine (acK); methyl Lysine (meK); methyl arginine (meR); phosphoryl serine (PS); ubiquitinated lysine (uK). *Adapted from Bennister A.J. and Kouzarides T. Cell Res 2011*.



The histone code hypothesis proposes that histone post-translational modifications act either singly or in combination to control distinct downstream pathways or processes on chromatin, ultimately defining the functional status of the underling DNA (19). The "letters" of this code are the modifications themselves, which are placed and removed by enzymes known as "writers" and "erasers", respectively. hPTMs exert their function on chromatin through two distinct mechanisms. In the first, higher orders of chromatin structure are altered via changes in inter-nucleosomal or histone-DNA interactions, thus

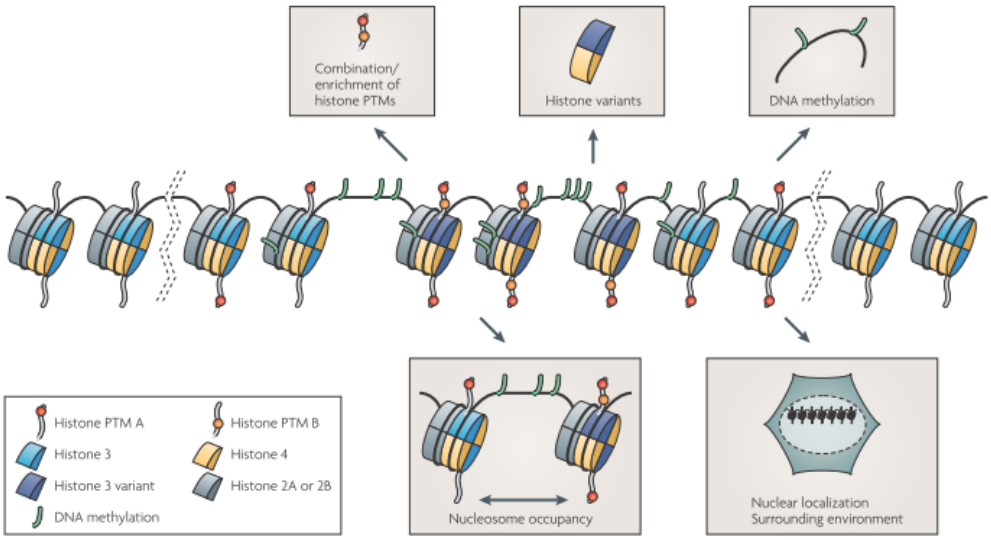
controlling the accessibility of DNA-binding proteins such as transcription factors (*cis* mechanisms). Alternatively, hPTMs can generate binding platforms for the recruitment of effector proteins containing specialized domains (*trans* mechanisms): the so-called "readers" of the code (Figure 2). The "readers" translate the information encoded by the modification patterns into specific biological outcomes (20-23).

Figure 2. Domains binding modified histones. Representation of some proteins with specific domains able to specifically bind modified histones. *Adapted from Kouzarides T. Cell 2007.*



In addition to hPTM patterns, chromatin is characterized by the local enrichment of a distinct set of histone variants; binding proteins, including various ATP-dependent chromatin remodelling complexes; and differential nucleosome density and position. Together, these components contribute to the establishment of specific "chromatin landscapes", defining the functional state of the genome in that territory (Figure 3) (24).

Figure 3. The distinct components contributing to define the functional state of chromatin domain. *Adapted from Margueron R. and Reinberg D. Nat Rev Genet 2010.*



Antibodies directed against specific hPTMs are traditionally used to study the language of histone modification in various assays. These include: immunofluorescence (IF) analyses of modifications at the single cell level; immunoblotting (WB), which allows profiling of PTMs in different samples and/or conditions; as well as chromatin immunoprecipitation (ChIP), which can be coupled to either PCR, DNA microarray (ChIP-on-chip) or deep sequencing (ChIP-Seq) for description of their local enrichments. The last two methods allow the genome-wide mapping of modifications, with a resolution of a few nucleosomes (25-27). Antibody-based assays are hampered by limitations in their specificity and efficiency when used to reveal the combinatorial aspect of the code. In fact, modifications can occur on adjacent or closely spaced residues within the same histone, making an epitope-masking effect more likely. To address this issue, a number of strategies have been developed to assess accurately the specificity of antibodies used in chromatin research. Peach *et al.* combine immunoprecipitation (IP) of native HPLC-purified H3 with mass spectrometry to detect PTMs co-enriched by a certain antibody on the same polypeptide. In

addition, Fuchs *et al.* have developed a peptide-array assay, based on a comprehensive library of modified peptides (28, 29).

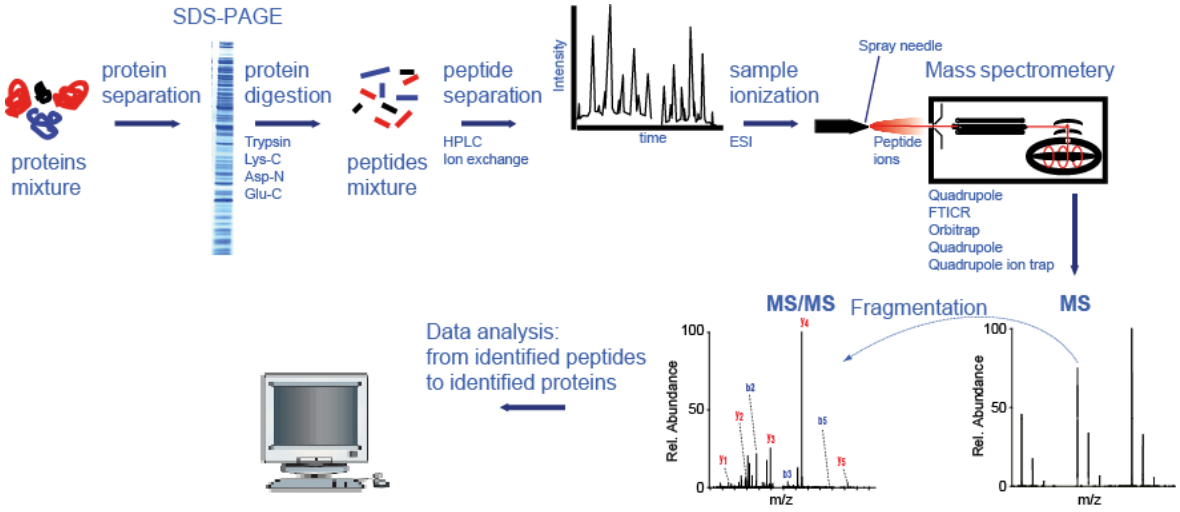
Mass spectrometry (MS) has emerged as a promising complementary analytical strategy to identify known and novel PTMs on proteins, as well as for the relative quantitation and detection of interactions between them (30). The recent advent of high-resolution mass spectrometry has increased the relevance of MS-based hPTM analysis by enabling the discrimination of near-isobaric modifications, either singly or in combinations, on very long polypeptides and even on intact histones (31-39). Finally, the use of different labeling strategies, both chemical and metabolic, has enabled the accurate quantitation of modifications, both in a relative and absolute manner (40).

The “epigenomics” and “chromatomics” fields share a common goal in studying chromatin structure, composition and features: to gain a comprehensive view, from genome to proteome, of the epigenetic phenomena underlying the establishment and inheritance of specific expression patterns (41, 42).

2.2 Mass Spectrometry analysis and MS based-proteomics

The steps of a typical proteomic experiment are shown in Figure 4. Briefly, after reducing the complexity of a protein preparation by electrophoretic speparation the proteins are subjected to enzymatic digestion, tipically using trypsin as protease. After MS analysis, by which the MS informations at the peptide level are obtained, the specific proteins are identified by software-assisted database searching, with which it is also possible to identify and localize PTMs within the peptide.

Figure 4. Proteomics experiment overview. A protein mixture is prepared from a biological sample and separated by SDS–PAGE. The generated peptides are separated by HPLC. Peptides are then ionized then analyzed by different mass spectrometers. Finally, the peptide-sequencing data that are obtained from the mass spectra are searched against protein databases using one of a number of database-searching programs.



Generally, the classical proteomic workflow for proteins identification is also applicable to the analysis of PTMs, although the analysis is inherently more difficult than simple protein/peptide identification for the following reasons:

1. PTMs of proteins are typically of low abundance. Hence high sensitivity of detection is required to identify the PTM and to assign its position within the peptide sequence.

2. PTMs are frequently labile: sometimes is difficult to maintain the peptide in its modified state during sample preparation and subsequent ionization in the mass spectrometer, since the covalent bond between the PTM and amino acid side chain in the peptide is typically labile.

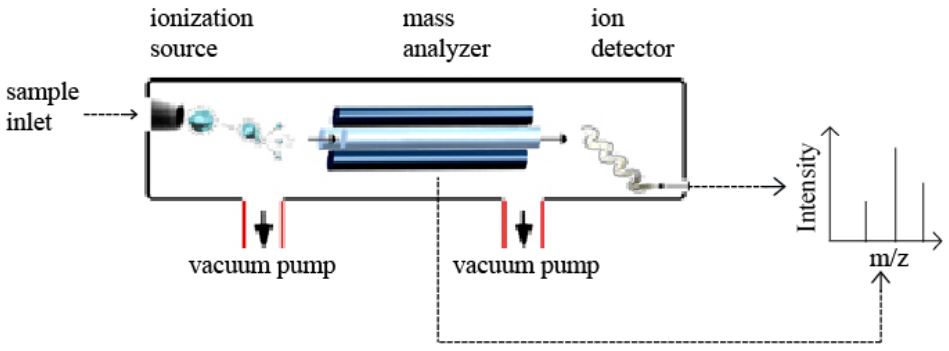
Before describing in detail the different MS approaches applied to in-depth investigation of histones and their PTMs, as well as non-histonic chromatin proteins, it is first necessary to provide an overview of the basic principles of mass spectrometry (43, 44).

2.2.1 Basic concepts of Mass Spectrometry

The mass spectrometer is a multistage instrument that measure the mass-to-charge ratio (m/z) of freely moving gas-phase ions in electric and/or magnetic fields. The elemental composition of the peptide can then be derived from his m/z ratio, whereas direct information about amino acid sequence cannot be deduced from m/z alone. To this aim, the desired peptide ion (precursor) needs first to be isolated and then fragmented, by a second cycle of MS, into its constituent amino acids. The fragments can be revealed in the MS/MS spectrum, providing unambiguous identification of the peptide sequence. Importantly, MS/MS enables the precise localization of modifications present on specific residues (45). A mass spectrometer consists of three main parts (Figure 5):

1. *an ion source*: it converts the peptides into gas-phase ions.
2. *a mass analyzer*: it separates the ions according to their mass/charge ratio (m/z).
3. *a detector*: it records the number of ions at each m/z value.

Figure 5. Schematic representation of the basic components of a mass spectrometer.



1. Ion Source and ionization of peptides

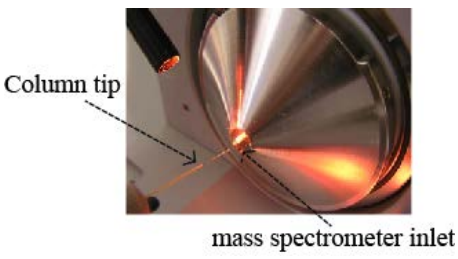
Proteins and peptides are polar, nonvolatile species that require an ionization method to transfer them into the gas phase, without extensive degradation. One of the most important developments in instrumentation has been the introduction of "soft-ionization" technology, which permits proteins and peptides to be analyzed by MS. Two techniques paved the way for the modern bench-top MS proteomics: Matrix-assisted laser desorption ionization (MALDI) (46, 47), and Electrospray ionization (ESI) (48). In a MALDI source, peptides are co-crystallized with a solid-phase matrix onto a metal plate. The matrix typically consists of a small organic molecule such as α-cyano-4-hydroxycinnamic acid or dihydrobenzoic acid (DHB). When laser pulses irradiate the resulting solid mixture, this absorbs the laser energy and transfers it to the acidified peptides. At the same time, the rapid heating causes desorption of both matrix and newly formed $[M+H]^+$ protonated peptides into the gas phase. Currently, MALDI ionization can support different types of mass analyzers, but the most common combination for proteomics studies is the MALDI/time-of-flight (TOF) setup (49). In recent mass analyzers, ions generated in the source are accelerated to a fixed amount of kinetic energy and travel down a flight tube. The small ions have a higher velocity and are recorded by a detector before the larger ones. The m/z value displayed in a TOF spectrum is proportional to the time, for a given analyte, required to reach the detector. Unlike MALDI, the ESI source produces ions from the solution. Peptides exist as ions in solution because they contain functional groups whose ionization is controlled by the pH of the solution. At acidic pH values, protonation of the amines will confer overall net positive charge to peptides and proteins, while at basic pH, de-protonation of the amines and carboxyl groups confers a more negative overall charge. Fragmentation of peptide ions is favored by positive charges on the peptide ions. For these reasons, ESI of peptides is most commonly done in the positive ion mode to analyze acidic samples.

Briefly, the ESI process consists of the formation of an electrically charged spray (Figure 6), driven by high voltage (2–6 kV), which then drives the desolvation of peptide/protein-solvent droplets. This process is aided by the high temperature provided by a heated capillary and, in some cases, by sheath gas flow at the mass spectrometer inlet. There are different theoretical models to describe ESI ion formation, however the important features are: formation of multiply charged species; sensitivity to analyte concentration and flow rate.

Figure 6. The process of electrospary ionization (ESI). The charged liquid exits the tip and forms a cone shape (known as a Taylor cone). After, the droplets burst away from each other into a fine spray.

Tipically, liquid chromatography (LC) instruments are usually coupled “on-line” with the ESI source to achieve continuous or high throughput analysis (Figure 7). After ionization, the peptide ions pass from the source into the mass analyzer where the ions are then separated in according to their mass/charge ratio.

Figure 7. Liquid chromatography directly coupled to the mass spectrometry (nanoLC-MS). After separation on a reverse-phase C18 nano-column, eluting peptides are directly electro-sprayed into the mass spectrometer.

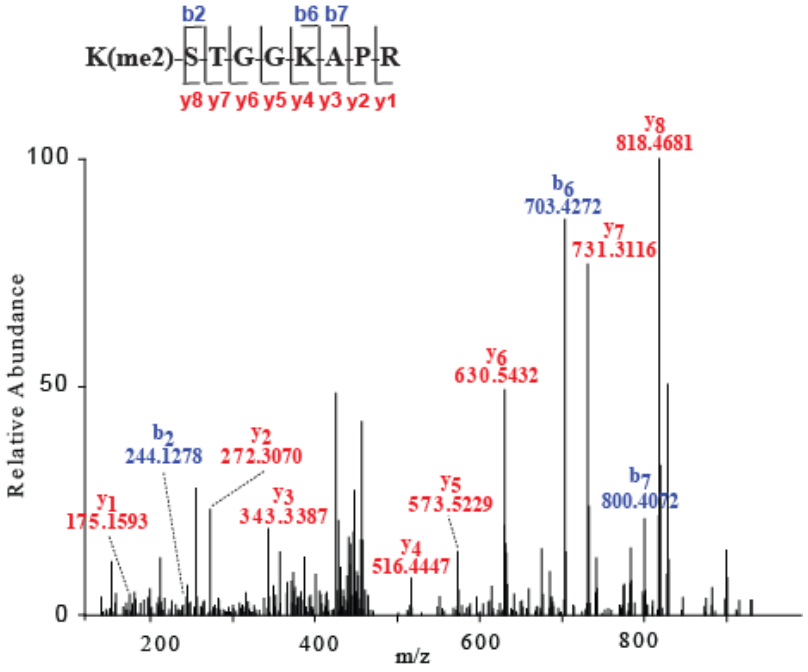


2. Mass Analyzer (and tandem mass analysis)

The mass analyzer is the heart of the instrument. Several types of tandem mass analyzers are commonly paired with ESI sources for proteomics work. These are: time of flight (TOF), quadrupole (Q), ion trap, orbitrap (OT) and Fourier transform-ion cyclotron resonance (FT-ICR). Although these mass analyzers differ in the details of how they work, they all perform the same type of analysis. From a mixture of peptide ions generated by the ion source, the tandem MS analyzers select a single m/z species. Collision-induced dissociation (CID) has been the most widely used MS/MS technique in proteomics research (50). Briefly, gas-phase peptide/protein cations are internally heated by multiple collisions with rare gas atoms. This leads to breakage of the C-N bond in the peptide backbone, resulting in a series of characteristic ions (b- and y- fragment ions). As a result of the slow-heating energetic aspect of this method, internal fragmentation and neutral-losses of water, ammonia and labile PTMs are common. This often results in limited sequence information for large peptides (>15 amino acids) and intact proteins.

This limitation has been addressed by the development of novel methods for ion-electron reactions to carry out peptide fragmentation: electron capture dissociation (ECD) and electron transfer dissociation (ETD) enable sequencing of larger peptides, providing an option to investigate combinatorial features of hPTMs (51-54). Both ECD and ETD are based on the transfer of electrons to the multi-protonated longer peptides (>2 kDa). In ECD, the electrons are generated from a heated filament, whereas in ETD they are transferred by gas-phase radical ions. Despite the similarity between the two techniques, ECD can be used only in combination with Fourier transform ion cyclotron resonance (FT-ICR) instruments, whereas ETD can be implemented in low-cost, high-capacity ion traps or new generation Orbitrap mass spectrometers and it has thus a wider applicability. The information contained in this tandem or MS-MS spectrum permits the sequence of the peptide to be deduced. Moreover, the nature and sequence location of peptide modification also can be established from an MS-MS spectrum (Figure 8).

Figure 8. Example of collision-induced dissociation (CID) MS/MS spectra. Representative MS/MS spectra using CID fragmentation. The b-ion and y-ion series allow to define the sequence of (9-17) peptide of histone H3 and to localize specifically the di-methylation on K9 residue.



3. Detector

A detector, as "Channeltron" or "Electron multiplier tubes (EMT)", is placed at the end of MS to record the number of ions at each m/z value. "Channeltron" is a dynode coated with semiconductor material. The ions that strike the inner walls cause the emission of electrons, which are recorded in a counting system. "EMT" consists in a series of dynodes that cause the same effect of "Channeltron", but it is able to amplify the signal of electric current until it is quantifiable. The final output is represented by data.RAW file.

Modern mass spectrometers provide high-quality data in combination with high MS/MS sequencing speed. Moreover, 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) represent other two key parameters. Nowadays, most mass spectrometers are so-

called hybrid instruments because they consist of a combination of two or more *m/z* separation devices of different types. Usually, they are built 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 50000 with matching mass accuracy. The dynamic ranges in single spectra are in the range of 1000 to 10000 for both types of instrument. A very recent 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 (55).

In the most common well-established analytical workflow, peptides are resolved over time via Reversed-Phase liquid chromatography (RP-LC) prior to MS analysis to achieve continuous or high throughput analysis. Tipically, peptide samples are separated based on their hydrophobicity using a long capillary column with a typical flow of ~200 nL/min (nanoflow). The column is packed with C-18 resin to capture the majority of the peptides that are loaded in acidified aqueous solution. A gradient of organic solvent (acetonitrile, ACN, usually) gradually elute the peptides, which are injected and analyzed by the mass spectrometers. For instance, Reverse Phase High-pressure Liquid Chromatography (RP-HPLC) has been widely adopted in proteomics to resolve very complex peptide mixtures prior to MS analysis (LC-MS), due to its high resolution, efficiency, reproducibility, and mobile phase compatibility with ESI. A further development of this technology is nano-ESI (56, 57), in which the flow rates are lowered to a nanoliter-per-minute regime to improve the method's sensitivity. Nano-ESI is compatible with capillary RP-HPLC columns (58) which allow users to perform higher sensitivity analyses than has been possible with standard analytical methods (59).

2.2.2 Different MS approaches for hPTMs analysis: from "Bottom Up" to "Top Down", via "Middle Down"

The "Bottom Up" approach is highly popular in proteomics studies for investigations of protein PTMs (30). It is a "peptide-centric" strategy, based on the enzymatic digestion of proteins into peptides prior to MS analysis (Figure 9). The "Bottom Up" approach has been demonstrated to detect known and novel modifications on histones, combining its sensitivity in detecting peptide m/z in full MS with its efficient MS/MS fragmentation *via* CID (31). The most common protease used in bottom up proteomics studies is trypsin, which cleaves at the C-terminal end of Arginine and Lysine residues (60). However, trypsin digestion is not ideal for the analysis of histones that are highly rich in these basic residues (expecially at the N-terminal regions, where the modifications accumulate), because the peptides produced are too short to be efficiently retained and separated in RP-HPLC and detected by mass spectrometer (61). Arg-C is a good alternative because of its unique specificity for the C-terminal region of Arginines, which produces peptides of optimal length for LC-MS, which are longer and easy-to-ionize (61, 62). In addition, Arg-C peptides are compatible for sequencing, as the C-terminal Arginine retains a positive charge, leading to a well-defined y-ion series (63-65). Alternatively, histones can be chemically derivatized using either propionic anhydride [$(C_3H_5O)_2O$] or deuterated acetic anhydride (D_6-acetic anhydride [$(CD_3CO)_2O$]), prior to trypsin digestion: these compounds alkylate Lysine residues preventing tryptic cleavages and resulting in an Arg-C-like digestion. The advantage of this approach is that it is possible to obtain the previously described benefits of an Arg-C-like digestion while using trypsin as the protease, which is well suited to in-gel digestion methods (66). The in-gel approach, commonly performed by SDS-PAGE, facilitates separation at the level of individual histone molecules (67).

An additional effect of derivatization is that it labels unmodified and mono-methylated Lysines with a deuterated acetyl moiety (showing a delta mass of 45.0294 Da) but does not react with di-methyl, tri-methyl and acetyl Lysines. This effect can be exploited to distinguish between isobaric modification-bearing peptides.

A limitation of the “Bottom Up” approach emerges when analyzing histone variants or combinations of histone modifications. In fact, the short tryptic and Arg-C-like peptides do not permit detection of simultaneously occurring, long-distance PTMs. Offline chromatography, to separate histone variants or differently modified versions of the same histone molecule prior to “Bottom Up” analysis, is one solution to this problem. For instance, the three mammalian variants of histone H3 (H3.1, H3.2 and H3.3) have the vast majority of their peptides in common after enzymatic digestion but the full-length proteins can be separated prior to digestion and LC-MS analysis using tap-tag purifications and/or RP-HPLC (68, 69). Alternatively, intact proteins or larger histone domains can be analyzed by mass spectrometry with the so-called “Top Down” and “Middle Down” strategies (Figure 9) (70, 71). Histones are basic proteins and, in the acidic conditions used in MS, they are typically highly charged and capable of producing multiply charged fragment ions in MS/MS. Consequently, non-ergodic fragmentation methods (72) such as ETD and ECD on high-resolution instruments (Orbitrap, FT-ICR) are feasible in “Top Down” analysis (53, 54). “Top Down” enables the user to distinguish between co-occurring histone variants and differently modified isoforms, with information about the relative abundances, thus providing a so-called “bird’s eye view” on the complete panel of histone isoforms present in a specific functional state (73). The approach however lacks the sensitivity of “Bottom Up” and, furthermore, the analysis of the spectra obtained is less straightforward. These two restraints have limited a broad application of this method so far, even though recent advances in online separation of intact proteins by ultra high-pressure (UPLC) liquid chromatography have made the approach more feasible. Further improvements in

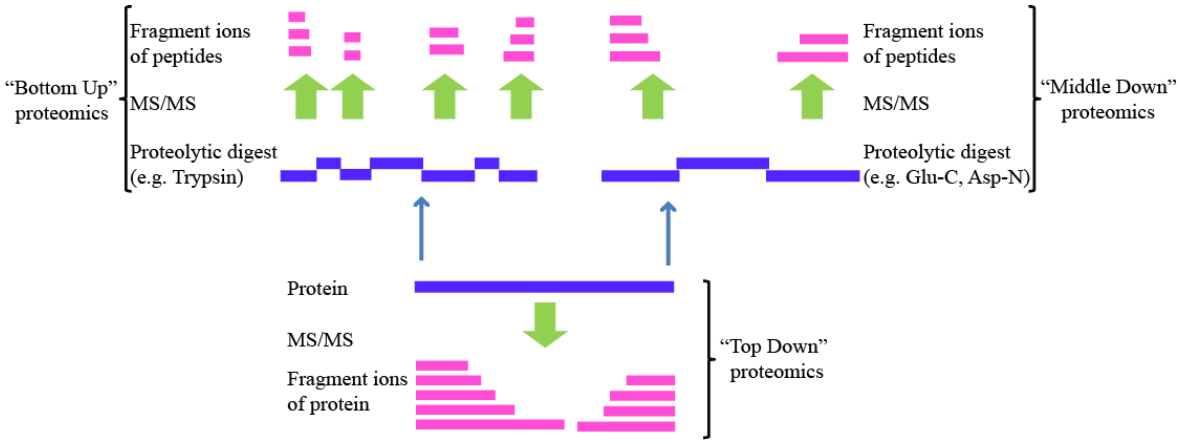
implementations are therefore still required to make “Top Down” analysis of intact histones, with variants and modified forms, a more routine approach (74, 75).

The “Middle Down” approach is an optimal compromise between “Top Down” and “Bottom Up” approaches, when the mass spectrometer is hyphenated to online liquid chromatography. In “Middle Down” approach, large histone peptides (>2 kDa) are analyzed upon the enzymatic digestion of histones with endoproteinases that have specificities to less frequently-occurring amino acids within histone sequences, such as Glu-C or Asp-N. In fact, since mammalian H3 contains the first Glutamic acid at position 50, Glu-C produces an N-terminal peptide (1-50) of 6 kDa that contains the majority of PTMs decorating this histone, as well as being suitable for MS analysis and sequencing by ETD MS/MS fragmentation. Similarly, Asp-N is useful for “Middle Down” analysis of histone H4, because it cleaves at the N-terminal side of Aspartic acid, present in position 24. Again, the resulting peptide (1-24) includes all modifications annotated for the H4 tail (32). The “Middle Down” approach therefore allows a more precise detection of PTM combinations on particular histone regions, especially when combined with pre-fractionation of the enzymatic digestion products. For instance, a combination of weak-cation exchange with hydrophilic interaction liquid chromatography (WCX-HILIC) prior to high-resolution MS, is a powerful analytical setup to resolve co-occurring and/or (near-) isobaric modified histone species (76), separating longer peptides first by their charge state and then by hydrophilicity. Based on this, Young *et al.* proposed a high-throughput approach using a gradient of decreasing organic solvent and decreasing pH on a commercial WCX-HILIC resin to separate and analyze by a “Middle Down” approach differentially modified histones domains (77) (See also paragraph 2.5).

In summary, an inconvenience of the “Top Down” and “Middle Down” approach is the need for specialized software to summarize the complex combinatorial networks existing among hPTMs. The main problems concern the complexity of the MS/MS spectra generated, either from intact histones or from large peptides, and the increased incidence of

internal peptide sequence fragments that further complicate the sequence annotation and consequently the PTM site-specific attribution in the MS/MS spectra (78-80). Improvements in computational approaches should enable more detailed comprehension and visualization of the inter-reliant relationships between unique modified forms.

Figure 9. Schematic representation of "peptide-centric" versus "protein-centric" MS analytical strategies.

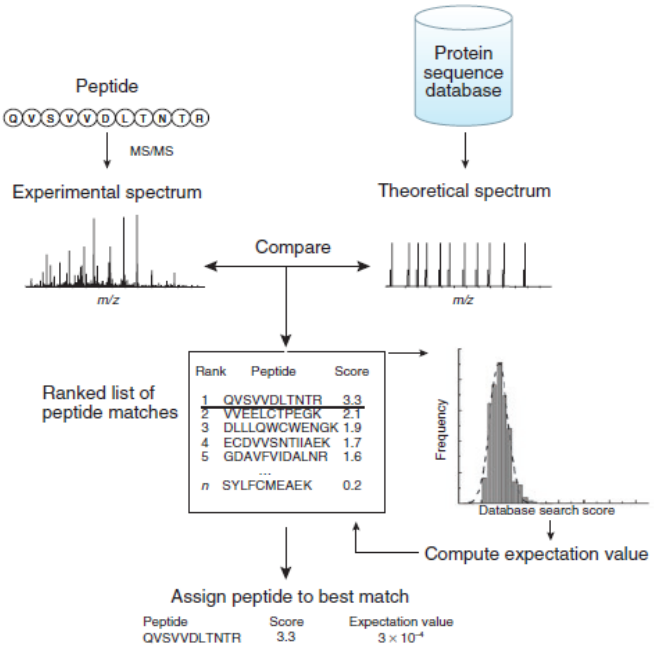


2.2.3 Data analysis and bioinformatics tools for hPTM analysis by MS

Efficient analysis of very large amounts of raw data is crucial in MS-based proteomics. Several MS/MS database search programs have been developed to interpret the large amount of data generated by modern mass spectrometers and their basic functionality is illustrated in Figure 10. Typically protein identification algorithms work as follows: they take the fragment ion spectrum of a peptide as input and score it against theoretical fragmentation patterns constructed for peptides from the searched database. The pool of candidate peptides is restricted based on user-specified criteria such as mass tolerance, proteolytic enzyme constraint and types of post-translational modification allowed. The output from the program is a list of fragment ions spectra matched to peptide sequences, ranked according to the search score. The search score measures the degree of

similarity between the experimental spectrum and the theoretical spectrum, and therefore serves as the primary discriminating parameter for separating correct from incorrect identifications.

Figure 10. Peptide identification by MS/MS database searching. *Adapted from Nesvizhskii A.I. et al. Nature Methods 2007.*



Of particular relevance for the analysis of the modifications that occur on histones are tools that enable identification of several different PTMs, often co-existing on the same peptide. Identification of PTM-bearing peptides in sequence databases, however, is more challenging than that of unmodified forms because the database search engine needs to take into account the diversity of modified forms that might exist. There are at present a number of computational methods available for the automated annotation of PTMs in peptides (Table 1). These methods analyze the MS and MS/MS data, taking into account the delta-mass values, and sometimes also neutral losses and other diagnostic ions for the PTM of interest (81).

Table 1. Software and search algorithms used to study hPTMs

Software	Freely Available	Reference
FindMod	+	Wilkins et al. (1999)
Mascot	-	Perkins et al. (1999)
MaxQuant	+	Cox et al. (2008)
Modificomb	+	Savitski et al. (2006)
OMSSA	+	Geer et al. (2004)
Phenyx	+	Colinge et al. (2004)
PILOT_PTM	+	Baliban et al. (2010)
ProSight PTM	+	Zamdborg et al. (2007)
ProteinProspector	+	Chalkley et al. (2008)
QuickMod	+	Ahrne et al. (2011)
SEQUEST	-	Eng et al. (1994)
VEMS	+	Matthiesen et al. (2005)
X!Tandem	+	Craig et al., 2004

The computational methods used to identify PTMs fall into two categories (31). In the first, the user selects a set of PTMs of interest prior to employ the bioinformatics tool for peptide and protein identification. This option is applied during the sequence database search, when PTMs are assigned to the relevant amino acids of a candidate peptide sequence. To limit the complexity required to search a very large set of possible modified forms, a restriction is usually imposed on the number of modifications that may be included in this search.

In the second approach, which is unbiased, PTMs are identified through a “blind” database search. In the initial step, a basic database search is performed, excluding the specification of PTMs of interest, but often specifying recurring/standard modifications such as oxidized Methionine, for example. The specification of this relatively common modification avoids false-positive PTM assignments later on. Once a set of peptides is identified in an MS/MS-based proteomics experiment the idea is that, since the PTM leads to a mass increment or deficit of the modified peptide relative to the form without the modification present, all unassigned MS/MS spectra can be searched to find those which might match a post-translationally modified form. The software therefore inspects

unassigned spectra, using information based on a list of known modifications such as delta-mass values and lists of predicted and observed peptide masses.

Computational methods that search for post-translational modifications are however associated with higher rates of false-positive identifications. The combinatorial issues associated with assigning the masses of included modifications can dramatically increase the number of peptide and protein candidates in the output. In this regard though, technological improvements that enable higher mass accuracies when generating the MS and/or MS/MS spectra have helped to address this issue (82).

High-resolution mass analyzers can resolve and identify peptides bearing modifications with similar delta-mass values as well as multiply charged ions in MS/MS spectra. Recent data analysis software therefore considers product ions with multiple charges either before or during database searching.

There are some constraints currently beyond the reach of current algorithms. The first is that some modifications may arise from *in vitro* artefacts rather than *in vivo* enzymatic activity. A well known example is the di-glycine (GG) tag which occurs on Lysine, and is used to determine ubiquitination sites: the elemental composition of this tag is identical to that of iodoacetamine, commonly used for the alkylation of Cysteines in standard shotgun MS proteomics workflows (83).

Another issue is that most of the available methods are sub-optimal for the analysis of MS/MS spectra deriving from long peptide sequences and intact proteins, which may result from a “Top Down” or “Middle Down” proteomics approach. As described in the recent review by Sidoli *et al.* (31), the complexity of these spectra require more specialised search algorithms, which can efficiently determine monoisoptic peaks, recognize ion charge states and deconvolute multiply-charged ion signals into singly-charged ion mass values. Currently, only a few software packages are available for this purpose.

2.3 Quantitation strategies in MS-based proteomics

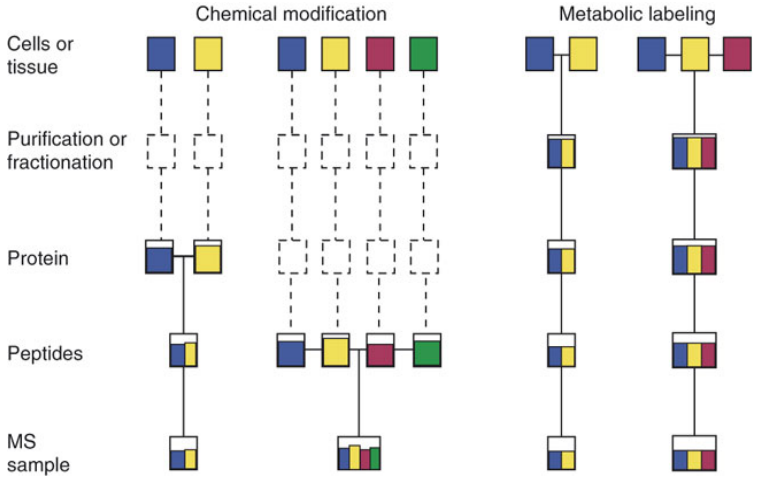
In the last few years, following the spread of mass spectrometry-based techniques, it became evident that qualitative identification of proteins is not sufficient for comparative analysis, which often require quantitative data. The direct comparison of the amount of a given protein across different samples is not often accurate because the intensity of MS and MS/MS peaks is poorly reproducible between different LC-MS runs. This is due to the variation in experimental conditions and the intrinsic variability of the selection of the precursor peptides in data dependent acquisition (DDA). Regarding the first factor, minimal variations in the LC system (e.g., buffer composition) or in sample preparation (e.g., salt or detergent content) can influence the retention time of the peptides and also the intensity and quality of the spectra. Minimal variations of the chromatographic run can affect the choice of the precursor ion to be fragmented, thus causing the acquisition of a MS/MS spectrum in one LC-MS run, but not in another. This phenomenon, called "undersampling", limits the possibility to profile a sample completely and principally affects low abundance peptides in complex mixtures.

Various strategies have been developed in MS-based proteomics for accurate protein quantitation, from single proteins up to global proteome profiling. These quantification strategies can be broadly divided in two groups. Label-free (LF) strategies use no labels, thus they can be used to compare the amount of proteins in different LC-MS runs without requiring particular sample handlings. Conversely, strategies based on isotope labeling require a specific preparation of the sample prior to LC-MS analysis. The rationale behind stable isotope labeling is to create a mass shift that distinguishes peptides deriving from different samples within a single MS analysis, thus avoiding run-to-run variations.

Generally, a labeled and an unlabeled sample are mixed to obtain a snapshot of concentration of proteins associated with different biological conditions. The signals deriving from the two samples can be distinguished due to a known mass shift between

labeled and unlabeled peptides. The ratios of the intensity of heavy and light signals allow the accurate quantification of the relative amounts of peptides and proteins originally present in the two samples. The isotope can be introduced in the peptides at different stages, depending on the labeling approach. It is possible distinguish two main strategies: metabolic labeling and chemical labeling (Figure 11).

Figure 11. Metabolic and chemical labeling methods. *Adapted from Ong S.E. and Mann M. MCP 2002.*



In the metabolic labeling, the isotope is added to growing cells as a metabolic precursor, to be incorporated uniformly in the proteome during protein biosynthesis. Major advantages of metabolic labeling techniques are that they can be applied to *in vivo* studies and they are compatible with complex purification procedures, since the samples can be mixed at the beginning of the workflow (even before the cell lysis) and handled together, thus minimizing the introduction of processing errors. Conversely, a clear drawback of these techniques is that it can be easily applied only to cultured cells.

Peptides can be chemically modified with an isotope labeled molecule that covalently binds the amino acids side chains or the peptide terminus. The labeling can be performed either before or after the proteolysis. The most common methods are isotope

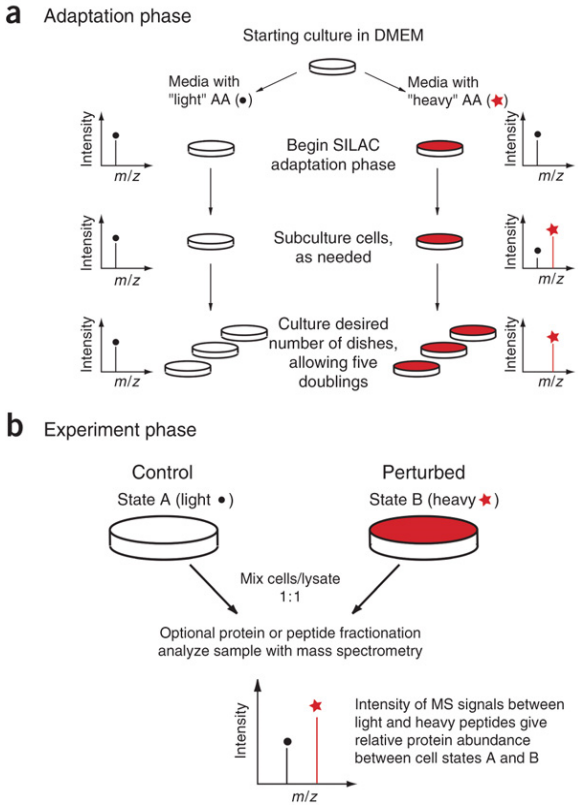
coded affinity tagging (ICAT) (84) and isobaric tags for relative and absolute quantitation (iTRAQ) (85). In ICAT, two samples to be compared are labeled with Cysteine-specific reagents containing an affinity biotin tag for purification and a differentially isotope-labeled chain displaying a known mass difference. Heavy- and light-labeled proteins are mixed and digested. Then, the tagged peptides are purified and analyzed by LC-MS. In iTRAQ technique, a reagent that reacts with the N-terminus and Lysine side-chains of peptides is used. Advantage of iTRAQ is represented by possibility of multiplexing, using up to eight different tags. This is of particular relevance to biological experiments in which multiple conditions or multiple time-points are being evaluated such as signalling networks. Overall, the quantification based on chemical derivatization can be applied to different types of sample (such as *ex-vivo* tissue and clinical biopsies) and is not limited to cultured cells, as metabolic labeling is. However, discrepancies in sample processing discrepancies can lead to the introduction of errors with these methods and side reactions can produce unwanted products. Moreover, since chemical labeling is often sequence-dependent, a complete incorporation of the isotope can be difficult to achieve, with consequent risk of biased quantification. One example of metabolic labeling strategy is represented by SILAC (stable isotope labeling with amino acids in cell culture, see next paragraph).

2.3.1 Stable isotope labeling with amino acids in cell culture (SILAC)

SILAC is simple, powerful and accurate procedure that can be used as a quantitative proteomic approach in any cell culture system (86). Since mammalian cells cannot synthesize a number of amino acid, therefore these “essential” amino acids must be supplied in cell culture medium as free amino acids for the medium to support cell growth. Isotopically labeled analogs of these amino acids, called “heavy” form, can be synthesized and are available commercially. Isotopes are variants of atoms of a particular chemical element, which have different numbers of neutrons. The number of protons and neutrons in

the nucleus, known as the mass number, is not the same for two isotopes of any element. For example, carbon-12, carbon-13 and carbon-14 are three isotopes of the element carbon with mass numbers 12, 13 and 14 respectively. If the “heavy” isotope of an amino acid is supplied instead of the natural (or “light” form) amino acid, it will be incorporated into each newly synthesized protein chain. After a certain number of cell doubling, this particular amino acid will be replaced by its isotopically labeled analog. Hence, the SILAC strategy requires that two populations of cells are grown in two separate medium formulations, the “light” medium containing the amino acid with the natural isotope abundance and the “heavy” medium containing non-radioactive stable isotope chose (Figure 12). When light and heavy cell populations are mixed, they remain distinguishable by MS, and protein abundances are determined from the relative MS signal intensities (86).

Figure 12. Overview of standard SILAC experiment. The SILAC experiment consists in two distinct phases: (a) an adaptation and (b) experiment phase. (a) Cell are grown in “light” and “heavy” media until fully incorporation (red star) (b) The two cell populations are mixed, protein are purified, digested and analyzed by MS. *Adapted from Ong SE, Mann M. MCP 2002.*



The possibility offered by this strategy to combine two cell populations from distinct media at a very early stage of the MS-proteomics workflow, significantly reduces the effects of experimental variations in sample preparation, thus leading to very accurate quantitation, which only takes into account changes caused by the different functional states.

The first SILAC experiments used $^{2}H_{3}$-Leu; the current protocols instead use labeled Arginine and Lysine (e.g. $^{13}C_{6}$-Lysine and $^{13}C_{6}$-Arginine), which offer the advantage to produce, upon digestion with Trypsin, most peptides containing a labeled amino acid. Typically, a cell line is cultured for at least 5 rounds of divisions in labeling medium, to achieve extensive proteins labeling (>95%): this procedure does not interfere with protein activity. Labeled and unlabeled cell lines or extracts are then mixed and analyzed together by LC-MS. SILAC has become a very popular technique and has been applied to the study of protein interactions (87, 88) and signaling networks (89, 90) but also in the investigation of post-translational modifications.

2.4 Quantitative MS-based approaches in epigenetic research

Different quantification strategies have been employed to the measurement of histone modification, variants and turnover. Chemical derivatization as a means to modify cleavable residues has been widely applied in epigenetic studies for their technical advantages, previously described (63, 64). In addition, the alkylation of Lysines with the deuterated acetic anhydride can also be used to quantitatively estimate the acetylation status of histones. For instance, distinct acetylated forms of H4 in *Drosophila melanogaster* and their developmental changes have been profiled using D_{6}-acetic anhydride prior digestion and MS-analysis (61). Similarly, propionylation of histones was used to observe the effect of G9a/Glp1 methyltransferase knockdown on global histone modifications (39).

Other chemical derivatization strategies, such as TMT (tandem mass tag) and iTRAQ, have only been employed on chromatin for protein-level profiling, with focus on PTM level changes (91-93).

In vivo metabolic labeling with isotope-encoded amino acids has emerged as the most powerful approach to accurately quantify changes of histones and their PTMs. In the last years, SILAC has gained wide popularity in proteomics and, more recently, also in chromatin studies (35, 94-96). SILAC is preferentially used to profile protein levels; however it has also been successfully applied to identify and quantify hPTMs, and in particular to profile modification dynamics during the cell cycle: Bonenfant *et al* showed increasing phosphorylation on histone H3 and H4 and decreasing methylation of H3K27/K36 during mitosis (97) while Pesavento *et al.* proved that H4K20 methylation degree was tightly linked to cell cycle progression (99). A SILAC-indipendent approach is used by Scharf *et al.* to demonstrate that H4K20 mono-methyaltion facilitates chromatin maturation ((98, 99). Using a SILAC MS-based experiment, Jung *et al.* showed that Polycomb repressive complex Suz-12 promotes the establishment of H3K27 di/tri-methylation in mouse embryonic stem cells, with a functional interplay between H3K27 tri-methylation and H3K27 acetylation, functioning as molecular switch in this system (100). Our group also used a modified version of the SILAC approach to determine breast cancer-specific histone PTM signatures. In this study, we focused on human breast cancer and comprehensively analyzed PTMs on histone H3 and H4 from four cancer cell lines (MCF7, MDA-MB231, MDA-MB453 and T-47D), in comparison with normal epithelial breast cells (MCF10). The SILAC-MS based approach enabled to quantitatively track the modification changes in cancer cells, as compared to their normal counterpart. With the accuracy of this strategy, it was possible to identify PTMs specifically associated to distinct type of breast cancer cell line with different properties (aggressiveness/prognosis). Among them some were already known as modifications linked to cancer, such as a

decrease of H4K20 tri-methylation, whereas some emerged as novel markers of breast cancer, such as decreased levels of H3K9 tri-methylation (101).

SILAC labeling in a pulse experiments were used to probe the turnover of both hPTMs and histone variants: Zee *et al.* showed that H2A.Z has higher turnover rates than canonical H2A variants and, more generally, that acetylated histone peptides appear to turn-over much faster than methylated ones (102).

Heavy-methyl SILAC (hmSILAC) is a variation of SILAC used for high confidence identification of protein methylation at Lysines and Arginines. In heavy methyl SILAC labeling, ^{13}CD3-methionine is added to Methionine-depleted media; upon uptake in the cell, the "heavy" Methionine is converted into *S*-adenosyl Methionine (SAM), the sole donor of methyl groups in enzymatic methylation reaction. As such, histone and all non-histonic proteins that contain methylations are enzymatically heavy-methyl labeled. Such isotopically methylated peptides are then identified with high confidence in MS, based on the presence of the specific 'light and heavy peak pair', markers of methylation, and subsequently quantified. Ong *et al.* first used this strategy to identify unambiguously methylated sites *in vivo* on both histones and non-histonic proteins (103). Afterwards, hmSILAC was applied to study the dynamic turnover at H3K9 tri-methylation in pericentric chromatin (103, 104). More recently, the same approach was applied to profile the dynamic turnover of histone Lysine methylation, revealing that mono-, di-, and tri-methylated residues generally have progressively slower rates of formation. Furthermore, methylations associated with active genes were found to have faster rates than methylations associated with silent genes (105).

A combination of both standard and heavy-methyl SILAC in pulse-chase experiments, carried out on synchronized cells, enabled Sweet *et al.* to track the progression of H3K79 methylations throughout the cell cycle (106). In addition, it was observed that H3K79 mono-methylations from newly-synthesized histone H3 proteins

have the same turnover rates as those in pre-existing histones, with no differences among the three histone H3 variants (106).

Label-free or ion intensity-based quantitation strategies have been applied in a few studies to profile differently modified, but isobaric histone isoforms, which have a special feature to present identical molecular weight/mass (isobars) but different PTMs configurations, so they are undistinguishable in full MS and can be hardly separated by standard LC. Since in MS/MS such isobaric species are distinguishable based on the positional selectivity of ion fragmentation, a relative quantitation is possible in a label-free MS/MS-based manner, using the relative ratios of their fragment ions. “Top Down” intact histone protein analyses were successfully used to quantify different modified forms of H3.2 and H4, in a label-free approach (107, 108) (See also paragraph 2.5).

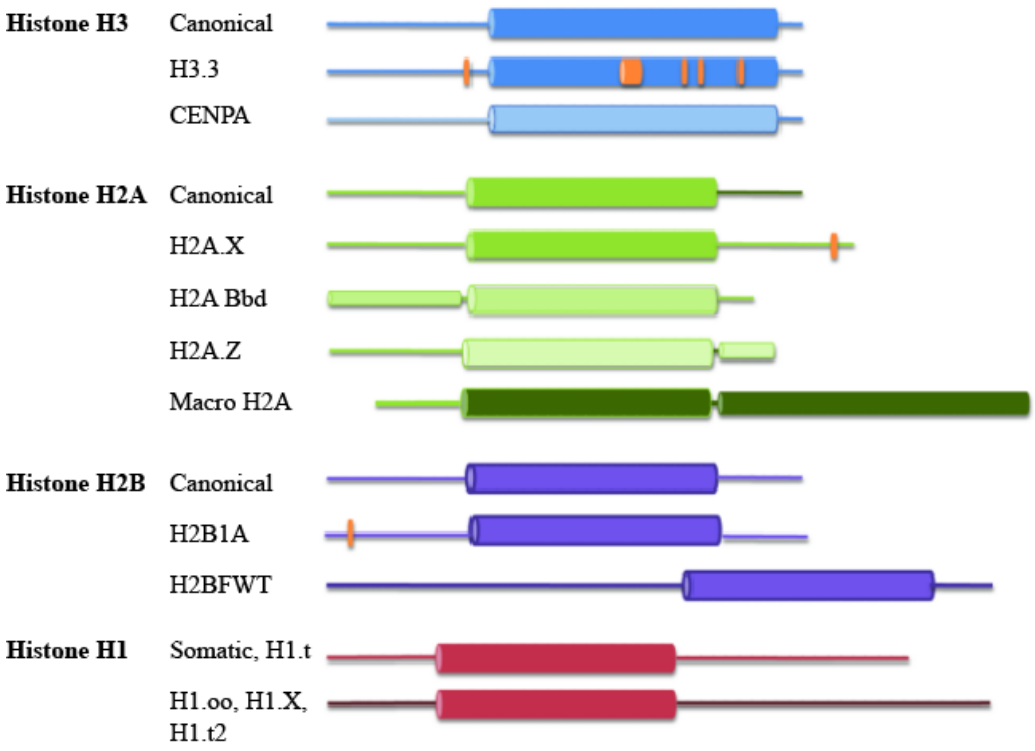
Lastly, synthetic, isotopically labeled peptides can be used as internal standards for both relative and absolute quantitation of histones and their PTMs, in “spike in” assays. Briefly, isotope-encoded peptides are synthesized with the same sequence of the modified histone peptide of interest, derived from the endoproteinase digestion used in the study. Relative quantitation is obtained when a known concentration of the standard peptide is “spiked into” each histone sample from the panel under investigation, and the intensity of the each native modified peptide is compared with that of the standard. With the same approach the absolute quantitation of modified peptides can be also achieved, when a calibration curve of the ion intensity *versus* the peptide standard, injected at distinct concentrations, is calculated. Typically, this approach is combined with single or multiple reaction monitoring MS (SRM/MRM), which allows a very sensitive detection of even sub-stoichiometric modifications. This technique benefits from the triple quadropole (QQQ) instrumentation. Briefly, targeted peptides are selected in the first mass analyzer (Q1), fragmented by CID (in Q2) and one or several of the fragment ions uniquely derived from the targeted peptide are measured by the third analyzer (Q3). In this way, each peptide is characterized by a specific “transition” which links both the precursor and

fragment ions, observed in both analyzers. The identity of each peptide can be inferred from the "transition" and the relative abundance can be estimated from the transition intensity relative to that of the standard (109). Darwanto and coworkers have successfully employed SRM upon spike in of isotopically encoded histone peptides in U937 lymphoma cells, expressing a mutated form of the hDot1a methyltransferase. They profiled changes in a set of hPTMs and observing that in these conditions the observed decrease of H3K79 methylation parallels a corresponding increase in H2B K120 ubiquitination (110).

2.5 Mass Spectrometry analysis of histone variants and their modifications

In addition to post-translational modifications, histone variants contribute to the epigenetic regulation of gene expression (111). Histone variants typically accumulate at specific genomic regions and show unique modification patterns, affecting a variety of chromatin-related processes. Some interpretative models propose that they represent an "extra layer" of the histone code (112), providing additional mechanisms to modulate chromatin structure. However, at least for the majority of variants, the processes by which specific variants accumulate at certain regions and are transmitted throughout the cell cycle remain unclear. Except for H4, all core histones and linker histones H1 have a number of variant counterparts (Figure 13), often differing in a few amino acids, which hampers their analysis via conventional approaches, such as antibody-based assays.

Figure 13. Comparison between histone variant sequences. Histone variants contain a highly conserved histone fold domain and differ mainly in their C and N-terminal sequences. Boxes represent the histone fold domain and orange lines represent site-specific sequence variations. Histones that are in different shades of the same color are from the same histone family but have large differences in sequence. *Adapted from Arnaudo A.M et al. Crit Rev Biochem Mol Biol. 2011 (113).*



Mammalian histone H3 has three major variants (H3.1, H3.2 and H3.3), in addition to a testis-specific variant (H3t) and a centromeric variant (CENP-A). The major variants are very similar in sequence composition. Histone H3.1 differs from H3.2 by a change in Cysteine 96 to Serine, while H3.3 differs from H3.1 by only 5 residues. However, they display differences in their expression, enrichment at specific chromatin domains, and in their post-translational modification signatures. Studies of the PTM patterns of H3 variants have been performed, profiting from all MS approaches described: “Bottom Up”, “Top Down” and “Middle Down”. “Bottom Up” analysis of mammalian, *Arabidopsis thaliana*, and *Drosophila melanogaster* H3 variants revealed that H3.3 is enriched in modifications associated with transcriptional activity (114-116). “Top Down” analysis of H3 variants

from rat brains showed comparable results using this complementary approach (117). Affinity purification of epitope-tagged H3.1 and H3.3 revealed a distinct set of modifications occurring on these two H3 variants before and after their assembly on chromatin, suggesting that pre-assembly modifications determine their final fate, as well as their PTM patterns on chromatin (68). A combinatorial view of modifications on H3.1 and H3.3 from asynchronous or colchicine-treated HeLa was achieved by "Top Down" revealing that, in asynchronous cells, only 5% of K4 was mono-methylated and about 50% of K9 was di-methylated in the H3.1 pool. In addition, more than 90% of the H3.1 pool was observed to be acetylated: K14 and K23 represent the major sites of acetylation. Upon colchicine treatment however the unmodified, mono- and di- phosphorylated S10 and S28 are detected in a 2:3:1 ratio, in addition to the K9 methylation and acetylations described. The absence of the K4 methylation in the colchicine-treated samples was probably due to the relatively small pool of molecules containing this modification (73). "Middle Down" analysis of H3 variants in a panel of rat tissues showed distinct patterns of H3.2 and H3.3 levels and modification status between various tissues (118). "Middle Down" was also successfully applied to the identification of more than 200 modifications in H3.2 and 70 modifications in H4 from human samples, including several that were not previously reported (77, 108).

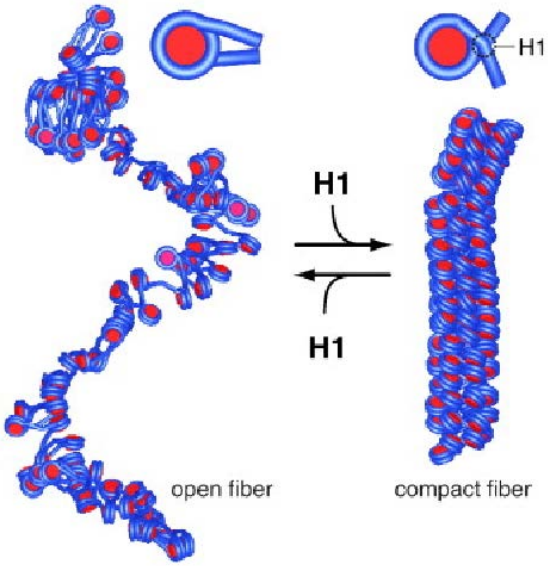
Canonical human histone H2A is encoded by sixteen genes that cluster on the genome. Kelleher and co-workers identified and characterized twelve unique sequences by using intact mass and fragmentation spectra (71). The modifications on the canonical H2A are incompletely characterized: only phosphorylation of S1 and acetylation on the N-terminal K5 are reproducibly reported (119), as well as mono-ubiquitination at K119, involved in gene silencing and mediated by Polycomb proteins (120). The non-canonical H2A variants include H2A.X, H2A.Bbd, H2A.Z and macro-H2A. H2A.X phosphorylated at S139 is the so-called γ-H2A.X, which localizes to sites of DNA double strand breaks (DSB) in response to DNA damage and thus represents a mark of the DNA damage

response (DDR). Acetylation and ubiquitination of H2A.X were also shown to be involved in this process: acetylation of K5 is a prerequisite for the poly-ubiquitination and the subsequent release of H2A.X from the DNA damage sites (121). H2A.Z is present at promoters where it is believed to maintain active chromatin in regions adjacent to silent ones. However, potential roles in gene silencing have also been proposed (122). Acetylation of K4 and K7 of this variant were identified by a "Middle Down" approach in Jurkat cells (119). Macro-H2A, the largest H2A variant, is generally enriched at transcriptionally silent regions. MS characterization of macro-H2A identified K115 ubiquitination and S137 phosphorylation. The former is implicated in X-inactivation whereas the latter is enriched in mitosis (123, 124). In addition, K17 mono-methylation, K122 di-methylation and Y128 phosphorylation are identified (123).

A combination of CID and ECD MS fragmentation at protein and peptide levels led to the characterization of several H2B variants and associated PTMs (119, 125): acetylation on K5, K12, K15 and K20, and ubiquitination on K120. These PTMs were confirmed by peptide mass fingerprinting (PMT), performed on bovine H2B, which revealed also K43 mono-methylation and K85 acetylation (126). "Bottom Up" approaches have also served to characterize modifications specific for the testis-specific variants of H2B (TH2B) (127). In addition, "Top Down" analysis using ECD fragmentation of the two major H2B variants of *Tetrahymena thermophila* led to the characterization of their primary sequences and modification patterns (128). Recently, mono-methylation and di-methylation at the N-terminal Proline of *Drosophila melanogaster* H2B have been identified using a combination of different strategies for sample preparation prior to MS analysis including D_6-acetic anhydride derivatization followed by Trypsin digestion and Asp-N digestion. The abundance of this Proline methylation seems be dependent of the developmental stage and is regulated by the enzyme dART8. The authors also observed predominant acetylation of H2B at K11 and K17 (129).

Histone H1 is commonly referred to as the linker histone. A single copy of this histone is proposed to bind near the entry/exit site of DNA on the nucleosome (the so called dyad), stabilizing the 30nm fiber and thus regulating higher order chromatin structure and stability (Figure 14) (130). Sequence divergence between histone H1 isoforms occurs mainly in the N- and C-terminal regions of the proteins, generating as many as eleven mammalian isoforms.

Figure 14. Binding of linker histone H1 to the nucleosomal string induces chromatin compaction. Chromatin fiber compaction induced by binding of linker histone H1. The left fiber represents an open conformation with straight linker DNA in the absence of linker histones, in which DNA access is facilitated for other proteins. The binding of linker histone H1 changes the local nucleosome geometry. This induces a transition to more compacted fiber conformations in which the DNA is less accessible. *Adapted from Wachsmuth M. et al. Biochimica et Biophysica Acta 2008.*



Mass spectrometry analysis, in combination with other experimental techniques, led to the identification of a number of PTMs specifically enriched on distinct linker histones, such as methylation, phosphorylation, acetylation, ubiquitination, formylation and ADP ribosylation (131-135). RP-HPLC of the different H1 variants, followed by chemical

derivatization of the protein with propionic anhydride and subsequent LC-MS/MS analysis revealed a K26 methylation and S27 phosphorylation on histone H1.4. Methylation on K26 appears to recruit heterochromatin protein1 (HP1), whereas phosphorylation at S27 appears to inhibit HP1 binding, so that these two adjacent PTMs are believed to function as a molecular switch for the modulation of gene expression silencing (133, 136). Moreover, the "Top Down" analysis of intact H1.2 and H1.4 molecules purified at distinct cell cycle stages suggested that S173 on H1.2 and S187 on H1.4 are phosphorylated only during interphase. Interphase phosphorylated H1.2 and H1.4 are associated with active rDNA and facilitate RNA pol I transcription. Finally, phosphorylation of H1 affects its chromatin dissociation and, in turn, chromatin accessibility to factors that regulate transcription and replication (137).

2.6 Interaction proteomics to study chromatin architecture

An improved knowledge of chromatin composition can contribute to a more comprehensive view of its higher-order structure and function. Until now, no purification method has emerged as a "gold standard" for chromatin purification and characterization due to the difficulty in enriching chromatin samples from specific functional regions in a quantity and purity sufficient for subsequent analysis. However, thanks to recent improvements in sensitivity and accuracy of MS-based quantitative proteomics a number of studies have demonstrated the high potential of this technology to characterize the chromatin proteome, with a specific focus on the histone code readers associated with specific functional states of chromatin (Figure 15).

The first attempt to characterize proteins associated with chromatin was the analysis of changes in protein levels in response to the overexpression of the oncoprotein MYC. This was done using differential detergent/salt extraction and chemical isotopic labeling by ICAT, in combination with multi-dimensional chromatography and mass

spectrometry (138). Subsequently, when *ad hoc* biochemical protocols were established for the purification of distinct chromosomes, MS proved to be a successful tool to characterize their protein composition: mitotic chromosomes were purified at different stages of the cell cycle (mitosis, metaphase and interphase) and co-associated non-histone proteins were characterized by MS (139-144). More recently, a multiclassifier combinatorial proteomics (MCCP) approach was developed where SILAC quantitative proteomics is integrated with a bioinformatics analysis pipeline. A statistical approach is applied to confirm which known and uncharacterized proteins are chromosomal, to obtain a more comprehensive collection of proteins associated at high confidence with mitotic chromosomes (145).

One elegant methodology to study the proteomic composition of telomeric regions was developed by the Kingston group using the PICh (*proteomics* of *isolated chromatin*) approach. In this approach, enrichment of cross-linked telomeric chromatin was achieved using DNA probes complementary to the telomeres, rich in repetitive sequences. The co-enriched proteins were characterized by MS and new telomere-associated proteins were observed (146). Yet, a drawback of PICh was the limited applicability to regions rich in repetitive DNA sequences.

All these methods provide a useful contribution to the knowledge of protein composition in large chromosomal regions or even intact chromosomes, but they are inadequate for obtaining information on chromatin locus-specific composition.

Recently, a number of interactomics assays combining affinity-interaction mapping with SILAC-quantitative MS read-out have been developed for the comprehensive characterization of hPTM "readers". Vermeulen *et al.* used pull-down assays with peptides that differ by a single post-translational modification to identify specific binders, either as individual interactors or as multiprotein complexes. Whit such approach, they discovered that TFIID binds H3K4 tri-methylation and recruits the entire transcription initiation complex, thereby providing a functional link between this modification and activation of transcription (88). The approach was extended further to screen all major tri-methylation

marks on histones and, in combination with ChIP-Seq and BAC-GFP pull-downs, to define the comprehensive Lysine trimethyl-interactome (147). As an additional elaboration of the strategy, a SNAP (*SILAC Nucleosome Affinity Purification*) approach was established where recombinant nucleosomes bearing combinations of hPTMs and methylated DNA were used as baits to provide a "modification binding profile" for proteins regulated by the contribution of both DNA and histone methylations (148). Similarly, a SILAC-based affinity purification assay was carried out with recombinant, uniformly modified chromatin templates (149). In addition, the CLASPI (*Cross-Linking Assisted and SILAC-based Protein Identification*) approach has been described, which combines SILAC with chemical proteomics using photo-crosslinking-based histone peptide probes, to detect weak but specific interactions that may escape standard pull-down approaches (150). Finally, peptide arrays and MS have been employed to systematically uncover methyl-Lysine and chromatin-binding module interactors, as well as to identify novel H3K23 mono-methylation marks, able to facilitate the recruitment of HP1β to the heterochromatin (151).

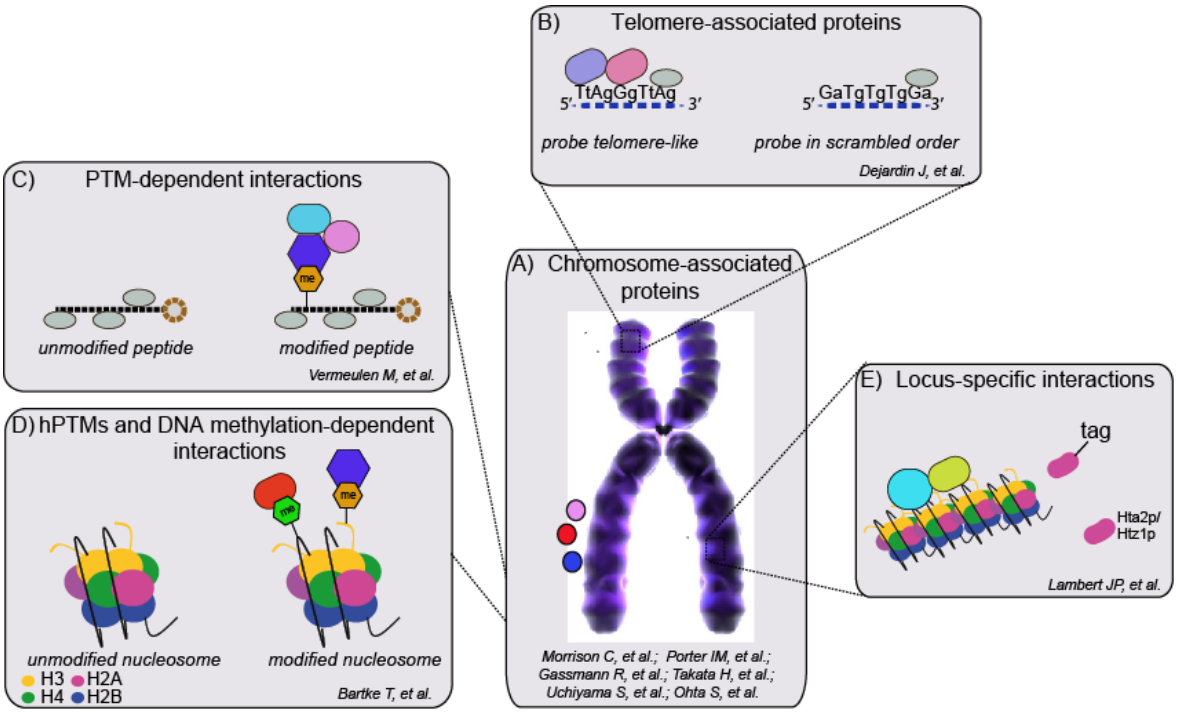
These *in vitro* studies are very powerful tools for screening the soluble binders of hPTMs, but fall short in extracting information on the relative PTM stoichiometry, combinations as well as their synergies with histone variants and chromatin modifiers, under physiological conditions. Hence, the locus-specific determination of hPTM patterns and their interactions with protein complexes remains a very attractive, partially unachieved goal.

A SILAC-based quantitative proteomics approach was employed to generate a differential profile of proteins associated with specific euchromatin and heterochromatin regions. This approach exploited the different accessibility of these regions to MNase, as a consequence of their differential nucleosome packaging. Upon limited MNase treatment, the two fractions of chromatin were separated by centrifugation, based on the differential density of the nucleosomal stretches, and SILAC was used to discriminate the proteins associated with these two functional chromatin regions (152). Another approach developed

for detection and characterization of proteins associated with specific chromatin domains was mChIP (153), where chromatin is isolated, sheared and then analysed by MS. mChIP was successfully applied to study the interactomes of H2A (Hta2p) and its variant Htz1p in *Saccharomyces cerevisiae*. However, this study did not provide quantitative information on binding proteins.

In this respect, global investigations of synergies between histone PTMs, variants, and chromatin-associated proteins in a locus-specific manner still remain a very attractive unachieved goal.

Figure 15. Different biochemical approaches for the proteomic characterization of chromatin composition and architecture. Some strategies address the chromosome as a whole (A), whereas others are focused on the characterization of specific chromatin regions: telomeres (B); regions enriched with distinct hPTMs (C, D), or certain histone variants (E).



3. AIM OF THE PROJECT

Chromatin architecture and its functional state are regulated by different associated proteins, post-translational modification of histone (hPTMs) and DNA methylation, which act in a concerted manner to create a “chromatin landscape”, with regulatory effect on gene expression. While extensive ChIP-Seq data are already available for all virtually known hPTM, and for a large set of chromatin binders, the investigation of the proteomic composition of chromatin started more recently and is still an ongoing effort. Recent studies focusing on protein components of chromatin were mainly focused on the characterization of histone code readers. These *in vitro* studies are very powerful tools for screening the soluble binders of hPTMs, but information on the relative PTM stoichiometry, combinations as well as their synergies with histone variants and chromatin modifiers remain still largely incomplete.

With these restraints in mind, and since I joined a group studying chromatin-mediated regulation of gene expression by quantitative proteomics, I developed a PhD project focused on the design and optimization of a novel biochemical and analytical method for the large-scale characterization of hPTM patterns and non-histonic proteins co-associated with specific functional regions of chromatin.

To this purpose, I combined chromatin immunoprecipitation (ChIP) and Mass Spectrometry (MS) analysis in order to establish an approach that could profit from the strengths of both methods, overcoming their respective limitations. ChIP using antibodies against specific modifications as “bait” were carried out to enrich specific functional chromatin domains at yield and purity sufficient for subsequent MS analysis. MS was then employed to characterize the co-associated hPTMs and to annotate all the proteins specifically interacting within the same chromatin domains. In particular native ChIP (N-ChIP), using unfixed chromatin selectively digested by micrococcal nuclease (MNase) was used in combination with MS to dissect PTM interactions, whereas cross-linking ChIP (X-

ChIP) in combination with SILAC quantitative proteomics was set to screen the specific interactions at the selected regions. It has already been widely demonstrated that in fact SILAC-based interactomics is particularly efficient in discriminating with high confidence specifically binders from backgrounds (see paragraph 2.5).

Since the method is completely novel, it was essential to demonstrate its robustness and reliability, which involved part of my work during these years. As proof of principle I used this approach to characterize silent and active chromatin regions, marked by H3K9me3 and H3K4me3, respectively. The accordance of the obtained results on hPTMs analysis and co-associated proteins with previous studies proved the robustness of our strategy. Basing on these achievements, the future perspective is to apply this strategy to extend the analysis on new modifications and variants enriched at regulatory regions, in a more biological and relevant model system (see paragraph 7).

4. MATERIALS and METHODS

4.1 General biochemistry buffers

(for buffers used in specific protocols, see the dedicated corresponding paragraphs)

1. **Phosphate-Buffered Saline (PBS), pH 7.4:**

   137 mM NaCl

   2.7 mM KCl

   8.1 mM $Na_2HPO_4(7H_2O)$

   1.76 mM KH_2PO_4

2. **Tris-Buffered Saline (TBS):**

   150 mM NaCl

   2.7 mM KCl

   25 mM Trizma HCl

3. **Urea lysis buffer:**

   20 mM Hepes, pH 8.0

   9 M Urea

4. **Running gel mix (1 gel at 17.5%):**

   Acrilamide/Bis-acrylamide stock solution - 37.5:1 ratio (3.48 mL)

   Tris-HCl, pH 8.8 (1.5 mL)

   ddH_2O (0.96 mL)

   SDS 20% (30 μL)

   APS 20% (30 μL)

   TEMED (3 μL)

5. Stacking gel mix (2 gel):

Acrilamide/Bis-acrylamide stock solution - 37.5:1 ratio (0.52 mL)

Tris-HCl, pH 6.8 (1.25 mL)

ddH_2O (3.10 mL)

SDS 20% (25 μL)

APS 20% (25 μL)

TEMED (5 μL)

6. LDS sample loading buffer (1X):

10% Glycerol

1% Lithium dodecyl sulfate (LDS)

1% Ficoll-400

0.2 M Triethanolamine-Cl pH 7.6

0.00625% Coomassie G250

0.5 mM EDTA disodium

7. SDS-PAGE running buffer:

25 mM Tris base

192 mM Glycine

2% SDS

8. Fixing gel solution:

50% Methanol

10% Acetic acid

9. Staining gel solution:

20% Methanol

20% Stainer A (Colloidal Blue Stain Invitrogen Kit)

5% Stainer B (Colloidal Blue Stain Invitrogen Kit)

10. Western transfer buffer:

25 mM Tris base

192 mM Glycine

20% Methanol

11. Stripping buffer:

62.5 mM Tris-HCl, pH 7.6

2% SDS

100 mM β-Mercaptoethanol

12. Tris-borate-EDTA (TBE):

89.2 mM Tris base, pH 7.6

88.95 mM Boric acid

2 mM EDTA, pH 8.0

4.2 Cell Culture and SILAC labeling

HeLaS3 cells were grown in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) supplemented with 10 % Fetal Bovine Serum (FBS, Invitrogen 10270-106), 1% Glutamine, 1% Pen/Strep and 10 mM HEPES pH 7.5. For metabolic labeling, HeLa S3 cells were grown in ‘‘Heavy’’ and ‘‘Light’’ SILAC media prepared adding to the SILAC DMEM (M-Medical FA30E15086), depleted of Lysine and Arginine, 10% dialyzed FBS (Invitrogen, 26400-044), 1% Glutamine, 1% Pen/Strep, 10 mM HEPES pH 7.5 and either the light isotope-coded amino acids $^{12}C_6$ $^{14}N_2$ L-Lysine (Lys 0, Sigma L8662) and $^{12}C_6$ $^{14}N_4$ L-Arginine (Arg0, Sigma A6969) or their heavy isotope-counterparts: $^{13}C_6$ $^{15}N_2$ L-Lysine (Lys 8, Sigma 68041) and $^{13}C_6$ $^{15}N_4$ L-Arginine (Arg10, Sigma 608033). Lys and Arg were added at a concentration of 73 mg/L and 42 mg/L, respectively. HeLaS3 were cultivated in SILAC media for 9 generations, with careful monitoring of growth rate, viability and overall morphology, to ensure that normal physiology is preserved.

NIH 3T3 mouse fibroblasts were grown in DMEM supplemented with 10 % Calf Serum (CS, Lonza 14-401F), 1% glutamine and 1% Pen/Strep.

4.3 Native chromatin immunoprecipitation (N-ChIP)

The protocol developed was modified from a previously described one (154). Two hundred millions HeLa S3 cells were homogenized in Lysis Buffer and nuclei were separated from cytoplasm, by centrifugation at 3750 rpm (4°C) for 30 minutes, putting cellular lysate on sucrose cushions. Nuclear pellets were washed twice with PBS, re-suspended in Digestion Buffer and digested with micrococcal nuclease (MNase, Roche) at a final concentration of 0.005 U/ml, at 37 °C for 60 minutes. The reaction was stopped by adding 1 mM EDTA and chilling on ice. The soluble fraction of chromatin (S1), comprising small fragments (mono-, di-nucleosomes), was collected as the supernatant obtained after centrifugation of re-suspended nuclei at 10000 rpm (4 °C) for 10 minutes. Pellets were instead re-suspended in Dialysis Buffer and dialyzed overnight at 4 °C in a dialysis tube (cut off 3.5 kDa). The second soluble fraction of chromatin (S2), comprising large fragments (tri- to epta-nucleosomes), was as the supernatant obtained after centrifugation at 10000 rpm (4 °C) for 10 minutes. DNA extracted by Qiaquick columns (QUIAGEN) was run on 1% agarose gel to evaluate fractions of chromatin. The S1 fraction is combined with a small aliquot of S2 fraction (1/100) in order to obtain an Input compose of about 95% of mono-nucleosomes. Chromatin was immunoprecipitated with 10 μg of the following antibodies: H3K9me3, H3K4me3 and H2AX. Antibodies were incubated overnight with chromatin; in parallel, 100 μl of G protein-coupled magnetic beads (Dynabeads, Invitrogen 100.04D) were blocked in BSA 0.5% PBS for an overnight. Blocked beads were washed and added to chromatin and incubated for 3 h at 4°C on a rotating wheel. Beads were washed four times (50 mM Tris-HCl pH 7.6, 10 mM EDTA) at increasing salt concentration (75, 125 and 175 mM NaCl). LDS Sample Buffer (Invitrogen

NP0007) supplemented with 50 mM DTT was added to the beads for 5 min at 70 °C to elute the immunoprecipitated proteins from the beads. Proteins were resolved on 4-12% Bis-Tris acrylamide SDS-PAGE pre-cast gels (Invitrogen NP0335BOX) on an Invitrogen system and visualized on the gel using Colloidal Comassie staining Kit (Invitrogen LC6025).

4.3.1 Buffers for N-ChIP:

1. Lysis Buffer: 10% sucrose, 0.5 mM EGTA pH 8.0, 15 mM NaCl, 60 mM KCl, 15 mM HEPES, 0.5% Triton, 0.5 mM PMSF, 1mM DTT, 5 mM NAF, 5 mM Na_3VO_4, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.
2. Digestion Buffer: 0.32 M sucrose, 50 mM Tris-HCl pH 7.6, 4 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM PMSF.
3. Dialysis Buffer: 10 mM Tris-HCl pH 7.6, 1 mM EDTA, 0.5 mM PMSF, 5 mM NAF, 5 mM Na_3VO_4, 5mM NaButyrate, protease inhibitors cocktail.

4.4 Cross-linking Chromatin immunoprecipitation (X-ChIP)

The protocol developed was a modification from a previously published one (155). Two hundred millions SILAC-labelled HeLa S3 cells were harvested; cell pellets were cross-linked in 0.75% formaldehyde PBS for 20 min at room temperature (RT) to stabilize protein-DNA and protein-protein interactions, with shaking on rotating wheel. Formaldehyde was quenched adding 125 mM Glycine for 5 min. After four washes with cold PBS, cells were suspended in Lysis Buffer for 10 min at 4°C. After centrifugation the nuclear pellets were washed once with Washing Buffer and then re-suspended in ChIP Incubation Buffer. Chromatin from nuclei was sonicated at 200 W for 15 min (cycles of 30 sec “on” and 1 min “off”, in a cooled Bioruptor (Diagenode). After sonication, 1% of Triton-100 was added to sonicated chromatin to pellet debris. Soluble nucleosomes,

contained in the soluble supernatant after centrifugation at 13000 rpm (4°C) for 10 minutes, were immunoprecipitated by adding 10 μg of the following antibodies: H3K9me3 and H3K4me3 and H2A.X. The immunoprecipitation procedure followed the steps as for the N-ChIP, except for the washes, which were carried out in 20 mM Tris-HCl pH 7.6, 2 mM EDTA, 0.1% SDS,1% Triton-100 and increasing NaCl concentration (150 and 300 mM). To reverse the crosslinking and elute the immunoprecipitated proteins, SDS-PAGE sample buffer (250 mM Tri-HCl pH 8.8, 0.5M β-mercaptoethanol, 2% SDS) was added to the beads for 25 min at 95 °C. Proteins were resolved on 4-12% Bis-Tris acrylamide SDS-PAGE pre-cast gradient gels Invitrogen system and visualized by Colloidal Comassie staining kit (Invitrogen). For DNA damage experiment the cells were treated with etoposide (Sigma E1383) at the concentration of 30 μM or DMSO (as control) for 1h, 2h, 4h and 8h to induce DSBs. After 3 h washing cells with fresh medium, cells were harvested and treated for a X-ChIP as previously described using antibodies against H3K9me3 and H3K4me3.

4.4.1 Buffers for X-ChIP:

1. Lysis Buffer: 50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton-100, 0.5 mM PMSF, 5 mM NAF, 5 mM Na_3VO_4, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.
2. Washing Buffer: 10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM PMSF, 5 mM NAF, 5 mM Na_3VO_4, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.
3. ChIP Incubation Buffer: 10 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% sodium lauroylsarcoside, 0.5 mM PMSF, 5 mM NAF, 5 mM Na_3VO_4, 5mM NaButyrate, 5 mg/ml Aprotinin, 5 mg/ml Pepstatin A, 5 mg/ml Leupeptin.

4.5 In-gel digestion of histones for MS analysis

Bands corresponding to the core histones were excised from the gel, de-stained with repeated washes in 50% acetonitrile (ACN) in ddH2O, alternated with dehydration steps in 100% ACN. Gel pieces were in gel chemically alkylated as previously described, by incubation with D_6-acetic anhydride (Sigma 175641) 1:9 in 1M NH_4HCO_3 and CH_3COONa solution as catalyzer (61). After 3h at 37 °C with high shaking in thermo mixer, chemically modified gel slices were washed increasing ACN % (50% and 100%). In-gel digestion was performed with 100 ng/µl trypsin (Promega V5113) in 50 mM NH_4HCO_3 at 37 °C overnight, in order to obtain an "in-gel"-like Arg-C digestion, which cleaves at the amide bond C-terminal to Arginine residues, producing peptides with an optimal length for MS analysis. Digested peptides were extracted, desalted and concentrated using a combination of reverse-phase C18/Carbon "sandwich" system and ion-exchange (SCX) chromatography, on hand-made nano-columns (StageTips) (156): digested peptides loaded on C_{18}/C and SCX StageTips were then eluted with high organic solvent (80% ACN) and NH_4OH, respectively. Eluted peptides were lyophilized, re-suspended in 0.1% TFA and 0.5% acetic acid in ddH2O, pooled and subjected to LC-MS/MS.

4.6 In-gel digestion of immunopurified proteins

Processing of gel- separated proteins prior MS analysis was carried out as previously described, with minor modifications (66). Briefly, slices were cut from gels and de-stained in 50% v/v acetonitrile (ACN)/50 mM NH_4HCO_3. Reduction was carried out with 10 mM DTT in 50 mM NH_4HCO_3, followed by alkylation with 55 mM iodoacetamide in 50 mM NH_4HCO_3. In-gel digestion was performed with 12.5 ng/µL trypsin (Promega V5113) in 50mM NH_4HCO_3, overnight at 37 °C.

Digested peptide were extracted with 3%TFA, 30%ACN and finally with 100% ACN, lyophilized, desalted and concentrated on C_{18} Stage Tips (156). Samples were loaded in 1% TFA and 5% ACN and eluted with high organic solvent (80% ACN). Eluted peptides were lyophilized, re-suspended in 0.1% TFA and 0.5% acetic acid in ddH2O, and subjected to LC-MS/MS analysis.

4.7 Liquid Chromatography and Tandem Mass Spectrometry (LC-MS/MS)

Peptide mixtures were separated by nano-liquid chromatography using Agilent 1100 Series (Agilent Technologies), coupled to a 7-Tesla LTQ-FT-ICR-Ultra mass spectrometer (ThermoFisher Scientific, Bremen, Germany). The nanoliter flow LC was operated in one column set-up with a 15 cm analytical column (75 μm inner diameter, 350 μm outer diameter) packed with C18 resin (ReproSil, Pur C18AQ 3 μm, Dr. Maisch, Germany). Solvent A was 0.1% FA and 5% ACN in ddH2O and solvent B was 95% ACN with 0.1% FA. Sample was injected in an aqueous solution at a flow rate of 500 nl/min. Peptides were separated with a gradient of 0-36% over 120 min followed by a gradient of 36-60% for 10 min 60-80% over 5 min at a flow rate of 250 nl/min. For histones, liquid chromatography separation was performed with a gradient of 0-40% solvent B over 90 min followed by a gradient of 40-60% for 10 min and 60-80% over 5 min at a flow rate of 250 nl/min. The nanoelectrospray ion source (Proxeon, Odense, Denmark) was used with a spray voltage of 2.4 kV. No sheath, sweep and auxiliary gasses were used and capillary temperature was set to 190 °C. The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. In the LTQ-FT Ultra, full scan MS spectra (200-1650 *m/z*) were acquired with a resolution of 100,000 (FWHM) at 400 *m/z* setting an AGC target of 1,000,000. The five most intense ions were isolated for fragmentation in the linear ion trap using collision-induced dissociation (CID) at a target value of 5,000. Singly charged precursor ions were excluded. Collision gas pressure was

1.3 millitorrs and normalized collision energy using wide band activation mode was 35%. Ion selection threshold was 250 counts with an activation q=0.25. The activation time of 30 ms was applied in MS2 acquisitions.

4.8 Quantitative MS analysis of hPTMs co-enriched in precipitated chromatin

Raw data from LTQ-FT Ultra were converted to mgf files using Raw2MSM software (version 1.10) (157). MS/MS spectra were searched with Mascot Daemon (version 2.2.2, Matrix Science) against the IPI human database (version 3.68, 87,061 entries). MS mass tolerance was set to 10 ppm and MS/MS mass tolerance was set to 0.5 Da. Chemical alkylation with D_6-acetic anhydride, which labels unmodified and mono-methylated Lysines with a deuterated acetyl moiety but does not react with di-methyl, tri-methyl and acetyl Lysines, results in a delta mass of 45.0294 amu for each D_3-acetyl group added and thus allows to unambiguously distinguish isobaric modified peptides by their different LC retention times. The search included variable modifications: Lysines D_3-acetylation (+45.0294 Da), Lysine mono-methylation (calculated as the sum of the masses of D_3-acetylation (+45.0294) and mono-methylation (+14.016 Da)), di-methylation (+ 28.031 Da), tri-methylation (+42.046 Da), Lysine acetylation (+42.010 Da), Methionine oxidation (+15.995 Da) and N-terminal protein acetylation (+42.010 Da). Low-confidence peptide identifications were filtered from Mascot results according to following criteria: peptides with either ion score lower than 15 or more than 5 putative PTMs were removed (100); redundant peptides with same ID were filtered out by selecting the peptide with the highest Mascot score. Extracted ion chromatograms (XIC) were constructed for each precursor based on the *m/z* value, using a mass tolerance 10 ppm and a mass precision up to 4 decimals. Histone PTMs were first quantified by calculating the area under the curve (AUC) of each peak corresponding to every specific modified peptide. Then, their relative abundance was estimated by dividing the area under the curve (AUC) of each peptide by

the sum of the areas corresponding to all observed modified forms of that peptide [(XIC modified peak/∑ XIC all peaks)x100)]. Modification enrichment was calculated as the ratio between the relative abundance in the ChIP-ed octamer and the corresponding relative abundance estimated from input. Peptides containing modifications were validated by manual annotation of MS/MS spectra (see Appendix I and II for peptide (3-8) and (9-17) of histone H3), using the QualBrowser version 2.0.7 (ThermoFisher Scientific).

4.8.1 Masses (in Da) of site-specific identification of PTMs on Histone H3

Peptide 3-8: TKQTAR, K4unmod=375.2208, K4me1= 382.2286, K4me2= 366.7218, K4me3=373.7296; peptide 9-17: KSTGGKAPR, K9unmod=496.2937, K9me1=503.3016, K9me2=487.7947, K9me3=494.8025, K9unmod/K14Ac=494.7843, K9me1/K14Ac=501.7921, K9me2/K14Ac=486.2853, K9me3/K14Ac=493.2931, K9Ac/K14Ac=493.2749; peptide 18-23: KQLATKAAR, K18/23unmod=538.8383, K18me1/K23unmod=545.8461, K18unmod/K23Ac or K18unmod/K23Ac=537.3289, K18me1/K23Ac=544.3367, K18Ac/K23Ac=535.8195; peptide 27-40: KSAPATGGVKKPHR, K27/36/37unmod=784.9645, K27me1 or K36me1=791.9723, K27me2 or K36me2=776.4655, K27me3 or K27me2/K36me1 or K36me2/K27me1=783.4733, K27me1/K36me1=798.9802, K27me2/K36me2=767.9664, K27me3/K36me1=790.4811; peptide 73-83: EIAQDFKTDLR, K79unmod=690.8635, K79me1=697.8713, K79me2=682.3644.

4.8.2. Masses (in Da) of site-specific identification of PTMs on Histone H2A

peptide 4-11: GKQGGKAR, K5/9unmod=446.2675, KAc=444.7581, Kdi-Ac=443.2487.

4.8.3. Masses (in Da) of site-specific identification of PTMs on Histone H4

peptide 4-17: GKGGKGLGKGGAKR, K5/K8/K12/K16unmod=725.9476, KAc=724.4381, Kdi-Ac=722.9287, Ktri-Ac =721.4193, Ktetra-Ac=719.9099.

4.9 Proteomic analysis of proteins co-associated within precipitated chromatin fractions

Protein interactors from the X-ChIPs were identified and quantified using MaxQuant software (158) (version 1.1.1.36). MS/MS spectra were recorded in “centroid” mode and the six most abundant peaks per 100 Da mass intervals were selected for search. Filtered MS/MS spectra were searched against the IPI human database (version 3.68, 87,061 entries), combined with the standard MaxQuant contaminants database, by Andromeda search engine (159, 160). MaxQuant analysis included an initial search with a precursor mass tolerance of 20 ppm, whose results were used for subsequent mass recalibration (161). Enzyme specificity was set to trypsin, allowing two misscleaveges and cleavage at the N-terminus of Proline. Peptide identification was based on a search with mass deviation of the precursor ion of 6 ppm and the fragment mass tolerance was set to 0.5 Da. The mass accuracy of the precursor ions was improved by the time-dependent recalibration. MaxQuant was employed to filter identifications at 1% false discovery rate (FDR) at three levels, namely: site, peptide, and protein. Carbamidomethylation of Cysteine was selected as a fixed modification whereas oxidation of Methionine, and acetylation of protein N-terminus were included as variable modifications. The modifications corresponding to Arginine and Lysine labeled with heavy stable isotopes are treated as fixed modifications in the Andromeda search. Additional peptides were identified by "the match between run" option in MaxQuant, which matches precursor masses in a 2-min retention time window (after realignment of the runs) based on the accurate mass measurement. Proteins were accepted if identified with at least two peptides one of which unique. Protein ratios were normalized by standard deviation; complete

output tables for the H3K9me3 and H3K4me3 chromatomes are provided in Appendix IV and V, respectively. Analysis and visualization of the data were performed using the open-source package R with in-house scripts and Perseus program (J. Cox, manuscript in preparation or www.maxquant.org).

4.10 Quantitative RT-PCR of immunopurified DNA

DNA from ChIP-ed material was eluted in TE (Tris-HCl pH 7.5, EDTA) containing 2% SDS for 15minutes at 65°C (for X-ChIP experiments DNA was also de-crosslinked at 65°C overnight) and DNA was then purified through Qiaquick columns (QUIAGEN). 1µl of purified DNA was used for substrate for amplification on Applied Biosystems 7500 Fast Real-time PCR system applying Biosystem Sybr-green).

3.10.1 Primers for quantitative PCR upon conventional ChIP:

AP945 (chr4:57142864(start)-57142927(end))

Forward primer: 5´-CGCTACTGTTGGGTGCTGG-3´

Reverse primer: 5´-GCCTGGAAAGCTGTATTTGCTG-3´

AP777 (chr2:198189648(start)-198189843(end))

Forward: 5´-TCCATCACGTGCGACGC-3´

Reverse: 5´-GAGGCGCGGTATCCCAG-3´

α-Repeats Regions

Forward primer: 5´-CTCAGTAACTTCCTTGTGTTGTG-3´

Reverse primer: 5´-ATTCTGTCTAGTTTCTATAAGAAG-3´

4.11 Immunoblot analysis

Input chromatin and immunoprecipitated histone octamers were separated in 17.5% SDS-PAGE and transferred to PVDF membranes (Millipore). Membranes were blocked 1 h in 5% milk in TBS supplemented with 0.1% Tween (TBS-T). After blocking, membranes were incubated at 4°C for an overnight with primary antibodies specific for histone modifications, diluted in TBS-T 5% milk. After three washes in TBS-T, binding was revealed by ECL Plus® Immunoblotting Detection System (Amersham Biosciences); antibodies against the unmodified version of both H3 and H4 were used as loading control.

For Western blot analysis the following antibodies were used, according to the manufacturer's instructions: H3K9me3 (Abcam 8898, dilution 1:1000), H3K36me2 (Abcam 9049, dilution 1:1000), H3K27me3 (Upstate Millipore 07-449, dilution 1:2500), H3K4me3 (Active Motif 39159, dilution 1:1000), H4K20me3 (Abcam 9053, dilution 1:1000), acetyl-Histone H4 (Upstate Millipore 06-866, dilution 1:2000) where K5/8/12/16 were acetylated, H3K79me2 (Abcam 3594-00, dilution 1:1000), acetyl-Histone H3 (Upstate 05-599, dilution 1:5000) where K9/14 were acetylated, H3K18me1 (Active Motif 39667, dilution 1:500), Histone H2A.X (phospho Tyr142) (Upstate Millipore 07-1590, dilution 1:100; Abcam 94602, dilution 1μg/μl), phospho-Histone H2A.X (Ser139) (Millipore 05-636, dilution 1:1000), histone H2A.X (Abcam 11175, dilution 1:5000), histone H3.3 (Abcam ab62642, dilution 1:1000), unmodified histone H3 (Abcam 1791, dilution 1:5000), unmodified histone H4 (Millipore Upstate 07-108, dilution 1:1000).

4.12 Immunofluorescence analysis

Cells grown on coverslips were washed twice with PBS, fixed with 4% para-formaldehyde for 20 min, permeabilized in 0.5% Triton X-100 in PBS for 5 min and then blocked in 10% BSA for 1 h. Cells were subsequently probed with the following mix of antibodies: WSTF (Sigma W3516, dilution 1:250) and HP1β (Millipore MAB34448,

dilution 1:500); H3K9me3 (Abcam 8898, dilution 1:500) and HP1β. After 1 h of primary antibody incubation, cells were washed three times with PBS and then incubated with either a-rabbit secondary antibody conjugated with Cy3 (diluted 1:800), or with a-rabbit secondary antibody conjugated with fluorescein isothiocyanate (FITC-conjugate), diluted 1:50. DNA was stained with DAPI, diluted 1:5000 in PBS, for 15 sec. Slides were mounted in Mowiol and images were acquired using a wide field Olympus Biosystem Microscope BX51.

4.13 ChIP-Sequencing: preparation of ChIP DNA libraries, sequencing and data analysis

ChIP DNA was prepared for Solexa 2G sequencing using a standard protocol. ChIP-ed DNA was treated to remove 3' overhangs and fill in 5' overhangs resulted in blunt ended DNA fragments. An A residue was added by terminal transferase to the 3' end and the resulting fragments were ligated with Illumina adapters. The resulting Adapter-modified DNA fragments were separated by agarose gel electrophoresis and the band between 120-200 bp was excised and the DNA fragments were extracted using a Qiaquick Gel Extraction Kit (Qiagen Inc). The specific DNA fragments were subjected to 18 cycles of PCR amplification; amplified fragments were then gel purified from an excess of PCR primers, using Qiagen columns. The DNA fragment library was quantified with Bioanalyzer using High Sensitivity Chip, diluted to a 10 nM working stock concentration for cluster generation. Finally, cluster generation was performed according to standard protocols of the manufacturer (Illumina) and loaded into individual lanes of a flow cell (4 picomoles/sample). ChIP-Seq data were acquired with the Illumina Genome Analyzer II, producing a fixed 36bp read length. After each base incorporation step, the flow cell surface was washed to remove reactants and then imaged by microscope objective. For the analysis of sequencing Illumina data, read tags passing standard Illumina quality filter

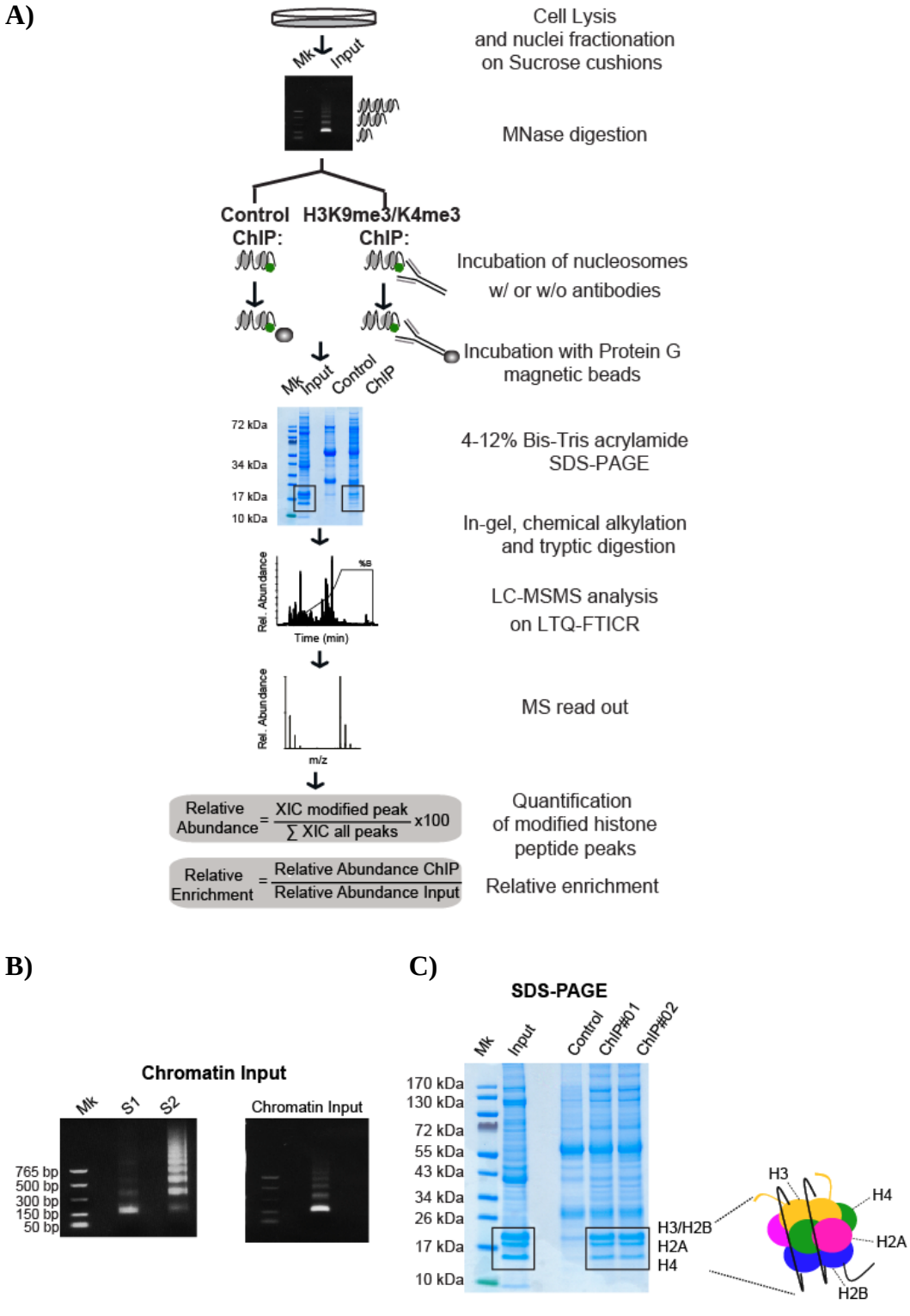
(Failed-chastity < 0.6) were aligned to hg18 genome using BWA 0.5.9 with default parameters (162). H2AX data and H3K9me3 and H3K4me3 ChIP-Seq data from (147), stored in the NCBI GEO SuperSeries GSE20303, were analyzed with dspchip 0.8.5 (http://code.google.com/p/dspchip). Non-duplicated tags with mapping quality higher than 15 were retained; normalized profile of IP data were subtracted from the Input; the resulting profile was processed using a Hanning Window low-pass filter (window size: 500 kbp) and negative values were removed. Correlations between ChIP profiles were calculated with wigCorrelate from UCSC Genome Browser utilities.

5. RESULTS

5.1 Characterization of hPTM patterns co-enriched at specific chromatin regions by N-ChIP combined with high resolution MS analysis

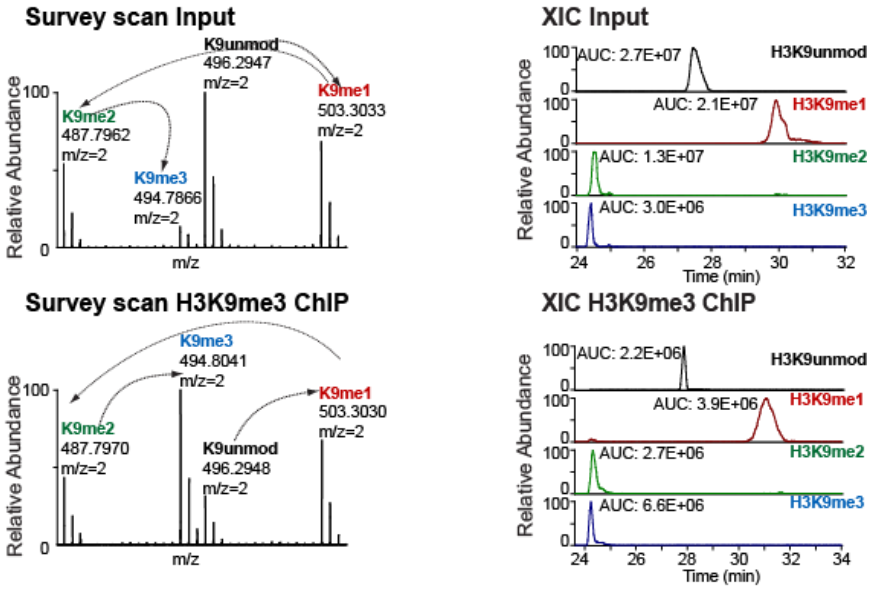
Chromatin Immunoprecipitation enables profiling of the localization of a protein or a histone modification along the genome by deep sequencing. In a proteomics equivalent of such an experiment, ChIP is followed by the analysis of the immunopurified proteins by MS. We established a preparative version of the classical "Native ChIP" (N-ChIP, namely ChIP without crosslinking) to purify chromatin regions enriched in a specific histone modification, in order to characterize by MS the hPTMs patterns and proteins associating with such regions. We investigated fractions enriched in silent or active chromatin by taking advantage of the well-characterized enrichment of H3K9me3 at both pericentric heterochromatin and repressed euchromatic genes, and of H3K4me3 in active promoters. The protocol is outlined in figure 16A. Briefly, chromatin from HeLaS3 cells was digested with MNase to obtain a first soluble fraction (S1), comprising small fragments (mono-, di-nucleosomes), and a second soluble fraction (S2), comprising large fragments (tri- to epta-nucleosomes) (Figure 16B, left panel). The fraction enriched in mono-nucleosomes (Figure 16B, right panel) was incubated with antibodies recognizing either H3K9me3 or H3K4me3. For immunoprecipitation we chose two antibodies widely employed by the community for ChIP-Seq studies, with proved specificity and efficiency (163, 164). Control ChIP was chromatin input not incubated with antibody. Immunopurified chromatin was captured on magnetic beads and the extracted proteins were separated by SDS-PAGE (Figure 16C). Each core histone from the fractionated nucleosomes was analyzed by LC-MS/MS.

Figure 16. Scheme of the N-ChroP strategy, combining N-ChIP and MS analysis. A) Scheme of the experimental approach. B) DNA isolated after digestion with micrococcal nuclease S7 (MNase), separated in fractions containing mono- (S1) or poly-nucleosome (S2) (left panel) and chromatin input (right panel), resolved on a 1% agarose gel stained with ethidium bromide. C) SDS-PAGE of ChIP-ed S1 fraction: core histones H3, H4, H2A and H2B are visible around and below the 17kDa band of the pre-stained protein marker, with H3 and H2B co-migrating.



Histone PTMs were first identified and then quantified by calculating the area under the curve (AUC) of each peak corresponding to every specific modified peptide (Figure 17). Then, their relative abundance was estimated, both in the Input and in the ChIP, by dividing the area under the curve (AUC) of each peptide by the sum of the areas corresponding to all observed modified forms of that peptide. Modification enrichment was calculated as the ratio between the relative abundance in the ChIP-ed octamer and the corresponding relative abundance estimated from input.

Figure 17. Representative scheme of hPTMs analysis. Zoomed mass spectra for the 2+ charge of the unmodified, mono-, di-, and tri-methylated K9, for Input and ChIP, and extracted ion chromatograms (XIC) constructed for each precursor at the corresponding *m/z* value for the same samples.

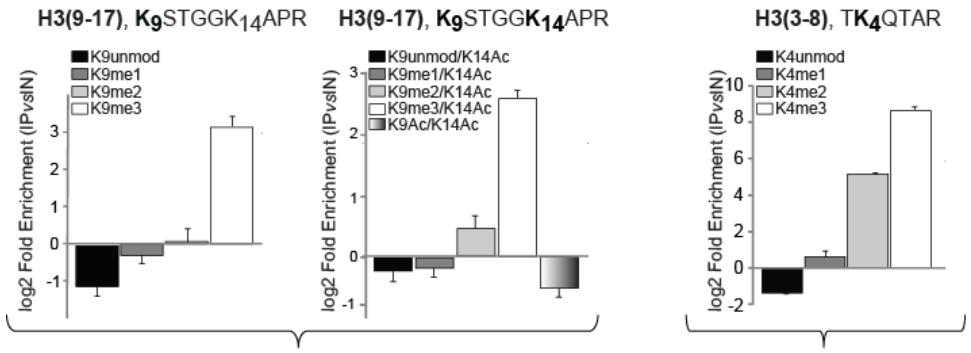


The analysis first focused on histone H3 (9-17) and (3-8) peptides: a specific enrichment of K9me3, as well as K4me3, associated with the depletion of unmodified H3 and H3 mono-methylated at K9 and K4, demonstrated the efficiency and specificity of the ChIP with the respective antibody (18A). The depletion of acetylated H3K9/K14 was also observed in

H3K9me3-nucleosomes (Figure 18A). Western blot (WB) analysis confirmed the enrichment of K9me3 and K4me3 obtained at MS level (Figure 18B).

Figure 18. Validation of the N-ChroP approach. A) Relative enrichment of K9me3 (co-existing or not with acetylated K14) and K4me3, in peptides (9-17) and (3-8) of H3, respectively. The results were obtained upon N-ChIP using antibodies α-H3K9me3 and α-H3K4me3 (#1 and #2 indicate two replicates). The enrichment is expressed as a log2 Ratio of the relative abundance of each methylation in the ChIP sample as compared to input and represents the averages ± SEM (Standard error of the mean) from three independent experiments. B) WB validation of H3K9me3 and H3K4me3 enrichment in the corresponding ChIPs: an aliquot of ChIP-ed material and of Input were loaded on SDS-PAGE, and immunoblotted with antibodies H3K9me3 and H3K4me3. 0.2 % and 0.04% of input were loaded for H3K9me3 and H3K4me3, respectively, for a semi-quantitative comparison with IP. Unmodified H3 is the loading control.

A)

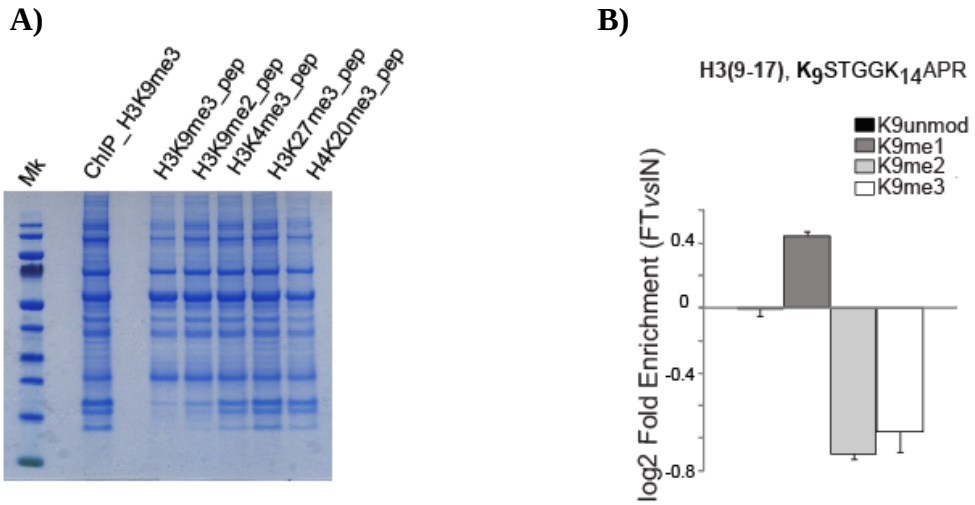


B)

The potential cross-reactivity of the anti-H3K9me3 antibody with other methyl Lysines was excluded with a competition assay where the chromatin was incubated with antibody α-H3K9me3 in the presence of an excess of different soluble peptides bearing the following modifications: K9me2, K4me3, K27me3 in H3 and K20me3 in H4. Unspecific

recognition of other Kme was excluded, except for K9me2, where a mild binding was detected (Figure 19A). The comparison between H3K9me3 abundance in flow-through and input demonstrated that about 50% of the chromatin regions of interest were captured in the ChIP, ruling out a potential bias due to the enrichment of only a sub-fraction of chromatin (Figure 19B).

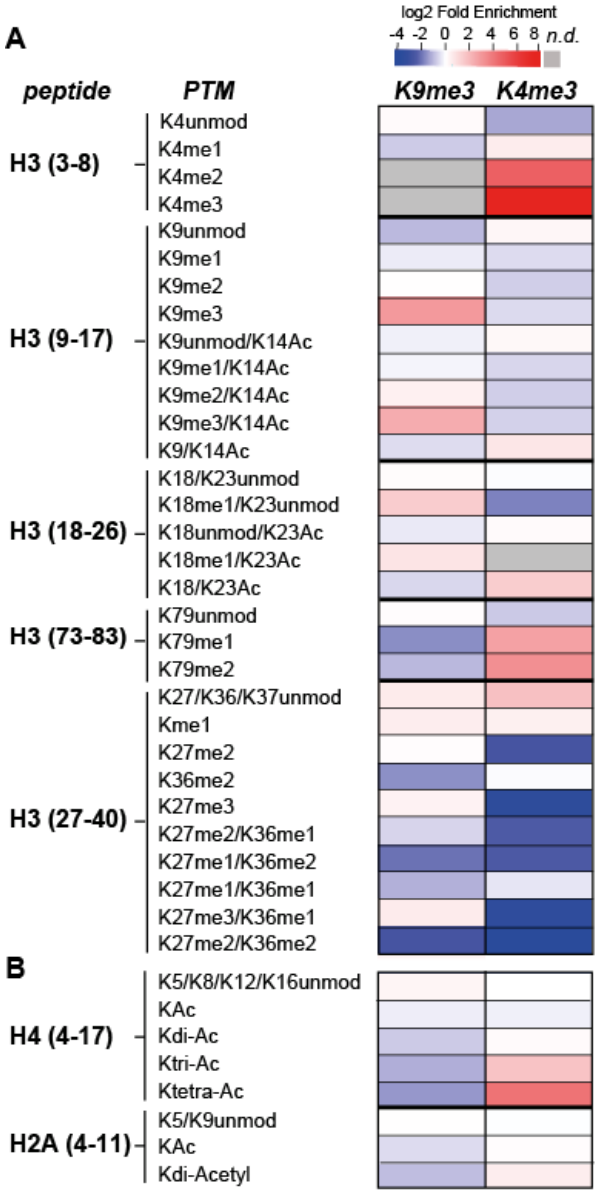
Figure 19. Evaluation of H3K9me3 antibody specificity and efficiency. A) Competition assays for antibody specificity test: after MNase digestion chromatin is incubated with H3K9me3 antibody with or without excess (150X) of different soluble peptides bearing methylations at distinct sites (H3K9me3, H3K9me2, H3K4me3, H3K27me3 and H4K20me3): immunoprecipitated materials are separated by SDS-PAGE and stained with Colloidal Comassie blue. B) Relative enrichment of unmodified, mono-, di-, and tri-methylated K9 in flow-through (FT) as compared to input (IN). Histogram represents the averages ± SEM from three independent experiments and results in significant depletion of K9me3 and K9me2 in FT.



5.2 Label free quantification of histone PTMs enriched in repressed and active chromatin domains

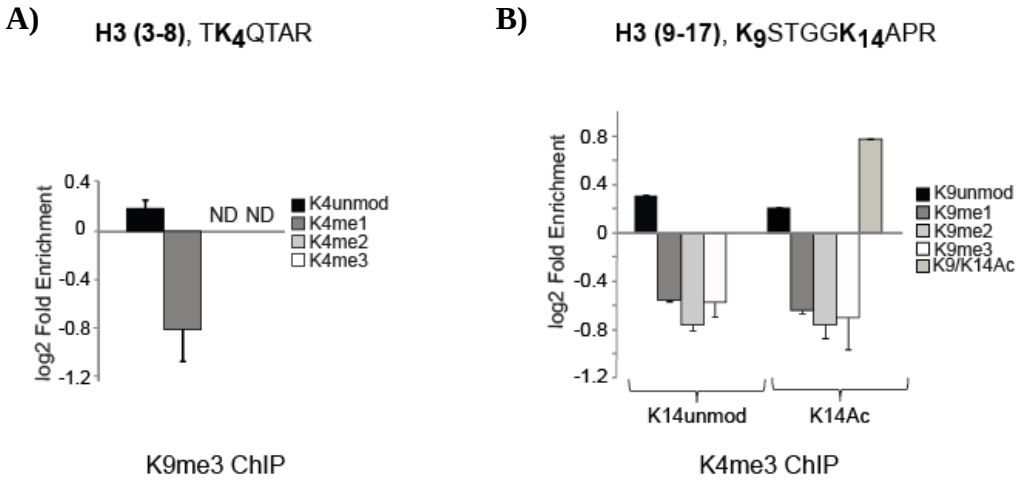
The hPTMs enriched in silent and active chromatin domains are summarized in the heatmap and histograms (Figure 20).

Figure 20. Relative enrichment of modifications in H3K9me3 and H3K4me3 mononucleosome. Heatmap of hPTMs enrichment for all the modified resides identified on H3, H4 and H2A histones. Each row corresponds to a distinct histone modification, whereas columns correspond to the different antibodies used for the ChIP (n.d. means not detected).



Focusing on the modifications associated on histone 3, our analysis on intact native nucleosomes validated the selectivity of our antibodies and overall recapitulated the "chromatin canonical states" described by Ernst and Kellis and based on a meta-analysis of a large set of data (165). In particular, H3K4me1 was found depleted in H3K9me3 nucleosomes (Figure 21A), since K4me1 is preferentially associated with hyperacetylated H3 and never observed in combination with H3K9me3; proving that H3K4me1, likewise H3K4me2 and H3K4me3, occurs mostly in transcriptionally active chromatin (108). In addition, we observed a depletion of K9 methylations and a parallel increase of K14ac and K9ac/K14ac in H3K4me3-euchromatin (Figure 21B), in line with genome-wide localization studies showing that H3K4me3 marks active promoters in human cells, with a pattern very similar to H3K9ac/K14ac (166).

Figure 21. Relative enrichment of modifications in (3-8) and (9-17) peptides of H3 in H3K9me3 and H3K4me3 ChIPs. A) Relative enrichment of K4 methylations with α-H3K9me3. B) Relative enrichment of K9 methylations with α-H3K4me3. Relative enrichment of each PTM is expressed as a log2 Ratio, as previously described.



H3R26me2, in association with K23Ac, was found enriched in H3K4me3 domains (Figure 22A and 22B): information on Arginine methylations is overall partial, however emerging evidence suggests that H3R26me2 may antagonize Polycomb repression due its proximity to K27 (167, 168).

We found H3K18me1 enriched in silent domains and depleted in active ones (Figure 23A and 23B). Indeed, a silencing role for this mark had been already proposed, based on two observations: first, its half-maximal life was reported as significantly lower than other mono-methylations with gene-activating function (105); second, it could be antagonizes K18ac, an active mark (169).

Figure 22. Identification of R26me2 in H3K4me3 regions, enriched by N-ChIP. A) Relative enrichment of R26me2 in (18-26) peptide of H3, upon N-ChIP using antibody α-H3K4me3. Relative enrichment is expressed as a log2 Ratio, as previously described. B) MS/MS spectra of H3 (18-26) peptide containing R26 di-methylated.

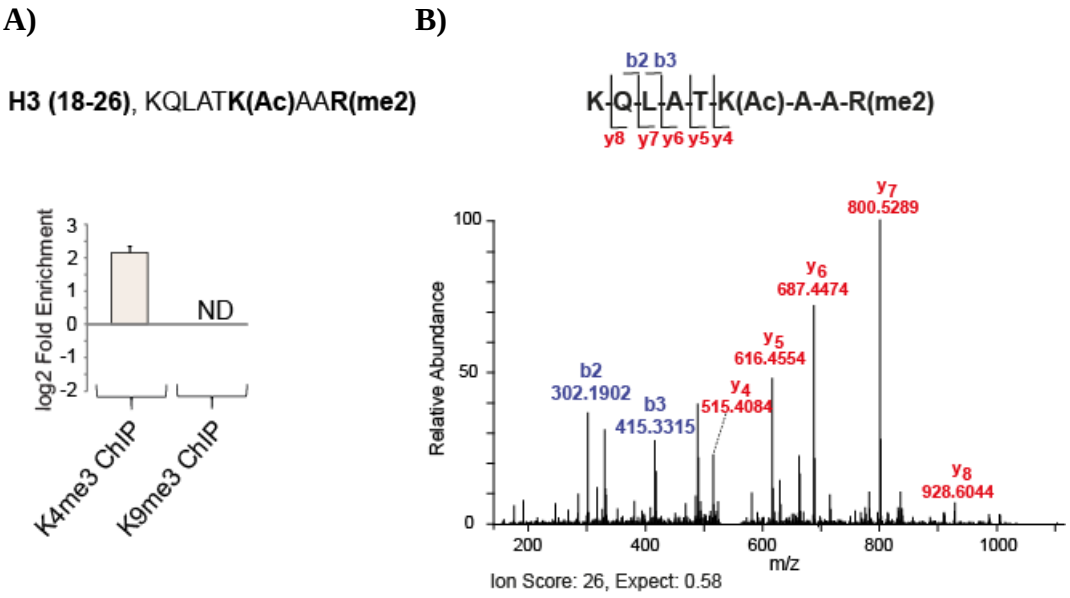
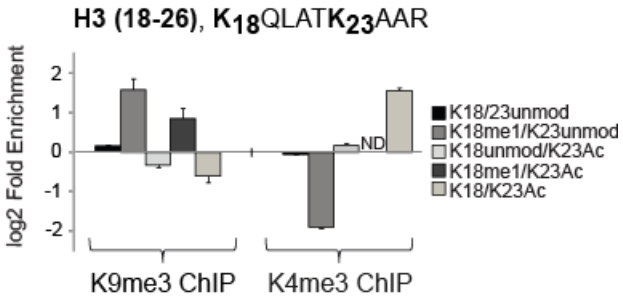
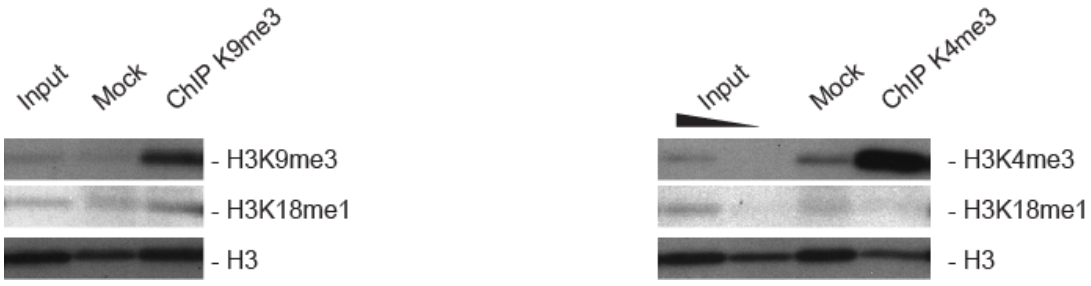


Figure 23. Relative enrichment of modifications on (18-23) peptides of H3 in H3K9me3 and H3K4me3 ChIPs. A) Relative enrichment of K18 and K23 methylation and acetylation in H3K9me3 and H3K4me3-enriched domains. Relative enrichment of each PTM is expressed as a log2 Ratio, as previously described. B) WB validation of H3K18me1 enrichment in the H3K9me3 ChIP. Antibodies against H3K9me3 and H3K4me3 are used to verify the effective enrichment of specific modifications in the corresponding IPs, and unmodified H3 is the loading control.

A)

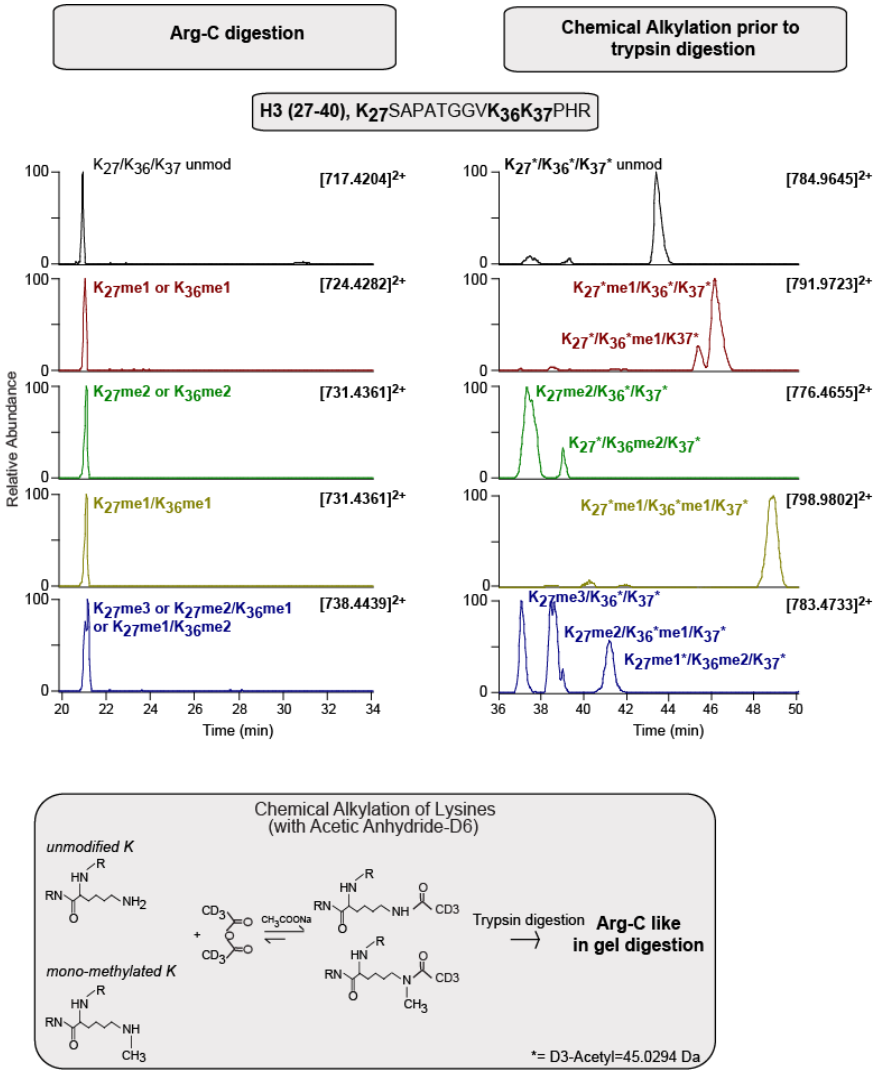


B)



Analysis of methylation profiles on peptide H3 (27-40) identified several distinct modified peptides. The chemical alkylation with D_6-acetic anhydride, which labels unmodified and mono-methylated Lysines with a deuterated acetyl moiety but does not react with di-methyl, tri-methyl and acetyl Lysines, allowed distinguishing different methylations on K27 and on K36/37. In a case where two of these Lysines are mono-methylated, it would be challenging to distinguish this species from an isobaric peptide containing a single di-methylation modification (Figure 24A). The derivatization approach however can remove this isobaric property (Figure 24B). The addition of the deuterated acetyl moiety to unmodified and mono-methylated Lysines leads to a mass difference between the two peptide-forms. Furthermore, the nature of this modification permits specific assignment to Lysines, for example in the case of K27 and K36. These properties are displayed in the shifted elution profile (Figure 24A and 24B).

Figure 24. Elution profile of H3 (27-40) modified peptide. A) Extracted ion chromatograms (XIC) of various 2+ charge modified forms relative to H3(27-40) peptide are reported, upon Arg-C digestion, for the time range corresponding to 20-34 min. Peptide ions at the specific *m/z* values correspond to unmodified, mono- (me1), di- (me2) and tri- (me3) methylated H3(27-40) peptides, respectively. B) Extracted ion chromatograms (XIC) of various 2+ charge modified forms relative to H3(27-40) peptide are reported, upon deuterated acetic anhydride alkylation, prior to trypsin digestion, for the time range corresponding to 36-50 min. Peptide ions at the specific *m/z* values correspond to unmodified, mono-(me1), di-(me2) and tri-(me3) methylated H3(27-40) peptides, respectively. Peptide ion at 798.9802 *m/z* is assigned to mono-methylations at K27 and K36. Based on the number of D_3-acetyl groups and methylations, distinct modification degrees at specific Lysines residues can be assigned unambiguously. Furthermore, with this strategy certain isobaric peptides (i.e. K27me2 and K36me2) can be efficiently resolved during chromatography by their distinct elution times.



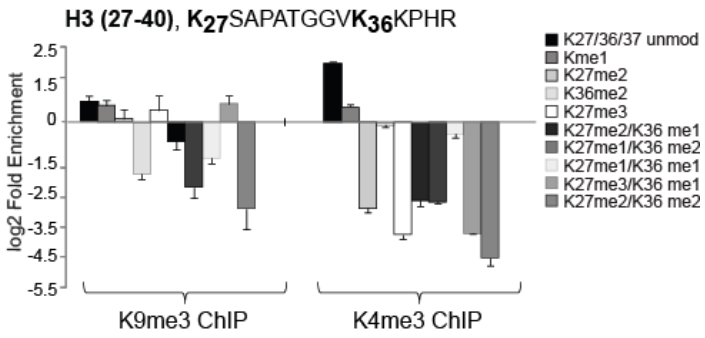
Quantitative analysis showed that H3K27me1 was enriched in heterochromatin (Figure 25A), confirming previous studies where constitutive heterochromatin was characterized by the focal enrichment of H3K9me3 and H3K27me1 (170). H3K27me3 was slightly enriched in heterochromatin and depleted in euchromatin, as previously described (Figure 25A and 25C) (42).

As expected, H3K79me1/me2 were depleted in H3K9me3 while significantly increased in H3K4me3 domains (Figure 25B and 25C); indeed H3K79 methylation is proposed to mark active, or at least accessible, genes and to function as a barrier to heterochromatin spreading (171).

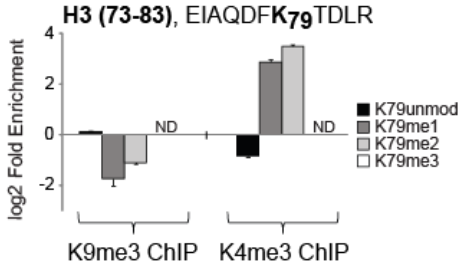
Overall, the confirmation of previously described or predicted PTMs synergisms confirmed the robustness of the newly established strategy and thus corroborated the discovery of novel associations among histone marks, both intra-molecularly and inter-molecularly within the immunopurified nucleosomes. For instance, we observed that di-methylated K36 is depleted in both chromatin domains (Figure 25A and 25C): the unexpected underrepresentation of this mark in H3K4me3 regions can be explained by reasoning that H3K4me3 is typically enriched at proximal promoters whereas H3K36me2/3 associate with gene elongation and thus are highly overrepresented at the 3´ of genes (172, 173): as such, this evidence further corroborates the high resolution of our analysis in dissecting PTMs clustering at a single nucleosome level.

Figure 25. Relative enrichment of modifications on (27-40) and (73-83) peptides of H3 in H3K9me3 and H3K4me3 ChIPs. A) Relative enrichment of K27 and K36 (A) as well as K79 (B) methylations in H3K9me3 and H3K4me3-enriched domains. Relative enrichment of each PTM is expressed aas log2 Ratio, as previously described. C) WB validation of specific Lysine methylations enrichment in the both ChIPs (#1 and #2 indicate two replicates): antibodies α-H3K27me3, K79me2 and K36me2 were used. Unmodified H3 is the loading control.

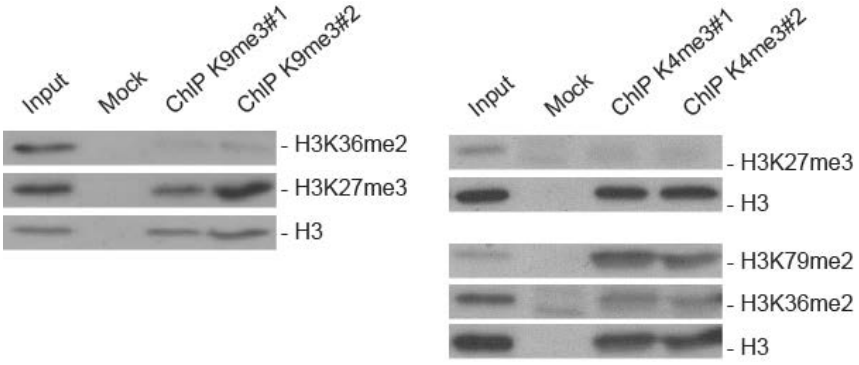
A)



B)



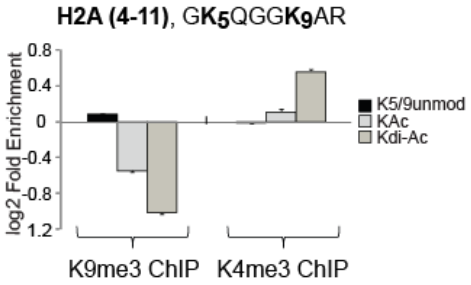
C)



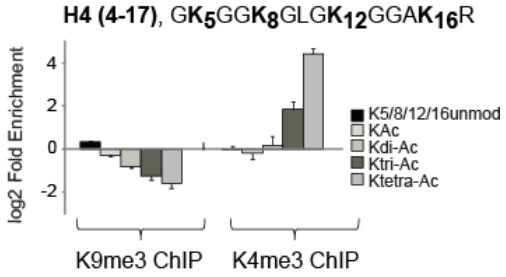
We then extended the study to the PTMs on the co-purified histone H2A and H4: acetylations on the H2A N-terminus were reduced in the H3K9me3-domains and enriched in H3K4me3-regions (Figure 26A); similarly, hyper acetylated H4 was found enriched in euchromatin (Figure 26B and 26C) (3). Methylation at Lysine 20 of H4 was over-represented in H3K9me3-domains while reduced in H3K4me3-domains, as detected by WB (Figure 26C): in fact H4K20me3 follows H3K9 methylation during initiation of a heterochromatic environment (174).

Figure 26. Relative enrichment of modifications on (4-11) peptide of H2A and on (4-17) peptide of H4 in H3K9me3 and H3K4me3 ChIPs. A) Relative enrichment of H2A (A) and H4 (B) Lysines acetylation in H3K9me3 and H3K4me3-enriched domains (#1 and #2 indicate two replicates). Relative enrichment of each PTM is expressed as a log2 Ratio Relative enrichment of each PTM is expressed as a log2 Ratio, as previously described. C) WB validation of H2A and H4 acetylation enrichment in H3K4me3 ChIP (right panel) and H4K20me3 enrichment in H3K9me3 ChIP (left panel).

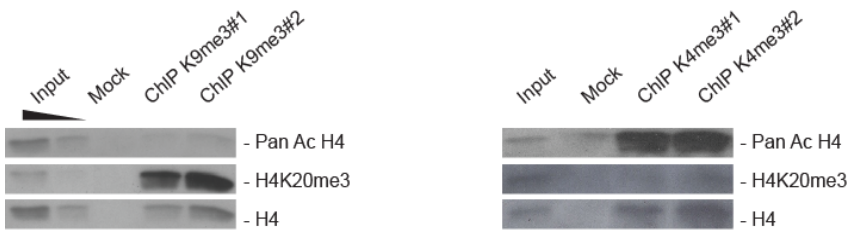
A)



B)



C)



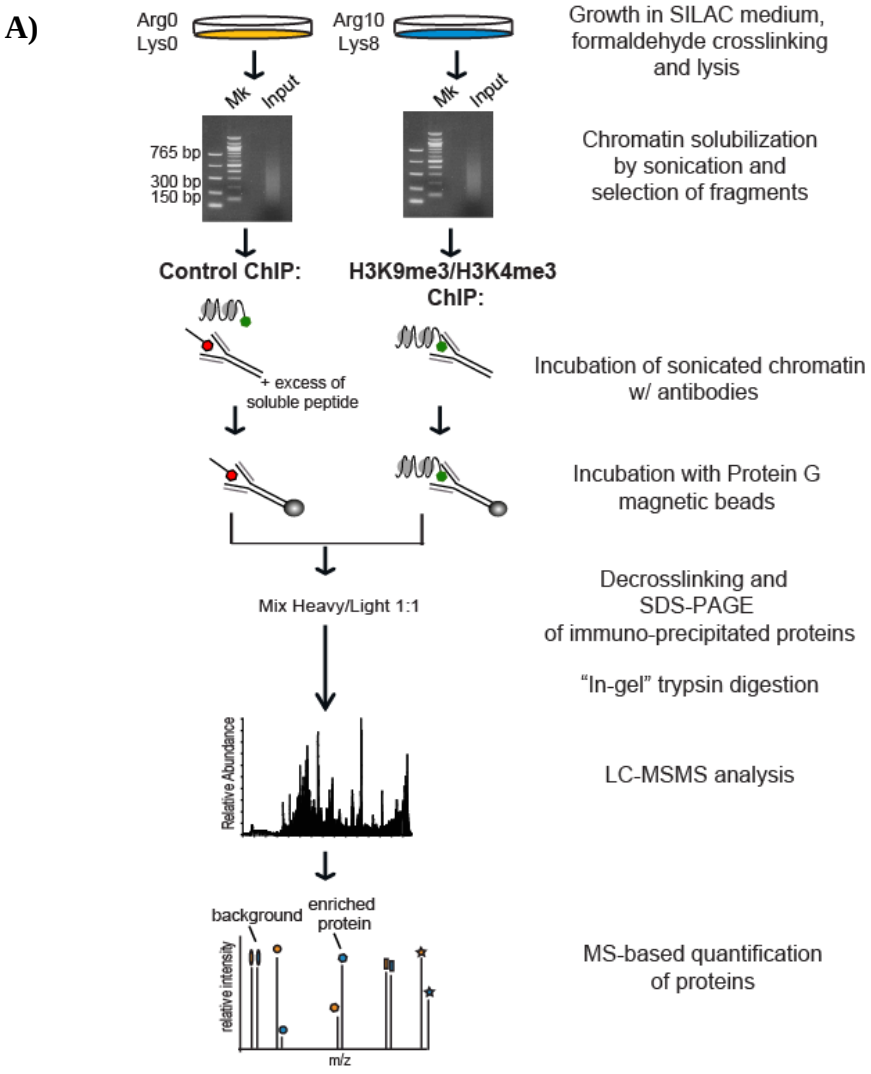
In conclusion, our approach indicates that ‘‘active’’ modifications (methylated K4, K36, K79, acetylated K9/K14, K18/K23 of H3 and hyper acetylated H2A and H4) are enriched in H3K4me3 domains, whereas ‘‘silent’’ modifications (methylated K9, K27 of H3 and K20 of H4) are overrepresented in H3K9me3 territories (42, 108, 175).

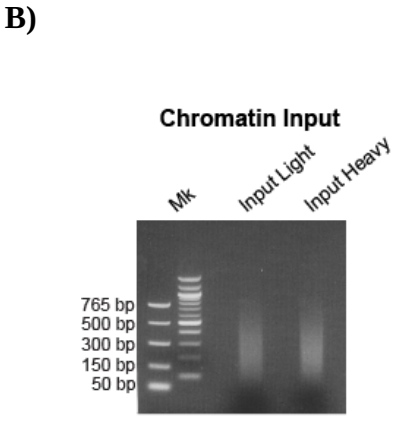
5.3 Large-scale study of chromatin-associated proteins by X-ChIP combined with high resolution MS analysis

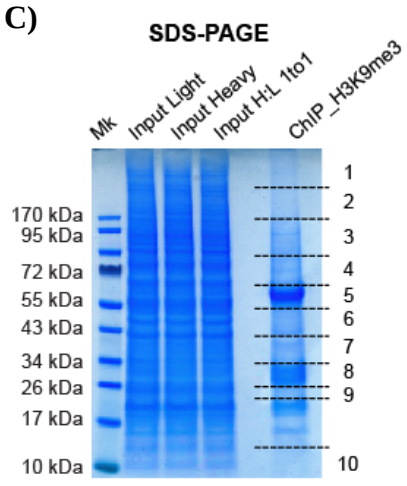
To characterize H3K9me3 and H3K4me3-interacting proteins we modified the crosslinking ChIP protocol to adapt it for SILAC-based quantitative interactomics (88). The protocol is outlined in figure 27A: HeLaS3 cells grown in “light” and “heavy” media were fixed with formaldehyde and chromatin was subjected to sonication to generate 300-500bp lengths of DNA fragments (Figure 27B). Both isotope-coded samples (heavy H and light L) were incubated with specific antibodies; in one of the two isotope-channels the antibody was saturated with an excess of soluble H3 peptide bearing the modification used as bait (either H3K9me3 or H3K4me3). All nucleosomes exposing those PTMs were competed out, as well as all co-associating proteins; unspecific proteins, instead, were not selectively competed. Two replicates of the X-ChIP were performed swapping the H and L channels in which the peptide-competition control was carried out (the so-called “forward” and “reverse” format), to increase the discriminating potential of the approach: in the forward experiment we incubated the specific antibody with a heavy (H)-labelled chromatin preparation while the excess of soluble peptide was added to the light (L)-labelled chromatin input; in the “reverse” format the two samples were inverted.

After incubation, the H and L immunopurified chromatins were pooled and extracted proteins were separated by SDS-PAGE (Figure 27C).

Figure 27. Scheme of the X-ChroP strategy, combining X-ChIP and SILAC quantitation. A) Scheme of the experimental approach. B) DNA isolated after shearing by sonication, resolved on a 1.3% agarose gel and stained with ethidium bromide. The sample prepared for X-ChIP usually contains fragments of 300-500bp. C) SDS-PAGE of co-immunoprecipitated proteins: lanes were sliced in ten pieces, digested with trypsin and analyzed by LC-MS/MS.



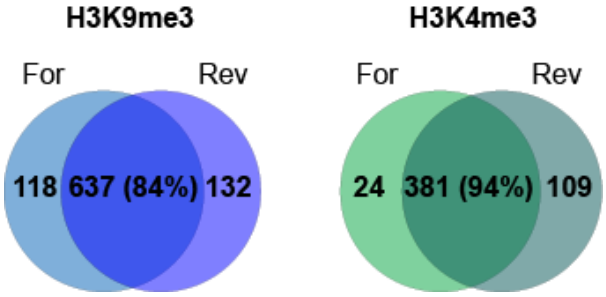




Samples were subjected to trypsin digestion, followed by MS analysis. Proteins were identified and quantified by the MaxQuant software (158). Using a confidence level of 99% (protein FDR 1%), 637 and 381 proteins were identified in the H3K9me3 and H3K4me3 X-ChIPs, respectively (Figure 28A and 28B).

Figure 28. Proteins identified and quantified in H3K9me3 and H3K4mee3 ChIPs. A) Venn Diagrams show the overlap of identified and quantified proteins (Ratio Count (RC)≥1) in two experimental replicates of ChIPs for the modification of interest: H3K9me3 (left) and H3K4me3 (right), respectively (For=Forward, Rev=Reverse). B) Table of identified and quantified proteins for H3K9me3 and H3K4me3 X-ChIPs.

A)



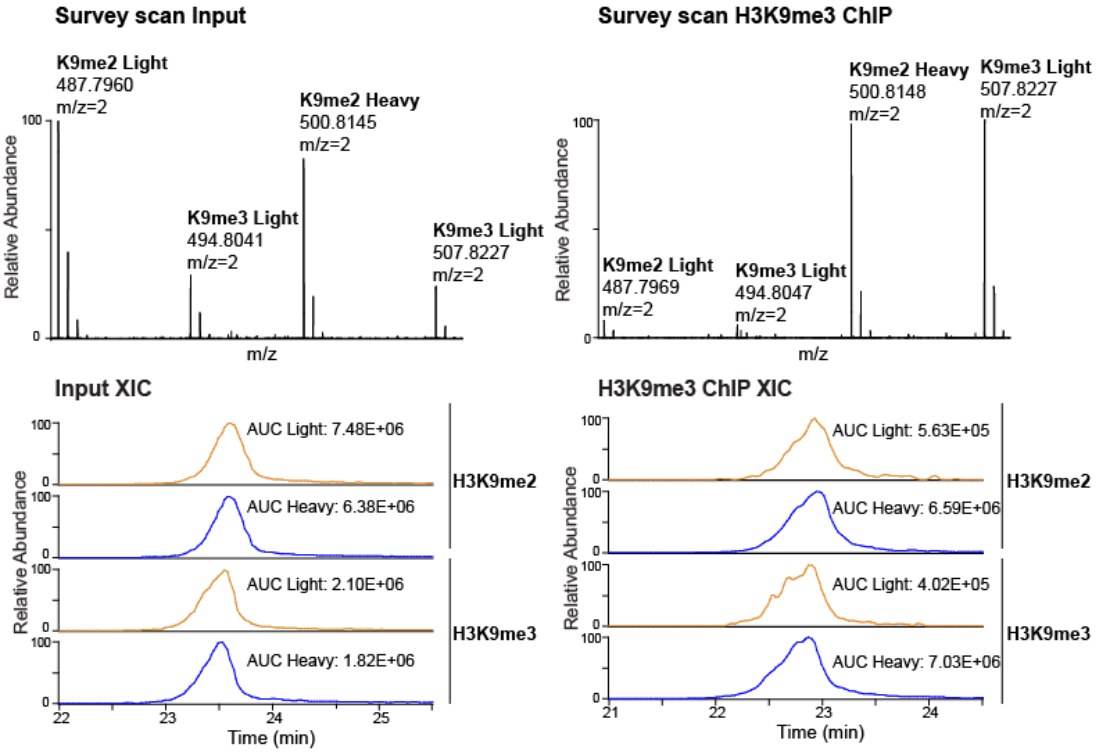
B)

	H3K9me3		H3K4me3	
	For	Rev	For	Rev
Proteins identified Unique peptides ≥ 1 Non unique peptides ≥ 2	776	790	426	519
Proteins identified Unique peptides ≥ 1 Non unique peptides ≥ 2 Ratio Count ≥ 1	755	769	405	490
Proteins identified Unique peptides ≥ 1 Non unique peptides ≥ 2 Ratio Count ≥ 2	711	705	362	422
Top 40% Protein Ratio	214		100	
False Discovery Rate (FDR)	1%			
Minimal peptide length (aa)	6			
Full MS mass tolerance	6 ppm			
MS/MS mass tolerance	0.5 Da			

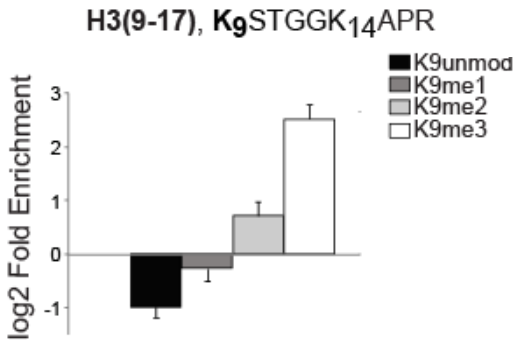
The specific enrichment of the modifications used as bait (H3K9me3 and H3K4me3), together with the corresponding depletion of the unmodified and mono-methylated forms, confirmed the efficiency and specificity of the X-ChIP (Figure 29A, 29B and 29C).

Figure 29. Validation of the X-ChroP approach. A) Zoomed mass spectra for the 2+ charge state of the H3K9me2 and H3K9me3 peptides (9-17), in both light and heavy forms (upper panel) and extracted ion chromatograms (XIC) constructed for each precursor *m/z* value (lower panel). B) Relative enrichment of K9me3 after H3K9me3 X-ChIP at MS level. C) Validation of H3K4me3 enrichment in the corresponding Forward (For) and Reverse (Rev) X-ChIPs: aliquot of ChIP-ed and input samples were probed with α-H3K4me3 antibody. Unmodified H3 is used as loading control.

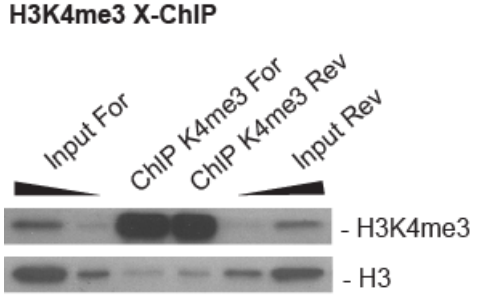
A)



B)



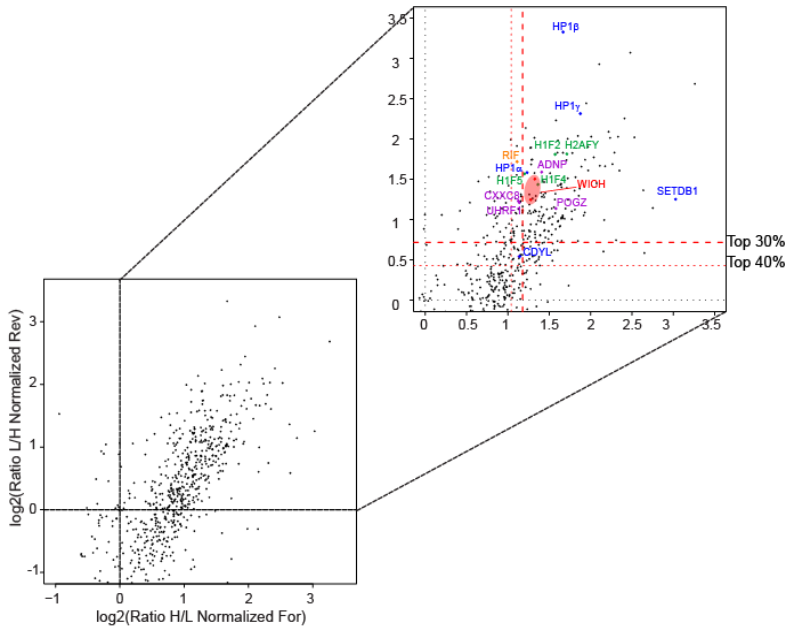
C)



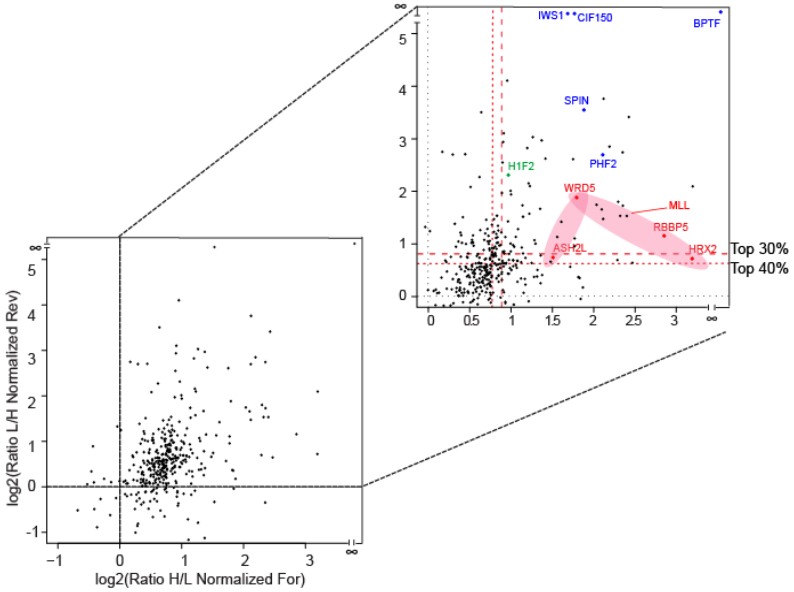
A $\log_2$ plot of the H/L ratios of the two sets of identified proteins from forward and reverse experiments allowed unambiguous identification of binders (Figure 30A and 30B for H3K9me3 and H3K4me3, respectively).

Figure 30. The heterochromatome and euchromatome identified with the X-ChroP. Proteins are plotted by their SILAC-ratios in the first (x axis) and second (y axis) SILAC experiment for H3K9me3 (A) and H3K4me3 (B) (For=Forward, Rev=Reverse). Specific interactors should lie in the upper right quadrant enlarged, close to the diagonal. Red, dotted lines represents the cutoffs, selecting the top 40% and 30% of protein. In Blue are highlighted already annotated specific interactors, in green the histone variants and inorange/violet proteins associated with heterochromatin.

A)



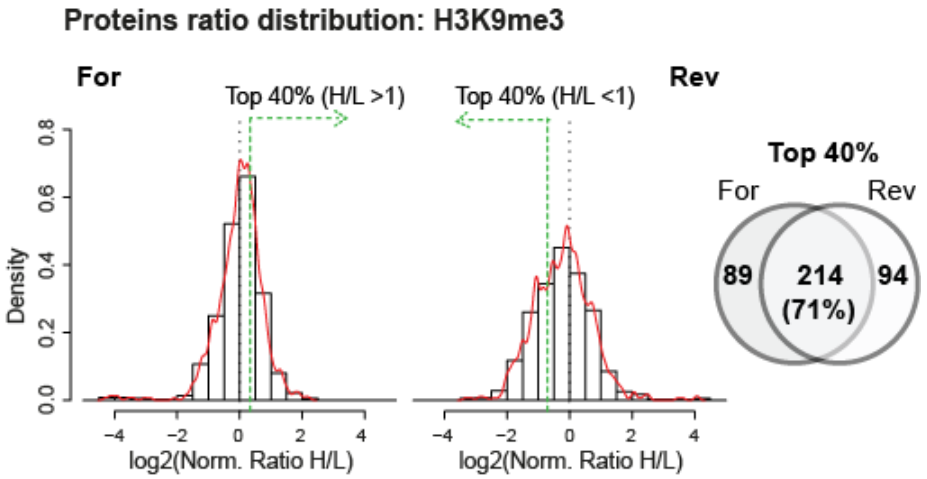
B)



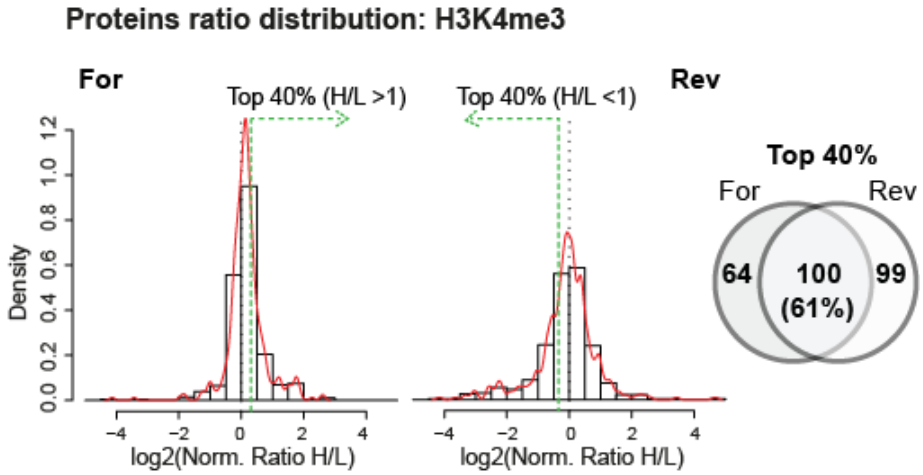
The overall spread distribution of the SILAC ratios observed in the scatter plots can be explained by the fact that completely independent cells and chromatin preparations were used as input for each X-ChIP in the "forward' and "reverse" experiments, which thus represent full biological replicates, resulting in increased variability. Yet, specific interactors could be selectively distinguished based on their SILAC H/L ratios (H/L>1 in the forward and H/L<1 in the reverse experiment, upper right quadrant of the scatter plot) from the background proteins, which have a constant ratio close to 1. In particular, we considered a protein to be specifically enriched when present in the top 40% of proteins with H/L ratios >1, a filter that, selected 214 and 100 interactors for H3K9me3 and H3K4me3, respectively (Figure 31A and 31B), among which we found several well known K9me3 and K4me3 binders. For the biological follow-up of novel interactors, we however chose even higher stringency and focused on proteins present in the top 20% of the ratios.

Figure 31. Features of the H3K4me3 and H3K9me3 interactomes. A) Protein ratio distribution for the H3K9me3 ChIP Forward (For) and Reverse (Rev) experiments. Venn Diagrams of Top 40% of protein ratios for two H3K9me3 X-ChIP replicates, with 71% overlap. B) Protein ratio distribution for the H3K4me3 ChIP For and Rev experiments. Venn Diagrams of Top 40% of protein ratios for two H3K4me3 replicates, with 65% overlap.

A)



B)


Proteins ratio distribution: H3K4me3
For
Top 40% (H/L >1)
Top 40% (H/L <1)
Rev
Density
0.0 0.2 0.4 0.6 0.8 1.0 1.2
−4 −2 0 2 4
log2(Norm. Ratio H/L)
−4 −2 0 2 4
log2(Norm. Ratio H/L)
Top 40%
For
Rev
64
100
(61%)
99

5.4 X-ChroP characterizes novel players in H3K9me3 and H3K4me3 chromatomes

Several of the top 40% heterochromatic interactors were previously ascribed to this region by independent biochemical experiments and were our positive controls (see table in Appendix IV): the three isoforms (α, β, γ) of heterochromatin protein 1 (HP1) (176), CDYL1 (177), SETDB1, the euchromatic HMT of the Suv39 family that tri-methylates H3K9 (178, 179), the DNMT1 (180); and its interactor UHRF1 (181), whose recognition of H3K9me3 is inhibited when by the acetylation of the same residue (181). RIF1, whose localization to a subset of replicating pericentromeric heterochromatin has been described (182). In addition to direct H3K9me3 binders, our approach also enabled the identification of novel and probably indirect interactions among heterochromatic proteins: for instance, we detected the HP1-associated proteins ADNP and POGZ (183, 184) and KDM2A, a H3K36 demethylase that associates indirectly to heterochromatin by HP1 binding (148, 185, 186).

In the H3K4me3 *chromatome* (see table in Appendix V) we identified Spindlin1 (148) and the subunits of the MLL-complexes WDR5, RbPB5 and ASH2 (187, 188). Furthermore, our screening confirmed various proteins also identified by Vermeulen with the SILAC peptide-pull downs, such as BPTF (189), IWS1 (190), PHF2 and various TBP-associated factors (TAFs) of the TFIID complex (88).

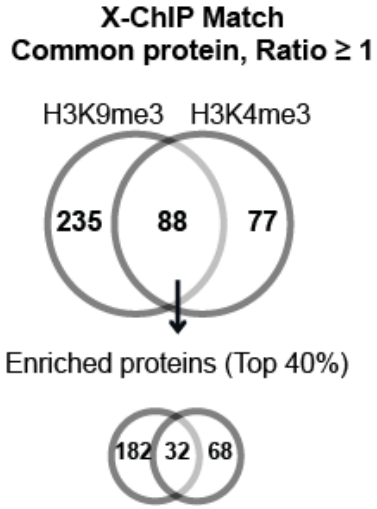
These results overall confirmed the robustness of our approach and allowed screening for novel interactors with higher confidence. Among the newly identified heterochromatin components (Table in Appendix IV), we found heterochromatin protein 1-binding protein 3 (HP1-BP74) (191). Since the role of HP1-BP74 is still elusive, our result suggests a potential role in formation and/or maintenance of the compacted chromatin structure, through HP1 binding. We also identified different scaffold attachment factors (SAFs) including SAFB1, SAFB2 and SAFA, which bind the AT-rich scaffold/matrix attachment regions (S/MARs) of DNA through their scaffold attachment factor-box (SAF-

Box) (192). Other members of the family found in the H3K9me3 interactome are: SAFB-like transcriptional modulator (SLTM) (193) and Matrin3 (MATR3), shown to bind to S/MARs and to localize in heterochromatin regions, similarly to SAFA (194, 195). With this as a basis, SAFB proteins could have a role in higher order heterochromatin organization, by interacting with chromatin remodeling complexes, to achieve a more compact structure.

Among novel H3K4me3-interactors we identified the FACT (*facilitates chromatin transcription*) complex, a heterodimer composed of two subunits of 140kDa and 80kDa (196). FACT is essential for Pol II-driven transcription on chromatin templates (196) and it has an intrinsic histone chaperone activity to reassemble histones onto DNA (197). In addition, FACT seems to be associated with actively transcribed Pol II genes in *Drosophila* (198) and in *Arabidopsis* FACT co-localizes with actively transcribed genes, whereas it seems excluded by heterochromatin and intragenic regions (199).

When directly comparing the *chromatomes* of silent and active chromatin we interestingly observed also minor subset of proteins equally, yet specifically, enriched in both fractions (Figure 32): for instance, SMARCC2 and SMARCA5, subunits of the ATP-dependent chromatin remodeling complex SWI/SNF (200) and HMGB2, a member of the high mobility-group family, (201). In addition, we identified various structural components of chromatin, with functions in chromosome condensation and separation, such as SCC112 that binds the Cohesin complex and associates with chromatin throughout the cell cycle, regulating sister chromatid cohesion during mitosis (202-204). Interestingly, the vast majority of these proteins homogeneously distributed exert more universal structural roles in chromatin homeostasis and dynamics, processes not linked to the transcriptional state. This may be suggestive of similar roles for those proteins in the same class whose function is still uncharacterized.

Figure 32. Proteins in common between the H3K9me3 and H3K4me3 interactomes. Chromatin purifications identify both distinct and common sets of proteins in the H3K4me3 and H3K9me3 interactomes: Venn diagrams show numbers of proteins identified and enriched (Ratio ≥ 1) in both data sets (upper) and present in the top 40% of proteins ratios (lower).



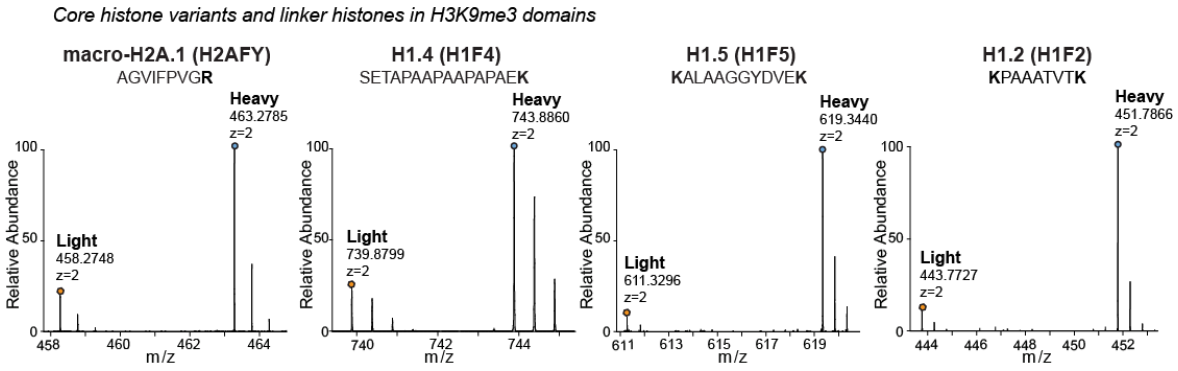
5.5 Accumulation of histone variants at specific chromatin domains

A great advantage offered by the ChroP approach in respect to other strategies is the possibility to characterize histone variants enriched at specific chromatin domains: histone macro-H2A.1 (mH2A1), was found among the top 20% interactors of H3K9me3 chromatin suggesting its prevalent heterochromatic localization (Figure 33A), in line with evidences of a putative structural role of this variant in organizing constitutive heterochromatin (205, 206).

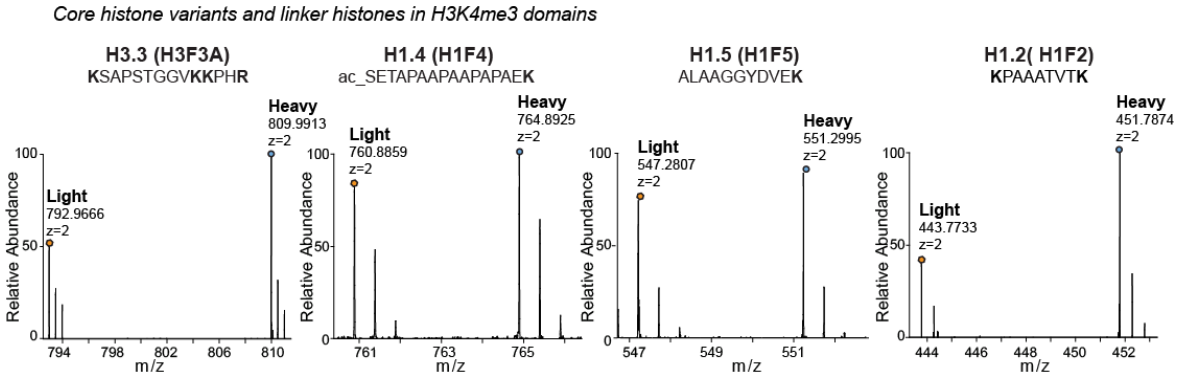
The H3.3 variant was instead overrepresented in the H3K4me3 regions (Figure 33B and Figure 33C). H3.3 is incorporated into chromatin in a replication-independent manner upon chromatin assembly following active transcription of genes; consequently it localizes mainly at the promoters of active genes, it is decorated by active marks and contributes to the epigenetic memory maintenance of active regions (68, 114, 115, 207). Interestingly, it has been suggested that H3.3 reduces the association of linker histone H1 to chromatin, thus causing a pronounced drop of H1 occupancy at TSSs and at more distant cis-regulatory sites of active genes (208). We confirmed this theory finding that H1.4 and H1.5 were increased in H3K9me3 chromatin, while not changed in H3K4me3 domains (Figure 33A and 33B, respectively) (209). Instead, the SILAC ratio of H1.2 suggested its ubiquitous distribution (Figure 33A and 33B), in apparent contradiction with previous reports proposing its euchromatic occurrence (210). The accumulation of individual H1 variants in either active or repressive chromatin suggests their specific contribution to establish or maintain the functional status of these regions. Intriguingly, when investigating the modification patterns of these variants, we found evidence of at least one novel mono-methylation on Lysine 90 of H1.2/H1.4, corresponding Lysine 93 of the H1.5 variant (Figure 34A and 34B). This preliminary evidence opens the path to the further characterization of PTMs of the H1 variants, at present less investigated, due to the lack of the adequate variant-specific antibodies.

Figure 33. Histone variants enriched in heterochromatin and euchromatin. A) Mass spectra of light and heavy SILAC peak pairs from H2AFY, H1.4, H1.5 and H1.2 demonstrating a specific enrichment of these proteins in the K9me3 ChIP-ed material (heavy), over the mock control (light). B) Mass spectra of light and heavy SILAC peak pairs from H3.3, H1.4, H1.5 and H1.2 demonstrating a specific enrichment of this variant in the K4me3-ChIPed material over the mock control. C) Confirmation of H3.3 enrichment in H3K4me3 domains by standard ChIP followed by WB.

A)



B)



C)

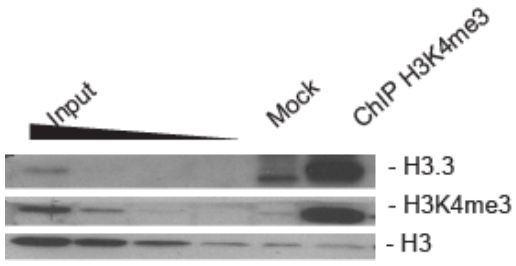
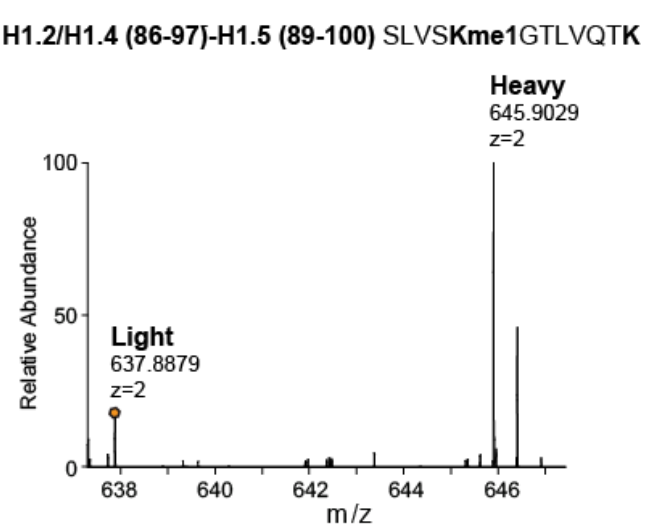
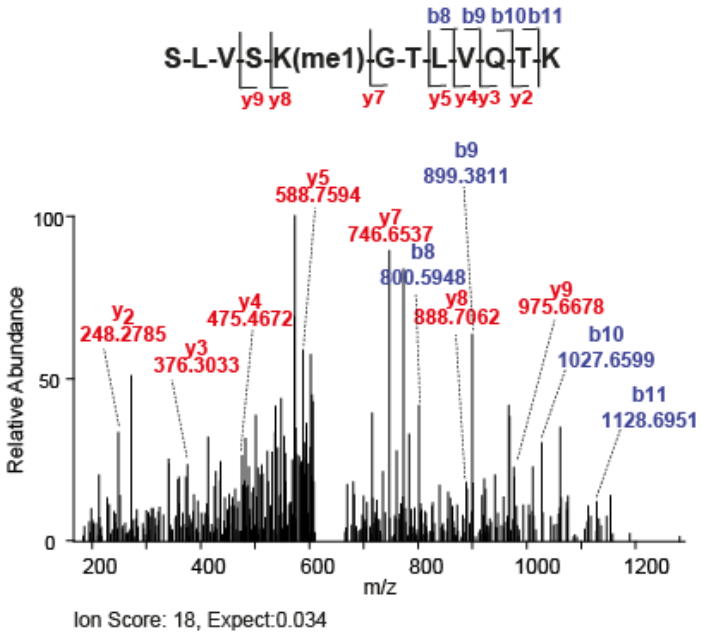


Figure 34. Novel Lysine 90 or Lysine 93 mono-methylation associate to H1.2/H1.4 and H1.5, respectively. A) Zoomed mass spectrum of precursor ions at m/z $[637.8879]^{2+}$ and $[645.9029]^{2+}$ corresponding to the peptide SLVSKme1GTLVQTK of linker H1.2, H1.4 and H1.5 species, in light and heavy forms. B) MS/MS spectrum of the peptide H1.2/H1.4 (86-97) or H1.5 (89-100), from which sequence and methylation was detected.

A)



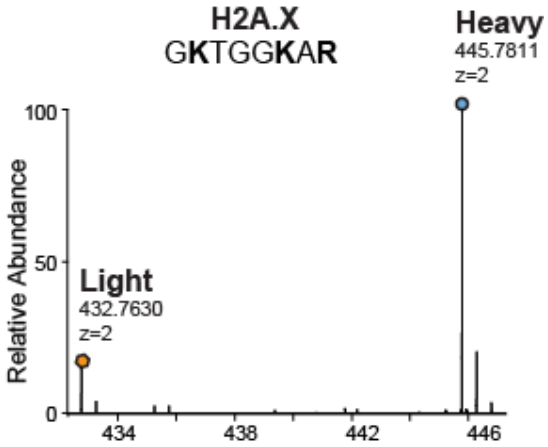
B)



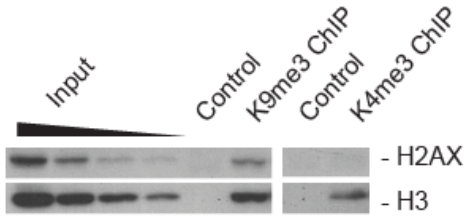
The enrichment of H2A.X in heterochromatin (Figure 35A) was the most unexpected finding, given the prevalent literature that focuses only on its Ser139-phosphorylated form (γ-H2A.X), typically described at the DNA damage foci (211, 212). Independent validation of the SILAC-based results by standard ChIP, followed by WB, confirmed the accumulation of H2A.X in chromatin regions marked by H3K9me3 (Figure 35B and 35C).

Figure 35. H2A.X heterochromatic enrichment, measured at the protein level. A) Mass spectra of light and heavy SILAC peak pairs from H2AX in the forward H3K9me3 X-ChIP: the H/L ratio >1 indicate specific enrichment of this protein. B) Western blot of unmodified H2A.X upon ChIP with α-K9me3 and α-K4me3. C) Western blot analysis of H3K9me3 and H3K4me3 upon ChIP with α-H2A.X (#1 and #2 indicate two replicates). Unmodified H3 is the loading control in both cases.

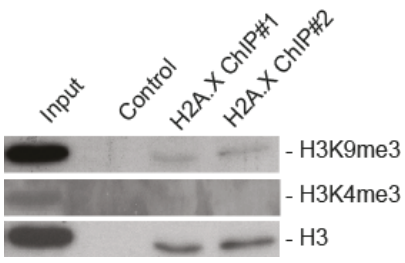
A)



B)

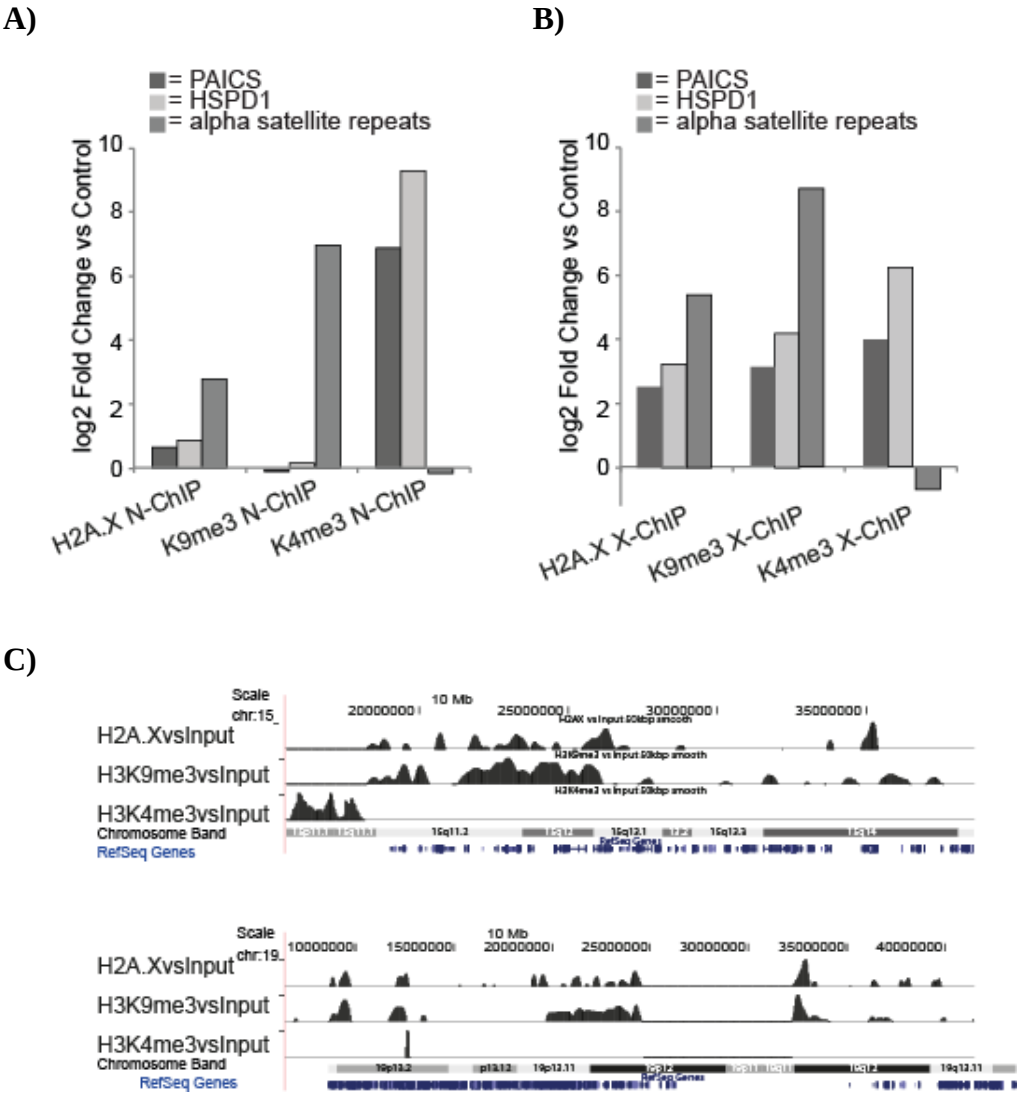


C)



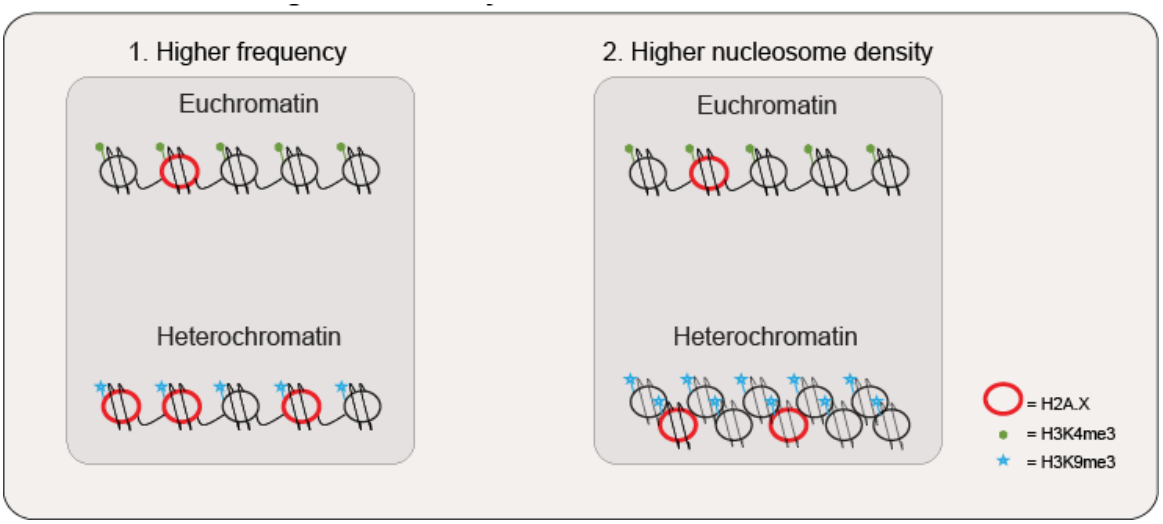
In addition, the measurement of α-satellite repeats and of two H3K4me3 target genes (213) by qPCR in α-H2A.X ChIPs indicated that this variant enriches for heterochromatic DNA regions but not for active genes (Figure 36A and 36B). Finally, H2A.X profiles in ChIP-Seq strongly correlated with H3K9me3, not with H3K4me3 (Pearson's r: 0.796 and 0.139, respectively) (Figure 36C).

Figure 36. H2A.X heterochromatic enrichment, measured at the gene level. qRT-PCR measures the levels of α-satellite repeat regions in H2A.X, H3K9me3 and H3K4me3 domains IP-ed by N-ChIP (A) and X-ChIP (B), over the mock control. The actively transcribed genes PAICS and HSPD1 are used as negative controls. C) ChIP-Seq profile of H2A.X, H3K9me3 and H3K4me3 compared to the input, across regions of human chromosomes 15 (chr15:15,396,089-38,421,489) and 19 (chr19:6,465,305-41,462,000).



In order to reconcile this compelling set of results with the general evidence that DNA damage foci marked by γ-H2A.X appear both in silenced and actively transcribed chromatin, we formulated a *higher local density* model, where H2A.X is present all along the genome, but its accumulation in repressed regions results either from a higher frequency of the variant in heterochromatin than in euchromatin within poly-nucleosomal stretches of the same length (the beads-on-string structure), or from a general higher density of nucleosomes in heterochromatin, as a consequence of narrower nucleosome spacing and/or a more compacted higher-order structure of chromatin (30nm fibers or loops) (Figure 37). Accurate measurement of H2A.X enrichment using as input nucleosome stretches of distinct and defined lengths, obtained via fine tuned MNAse digestion, followed by CsCl purification, may enable to dissect the model (214, 215).

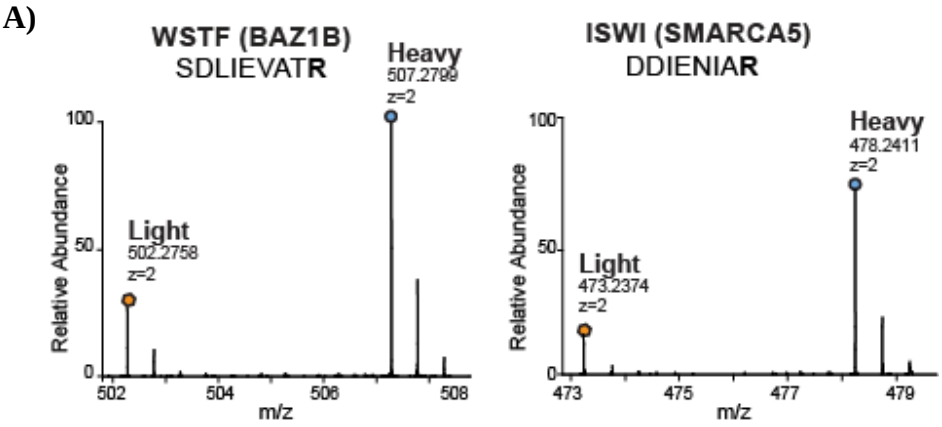
Figure 37. Model of "higher local density" of H2A.X in heterochromatin

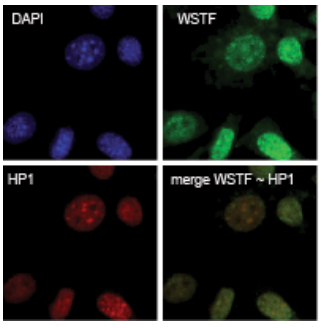


5.6 Heterochromatic enrichment of the WICH complex, recruited to H2A.X

An in depth analysis of the H3K9me3-interactome revealed a reproducible overrepresentation (top 20%) of the WICH (*WSTF-ISWI Chromatin remodeling*) complex (Figure 38A), an observation that we corroborated by immunofluorescence (Figure 38B). WICH consists of two subunits: ISWI, the ATPase subunit common to several remodeling complexes that mediate nucleosome positioning (216), and WSTF, encoded by the *BAZ1B* gene. Descriptive studies by immunofluorescence had already observed the accumulation of WICH in replicating pericentric heterochromatin (217), however these observation were not followed by studies elucidating the mechanism of its recruitment and function in this region.

Figure 38. WICH involvement in heterochromatin. A) Mass spectra of light and heavy SILAC peak pairs from WSTF and ISWI: the H/L ratios >1 indicate specific enrichment of these proteins. B) WSTF co-localize with HP1β, marker of pericentromeric heterochromatin (merge) (1 pixel=0.172 mm; original magnification 60X).



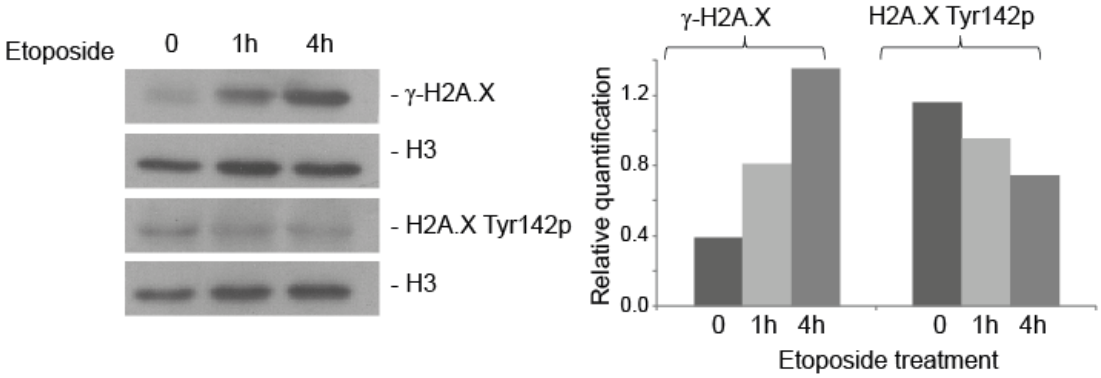


The interaction between H2A.X and the WICH complex has been recently proven, together with the phosphorylation of Tyrosine 142, in bulk extract, of the same variant by WSTF. Phosphorylation of Tyr142 seems to be present in basal conditions, but decreases upon double strand breaks (DBSs) (218) (Figure 39A), by EYA-mediated de-phosphorylation (219). These observations indicate that Tyr142p loss may correlate with γ-H2A.X appearance; in line with this idea it has been proposed that the two neighboring phosphorylations on Try142 and Ser139 may function as a molecular switch that acts as an additional level of regulation of the DNA damage response (DDR) (218). Additionally, we discovered an enrichment of this Tyr142p in H3K9me3 domains (Figure 39B), opening the possibility that this modification could be associated with heterochromtin.

Taking together these observations, we thus hypothesize that in basal conditions the higher local density of H2A.X may lead to preferential Tyr142 phosphorylation in heterochromatin by the accumulating WICH complex and that a delayed DDR, marked by γ-H2A.X, may be elicited in this region upon DSBs-induction as a consequence of the additional required step of Tyr142 de-phosphorylation (Figure 40).

Figure 39. H2A.X Y142p enrichment in H3K9me3-enriched domains. A) Left: WB analysis of γ-H2A.X and H2A.X Tyr142p in the Input used for X-ChIP in untreated cells and 1 and 4 hours after etoposide treatment, with chat expressing relative quantitation upon normalization with loading control H3 (right). B) WB of unmodified H2A.X and H2A.X-Tyr142p (red arrow), upon ChIP with α-K9me3 and α-K4me3. Unmodified H3 is the loading control. Black arrow indicates an unspecific band at lower molecular weight, also detected in control. Red upper arrow indicates the specific signal for Tyr142p.

A)



B)

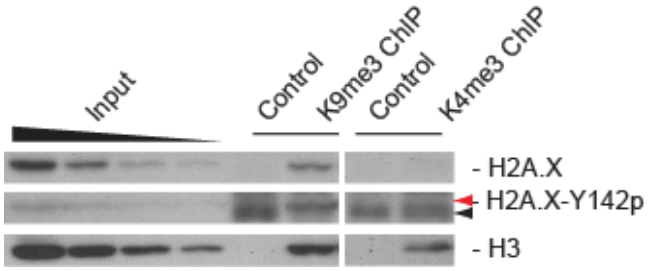
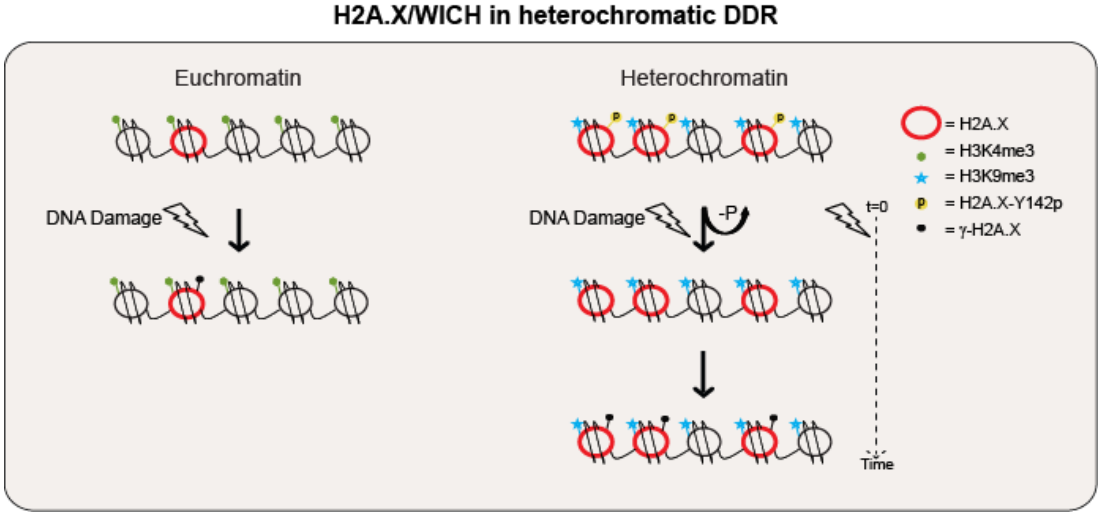


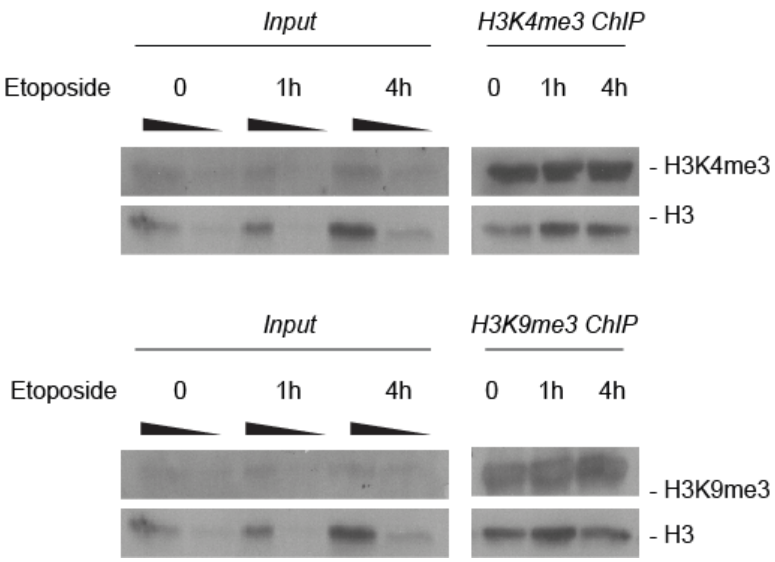
Figure 40. Model of H2A.X/WICH involvement in DDR, with temporal shift of γ-H2A.X appearance upon DSBs between eu- and heterochromatin.



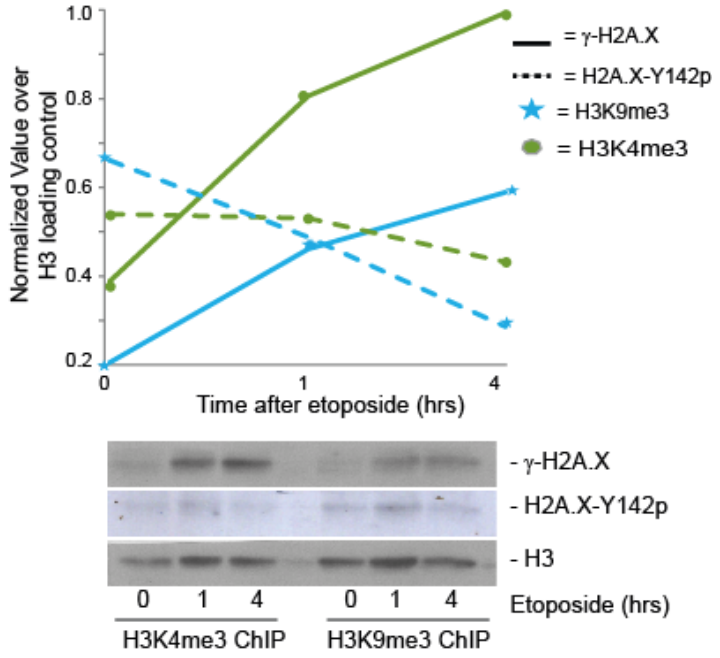
ChIP experiments with α-H3K9me3 and α-H3K4me3 (Figure 41A) followed by WB to monitor the phosphorylation state of H2A.X confirmed the enrichment of H2A.X Tyr142p in heterochromatin versus euchromatin in basal conditions (time point 0 hrs). The same assay carried out at two time-points after etoposide treatment suggested that the decline of Tyr142p in silenced chromatin slightly precedes the appearance of γ-H2AX mark (Figure 41B). To confirm this result we repeated the experiment, extending the kinetics of DNA damage until 16h (Figure 41C).

Figure 41. Evaluation of γ-H2A.X and H2A.X Y142p levels upon DNA damage. A) WB validation of H3K9me3 and H3K4me3 enrichment in the corresponding ChIPs: an aliquot of both input and ChIP samples were loaded on SDS-PAGE, transferred on PVDF and probed with α-H3K4me3 (up) α-H3K9me3 (down). B) Upper chart: normalised level of γ-H2AX and H2AX Y142p for H3K4me3 and H3K9me3 chromatin plotted at basal level and after at two time points (1 and 4 hrs) upon etoposide treatment and after X-ChIP enrichment, as monitored by WB (bottom panel). The signal for unmodified H3 was used as the control to normalize for loading variation. C) WB validation of γ-H2AX and H2AX Y142p levels in H3K4me3 and H3K9me3 domains at basal level and after four time points (2, 4, 8 and 16 hrs) upon etoposide treatment.

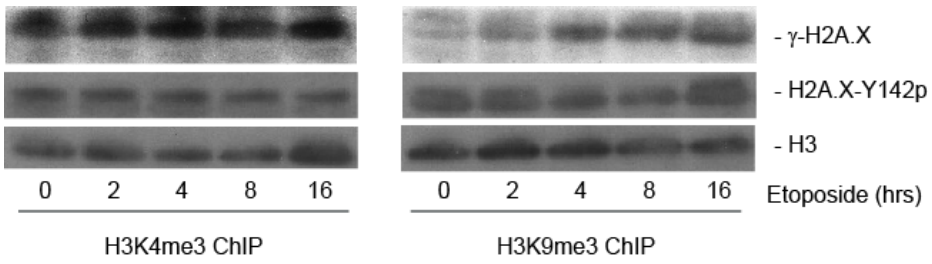
A)



B)


1.0
0.8
0.6
0.4
0.2
Normalized Value over H3 loading control
0
1
4
Time after etoposide (hrs)
= γ-H2A.X
= H2A.X-Y142p
= H3K9me3
= H3K4me3
- γ-H2A.X
- H2A.X-Y142p
- H3
0 1 4 0 1 4
Etoposide (hrs)
H3K4me3 ChIP
H3K9me3 ChIP

C)


- γ-H2A.X
- H2A.X-Y142p
- H3
0 2 4 8 16
0 2 4 8 16
Etoposide (hrs)
H3K4me3 ChIP
H3K9me3 ChIP

6. DISCUSSION

Chromatin is a highly structured nucleoprotein complex made of histone proteins and DNA that controls nearly all DNA-dependent processes. Chromatin plasticity is regulated by different associated proteins, post-translational modifications on histones (hPTMs) and DNA methylation, which act in a concerted manner to enforce a specific "chromatin landscape", with a regulatory effect on gene expression. Mass spectrometry-based proteomics has emerged as a powerful analytical method for the analysis of histone proteins, their post-translational modifications and variants, as well as their associated "writers" and "readers". ChroP approach represents a novel analytical strategy for the proteomic investigation of chromatin at the resolution of a few nucleosomes. ChroP uses a modified version of ChIP to isolate native mononuclesomes up to crosslinked 500bp oligonucleosomal stretches derived from distinct chromatin domains. Amount, purity and quality of the isolated chromatin enable the subsequent mass spectrometric examination of the protein component, providing information on histone modification patterns, variants and interactors. For the optimization of the method, we focused on two distinct, non overlapping chromatin compartments: on the one hand inactive pericentromeric heterochromatin and transcriptionally repressed patches of chromatin in actively transcribed areas, marked by H3K9me3 and on the other hand active promoters in euchromatin, characterized by H3K4me3.

The analysis of the PTMs enrichment on histone 3 in the H3K9me3 and H3K4me3 domains confirmed the results recently described using IP of HPLC-purified H3 (28), thus verifying the selectivity of the antibodies used in our study. Overall the annotation of the H3K9me3 *modificome* revealed a significant enrichment of known heterochromatic marks, with the corresponding depletion of active modifications, whereas an opposite trend was observed in the H3K4me3 domains, as expected. The accordance of our results with previous studies describing PTMs patterning within the same histone molecule proved the

robustness of the strategy, for the following investigation of novel PTMs; more importantly, ChroP exhibits a unique strength in revealing PTMs associations also between the different core histones within the same mono-nucleosome, enriched at quasi-purity in N-ChIP.

The SILAC-based investigation of the corresponding *interactomes* by X-ChIP confirmed several previously described interactions, thereby validating our method; in addition, we identified numerous novel candidate interactors, which have not been experimentally described previously.

The remarkable advantage offered by ChroP consists in the fact that homogeneous 300-500bp nucleosomal stretches, in which weak protein–protein interactions are stabilized by formaldehyde crosslinking, are purified and MS-analyzed: this enables the dissection of not only modification "readers", but also more complex architectural structures, resulting from both direct and indirect interactions within intact chromatin domains. For instance, the selected H3K9me3 interactome revealed the enrichment of both KDM2A and HP1, which previously could be functionally associated only by the intersection of distinct independent assays, such as peptide- and GFP- pull downs (147). This gain comes at the cost of simplicity of interpretation of the proteomic readout; hence these new composite hierarchical protein architectures need to be further dissected (220).

The two facets of ChroP, N- and X-ChIP, are highly complementary, with the possibility of arranging in a unique puzzle the different pieces composing chromatin architecture: for instance, the analysis of the H3K9me3 *modificome* reveals an unexpected depletion of H3K36me2/me3, which is explained by the recruitment of KDM2A found by the interactomics investigation of the same region.

Besides recapitulating known interactors, ChroP indicates putative localization in distinct chromatin compartments for histone variants and binders not previously described; a representative case is offered by the H1 isoforms: the accumulation of individual H1 variants in either active or repressive chromatin suggests their specific contribution to

establish or maintain the functional status of these regions. Intriguingly, when investigating the modification patterns of these variants, we found evidence of at least one novel mono-methylation. This preliminary evidence opens the path to the further characterization of PTMs on H1 variants, at present less investigated, due to the lack of the adequate variant-specific antibodies.

The presence of the WICH complex and H2A.X in the H3K9me3 *chromatome* is particularly exciting: the "higher local density" of H2A.X in heterochromatin can be further developed including the evidence of WICH heterochromatic enrichment, thus leading to the hypothesis that H2A.X accumulating in heterochromatin is preferentially phosphorylated on Tyr142 by WSTF. Since it has already been proposed that Tyr142p and Ser139p function as a molecular switch, we believe that WICH and Tyr142p provide an additional regulatory step of DDR in heterochromatin, with an impact on the genomic stability in this chromatin region. Some evidence seems to support our model: first, γ-H2A.X foci are largely excluded from heterochromatin as compared to active euchromatic compartments both in human cells and budding yeast (221-223); second, in both human and mouse embryonic fibroblasts, radiation-induced DSBs associated with condensed chromatin are repaired more slowly than in euchromatin (224). Despite these suggestions, the mechanism underlying the delayed and diminished γ-H2A.X signal in heterochromatin remains unclear. According to our model, the delay in γ-H2A.X occurrence in heterochromatin may be due to the presence of basal H2A.X Tyr142p that must be erased in a regulated manner, preceding γ-H2A.X appearance. In euchromatin, γ-H2A.X takes place more quickly, in early DDR, since this intermediate regulatory step is not present. Our results represent a very preliminary validation of this model and the first step towards the understanding of how the cells may adjust the DNA repair process in relation to chromatin compartmentalization (225). In this perspective, the observed correlation between aberrant expression of Su(var)3-9, with the consequent alteration of

H3K9me3 pattern and HP1 binding and the genomic instability of heterochromatin during mammalian development, is highly suggestive (226-228).

The recruitment of WICH to heterochromatin cannot be explained solely by the described binding of WSTF to H2A.X; in fact, although our results suggest a higher local density of this variant in heterochromatin, the appearance of the Ser139-phosphorylated isoform *in both* silent and active regions upon DNA damage is not deniable; hence, additional mechanisms, including either the binding of the complex to co-clustering hPTMs or its interaction with co-enriching proteins, must be postulated. Further studies on WSTF *interactomes*, both soluble and chromatin-bound, will address this open question

Overall the ChroP approach described here offers the possibility to dissect the synergism of hPTMs, variants and non-histonic interactors at functionally distinct chromatin domains, with a resolution of mono- to oligo-nucleosomes. The approach is relatively easy to setup and to be implemented in the epigenetics groups, given the limited changes made to the conventional N- and X- ChIP protocols, used for ChIP-Seq studies. Hence, for the relatively straightforward optimization, we predict that ChroP will be amenable to numerous applications in more functional studies, for instance to study the dynamic changes of the *modificome* and *intractome* of specific chromatin regions upon various perturbations, such as global transcriptional activation, differentiation, depletion of distinct chromatin components, or treatment with epigenetic drugs. Therefore our ChroP strategy emerges as an additional and valuable tool in the arsenal of analytical strategies available to study chromatin composition and dynamics in a system-wide fashion.

7. CONCLUSIONS AND PRESPECTIVES

During these years I have established a novel proteomics approach combining two common and complementary methods, typically employed separately by the researchers to investigate the epigenetic components of chromatin: chromatin immunoprecipitation and Mass Spectrometry analysis. On the one hand MS served to characterize the PTMs co-associated to enriched sub-domains and on the other hand to annotate the corresponding "chromatin interactome", representing all the non-histonic proteins specifically interacting within the same chromatin domains. The results that I have obtained characterizing the composition of heterochromatin and euchromatin support the vision that chromatin is a composite puzzle, made of many pieces, which coordinately act to regulate the functional state of the underlying DNA. Due to this complexity, novel approaches such as ChroP, based on the combination of different analytical techniques, may represent the solution to better understand how the structural and functional interactions among different epigenetic components of chromatin enforce specific functional transition on the genome.

Although ChroP proved to be a very robust approach and highly complementary to more conventional strategies to analyze chromatin, I have envisaged two minor limitations of the project: the first is more methodological, and the second regards the application of the method to functional studies.

Concerning the methodological aspect, although N-ChroP followed by "Bottom Up" MS analysis enables to characterize the modifications associated with distinct chromatin domains, the detection of the combinatorial aspect of the histone code has been only partially addressed in my study. As discussed in the introduction (see paragraph 2.2.2), the "Bottom Up" analysis with CID fragmentation offers only a partial view of the complex cross-talks among different hPTMs; therefore I would be interested in analyzing the nucleosomes purified by N-ChroP with a combination of "Bottom Up" and "Middle and Top Down" approaches, to effectively achieve a more comprehensive view on the

hPTMs stochiometry and abundances at specific regions and to finally “crack” the code. In fact, “Middle and Top Down” methods are employed to study longer peptides up to intact proteins, respectively. The longer peptides can be generated either spontaneously by missed cleavage of trypsin or by using different proteases, such as Asp-N and Glu-C. The fragmentation technique employed to analyze those longer and highly charged peptides is electron transfer dissociation (ETD). Recently, ETD was also employed to improve the detection of Arginine methylation; hence I plan to focus on histone Arginine methylation, a modification that is less characterized than Lysine methylation, but instead equally biological relevant.

The second limitation of my study regards the fact that I restricted the use of ChroP for the analysis to chromatin domains of HeLa cells, which makes my study an optimal proof of principle but not particularly relevant in a functional perspective. In fact HeLa were a useful cell line for the set up of the method because they are easy to culture and to scale up; furthermore the hPTMs patterns are largely characterized in bulk. However, they are not representative of specific cell type or specific function. In this respect, I think that applying this approach to a different model system and in a dynamic manner will be useful to profile the epigenetic changes during functional transitions, such as a global wave of transcriptional activation or a change from healthy condition to disease. In the follow up of my project, we started to explore distinct chromatin domains using macrophages as a model system, upon LPS stimulation. Moreover, we intend to use ChroP for a more in-depth dissection of euchromatin, investigating *cis*-regulatory regions characterized by H3K4me1 (enhancers) and H3K4me3 (transcription start sites, TSSs). Taking advantage of these two modifications as bait, we started to use N-ChIP and X-ChIP to enrich these regions and to use MS to analyse hPTMs, variants and protein complexes, before and after LPS stimulation. As such we will be able to quantitatively profile how they dynamically change in a coordinated manner, upon inflammation.

8. REFERENCES

1. Kornberg, R. D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868-871.
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389, 251-260.
3. Dillon, N. (2004) Heterochromatin structure and function. *Biol Cell* 96, 631-637.
4. Beisel, C., and Paro, R. (2011) Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet* 12, 123-135.
5. Grewal, S. I., and Jia, S. (2007) Heterochromatin revisited. *Nat Rev Genet* 8, 35-46.
6. Probst, A. V., Dunleavy, E., and Almouzni, G. (2009) Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol* 10, 192-206.
7. Waddington, C. H. (2011) The epigenotype. 1942. *Int J Epidemiol* 41, 10-13.
8. Berger, S. L. (2007) The complex language of chromatin regulation during transcription. *Nature* 447, 407-412.
9. Nightingale, K. P., O'Neill, L. P., and Turner, B. M. (2006) Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr Opin Genet Dev* 16, 125-136.
10. Turner, B. M. (2007) Defining an epigenetic code. *Nat Cell Biol* 9, 2-6.
11. Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009) An operational definition of epigenetics. *Genes Dev* 23, 781-783.
12. Klose, R. J., and Bird, A. P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31, 89-97.
13. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6-21.
14. Amaral, P. P., Dinger, M. E., Mercer, T. R., and Mattick, J. S. (2008) The eukaryotic genome as an RNA machine. *Science* 319, 1787-1789.
15. Mattick, J. S., Amaral, P. P., Dinger, M. E., Mercer, T. R., and Mehler, M. F. (2009) RNA regulation of epigenetic processes. *Bioessays* 31, 51-59.
16. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* 128, 693-705.
17. Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res* 21, 381-395.
18. Campos, E. I., and Reinberg, D. (2009) Histones: annotating chromatin. *Annu Rev Genet* 43, 559-599.
19. Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications. *Nature* 403, 41-45.
20. Spotswood, H. T., and Turner, B. M. (2002) An increasingly complex code. *J Clin Invest* 110, 577-582.
21. Jenuwein, T., and Allis, C. D. (2001) Translating the histone code. *Science* 293, 1074-1080.
22. 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.
23. Martin, C., and Zhang, Y. (2005) The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 6, 838-849.
24. Margueron, R., and Reinberg, D. (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* 11, 285-296.
25. Park, P. J. (2008) Epigenetics meets next-generation sequencing. *Epigenetics* 3, 318-321.

26. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.
27. Ku, C. S., Naidoo, N., Wu, M., and Soong, R. (2011) Studying the epigenome using next generation sequencing. *J Med Genet* 48, 721-730.
28. Peach, S. E., Rudomin, E. L., Udeshi, N. D., Carr, S. A., and Jaffe, J. D. (2012) Quantitative Assessment of Chromatin Immunoprecipitation Grade Antibodies Directed against Histone Modifications Reveals Patterns of Co-occurring Marks on Histone Protein Molecules. *Mol Cell Proteomics* 11, 128-137.
29. Fuchs, S. M., Krajewski, K., Baker, R. W., Miller, V. L., and Strahl, B. D. (2011) Influence of combinatorial histone modifications on antibody and effector protein recognition. *Curr Biol* 21, 53-58.
30. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198-207.
31. Sidoli, S., Cheng, L., and Jensen, O. N. (2012) Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *J Proteomics* 75, 3419-3433.
32. Taverna, S. D., Ueberheide, B. M., Liu, Y., Tackett, A. J., Diaz, R. L., Shabanowitz, J., Chait, B. T., Hunt, D. F., and Allis, C. D. (2007) Long-distance combinatorial linkage between methylation and acetylation on histone H3 N termini. *Proc Natl Acad Sci U S A* 104, 2086-2091.
33. Beck, H. C. (2010) Mass spectrometry in epigenetic research. *Methods Mol Biol* 593, 263-282.
34. Britton, L. M., Gonzales-Cope, M., Zee, B. M., and Garcia, B. A. (2011) Breaking the histone code with quantitative mass spectrometry. *Expert Rev Proteomics* 8, 631-643.
35. Eberl, H. C., Mann, M., and Vermeulen, M. (2011) Quantitative proteomics for epigenetics. *Chembiochem* 12, 224-234.
36. 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.
37. Villar-Garea, A., and Imhof, A. (2006) The analysis of histone modifications. *Biochim Biophys Acta* 1764, 1932-1939.
38. Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durbin, K. R., and Kelleher, N. L. (2011) Analysis of intact protein isoforms by mass spectrometry. *J Biol Chem* 286, 25451-25458.
39. Plazas-Mayorca, M. D., Zee, B. M., Young, N. L., Fingerman, I. M., LeRoy, G., Briggs, S. D., and Garcia, B. A. (2009) One-pot shotgun quantitative mass spectrometry characterization of histones. *J Proteome Res* 8, 5367-5374.
40. Zee, B. M., Young, N. L., and Garcia, B. A. (2011) Quantitative proteomic approaches to studying histone modifications. *Curr Chem Genomics* 5, 106-114.
41. Imhof, A., and Bonaldi, T. (2005) "Chromatomics" the analysis of the chromatome. *Mol Biosyst* 1, 112-116.
42. Bernstein, B. E., Meissner, A., and Lander, E. S. (2007) The mammalian epigenome. *Cell* 128, 669-681.
43. Yates, J. R., Ruse, C. I., and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 11, 49-79.
44. Walther, T. C., and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology. *J Cell Biol* 190, 491-500.
45. Steen, H., and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5, 699-711.
46. Hillenkamp, F., and Karas, M. (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. *Methods Enzymol* 193, 280-295.

47. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* 63, 1193A-1203A.
48. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71.
49. Lagarrigue, M., Lavigne, R., Guevel, B., Com, E., Chaurand, P., and Pineau, C. (2012) Matrix-assisted laser desorption/ionization imaging mass spectrometry: a promising technique for reproductive research. *Biol Reprod* 86, 74.
50. Shukla, A. K., and Futrell, J. H. (2000) Tandem mass spectrometry: dissociation of ions by collisional activation. *J Mass Spectrom* 35, 1069-1090.
51. Kelleher, N. L., Zubarev, R. A., Bush, K., Furie, B., Furie, B. C., McLafferty, F. W., and Walsh, C. T. (1999) Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem* 71, 4250-4253.
52. Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J., and Hunt, D. F. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* 1764, 1811-1822.
53. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 101, 9528-9533.
54. Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K., and McLafferty, F. W. (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* 72, 563-573.
55. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* 8, 2759-2769.
56. Wilm, M., and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem* 68, 1-8.
57. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379, 466-469.
58. Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal Chem* 74, 4235-4249.
59. Shen, Y., Moore, R. J., Zhao, R., Blonder, J., Auberry, D. L., Masselon, C., Pasa-Tolic, L., Hixson, K. K., Auberry, K. J., and Smith, R. D. (2003) High-efficiency on-line solid-phase extraction coupling to 15-150-microm-i.d. column liquid chromatography for proteomic analysis. *Anal Chem* 75, 3596-3605.
60. 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.
61. Bonaldi, T., Imhof, A., and Regula, J. T. (2004) A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics* 4, 1382-1396.
62. Jufvas, A., Stralfors, P., and Vener, A. V. (2011) Histone variants and their post-translational modifications in primary human fat cells. *PLoS One* 6, e15960.
63. Smith, C. M., Haimberger, Z. W., Johnson, C. O., Wolf, A. J., Gafken, P. R., Zhang, Z., Parthun, M. R., and Gottschling, D. E. (2002) Heritable chromatin structure: mapping "memory" in histones H3 and H4. *Proc Natl Acad Sci U S A* 99 Suppl 4, 16454-16461.

64. Garcia, B. A., Mollah, S., Ueberheide, B. M., Busby, S. A., Muratore, T. L., Shabanowitz, J., and Hunt, D. F. (2007) Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* 2, 933-938.
65. Bonaldi, T., Regula, J. T., and Imhof, A. (2004) The use of mass spectrometry for the analysis of histone modifications. *Methods Enzymol* 377, 111-130.
66. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1, 2856-2860.
67. Trelle, M. B., Salcedo-Amaya, A. M., Cohen, A. M., Stunnenberg, H. G., and Jensen, O. N. (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum. *J Proteome Res* 8, 3439-3450.
68. Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006) PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol Cell* 24, 309-316.
69. Trelle, M. B., and Jensen, O. N. (2007) Functional proteomics in histone research and epigenetics. *Expert Rev Proteomics* 4, 491-503.
70. Kelleher, N. L. (2004) Top-down proteomics. *Anal Chem* 76, 197A-203A.
71. Boyne, M. T., 2nd, Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2006) Precise characterization of human histones in the H2A gene family by top down mass spectrometry. *J Proteome Res* 5, 248-253.
72. Breuker, K., Oh, H., Lin, C., Carpenter, B. K., and McLafferty, F. W. (2004) Nonergodic and conformational control of the electron capture dissociation of protein cations. *Proc Natl Acad Sci U S A* 101, 14011-14016.
73. Thomas, C. E., Kelleher, N. L., and Mizzen, C. A. (2006) Mass spectrometric characterization of human histone H3: a bird's eye view. *J Proteome Res* 5, 240-247.
74. Eliuk, S. M., Maltby, D., Panning, B., and Burlingame, A. L. (2010) High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation: determination of abundant histone isoforms and post-translational modifications. *Mol Cell Proteomics* 9, 824-837.
75. Tian, Z., Zhao, R., Tolic, N., Moore, R. J., Stenoien, D. L., Robinson, E. W., Smith, R. D., and Pasa-Tolic, L. (2010) Two-dimensional liquid chromatography system for online top-down mass spectrometry. *Proteomics* 10, 3610-3620.
76. Garcia, B. A. (2010) What does the future hold for Top Down mass spectrometry? *J Am Soc Mass Spectrom* 21, 193-202.
77. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A., and Garcia, B. A. (2009) High throughput characterization of combinatorial histone codes. *Mol Cell Proteomics* 8, 2266-2284.
78. DiMaggio, P. A., Jr., Young, N. L., Baliban, R. C., Garcia, B. A., and Floudas, C. A. (2009) A mixed integer linear optimization framework for the identification and quantification of targeted post-translational modifications of highly modified proteins using multiplexed electron transfer dissociation tandem mass spectrometry. *Mol Cell Proteomics* 8, 2527-2543.
79. Frank, A. M., Pesavento, J. J., Mizzen, C. A., Kelleher, N. L., and Pevzner, P. A. (2008) Interpreting top-down mass spectra using spectral alignment. *Anal Chem* 80, 2499-2505.
80. Guan, S., and Burlingame, A. L. (2010) Data processing algorithms for analysis of high resolution MSMS spectra of peptides with complex patterns of posttranslational modifications. *Mol Cell Proteomics* 9, 804-810.
81. Matthiesen, R., Trelle, M. B., Hojrup, P., Bunkenborg, J., and Jensen, O. N. (2005) VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J Proteome Res* 4, 2338-2347.

82. Marshall, A. G., and Hendrickson, C. L. (2008) High-resolution mass spectrometers. *Annu Rev Anal Chem (Palo Alto Calif)* 1, 579-599.
83. Nielsen, M. L., Vermeulen, M., Bonaldi, T., Cox, J., Moroder, L., and Mann, M. (2008) Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods* 5, 459-460.
84. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994-999.
85. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154-1169.
86. 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.
87. de Hoog, C. L., Foster, L. J., and Mann, M. (2004) RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell* 117, 649-662.
88. 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.
89. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22, 1139-1145.
90. Kruger, M., Kratchmarova, I., Blagoev, B., Tseng, Y. H., Kahn, C. R., and Mann, M. (2008) Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc Natl Acad Sci U S A* 105, 2451-2456.
91. Montes de Oca, R., Shoemaker, C. J., Gucek, M., Cole, R. N., and Wilson, K. L. (2009) Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. *PLoS One* 4, e7050.
92. Yuan, X., Gu, X., Crabb, J. S., Yue, X., Shadrach, K., Hollyfield, J. G., and Crabb, J. W. (2010) Quantitative proteomics: comparison of the macular Bruch membrane/choroid complex from age-related macular degeneration and normal eyes. *Mol Cell Proteomics* 9, 1031-1046.
93. Salim, K., Kehoe, L., Minkoff, M. S., Bilsland, J. G., Munoz-Sanjuan, I., and Guest, P. C. (2006) Identification of differentiating neural progenitor cell markers using shotgun isobaric tagging mass spectrometry. *Stem Cells Dev* 15, 461-470.
94. Stunnenberg, H. G., and Vermeulen, M. (2011) Towards cracking the epigenetic code using a combination of high-throughput epigenomics and quantitative mass spectrometry-based proteomics. *Bioessays* 33, 547-551.
95. Ong, S. E., and Mann, M. (2007) Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods Mol Biol* 359, 37-52.
96. Pimienta, G., Chaerkady, R., and Pandey, A. (2009) SILAC for global phosphoproteomic analysis. *Methods Mol Biol* 527, 107-116, x.
97. Bonenfant, D., Towbin, H., Coulot, M., Schindler, P., Mueller, D. R., and van Oostrum, J. (2007) Analysis of dynamic changes in post-translational modifications of human histones during cell cycle by mass spectrometry. *Mol Cell Proteomics* 6, 1917-1932.
98. Pesavento, J. J., Yang, H., Kelleher, N. L., and Mizzen, C. A. (2008) Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. *Mol Cell Biol* 28, 468-486.

99. Scharf, A. N., Meier, K., Seitz, V., Kremmer, E., Brehm, A., and Imhof, A. (2009) Monomethylation of lysine 20 on histone H4 facilitates chromatin maturation. *Mol Cell Biol* 29, 57-67.
100. Jung, H. R., Pasini, D., Helin, K., and Jensen, O. N. (2010) Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Mol Cell Proteomics* 9, 838-850.
101. Cuomo, A., Moretti, S., Minucci, S., and Bonaldi, T. (2011) SILAC-based proteomic analysis to dissect the "histone modification signature" of human breast cancer cells. *Amino Acids* 41, 387-399.
102. Zee, B. M., Levin, R. S., Dimaggio, P. A., and Garcia, B. A. (2010) Global turnover of histone post-translational modifications and variants in human cells. *Epigenetics Chromatin* 3, 22.
103. Ong, S. E., Mittler, G., and Mann, M. (2004) Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods* 1, 119-126.
104. Fodor, B. D., Kubicek, S., Yonezawa, M., O'Sullivan, R. J., Sengupta, R., Perez-Burgos, L., Opravil, S., Mechtler, K., Schotta, G., and Jenuwein, T. (2006) Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells. *Genes Dev* 20, 1557-1562.
105. Zee, B. M., Levin, R. S., Xu, B., LeRoy, G., Wingreen, N. S., and Garcia, B. A. (2010) In vivo residue-specific histone methylation dynamics. *J Biol Chem* 285, 3341-3350.
106. Sweet, S. M., Li, M., Thomas, P. M., Durbin, K. R., and Kelleher, N. L. (2010) Kinetics of re-establishing H3K79 methylation marks in global human chromatin. *J Biol Chem* 285, 32778-32786.
107. Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2006) Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal Chem* 78, 4271-4280.
108. Garcia, B. A., Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2007) Pervasive combinatorial modification of histone H3 in human cells. *Nat Methods* 4, 487-489.
109. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 9, 555-566.
110. Darwanto, A., Curtis, M. P., Schrag, M., Kirsch, W., Liu, P., Xu, G., Neidigh, J. W., and Zhang, K. (2010) A modified "cross-talk" between histone H2B Lys-120 ubiquitination and H3 Lys-79 methylation. *J Biol Chem* 285, 21868-21876.
111. Yuan, G., and Zhu, B. (2012) Histone variants and epigenetic inheritance. *Biochim Biophys Acta* 1819, 222-229.
112. Hake, S. B., and Allis, C. D. (2006) Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc Natl Acad Sci U S A* 103, 6428-6435.
113. Arnaudo, A. M., Molden, R. C., and Garcia, B. A. (2011) Revealing histone variant induced changes via quantitative proteomics. *Crit Rev Biochem Mol Biol* 46, 284-294.
114. Hake, S. B., Garcia, B. A., Duncan, E. M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D. P., Allis, C. D., and Hunt, D. F. (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem* 281, 559-568.
115. McKittrick, E., Gafken, P. R., Ahmad, K., and Henikoff, S. (2004) Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A* 101, 1525-1530.
116. Johnson, L., Mollah, S., Garcia, B. A., Muratore, T. L., Shabanowitz, J., Hunt, D. F., and Jacobsen, S. E. (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res* 32, 6511-6518.

117. Garcia, B. A., Siuti, N., Thomas, C. E., Mizzen, C. A., and Kelleher, N. L. (2007) Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry. *International Journal of Mass Spectrometry* 259, 184-196.
118. 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.
119. 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.
120. Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., and Zhang, Y. (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873-878.
121. Ikura, T., Tashiro, S., Kakino, A., Shima, H., Jacob, N., Amunugama, R., Yoder, K., Izumi, S., Kuraoka, I., Tanaka, K., Kimura, H., Ikura, M., Nishikubo, S., Ito, T., Muto, A., Miyagawa, K., Takeda, S., Fishel, R., Igarashi, K., and Kamiya, K. (2007) DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics. *Mol Cell Biol* 27, 7028-7040.
122. 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.
123. Chu, F., Nusinow, D. A., Chalkley, R. J., Plath, K., Panning, B., and Burlingame, A. L. (2006) Mapping post-translational modifications of the histone variant MacroH2A1 using tandem mass spectrometry. *Mol Cell Proteomics* 5, 194-203.
124. Bernstein, E., Muratore-Schroeder, T. L., Diaz, R. L., Chow, J. C., Changolkar, L. N., Shabanowitz, J., Heard, E., Pehrson, J. R., Hunt, D. F., and Allis, C. D. (2008) A phosphorylated subpopulation of the histone variant macroH2A1 is excluded from the inactive X chromosome and enriched during mitosis. *Proc Natl Acad Sci U S A* 105, 1533-1538.
125. Siuti, N., Roth, M. J., Mizzen, C. A., Kelleher, N. L., and Pesavento, J. J. (2006) Gene-specific characterization of human histone H2B by electron capture dissociation. *J Proteome Res* 5, 233-239.
126. Zhang, L., Eugeni, E. E., Parthun, M. R., and Freitas, M. A. (2003) Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* 112, 77-86.
127. Lu, S., Xie, Y. M., Li, X., Luo, J., Shi, X. Q., Hong, X., Pan, Y. H., and Ma, X. (2009) Mass spectrometry analysis of dynamic post-translational modifications of TH2B during spermatogenesis. *Mol Hum Reprod* 15, 373-378.
128. Medzihradszky, K. F., Zhang, X., Chalkley, R. J., Guan, S., McFarland, M. A., Chalmers, M. J., Marshall, A. G., Diaz, R. L., Allis, C. D., and Burlingame, A. L. (2004) Characterization of Tetrahymena histone H2B variants and posttranslational populations by electron capture dissociation (ECD) Fourier transform ion cyclotron mass spectrometry (FT-ICR MS). *Mol Cell Proteomics* 3, 872-886.
129. Villar-Garea, A., Forne, I., Vetter, I., Kremmer, E., Thomae, A., and Imhof, A. (2012) Developmental regulation of N-terminal H2B methylation in Drosophila melanogaster. *Nucleic Acids Res* 40, 1536-1549.
130. Wachsmuth, M., Caudron-Herger, M., and Rippe, K. (2008) Genome organization: balancing stability and plasticity. *Biochim Biophys Acta* 1783, 2061-2079.
131. Talasz, H., Sarg, B., and Lindner, H. H. (2009) Site-specifically phosphorylated forms of H1.5 and H1.2 localized at distinct regions of the nucleus are related to different processes during the cell cycle. *Chromosoma* 118, 693-709.
132. Wisniewski, J. R., Zougman, A., Kruger, S., and Mann, M. (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations,

methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol Cell Proteomics* 6, 72-87.
133. Garcia, B. A., Busby, S. A., Barber, C. M., Shabanowitz, J., Allis, C. D., and Hunt, D. F. (2004) Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. *J Proteome Res* 3, 1219-1227.
134. Weiss, T., Hergeth, S., Zeissler, U., Izzo, A., Tropberger, P., Zee, B. M., Dundr, M., Garcia, B. A., Daujat, S., and Schneider, R. (2010) Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D. *Epigenetics Chromatin* 3, 7.
135. Villar-Garea, A., and Imhof, A. (2008) Fine mapping of posttranslational modifications of the linker histone H1 from Drosophila melanogaster. *PLoS One* 3, e1553.
136. Daujat, S., Zeissler, U., Waldmann, T., Happel, N., and Schneider, R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J Biol Chem* 280, 38090-38095.
137. Zheng, Y., John, S., Pesavento, J. J., Schultz-Norton, J. R., Schiltz, R. L., Baek, S., Nardulli, A. M., Hager, G. L., Kelleher, N. L., and Mizzen, C. A. (2010) Histone H1 phosphorylation is associated with transcription by RNA polymerases I and II. *J Cell Biol* 189, 407-415.
138. Shiio, Y., Eisenman, R. N., Yi, E. C., Donohoe, S., Goodlett, D. R., and Aebersold, R. (2003) Quantitative proteomic analysis of chromatin-associated factors. *J Am Soc Mass Spectrom* 14, 696-703.
139. Porter, I. M., McClelland, S. E., Khoudoli, G. A., Hunter, C. J., Andersen, J. S., McAinsh, A. D., Blow, J. J., and Swedlow, J. R. (2007) Bod1, a novel kinetochore protein required for chromosome biorientation. *J Cell Biol* 179, 187-197.
140. Gassmann, R., Henzing, A. J., and Earnshaw, W. C. (2005) Novel components of human mitotic chromosomes identified by proteomic analysis of the chromosome scaffold fraction. *Chromosoma* 113, 385-397.
141. Uchiyama, S., Kobayashi, S., Takata, H., Ishihara, T., Hori, N., Higashi, T., Hayashihara, K., Sone, T., Higo, D., Nirasawa, T., Takao, T., Matsunaga, S., and Fukui, K. (2005) Proteome analysis of human metaphase chromosomes. *J Biol Chem* 280, 16994-17004.
142. Takata, H., Uchiyama, S., Nakamura, N., Nakashima, S., Kobayashi, S., Sone, T., Kimura, S., Lahmers, S., Granzier, H., Labeit, S., Matsunaga, S., and Fukui, K. (2007) A comparative proteome analysis of human metaphase chromosomes isolated from two different cell lines reveals a set of conserved chromosome-associated proteins. *Genes Cells* 12, 269-284.
143. Morrison, C., Henzing, A. J., Jensen, O. N., Osheroff, N., Dodson, H., Kandels-Lewis, S. E., Adams, R. R., and Earnshaw, W. C. (2002) Proteomic analysis of human metaphase chromosomes reveals topoisomerase II alpha as an Aurora B substrate. *Nucleic Acids Res* 30, 5318-5327.
144. Khoudoli, G. A., Gillespie, P. J., Stewart, G., Andersen, J. S., Swedlow, J. R., and Blow, J. J. (2008) Temporal profiling of the chromatin proteome reveals system-wide responses to replication inhibition. *Curr Biol* 18, 838-843.
145. Ohta, S., Bukowski-Wills, J. C., Sanchez-Pulido, L., Alves Fde, L., Wood, L., Chen, Z. A., Platani, M., Fischer, L., Hudson, D. F., Ponting, C. P., Fukagawa, T., Earnshaw, W. C., and Rappsilber, J. (2010) The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* 142, 810-821.
146. Dejardin, J., and Kingston, R. E. (2009) Purification of proteins associated with specific genomic Loci. *Cell* 136, 175-186.
147. 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.

148. 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.
149. Nikolov, M., Stutzer, 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, M110 005371.
150. Li, X., Foley, E. A., Molloy, K. R., Li, Y., Chait, B. T., and Kapoor, T. M. (2012) Quantitative chemical proteomics approach to identify post-translational modification-mediated protein-protein interactions. *J Am Chem Soc* 134, 1982-1985.
151. Liu, H., Galka, M., Iberg, A., Wang, Z., Li, L., Voss, C., Jiang, X., Lajoie, G., Huang, Z., Bedford, M. T., and Li, S. S. (2010) Systematic identification of methyllysine-driven interactions for histone and nonhistone targets. *J Proteome Res* 9, 5827-5836.
152. Torrente, M. P., Zee, B. M., Young, N. L., Baliban, R. C., LeRoy, G., Floudas, C. A., Hake, S. B., and Garcia, B. A. (2011) Proteomic interrogation of human chromatin. *PLoS One* 6, e24747.
153. Lambert, J. P., Mitchell, L., Rudner, A., Baetz, K., and Figeys, D. (2009) A novel proteomics approach for the discovery of chromatin-associated protein networks. *Mol Cell Proteomics* 8, 870-882.
154. O'Neill, L. P., and Turner, B. M. (2003) Immunoprecipitation of native chromatin: NChIP. *Methods* 31, 76-82.
155. Lee, T. I., Johnstone, S. E., and Young, R. A. (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 1, 729-748.
156. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2, 1896-1906.
157. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 4, 2010-2021.
158. 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.
159. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10, 1794-1805.
160. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10, 1794-1805.
161. Cox, J., Michalski, A., and Mann, M. (2011) Software lock mass by two-dimensional minimization of peptide mass errors. *J Am Soc Mass Spectrom* 22, 1373-1380.
162. Li, H., and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589-595.
163. Austenaa, L., Barozzi, I., Chronowska, A., Termanini, A., Ostuni, R., Prosperini, E., Stewart, A. F., Testa, G., and Natoli, G. (2012) The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. *Immunity* 36, 572-585

164. Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., Durham, T., Zhang, X., Donaghey, J., Epstein, C. B., Regev, A., and Bernstein, B. E. (2011) Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* 147, 1628-1639.

165. Ernst, J., and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28, 817-825.
166. 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., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-318.
167. Min, J., Zhang, Y., and Xu, R. M. (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* 17, 1823-1828.
168. 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., Pirrotta, V., Reinberg, D., and Gamblin, S. J. (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762-767.
169. Fischle, W., Wang, Y., and Allis, C. D. (2003) Binary switches and modification cassettes in histone biology and beyond. *Nature* 425, 475-479.
170. Peters, A. H., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J. H., and Jenuwein, T. (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* 12, 1577-1589.
171. Varga-Weisz, P. D., and Dalgaard, J. Z. (2002) A mark in the core: silence no more! *Mol Cell* 9, 1154-1156.
172. Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005) Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* 280, 17732-17736.
173. Li, B., Jackson, J., Simon, M. D., Fleharty, B., Gogol, M., Seidel, C., Workman, J. L., and Shilatifard, A. (2009) Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. *J Biol Chem* 284, 7970-7976.
174. Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* 18, 1251-1262.
175. Garcia, B. A., Hake, S. B., Diaz, R. L., Kauer, M., Morris, S. A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B. D., Allis, C. D., and Hunt, D. F. (2007) Organismal differences in post-translational modifications in histones H3 and H4. *J Biol Chem* 282, 7641-7655.
176. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120-124.
177. 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.
178. Li, H., Rauch, T., Chen, Z. X., Szabo, P. E., Riggs, A. D., and Pfeifer, G. P. (2006) The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J Biol Chem* 281, 19489-19500.
179. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., and Rauscher, F. J., 3rd (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* 16, 919-932.
180. Smallwood, A., Esteve, P. O., Pradhan, S., and Carey, M. (2007) Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev* 21, 1169-1178.
181. Rottach, A., Frauer, C., Pichler, G., Bonapace, I. M., Spada, F., and Leonhardt, H. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res* 38, 1796-1804.

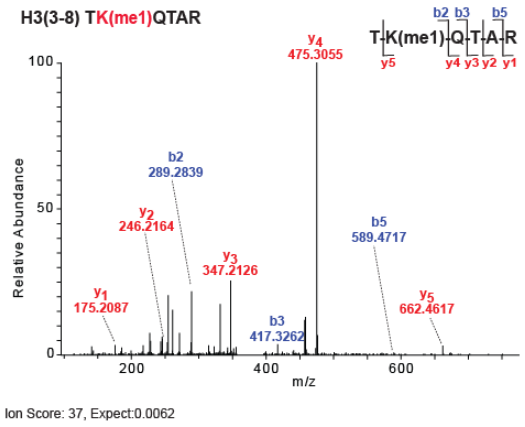
182. Buonomo, S. B. Heterochromatin DNA replication and Rif1. *Exp Cell Res* 316, 1907-1913.
183. Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G., and Rauscher, F. J., 3rd (2005) The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochem Biophys Res Commun* 331, 929-937.
184. Mosch, K., Franz, H., Soeroes, S., Singh, P. B., and Fischle, W. (2011) HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats. *PLoS One* 6, e15894.
185. Frescas, D., Guardavaccaro, D., Kuchay, S. M., Kato, H., Poleshko, A., Basrur, V., Elenitoba-Johnson, K. S., Katz, R. A., and Pagano, M. (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle* 7, 3539-3547.
186. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439, 811-816.
187. Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* 13, 713-719.
188. Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H., and Allis, C. D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* 121, 859-872.
189. 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.
190. Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J. F. (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol* 22, 6979-6992.
191. Hayashihara, K., Uchiyama, S., Shimamoto, S., Kobayashi, S., Tomschik, M., Wakamatsu, H., No, D., Sugahara, H., Hori, N., Noda, M., Ohkubo, T., Zlatanova, J., Matsunaga, S., and Fukui, K. (2010) The middle region of an HP1-binding protein, HP1-BP74, associates with linker DNA at the entry/exit site of nucleosomal DNA. *J Biol Chem* 285, 6498-6507.
192. Townson, S. M., Dobrzycka, K. M., Lee, A. V., Air, M., Deng, W., Kang, K., Jiang, S., Kioka, N., Michaelis, K., and Oesterreich, S. (2003) SAFB2, a new scaffold attachment factor homolog and estrogen receptor corepressor. *J Biol Chem* 278, 20059-20068.
193. Chan, C. W., Lee, Y. B., Uney, J., Flynn, A., Tobias, J. H., and Norman, M. (2007) A novel member of the SAF (scaffold attachment factor)-box protein family inhibits gene expression and induces apoptosis. *Biochem J* 407, 355-362.
194. Malyavantham, K. S., Bhattacharya, S., Barbeitos, M., Mukherjee, L., Xu, J., Fackelmayer, F. O., and Berezney, R. (2008) Identifying functional neighborhoods within the cell nucleus: proximity analysis of early S-phase replicating chromatin domains to sites of transcription, RNA polymerase II, HP1gamma, matrin 3 and SAF-A. *J Cell Biochem* 105, 391-403.
195. Lobov, I. B., Tsutsui, K., Mitchell, A. R., and Podgornaya, O. I. (2001) Specificity of SAF-A and lamin B binding in vitro correlates with the satellite DNA bending state. *J Cell Biochem* 83, 218-229.
196. Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S., and Reinberg, D. (1998) FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* 92, 105-116.

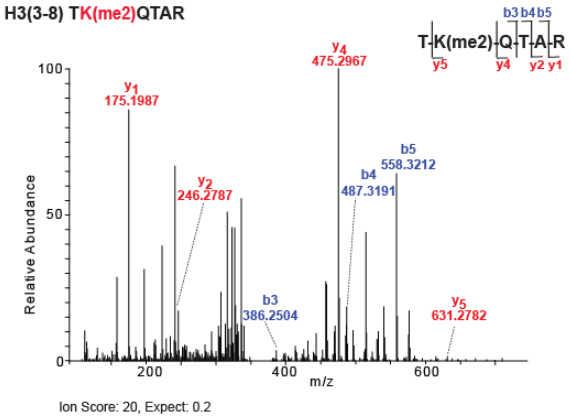
197. Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., and Reinberg, D. (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science* 301, 1090-1093.
198. Saunders, A., Werner, J., Andrulis, E. D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J. T. (2003) Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science* 301, 1094-1096.
199. Duroux, M., Houben, A., Ruzicka, K., Friml, J., and Grasser, K. D. (2004) The chromatin remodelling complex FACT associates with actively transcribed regions of the Arabidopsis genome. *Plant J* 40, 660-671.
200. Phelan, M. L., Sif, S., Narlikar, G. J., and Kingston, R. E. (1999) Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Mol Cell* 3, 247-253.
201. Itou, J., Taniguchi, N., Oishi, I., Kawakami, H., Lotz, M., and Kawakami, Y. (2011) HMGB factors are required for posterior digit development through integrating signaling pathway activities. *Dev Dyn* 240, 1151-1162.
202. Hagemann, C., Weigelin, B., Schommer, S., Schulze, M., Al-Jomah, N., Anacker, J., Gerngras, S., Kuhnel, S., Kessler, A. F., Polat, B., Ernestus, R. I., Patel, R., and Vince, G. H. (2011) The cohesin-interacting protein, precocious dissociation of sisters 5A/sister chromatid cohesion protein 112, is up-regulated in human astrocytic tumors. *Int J Mol Med* 27, 39-51.
203. Kumar, D., Sakabe, I., Patel, S., Zhang, Y., Ahmad, I., Gehan, E. A., Whiteside, T. L., and Kasid, U. (2004) SCC-112, a novel cell cycle-regulated molecule, exhibits reduced expression in human renal carcinomas. *Gene* 328, 187-196.
204. Zheng, M. Z., Zheng, L. M., and Zeng, Y. X. (2008) SCC-112 gene is involved in tumor progression and promotes the cell proliferation in G2/M phase. *J Cancer Res Clin Oncol* 134, 453-462.
205. Mietton, F., Sengupta, A. K., Molla, A., Picchi, G., Barral, S., Heliot, L., Grange, T., Wutz, A., and Dimitrov, S. (2009) Weak but uniform enrichment of the histone variant macroH2A1 along the inactive X chromosome. *Mol Cell Biol* 29, 150-156.
206. Ladurner, A. G. (2003) Inactivating chromosomes: a macro domain that minimizes transcription. *Mol Cell* 12, 1-3.
207. Chow, C. M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005) Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep* 6, 354-360.
208. Braunschweig, U., Hogan, G. J., Pagie, L., and van Steensel, B. (2009) Histone H1 binding is inhibited by histone variant H3.3. *EMBO J* 28, 3635-3645.
209. Oberg, C., Izzo, A., Schneider, R., Wrange, O., and Belikov, S. (2011) Linker histone subtypes differ in their effect on nucleosomal spacing in vivo. *J Mol Biol* 419, 183-197.
210. Parseghian, M. H., Newcomb, R. L., Winokur, S. T., and Hamkalo, B. A. (2000) The distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin: distribution in human fetal fibroblasts. *Chromosome Res* 8, 405-424.
211. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., and Bonner, W. M. (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273, 5858-5868.
212. Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004) H2AX: the histone guardian of the genome. *DNA Repair (Amst)* 3, 959-967.
213. Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall' Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. (2006) Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat Cell Biol* 8, 764-770.
214. Strutt, H., and Paro, R. (1999) Mapping DNA target sites of chromatin proteins in vivo by formaldehyde crosslinking. In: Becker, P. B., ed. *Methods in Molecular Biology, Chromatin Protocols*, pp. 455-467, Humana Press.

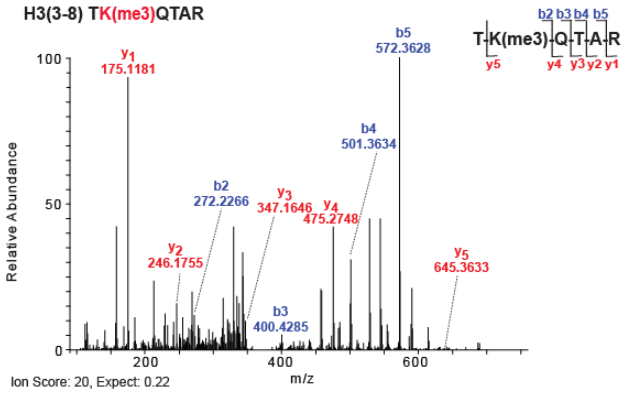
215. Orlando, V., Strutt, H., and Paro, R. (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* 11, 205-214.
216. Corona, D. F., and Tamkun, J. W. (2004) Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim Biophys Acta* 1677, 113-119.
217. Bozhenok, L., Wade, P. A., and Varga-Weisz, P. (2002) WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J* 21, 2231-2241.
218. Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L. A., Erdjument-Bromage, H., Ishibe-Murakami, S., Wang, B., Tempst, P., Hofmann, K., Patel, D. J., Elledge, S. J., and Allis, C. D. (2009) WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature* 457, 57-62.
219. Cook, P. J., Ju, B. G., Telese, F., Wang, X., Glass, C. K., and Rosenfeld, M. G. (2009) Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature* 458, 591-596.
220. Poser, I., Sarov, M., Hutchins, J. R., Heriche, J. K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K. P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A. F., Peters, J. M., Buchholz, F., and Hyman, A. A. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat Methods* 5, 409-415.
221. Kim, J. A., Kruhlak, M., Dotiwala, F., Nussenzweig, A., and Haber, J. E. (2007) Heterochromatin is refractory to gamma-H2AX modification in yeast and mammals. *J Cell Biol* 178, 209-218.
222. Cowell, I. G., Sunter, N. J., Singh, P. B., Austin, C. A., Durkacz, B. W., and Tilby, M. J. (2007) gammaH2AX foci form preferentially in euchromatin after ionising-radiation. *PLoS One* 2, e1057.
223. Vasireddy, R. S., Karagiannis, T. C., and El-Osta, A. (2010) gamma-radiation-induced gammaH2AX formation occurs preferentially in actively transcribing euchromatic loci. *Cell Mol Life Sci* 67, 291-294.
224. Goodarzi, A. A., Jeggo, P., and Lobrich, M. (2010) The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax. *DNA Repair (Amst)* 9, 1273-1282.
225. Sullivan, B., and Karpen, G. (2001) Centromere identity in Drosophila is not determined in vivo by replication timing. *J Cell Biol* 154, 683-690.
226. Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., and Jenuwein, T. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107, 323-337.
227. Ayoub, N., Jeyasekharan, A. D., Bernal, J. A., and Venkitaraman, A. R. (2008) HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. *Nature* 453, 682-686.
228. Peng, J. C., and Karpen, G. H. (2009) Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet* 5, e1000435.

9. APPENDIX

APPENDIX I. MS/MS spectra of H3 (3-8) peptide, containing the Lysine 4. Fragmentation spectra were used for the site-specific assignment of modifications within the peptides; MASCOT search, with the most intense ions identified in the MSMS spectra and relative calculated score: experimental spectra are displayed.

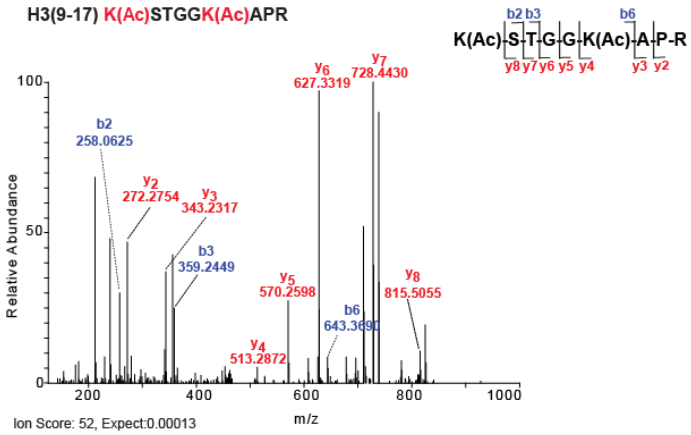


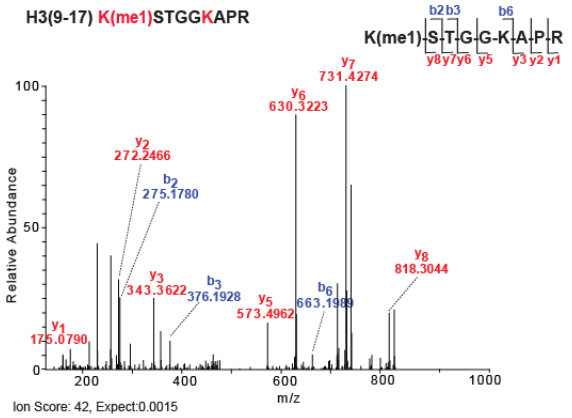


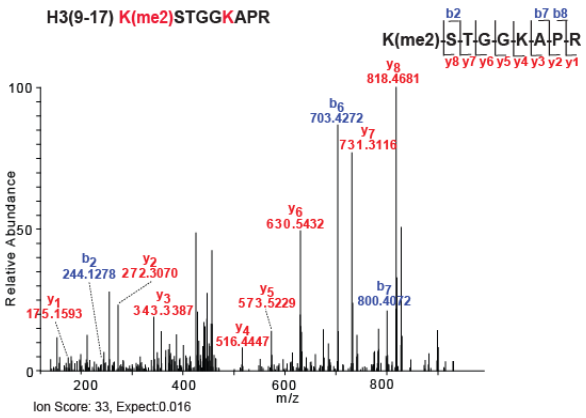


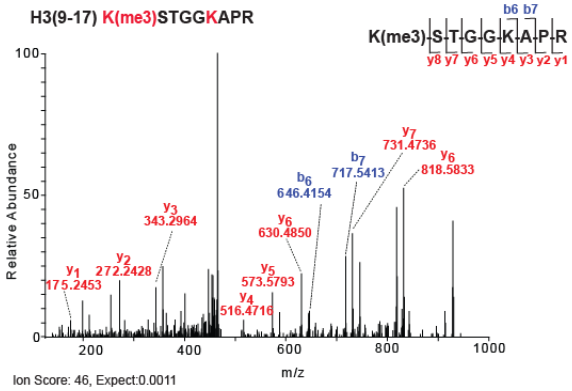
APPENDIX II. MS/MS spectra of H3 (9-17) peptide, containing the Lysine 9 and 14.

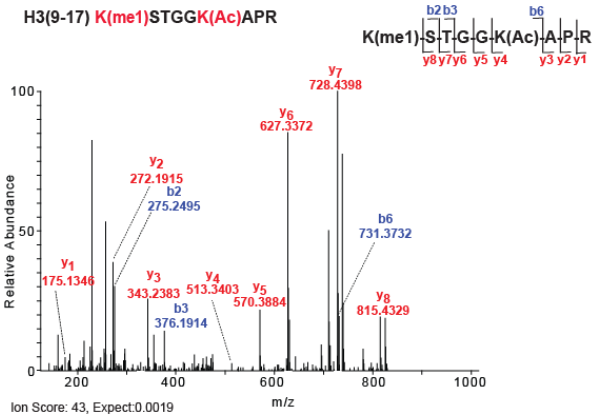
Fragmentation spectra were used for the site-specific assignment of modifications within the peptides; MASCOT search, with the most intense ions identified in the MSMS spectra and relative calculated score: experimental spectra are displayed.

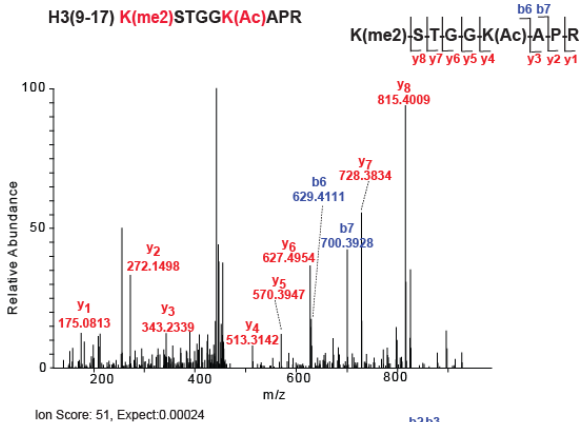


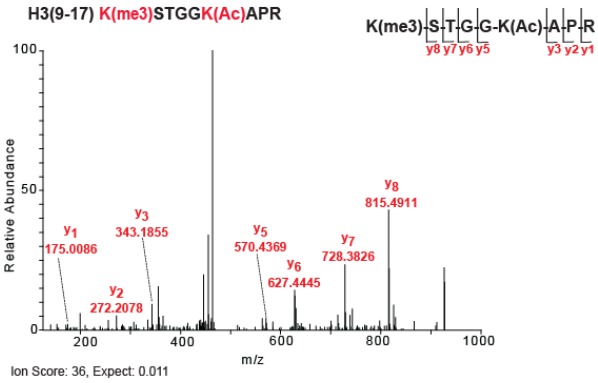


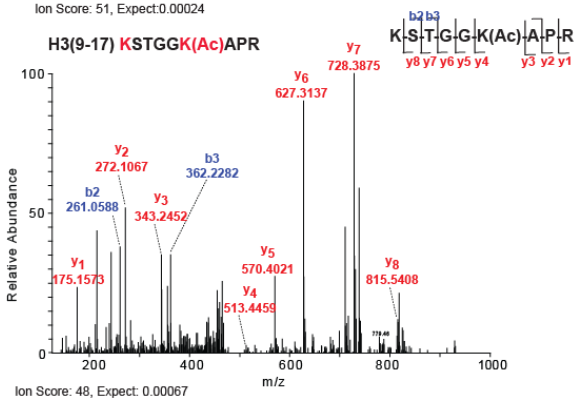





H3(9-17) K(me1)STGGK(Ac)APR
K(me1)-S-T-G-G-K(Ac)-A-P-R
Relative Abundance
m/z
Ion Score: 43, Expect:0.0019


H3(9-17) K(me2)STGGK(Ac)APR
K(me2)-S-T-G-G-K(Ac)-A-P-R
Relative Abundance
m/z
Ion Score: 51, Expect:0.00024


H3(9-17) K(me3)STGGK(Ac)APR
K(me3)-S-T-G-G-K(Ac)-A-P-R
Relative Abundance
m/z
Ion Score: 36, Expect: 0.011


H3(9-17) KSTGGK(Ac)APR
K-S-T-G-G-K(Ac)-A-P-R
Relative Abundance
m/z
Ion Score: 48, Expect: 0.00067

APPENDIX III. Informations contained in the interactome tables.

Gene Names: Name(s) of the gene(s) associated to the protein(s).

Uniprot: UniProt (http://www.uniprot.org) ID(s) of the protein(s).

Peptides: The total number of peptide sequences associated with the protein group (i.e. for all the proteins in the group) (For and Rev, indicate Forward and Reverse experiments, respectively).

Unique Peptides: The total number of unique peptides associated with the protein group (e.g. these peptides are not shared with another protein group). (For and Rev, indicate Forward and Reverse experiments, respectively).

Sequence Coverage [%]: Percentage of the sequence that is covered by the identified peptides of the best protein sequence contained within the group.

Mol. Weight [kDa]: Molecular weight of the best protein sequence contained within the protein group.

Sequence Length: The total length of the best protein sequence contained within the group.

Ratio H/L For: Ratio H/L Normalized divided to the standard deviation in the Forward experiment.

Ratio H/L Count For: Number of redundant peptides used for quantitation in Forward experiment.

Ratio H/L Rev: Ratio H/L Normalized divided to the standard deviation in the Reverse experiment.

Ratio H/L Count Rev: Number of redundant peptides used for quantitation in the Reverse experiment.

APPENDIX IV. Proteins identified and quantified with a least two peptides, one of which unique, in the H3K9me3 interactome. Protein Group output from MaxQuant software, representing the Top40% of proteins binders.

Gene Names	Uniprot	Peptides For	Peptides Rev	Unique Pep. For	Unique Pep. Rev	Sequence Coverage [%]	Mol. Weight [kDa]	Sequence Length	Ratio H/L For	Ratio H/L Count For	Ratio H/L Rev	Ratio H/L Count Rev
CBX1	B5MD17	3	2	2	1	18.5	21.915	189	**3.2932**	**3**	**0.0996**	**1**
HIST2H3PS2	Q5TEC6	6	5	1	1	27.2	15.43	136	**4.4974**	**13**	**0.1133**	**6**
BAF	O75531	3	2	3	2	56.2	10.058	89	**4.3362**	**4**	**0.1347**	**2**
H2AFA	P04908	6	4	1	1	57.7	14.135	130	**7.8716**	**6**	**0.1578**	**1**
SFRS10	Q59GA1	6	5	6	5	20.4	33.571	289	**3.9058**	**10**	**0.1863**	**16**
DEK	P35659	12	9	12	9	30.1	42.674	375	**3.0638**	**29**	**0.2130**	**26**
SP140L	Q9H930-1	2	2	1	1	3.4	67.005	580	**2.5578**	**2**	**0.2315**	**1**
SFRS3	P84103	6	5	5	4	38.4	19.329	164	**2.7555**	**9**	**0.2460**	**10**
NP220	Q14966-1	21	13	21	13	12.9	220.62	1978	**4.6609**	**22**	**0.2464**	**16**
SFRS6	Q13247-1	8	6	7	5	24.4	39.586	344	**3.6453**	**12**	**0.2479**	**16**
ASF	Q07955-1	11	10	10	10	47.6	27.744	248	**4.0771**	**29**	**0.2491**	**35**
TRA2A	Q13595-1	2	2	2	2	8.2	32.688	282	**5.1072**	**3**	**0.2492**	**3**
HIST2H2AB	Q8IUE6	4	2	1	1	40	13.995	130	**3.3613**	**3**	**0.2505**	**1**
SFRS7	Q16629-1	7	5	6	4	29.4	27.366	238	**3.0631**	**18**	**0.2587**	**16**
SAFB	A0AV56	17	16	8	6	22.4	102.85	917	**5.4024**	**16**	**0.2606**	**11**
HNRNPU	Q00839-1	33	30	33	30	41.3	90.583	825	**3.6163**	**125**	**0.2642**	**122**
SFRS2	Q01130	7	6	7	6	37.1	25.476	221	**3.8919**	**10**	**0.2721**	**19**
CBX3	Q13185	7	5	5	4	36.6	20.811	183	**3.5583**	**18**	**0.2739**	**15**
KIAA0138	Q14151	25	26	16	16	31	107.47	953	**5.8520**	**42**	**0.2744**	**41**
YLPM1	B4DMQ9	10	4	10	4	5.3	241.64	2146	**4.5553**	**9**	**0.2813**	**4**
RNPU1Z	P08621-1	9	10	9	10	29.5	51.556	437	**3.2193**	**10**	**0.2819**	**19**
	Q1AHP8	4	2	1	1	17.6	28.019	245	**3.5338**	**5**	**0.2840**	**1**
RBM12B	Q8IXT5	13	11	13	11	18.5	118.1	1001	**5.3077**	**26**	**0.2843**	**11**
H4/A	P62805	11	10	11	10	54.4	11.367	103	**3.4679**	**212**	**0.2869**	**122**
H1F2	P16403	18	13	5	3	46	21.364	213	**3.0975**	**108**	**0.2935**	**94**
H2AFY	O75367-1	15	16	15	16	51.9	39.617	372	**3.2085**	**68**	**0.2955**	**37**

RBAP48	Q09028	5	4	1	1	12	47.655	425	**4.3997**	**1**	**0.2985**	**1**
TOP2	P11388-4	34	34	25	25	22	182.68	1612	**2.4628**	**71**	**0.3030**	**77**
AHCTF1	Q8WYP5-2	12	12	12	12	7.2	256.25	2304	**2.4872**	**18**	**0.3033**	**8**
ATAD2	Q6PL18-1	16	14	16	14	17.9	158.55	1390	**2.4075**	**26**	**0.3039**	**17**
HRS	Q13243-1	5	3	4	2	19.9	31.263	272	**3.3101**	**6**	**0.3051**	**2**
H2BFD	B4DR52	13	12	2	2	50	18.041	166	**2.3233**	**27**	**0.3057**	**20**
RIF1	C9JI70	2	4	2	4	2.9	274.48	2472	**2.1838**	**3**	**0.3057**	**4**
HNRNPA1	P09651-1	18	18	12	11	51.9	38.845	372	**3.7049**	**123**	**0.3148**	**66**
INCENP	Q9NQS7-1	8	4	8	4	9	105.43	918	**2.5677**	**9**	**0.3149**	**3**
SFRS9	Q13242	8	6	7	6	35.3	25.542	221	**4.5666**	**12**	**0.3209**	**4**
MET	Q9NWH9	15	14	15	14	17.2	117.15	1034	**4.5699**	**28**	**0.3222**	**13**
BRE1A	Q5VTR2	3	3	2	2	5.9	113.98	977	**4.2904**	**2**	**0.3236**	**2**
HNRPG	P38159	16	14	5	5	39.1	42.331	391	**4.7885**	**50**	**0.3251**	**41**
BRE1B	O75150-1	2	2	1	1	2.5	113.68	1001	**2.5983**	**1**	**0.3254**	**1**
FUSIP1	O75494-1	5	4	5	4	23.3	31.3	262	**5.1612**	**7**	**0.3338**	**8**
H2BFH	P23527	13	12	2	2	65.9	13.906	126	**2.4963**	**158**	**0.3350**	**92**
ADNP	Q9H2P0	2	2	2	2	2.6	123.56	1102	**2.5284**	**3**	**0.3371**	**2**
CBX5	P45973	9	8	8	8	53.4	22.225	191	**2.3396**	**17**	**0.3393**	**11**
FTP3	P55795	6	5	1	1	17.1	49.263	449	**6.2432**	**1**	**0.3444**	**1**
BTF3	P20290-1	4	3	4	3	20.9	22.168	206	**2.9262**	**5**	**0.3457**	**3**
H1F5	P16401	12	10	6	7	34.5	22.58	226	**2.6446**	**15**	**0.3503**	**16**
WIZ	O95785-1	6	5	6	5	7.7	178.67	1651	**4.4522**	**5**	**0.3515**	**3**
CGI-55	Q8NC51-1	5	6	5	6	17.4	44.965	408	**6.5009**	**7**	**0.3517**	**9**
HNRPDL	O14979-1	4	5	3	4	14.5	46.437	420	**3.0269**	**8**	**0.3520**	**7**
HNRNPA3	P51991-1	6	8	5	6	31.2	39.594	378	**3.6357**	**21**	**0.3568**	**14**
H2AFZ	P0C0S5	4	4	2	2	22.1	18.414	181	**3.2308**	**11**	**0.3572**	**7**
H1F4	P10412	14	11	1	1	42.5	21.865	219	**2.6054**	**13**	**0.3583**	**4**
BAZ1B	Q9UIG0-1	36	31	36	31	29.9	170.9	1483	**2.5349**	**56**	**0.3594**	**39**

ZFR	B5MEH6	8	10	8	10	15.4	115.15	1056	**2.7555**	**9**	**0.3645**	**9**
C3orf63	Q9UK61-1	2	11	2	11	7	189.03	1670	**2.4775**	**2**	**0.3671**	**8**
PHIP	Q8WWQ0	9	6	9	6	6.2	206.64	1821	**2.7464**	**10**	**0.3688**	**7**
RCC1	Q16269	9	8	9	8	27.9	48.145	452	**3.0074**	**24**	**0.3716**	**12**
U2AF2	P26368-1	3	3	3	3	15.8	53.5	475	**3.1405**	**3**	**0.3721**	**4**
HCC1	Q14498-1	6	6	6	6	23.2	59.379	530	**2.2574**	**12**	**0.3726**	**17**
HP1BP3	Q5SSJ5-1	19	15	19	15	30.9	61.206	553	**3.2489**	**44**	**0.3764**	**29**
KIAA0650	A6NHR9-1	36	29	36	29	20.7	226.37	2005	**2.2309**	**53**	**0.3804**	**31**
CENPV	Q7Z7K6-1	5	5	1	1	29.1	29.946	275	**2.6043**	**8**	**0.3856**	**7**
ORC1	Q13415	2	3	2	3	4.4	97.349	861	**3.6960**	**2**	**0.3901**	**3**
TOP2B	Q02880-1	18	18	9	9	14.1	183.26	1626	**3.2180**	**11**	**0.3924**	**8**
HDGF	P51858	7	7	7	7	34.6	26.788	240	**3.0906**	**14**	**0.3961**	**9**
TRIP12	Q14CA3	8	8	8	8	6.3	225.52	2040	**2.4182**	**7**	**0.4025**	**8**
NONO	Q15233	21	20	18	17	46.5	54.231	471	**3.6094**	**69**	**0.4036**	**41**
HMG1	P09429	9	7	6	4	40	24.893	215	**2.8833**	**21**	**0.4052**	**8**
GEMIN5	Q8TEQ6	2	2	2	2	1.1	168.56	1508	**3.0161**	**1**	**0.4074**	**1**
hCG_15646	B1AMT5	10	11	10	11	10.4	145.75	1268	**2.5166**	**11**	**0.4083**	**9**
	C9JMX5	5	4	1	1	24.6	27.378	240	**2.7292**	**5**	**0.4105**	**6**
ITM1	P46977	2	4	2	4	5.8	80.529	705	**2.2627**	**3**	**0.4139**	**5**
CHD4	Q14839-2	14	14	14	14	8.7	220.83	1940	**2.1331**	**19**	**0.4143**	**19**
CREAP1	O95232-1	4	2	4	2	12	51.466	432	**2.2664**	**4**	**0.4156**	**3**
U2AF1	Q01081	4	5	4	5	17.9	27.872	240	**2.5454**	**8**	**0.4166**	**9**
ANP32B	Q92688-1	6	2	4	2	28.7	28.787	251	**2.4901**	**6**	**0.4173**	**2**
PC4	P53999	3	2	3	2	26	14.395	127	**2.7610**	**7**	**0.4185**	**6**
SMT3B	P61956	2	2	1	1	23.2	10.871	95	**2.6900**	**12**	**0.4187**	**13**
KIAA0067	Q15047-1	7	8	7	8	8	143.16	1291	**8.4735**	**4**	**0.4235**	**1**
USP31	Q86UV5-1	3	2	3	2	5.3	120.52	1047	**3.0686**	**2**	**0.4255**	**1**
PBSCG	A8MWD9	5	3	5	3	36.8	8.544	76	**2.4652**	**6**	**0.4290**	**4**

CXXC8	Q9Y2K7-1	8	7	8	7	8.8	132.79	1162	**2.1987**	**17**	**0.4304**	**7**
AUF1	Q14103-1	9	12	8	11	32.1	38.434	355	**3.2811**	**29**	**0.4307**	**32**
RPL23A	P62750	11	11	11	11	43.6	17.695	156	**2.7402**	**29**	**0.4362**	**36**
SMARCA5	O60264	39	36	39	36	36.7	121.9	1052	**2.4841**	**60**	**0.4377**	**45**
UHRF1	A8K024	15	12	15	12	29.3	91.099	806	**2.2292**	**21**	**0.4393**	**13**
LMN2	Q03252	25	20	19	15	42.3	69.948	620	**2.2876**	**36**	**0.4394**	**16**
CDC46	P33992	25	22	25	22	45.1	82.285	734	**2.2345**	**50**	**0.4400**	**31**
HNRNPH3	P31942-1	6	6	5	5	26.3	36.926	346	**3.2615**	**11**	**0.4445**	**9**
UBA52	P62987	9	8	9	8	53.9	14.728	128	**2.8335**	**97**	**0.4455**	**89**
CENPV	Q7Z7K6-3	5	5	1	1	28.3	29.73	272	**2.3588**	**1**	**0.4473**	**1**
AIM	P26358-1	35	28	35	28	26.3	189.56	1678	**2.2571**	**40**	**0.4533**	**29**
KIAA1637	Q9HCD5	3	3	3	3	7.8	65.536	579	**6.9046**	**2**	**0.4570**	**2**
ARGLU1	Q9NWB6-1	3	2	3	2	15	33.216	273	**2.5414**	**4**	**0.4611**	**4**
KIAA0461	Q7Z3K3-1	2	3	2	3	3.3	155.34	1410	**3.0215**	**2**	**0.4618**	**2**
HNRNPF	P52597	8	7	6	5	25.3	45.671	415	**2.9972**	**14**	**0.4620**	**11**
HNRNPC	P07910-2	13	12	13	12	44.4	32.337	293	**2.5864**	**86**	**0.4624**	**50**
HNRNPA2B1	P22626-1	21	21	20	20	63.5	37.429	353	**3.4488**	**114**	**0.4657**	**93**
CDCA8	Q53HL2	7	2	7	2	30.4	31.323	280	**2.3632**	**8**	**0.4659**	**2**
SET	Q01105-1	3	4	3	4	22.8	33.488	290	**2.4842**	**6**	**0.4661**	**5**
hCG_22498	B2R7C5	31	29	31	29	44.2	95.907	853	**2.2148**	**66**	**0.4675**	**36**
LMN1	P02545-1	50	49	3	3	64.8	74.139	664	**2.2876**	**345**	**0.4685**	**236**
ZNF326	Q5BKZ1-1	6	4	6	4	12	65.653	582	**2.6638**	**8**	**0.4720**	**6**
FANCI	Q9NVI1-3	8	7	8	7	11.3	149.32	1328	**2.6239**	**6**	**0.4730**	**6**
KIAA1291	Q9HAV4-1	2	2	2	2	1.7	136.31	1204	**2.4632**	**2**	**0.4730**	**2**
HNRNPL	P14866	15	14	14	13	29.4	64.132	589	**3.5264**	**43**	**0.4735**	**35**
HNRPCL2	Q9UKM9-1	11	11	11	11	29.3	32.55	307	**2.7391**	**17**	**0.4741**	**15**
ILF2	Q12905	10	9	10	9	33.1	43.062	390	**2.4864**	**19**	**0.4741**	**15**
KHDRBS1	Q07666-1	6	5	6	5	14	48.227	443	**3.9629**	**11**	**0.4752**	**10**

CGI-59	Q9Y383-1	7	5	3	1	16.8	54.223	458	**2.5075**	**12**	**0.4761**	**14**
LMN2	P20700	38	35	32	30	59.7	66.408	586	**2.3030**	**114**	**0.4785**	**79**
RBM14	Q96PK6-1	15	14	13	11	27.5	69.491	669	**2.8425**	**39**	**0.4792**	**22**
DDX46	Q7L014	19	13	19	13	23	117.46	1032	**2.1652**	**22**	**0.4807**	**24**
H1F0	P07305	4	4	4	4	20.6	20.863	194	**3.0118**	**6**	**0.4814**	**4**
TARDBP	Q13148	4	4	4	4	12.3	44.991	416	**2.5003**	**6**	**0.4814**	**5**
KIAA0648	Q29RF7-1	10	14	9	12	15.8	150.83	1337	**2.2261**	**11**	**0.4861**	**14**
CG1	Q86UP2-1	9	20	9	20	18.3	156.27	1357	**2.5712**	**11**	**0.4894**	**24**
SPNR	Q96SI9-1	7	8	4	5	13.7	73.652	672	**3.5726**	**4**	**0.4899**	**5**
HNRNPA0	Q13151	9	7	9	7	39.3	30.84	305	**3.1982**	**20**	**0.4906**	**17**
ANP32A	P39687	7	3	3	2	26.5	28.585	249	**2.5897**	**8**	**0.4942**	**3**
PSF	P23246-1	22	24	20	22	38.3	76.149	707	**3.1636**	**74**	**0.4950**	**68**
H1FX	Q92522	10	9	10	9	41.3	22.487	213	**2.9653**	**29**	**0.4958**	**15**
MLN70	P31949	5	4	5	4	52.4	11.74	105	**3.0840**	**8**	**0.4966**	**5**
DDX5	P17844	29	28	22	21	41.5	69.147	614	**2.6988**	**73**	**0.4968**	**54**
FUS	P35637-1	6	6	4	4	11.2	53.376	528	**3.6193**	**18**	**0.4997**	**18**
DNAJC9	Q8WXX5	9	4	9	4	28.8	29.909	260	**3.3611**	**10**	**0.5003**	**3**
CPSF1	Q10570	6	6	6	6	6.8	160.88	1443	**3.1944**	**7**	**0.5028**	**5**
DBP2	O60231	4	2	4	2	4.3	119.26	1041	**2.6248**	**4**	**0.5038**	**2**
RBAP46	Q16576	6	4	2	1	15.3	47.82	425	**4.5516**	**7**	**0.5084**	**5**
KIAA0144	Q14157-2	3	6	3	6	6.5	114.53	1087	**2.6252**	**3**	**0.5105**	**8**
HNRNPK	P61978-3	19	18	19	18	53	48.562	440	**3.2319**	**77**	**0.5105**	**79**
HNRNPR	O43390-1	20	20	1	1	37.9	70.942	633	**3.0983**	**66**	**0.5136**	**62**
MATR3	A8MXP9	21	21	21	21	33.4	99.966	895	**2.6669**	**55**	**0.5152**	**70**
HELLS	Q9NRZ9-1	14	7	14	7	18.6	97.073	838	**2.7513**	**16**	**0.5214**	**5**
ABBP1	Q99729-2	10	9	10	9	29.8	35.967	332	**2.9267**	**28**	**0.5251**	**23**
SNRPE	P62304	4	2	4	2	65.2	10.803	92	**3.0900**	**4**	**0.5256**	**1**
BAF170	Q8TAQ2-1	2	5	1	2	5.4	132.88	1214	**3.1546**	**2**	**0.5285**	**6**

CTDSPL2	Q05D32-1	3	3	3	3	5.8	52.998	466	**3.2513**	**3**	**0.5334**	**2**
DXS423E	Q14683	33	35	33	35	34	143.23	1233	**2.1855**	**56**	**0.5385**	**41**
HNRNPH1	P31943	11	9	5	4	35.2	51.229	472	**2.3682**	**47**	**0.5432**	**28**
CFIM25	O43809	9	8	9	8	39.2	26.227	227	**2.8758**	**15**	**0.5444**	**9**
APE	P27695	9	8	9	8	46.2	35.554	318	**3.0723**	**16**	**0.5484**	**10**
MCM6	Q14566	30	26	30	26	43	92.888	821	**2.1212**	**49**	**0.5492**	**31**
HNRNPUL2	Q1KMD3	13	17	13	17	26.4	85.104	747	**3.0236**	**20**	**0.5498**	**20**
RPS11	P62280	11	11	5	4	57.6	18.431	158	**2.2376**	**26**	**0.5533**	**25**
FUBP2	Q92945-1	16	14	13	12	31.4	73.114	711	**2.7493**	**17**	**0.5550**	**15**
RPS24	P62847-1	4	3	4	3	16.3	32.43	289	**2.4187**	**5**	**0.5578**	**6**
KIF4	O95239-1	3	2	3	2	2.4	139.88	1232	**3.0106**	**3**	**0.5589**	**1**
EIF4B	P23588	3	2	3	2	6.5	69.725	616	**2.4866**	**4**	**0.5598**	**3**
HMG2	P26583	8	6	6	4	33	24.033	209	**2.8399**	**17**	**0.5607**	**4**
CSTF3	Q12996	5	4	5	4	12.7	82.921	717	**2.7508**	**5**	**0.5611**	**3**
FUBP1	Q96AE4-1	21	17	17	15	43.9	67.689	645	**2.7642**	**38**	**0.5618**	**31**
CDC47	P33993-1	25	18	25	18	47.8	81.307	719	**2.1318**	**52**	**0.5618**	**25**
RPS7	P62081	12	10	12	10	50	22.127	194	**2.4593**	**25**	**0.5632**	**21**
CDC21	P33991	24	22	24	22	36.7	96.557	863	**2.4214**	**63**	**0.5673**	**27**
hCG_20560	Q9BUQ0	13	9	13	9	35	59.632	557	**2.7678**	**39**	**0.5734**	**27**
KIAA1470	Q9P258	14	12	14	12	34.7	56.084	522	**2.9194**	**30**	**0.5746**	**26**
E1BAP5	Q9BUJ2-1	5	6	5	6	10.6	95.737	856	**2.5961**	**9**	**0.5831**	**6**
WDR76	Q9H967	8	9	8	9	14.7	69.752	626	**2.4178**	**12**	**0.5854**	**8**
DDX17	Q59F66	23	21	16	14	35.2	80.457	731	**2.6552**	**43**	**0.5856**	**36**
RPL38	P63173	6	4	6	4	52.9	8.2178	70	**3.6526**	**9**	**0.5868**	**6**
RPL6	B2R4K7	13	14	13	14	49.5	32.872	289	**2.4866**	**27**	**0.5887**	**37**
DDX38	Q92620	4	3	4	3	5.5	140.5	1227	**2.2700**	**3**	**0.5900**	**3**
LEMD3	Q9Y2U8	6	4	6	4	9.7	99.996	911	**2.1906**	**6**	**0.5947**	**4**
ERH	P84090	3	3	3	3	37.5	12.259	104	**2.2329**	**3**	**0.5985**	**3**

VRK1	Q99986	2	2	2	2	11.6	45.476	396	**5.0153**	**2**	**0.6002**	**2**
SRP9	P49458	4	4	4	4	43	10.112	86	**2.5566**	**5**	**0.6041**	**5**
RPL17	P18621	9	9	9	9	51.6	21.397	184	**2.4736**	**24**	**0.6090**	**27**
HAUSP	Q93009	14	12	14	12	16.2	128.3	1102	**2.1417**	**17**	**0.6108**	**13**
EEF1D	P29692-1	9	8	1	1	55.2	31.121	281	**3.1740**	**1**	**0.6108**	**4**
DRBF	Q12906-4	21	19	18	16	26.8	95.807	898	**2.4906**	**62**	**0.6151**	**36**
SRP14	P37108	7	5	7	5	44.1	14.57	136	**2.8571**	**9**	**0.6166**	**6**
HSD48	Q13765	4	4	4	4	5.9	94.68	925	**2.1525**	**10**	**0.6185**	**13**
RBP56	Q92804-1	3	4	1	2	8.3	61.829	592	**3.0270**	**1**	**0.6228**	**2**
KIAA1401	Q2NL82	13	11	13	11	21	91.809	804	**2.2232**	**18**	**0.6253**	**12**
C6orf150	Q8N884-1	7	6	7	6	18.4	58.814	522	**2.8504**	**9**	**0.6308**	**4**
HNRNPM	P52272-1	31	26	31	26	50.1	77.515	730	**2.6294**	**87**	**0.6336**	**73**
WBSCR22	C9K060	3	2	3	2	16.4	33.845	298	**2.5071**	**3**	**0.6337**	**3**
AAG	P29372-1	10	4	10	4	41.3	32.868	298	**2.9860**	**10**	**0.6394**	**5**
ELAVL1	B4DVB8	12	9	12	9	36.5	38.996	353	**2.9276**	**17**	**0.6398**	**17**
EMC19	P63208-1	3	3	3	3	20.2	18.658	163	**2.1302**	**3**	**0.6436**	**3**
ARS2	Q9BXP5-1	6	4	6	4	10.2	100.67	876	**2.2638**	**6**	**0.6503**	**6**
UHX1	P51784	2	2	1	2	2.8	109.82	963	**2.8429**	**2**	**0.6527**	**2**
FEN1	P39748	9	9	9	9	31.8	42.592	380	**2.3008**	**13**	**0.6546**	**8**
BAM	Q9UQE7	22	25	22	25	25.6	141.54	1217	**2.1525**	**34**	**0.6549**	**26**
DBC1	Q8N163-1	11	12	11	12	20.3	102.9	923	**2.1721**	**13**	**0.6554**	**18**
RPL1	P36578	16	17	16	17	38.9	47.697	427	**2.4236**	**37**	**0.6595**	**33**
FLYWCH1	Q4VC44-3	7	5	1	1	17.1	85.152	765	**2.1598**	**9**	**0.6618**	**2**
HCA90	Q96RR5	3	2	3	2	5.5	89.392	783	**2.8632**	**4**	**0.6626**	**2**
LBR	Q14739	2	2	2	2	3.1	70.702	615	**2.3942**	**3**	**0.6666**	**7**
EIF2G	P41091	6	5	6	5	12.9	51.109	472	**4.6697**	**8**	**0.6686**	**6**
LAP2	P42167-1	9	10	4	4	29.3	50.67	454	**6.3721**	**5**	**0.6727**	**4**
DFS70	O75475-1	5	4	5	4	13.2	60.103	530	**2.6389**	**6**	**0.6767**	**6**

TKT	B4DE31	25	23	25	23	45.3	68.741	631	**2.2096**	**72**	**0.6891**	**46**
CDYL	Q9Y232-1	2	2	2	2	4.2	66.481	598	**2.3043**	**2**	**0.6921**	**1**
hCG_2016483	B3KUY2	3	3	3	3	31.5	19.448	165	**2.2226**	**4**	**0.6949**	**3**
RPL7A	P62424	16	15	16	15	45.5	29.995	266	**2.3846**	**31**	**0.6972**	**33**
POLR2B	P30876	4	4	4	4	6.5	133.9	1174	**2.4870**	**4**	**0.6980**	**4**
AEG1	Q86UE4	3	5	3	5	13.9	63.836	582	**2.2140**	**3**	**0.6980**	**6**
CFIM68	Q16630-2	3	4	3	4	8.8	63.47	588	**2.6267**	**5**	**0.6988**	**4**
DAD1	P61803	3	3	3	3	26.5	12.497	113	**2.2949**	**4**	**0.6991**	**3**
RPL12	P30050-1	8	7	8	7	68.5	17.818	165	**2.3098**	**10**	**0.6999**	**11**
EEF1G	P26641	12	12	12	12	37.8	50.118	437	**2.8496**	**30**	**0.7096**	**20**
NSUN2	Q08J23	18	13	18	13	30.2	86.47	767	**2.1227**	**26**	**0.7124**	**15**
GTF3C4	Q9UKN8	4	2	4	2	4.5	91.981	822	**2.4275**	**4**	**0.7139**	**2**
RPS10	P46783	6	7	6	7	37.6	18.898	165	**2.2386**	**12**	**0.7148**	**13**
PA2G4	Q05D08	9	9	9	9	33.7	45.151	406	**2.5044**	**11**	**0.7182**	**15**
TIAL1	A8K4L9	2	3	1	1	11	43.448	392	**2.9172**	**3**	**0.7206**	**4**
KIAA0664	O75153	6	13	6	13	11.1	146.67	1309	**2.4806**	**6**	**0.7237**	**12**
CACY	P06703	2	2	2	2	16.7	10.18	90	**2.6325**	**3**	**0.7265**	**2**

APPENDIX V. Proteins identified and quantified with a least two peptides, one of which unique, in the H3K4me3 interactome. Protein Group output from MaxQuant software, representing the Top40% of proteins binders.

Gene Names	Uniprot	Peptides For	Peptides Rev	Unique Pep. For	Unique Pep. Rev	Sequence Coverage [%]	Mol. Weight [kDa]	Sequence Length	Ratio H/L For	Ratio H/L Count For	Ratio H/L Rev	Ratio H/L Count Rev
MKI67	P46013-1	43	52	43	52	23	358.69	3256	**1.7868**	**83**	**0.4569**	**103**
H4/A	P62805	10	11	10	11	53.4	11.367	103	**2.1203**	**91**	**0.4112**	**99**
ADPRT	P09874	33	32	33	32	38.3	113.08	1014	**1.7925**	**55**	**0.4903**	**78**
H3F2	Q71DI3	8	8	1	1	60.3	15.388	136	**2.3171**	**79**	**0.2245**	**71**
UBA52	P62987	9	9	9	9	53.9	14.728	128	**1.8880**	**67**	**0.5509**	**69**
H1F2	P16403	9	13	3	5	41.8	21.364	213	**1.9563**	**71**	**0.2012**	**66**
DDX21	Q9NR30-1	29	30	27	28	44.7	87.343	783	**1.7614**	**53**	**0.5866**	**60**
TOP2	P11388-4	39	35	28	23	25	182.68	1612	**1.9910**	**63**	**0.4972**	**53**
BAP135	P78347-1	21	26	21	26	28.4	112.42	998	**1.7741**	**37**	**0.4932**	**38**
H2AFY	O75367-1	13	16	13	16	51.9	39.617	372	**1.7913**	**27**	**0.3936**	**32**
SMARCA5	O60264	20	26	20	26	26	121.9	1052	**1.9681**	**30**	**0.5696**	**32**
CHD4	Q14839-2	25	29	25	29	14.9	220.83	1940	**2.0812**	**38**	**0.4547**	**30**
KIAA1470	Q9P258	7	13	7	13	34.7	56.084	522	**2.4348**	**12**	**0.4757**	**25**
OK/SW-cl.32	P39023	12	12	12	12	24.8	46.108	403	**1.8436**	**22**	**0.4282**	**24**
NCL	P19338	16	15	16	15	28.5	76.613	710	**2.1770**	**21**	**0.6351**	**20**
CATX11	O76021	15	16	15	16	30	54.972	490	**2.0376**	**18**	**0.5078**	**20**
POLR2B	P30876	18	17	18	17	25.1	133.9	1174	**4.3296**	**19**	**0.3597**	**20**
SIN3A	Q96ST3	20	18	20	18	19.6	145.17	1273	**4.9947**	**27**	**0.3457**	**19**
H1FX	Q92522	4	5	4	5	26.3	22.487	213	**1.7493**	**15**	**0.5633**	**18**
H2BFD	B4DR52	7	9	2	2	42.8	18.041	166	**2.3429**	**22**	**0.2335**	**18**
RPL7	A8MVV7	10	11	10	11	46.3	32.275	272	**1.7305**	**16**	**0.5883**	**17**
HMGB1	Q5T7C6	6	8	5	7	30.6	25.814	216	**4.5691**	**11**	**0.1385**	**16**
IFI16	Q16666-1	14	13	14	13	21.5	88.255	785	**1.8129**	**16**	**0.5098**	**16**
FACT140	Q9Y5B9	11	14	11	14	16.5	119.91	1047	**2.7876**	**16**	**0.6320**	**16**
DEK	P35659	8	8	8	8	24.5	42.674	375	**2.4557**	**13**	**0.4984**	**15**
hCG_1744585	B5MDF5	8	8	8	8	28.4	26.409	236	**1.8618**	**17**	**0.6449**	**15**

CCG2	P62701	12	10	12	10	42.6	29.597	263	**1.7695**	**21**	**0.5316**	**15**
DDB1	Q16531	9	13	9	13	12	126.97	1140	**2.0345**	**10**	**0.6181**	**15**
POLR2	P24928	20	19	20	19	12.8	217.17	1970	**3.4119**	**27**	**0.5150**	**15**
HCFC1	A6NEM2	13	14	12	13	9.3	213.47	2080	**4.0966**	**17**	**0.2985**	**14**
HRX2	Q9UMN6-1	25	34	23	30	17.4	293.51	2715	**9.1265**	**7**	**0.6063**	**14**
SPT5	O00267-1	9	14	9	14	17.8	121	1087	**4.2804**	**7**	**0.3172**	**13**
BRIX	Q8TDN6	7	8	7	8	25.2	41.401	353	**1.8298**	**11**	**0.5072**	**12**
DDX48	P38919	11	10	9	8	29.4	46.871	411	**1.9080**	**8**	**0.4565**	**12**
FACT80	Q08945	11	11	11	11	20.5	81.074	709	**2.9439**	**11**	**0.4560**	**12**
FBL	P22087	7	8	6	7	29	33.784	321	**2.6688**	**10**	**0.1622**	**11**
KIAA0648	Q29RF7-1	11	13	9	11	11.3	150.83	1337	**1.7513**	**12**	**0.5711**	**11**
H2AFZ	P0C0S5	4	4	2	2	22.1	18.414	181	**3.4120**	**10**	**0.4642**	**10**
OK/SW-cl.103	P32969	5	6	5	6	32.3	21.863	192	**1.7475**	**7**	**0.6451**	**10**
PA2G4	Q05D08	4	7	4	7	20.2	45.151	406	**1.9197**	**4**	**0.6417**	**10**
HDGF	P51858	6	7	5	6	32.5	26.788	240	**2.2109**	**7**	**0.3490**	**10**
HNRNPA3	P51991-1	5	6	5	6	21.2	39.594	378	**1.9481**	**6**	**0.3712**	**9**
HIST2H3PS2	Q5TEC6	6	6	1	1	36.8	15.43	136	**5.3761**	**9**	**0.0939**	**9**
EBNA1BP2	Q99848	7	6	7	6	19.9	40.684	361	**2.4007**	**9**	**0.1223**	**9**
ILF2	Q12905	6	7	6	7	21.5	43.062	390	**1.8692**	**6**	**0.5533**	**9**
DDX27	Q96GQ7	9	10	9	10	15.2	89.834	796	**1.8093**	**10**	**0.6135**	**9**
HNRNPA0	Q13151	2	2	2	2	5.2	30.84	305	**1.7985**	**7**	**0.4188**	**8**
RPL38	P63173	5	5	5	5	50	8.2178	70	**2.2918**	**7**	**0.1413**	**8**
NPM	P06748-1	3	6	3	6	35.7	32.575	294	**1.7325**	**6**	**0.4265**	**8**
RPL6	B2R4K7	8	8	8	8	31.5	32.872	289	**2.0003**	**8**	**0.2603**	**8**
APE	P27695	3	7	3	7	30.2	35.554	318	**2.5524**	**3**	**0.3147**	**7**
GTBP	P52701-1	4	7	4	7	5.8	152.78	1360	**1.8848**	**5**	**0.6428**	**7**
RECQ1	P46063	7	7	7	7	15.4	73.457	649	**2.6266**	**9**	**0.4508**	**7**
UBF	P17480-1	5	8	5	8	11.4	89.405	764	**2.5658**	**5**	**0.5814**	**7**

VCP	P55072	11	9	11	9	23.3	89.321	806	**5.2788**	**7**	**0.3455**	**7**
SMT3B	P61956	2	2	2	2	23.2	10.871	95	**2.1903**	**8**	**0.5237**	**6**
DB1	Q14119	3	3	2	2	5.6	56.931	521	**5.0997**	**5**	**0.1494**	**6**
RPL13A	P40429	3	4	3	4	17.7	23.577	203	**2.0527**	**4**	**0.5182**	**6**
EEF1G	P26641	3	5	3	5	16	50.118	437	**1.8519**	**3**	**0.2553**	**6**
CDABP0061	Q9Y5J1	5	5	5	5	14.3	64.063	573	**1.7505**	**4**	**0.5171**	**6**
BAF170	Q8TAQ2-1	3	6	1	4	6	132.88	1214	**1.8881**	**4**	**0.5652**	**6**
HR21	O60216	4	7	4	7	13.5	71.689	631	**2.2118**	**4**	**0.5660**	**6**
KIAA0162	Q7KZ85-1	10	10	10	10	11.7	199.07	1726	**3.0494**	**10**	**0.3732**	**6**
SRP14	P37108	4	4	4	4	55.1	14.57	136	**1.8638**	**6**	**0.1708**	**5**
CYPA	P62937	4	5	3	4	23	18.012	165	**2.1145**	**3**	**0.6245**	**5**
RBBP5	Q15291-1	6	7	6	7	17.1	63.303	573	**7.2304**	**2**	**0.4485**	**5**
SAMD1	Q6SPF0	8	8	8	8	13.9	56.051	538	**5.1117**	**6**	**0.3015**	**5**
BAF180	Q86U86-1	8	10	8	10	6.5	192.95	1689	**2.1068**	**5**	**0.6228**	**5**
CTCF	P49711	3	3	3	3	3.6	82.785	727	**3.3665**	**4**	**0.1636**	**4**
RCC1	Q16269	4	3	4	3	11.1	48.145	452	**2.3307**	**6**	**0.3352**	**4**
FEN1	P39748	4	3	4	3	9.2	42.592	380	**1.7766**	**4**	**0.3441**	**4**
GLYR1	Q49A26-1	4	4	4	4	10.1	60.556	553	**1.7678**	**4**	**0.4382**	**4**
CGI-37	Q9Y221-1	2	2	2	2	13.3	20.462	180	**1.9645**	**3**	**0.6161**	**3**
EIF2A	P05198	2	2	2	2	8.6	36.112	315	**1.8728**	**1**	**0.1305**	**3**
BZAP45	B4DLZ8	2	3	2	2	6	51.281	451	**2.2579**	**2**	**0.5996**	**3**
BIG3	P61964	3	3	3	3	15.3	36.588	334	**3.4671**	**4**	**0.2711**	**3**
BXDC1	Q9H7B2	3	3	3	3	13.4	35.582	306	**1.8820**	**4**	**0.1162**	**3**
ASH2L	Q9UBL3-1	5	3	5	3	12.3	68.722	628	**2.8407**	**5**	**0.5975**	**3**
DFS70	O75475-1	3	4	2	3	8.5	60.103	530	**1.9358**	**4**	**0.0581**	**3**
hCG_1811093	Q96AN2	4	5	4	5	10.7	55.673	522	**9.1732**	**3**	**0.2341**	**3**
SAP130	Q9H0E3-2	2	2	2	2	1.9	113.97	1083	**5.5375**	**1**	**0.6393**	**2**
GIG38	O00422	2	2	2	2	12.2	19.526	172	**1.9946**	**2**	**0.5518**	**2**

SRPR	P08240	2	2	2	2	3	69.81	638	**1.8588**	**2**	**0.6246**	**2**
LIG3	P49916-1	2	2	2	2	3	112.91	1009	**1.7615**	**2**	**0.4606**	**2**
HMG2	P26583	2	2	1	1	12	24.033	209	**1.7286**	**2**	**0.5274**	**2**
RIG	P62841	2	3	2	3	13.8	17.04	145	**2.5875**	**2**	**0.1277**	**2**
KIAA0662	O75151	3	3	3	2	4.9	121.06	1099	**4.3333**	**2**	**0.1540**	**2**
MIG10	P00558	4	3	4	3	14.9	44.614	417	**1.9933**	**3**	**0.2479**	**2**
SMARCA4	B9EGQ8	3	4	3	4	4.3	189.45	1681	**1.7334**	**2**	**0.3606**	**2**
WDR76	Q9H967	4	5	4	5	8.9	69.752	626	**2.0364**	**4**	**0.6207**	**2**
C20orf158	Q9BTC0-4	2	6	2	6	3.5	243.87	2240	**2.1987**	**2**	**0.3005**	**2**
EZR	P15311	5	8	1	3	13.7	69.412	586	**2.3783**	**1**	**0.6425**	**2**
OCR	Q9Y657	2	2	2	2	13	29.6	262	**2.4112**	**1**	**0.0821**	**1**
BANP	Q8N9N5-1	3	2	3	2	8.5	56.484	519	**4.8794**	**1**	**0.6173**	**1**
ZBP89	Q9UQR1	3	4	3	4	8.3	88.975	794	**4.9128**	**2**	**0.2865**	**1**
PPP1CB	P62140	4	6	1	2	17.4	37.186	327	**4.3475**	**1**	**0.0738**	**1**
CIF150	Q6P1X5	2	1	2	1	1.3	136.97	1199	**3.1018**	**1**		**0**
IWS1	Q96ST2-1	3	2	3	2	4	91.954	819	**2.2656**	**3**		**0**
BPTF	Q12830-1	1	3	1	3	1.1	338.26	3046		**0**		**0**
BRD4	O60885-1	2	3	2	2	2.5	152.22	1362		**0**		**0**

ACKNOWLEDGEMENTS

First of all, I am grateful to my PhD supervisor, Dr. Tiziana Bonaldi, for giving me the opportunity to work in her laboratory and for her support and guidance during the years of my PhD study.

I would also to express my gratitude to the internal supervisor Dr. Gioacchino Natoli and external supervisor Dr. Tobias Straub for their great suggestions and comments.

A special thanks goes to people of TB group for their teaching, assistant and support. Thank you very much!

I would also to thank Serena Ghisletti, Sara Polletti and Paolo Soffientini.

(In Italian!) Un grazie speciale a mamma e papà perché mi avete sempre sostenuto anche quando (quasi sempre!) ho fatto di testa mia!

…e adesso un grazie infinito a Paolo ed Andrea, la mia FAMIGLIA!...il grazie più grande va a Paolo con cui ho condiviso a pieno questi anni, non sempre facili. Grazie perché ci sei sempre stato e mi hai sempre supportato e capito anche quando, me ne rendo conto, non era semplice!!!

I would like to thank the many peoples who made this thesis possible!!!