Published in final edited form as:

Methods Enzymol. 2017 ; 586: 359–378. doi:10.1016/bs.mie.2016.10.021.

Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology

Kevin A. Janssen, Simone Sidoli, and Benjamin A. Garcia*

Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

Functional epigenetic regulation occurs by dynamic modification of chromatin, including genetic material (i.e. DNA methylation), histone proteins, and other nuclear proteins. Due to the highly complex nature of the histone code, mass spectrometry (MS) has become the leading technique in identification of single and combinatorial histone modifications. MS has now overcome antibody based strategies due to its automation, high resolution and accurate quantitation. Moreover, multiple approaches to analysis have been developed for global quantitation of post-translational modifications (PTMs), including large-scale characterization of modification co-existence (middle-down and top-down proteomics), which is not currently possible with any other biochemical strategy. Recently, our group and others have simplified and increased the effectiveness of analyzing histone PTMs by improving multiple MS methods and data analysis tools. This review provides an overview of the major achievements in the analysis of histone PTMs using MS with a focus on the most recent improvements. We speculate that the workflow for histone analysis at its state-of-the-art is highly reliable in terms of identification and quantitation accuracy, and it has the potential to become a routine method for systems biology thanks to the possibility of integrating histone MS results with genomics and proteomics datasets.

1. Introduction

The nucleosome is formed by two each of histone proteins H2A, H2B, H3, and H4, or histone variants. DNA is wrapped around the nucleosome twice in a supercoil, which is the first level of condensation of genetic material into chromatin. Histone proteins are critical for controlling the conformation of DNA in the nucleus: they are dynamically modified by post-translational modifications (PTMs), and many of the modifications have been correlated with regulatory cellular processes and chromatin structure. These observations led to the theory that histone modification is a regulatory event that leads observed cellular functions; for example, acetyl groups on lysine residues neutralize positive charges that attract the phosphate backbone of DNA, suggesting that the DNA can be more easily freed from the nucleosome and accessed by proteins involved in transcription, repair, and other genetic functions (Hebbes, Thorne, & Crane-Robinson, 1988). Since the inception of this

*Corresponding author: Benjamin A. Garcia, Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Room 9-124, 3400 Civic Center Blvd, Bldg 421, Philadelphia, PA 19104, USA. bgarci@mail.med.upenn.edu; Phone: 1-215-573-9423; Fax: 215-573-4764.

hypothesis, different histone modification states have been implicated in a wide breadth of cellular states and functions, including differentiation, cell type, cell cycle, cell signaling, cell shape, and disease states, to name a few. Indeed, to date, four inhibitors of histone deacetylases (HDACs) have been approved by the FDA for treatment of lymphomas and myelomas, and many are in clinical trials to treat cervical, lung, breast, and liver cancer, among others. The advancement of knowledge of the roles of dynamic histone PTMs is critical to understanding cellular biology, and will inevitably lead to improvements in disease outcomes.

1.1 Modifications Overview

Most histone PTMs are localized on the N-terminal tail of the histone sequence. The structure of the nucleosome leaves the N-terminal tails of the histones exposed, allowing them to be readily accessed by members of histone-related protein families. These families are grouped in three primary categories: writers, readers, and erasers. Writers include proteins that add modifications, such as histone acetyltransferases (HATs), histone methyltransferases (HMTs), kinases, and others. Readers are usually not enzymes, but proteins containing domains that recognize specific modifications, potentially recruiting complexes like transcription factors or DNA repair machinery. Erasers remove modifications, including acetyl groups (HDACs) and methyl groups (HDMs). The presence of PTMs can facilitate cellular outcomes by altering the chemical environment of the nucleosome, which can recruit readers to the chromatin or disrupt interactions between complexes comprised of nucleic acids and/or proteins (Sidoli, Cheng, & Jensen, 2012). A diverse array of modifications have been identified, including methylation (Oliver & Denu, 2011), phosphorylation (Price & D'Andrea, 2013), SUMOylation (Shiio & Eisenman, 2003), ubiquitination (Li, Nagaraja, Delcuve, Hendzel, & Davie, 1993), glycosylation (Dehennaut, Leprince, & Lefebvre, 2014), and acyl modifications such as acetylation, propionylation, butyrylation, and crotonylation (Rousseaux & Khochbin, 2015). These modifications are mostly found on the amino group of lysine residues, although exceptions are common; for instance, arginine residues may be methylated or citrullinated (deiminated) (Cuthbert et al., 2004), while serine and threonine residues are subject to phosphorylation and O-linked glycosylation. Serine and threonine residues have also been observed to be acetylated (Britton et al., 2013), but their abundance is too low to currently establish a biological role. In general terms, almost any protein PTM detected in cells can be found on histones as well. Whether the rare marks are functionally active or “chemical noise” is still argument of debate. In fact, CoAs are reactive molecules; acetylation led by chemical reactivity of acetyl-CoA has been observed in histones and other proteins in the absence of acetyltransferases. To date, the complex equilibria between writing and erasing PTMs from chromatin are still not fully understood. Another reason it is hard to assign functions to histone modifications is their combinatorial nature; a given PTM co-exists with myriad others on the same histone protein, and it is not rare for them to be found to cross-talk with one another (Sadakierska-Chudy & Filip, 2015). However, it is not yet possible to identify all known modifications in a single experiment. Mass spectrometry (MS) is currently the most suitable technique to address this challenge. MS is the ideal approach for analysis of histones because it overcomes many of the limitations of alternative approaches such as western blots and DNA sequencing. Antibodies used for western blots are expensive to produce, and can yield

inadequate specificity (Egelhofer et al., 2011). Use of an antibody generalized to any acetylations on histone H3 is common, but data cannot discern the location or quantity of acetyl groups, and false positives can be caused by acetylation of histone H4. Other antibodies share similar, if not more significant, drawbacks. Furthermore, discovery of unexpected histone PTM changes and combinatorial marks is limited because highly specific hypotheses must be generated with limited prior knowledge before any analysis is performed. Conversely, MS allows for unbiased identification of histone modifications, meaning that the entire modification profile can be related to cell functions rather than specific selected modifications. Additionally, MS provides accurate quantitation of the relative abundance of histone modifications within a single sample using a little as a single MS run. However, compared to traditional proteomics workflows, histone analysis requires careful specific procedures, as we will discuss in the next section.

2. Sample Preparation

2.1 Background and Standard Approaches

There are three approaches to using mass spectrometry to analyze histones: bottom-up, top-down, and middle-down. The techniques vary in the type of information they produce and are selected based on what information is needed. Bottom-up involves extensive digestion of proteins; peptides can be more easily separated by reversed-phase HPLC, providing higher sensitivity, but the nearly all information about co-existence of combinatorial marks cannot be obtained due to the fact that most PTMs are split on different peptides. Top-down is the analysis of intact proteins; with extensive fragmentation, the PTMs present in a single isoform can be detected. However, isoforms of histones do not separate as well as peptides by HPLC, and thus the complexity of the spectra generated by co-isolating a large number of proteoforms harms identification and differential quantitation of isobaric species. Moreover, many isoforms can be missed due to the low sensitivity associated with large protein analysis, caused by issues such as charge state distribution, the height of MS resolution required, and poor resolution of LC peaks. Middle-down is a compromise between the other two techniques: proteins are digested into large peptides, allowing for medium quality separation, most commonly by hybrid hydrophilic HPLC, and a better overview of co-existing PTMs on histone tails. However, middle-down cannot currently be used with the same throughput and sensitivity of bottom-up, as chromatographic separations are longer and low abundance PTMs are frequently missed. Each of these strategies has a different sample preparation procedure. While middle-down and top-down strategies have a very simple procedures, bottom-up requires special precautions due to the generation of hydrophilic peptides.

2.2 Recent Improvements in Histone Sample Preparation

Sample preparation for middle-down analysis adopts GluC or AspN digestion, as they cleave at specific acidic amino acid residues (glutamic and aspartic acid, respectively). Those amino acids are rare on histone sequences, and absent on the N-terminal tails, so this type of digestion leads to the generation of large N-termini polypeptides roughly 50 to 60 amino acids in length. In bottom-up analysis, the most adopted method applies derivatization of lysine residues followed by trypsin digestion. In particular, propionic anhydride is used for

propionylation of lysine side chains because trypsin would otherwise cleave at this amino acid and generate excessively short hydrophilic peptides that would be difficult to retain and separate by RP-HPLC (Garcia et al., 2007). Though propionylation must be performed with care, the protocol has also been optimized for derivatization and digestion in a 96-well plate format (not published). This improvement has the potential to dramatically increase the high throughput in the histone preparation process. However, it is critical to keep all solutions within the pH range of 7–9; propionic anhydride reacts with free amines and releases propionic acid, which lowers pH. At low pH, propionic anhydride can hydrolyze without reacting with an amine, resulting in underivatized lysine residues and N-termini (Lin & Garcia, 2012). If pH is above 10, amino acids with high pK_a can be propionylated, providing artifacts that are problematic for data analysis. Because some samples may contain basic proteins in addition to histones that could cause pH changes during propionylation, it is critical to monitor the pH of the reactions, even if the protocol is strictly adhered to (Sidoli, Bhanu, Karch, Wang, & Garcia, 2016), thus, the 96-well plate format requires even more care than single sample preparation. Recently, a few publications highlighted the issues related to unspecific derivatization of histones, which can potentially lead to incorrect PTM quantitation (Meert et al., 2016; Meert, Govaert, Scheerlinck, Dhaenens, & Deforce, 2015; Paternoster et al., 2016; Sidoli & Garcia, 2015; Soldi, Cuomo, & Bonaldi, 2016). However, the general conclusion from all these publications is that a properly followed protocol leads to minimal side reactions that do not hamper the final quantitation.

Other improvements have been suggested, some of them proving useful, while others have not left lasting impact. For instance, Liao et al. presented Propionate-NHS derivatization of peptide N-termini to increase peptide hydrophobicity and improve their retention on reversed phase chromatography (Liao et al., 2013). In addition, Maile et al. used phenyl isocyanate for the same purpose (Maile et al., 2015). Our lab has investigated the use of different anhydrides, including modifications more hydrophobic than propionate, to improve chromatographic separation. However, those other chemicals proved to be less effective or provide lower overall efficiency such that quantitative results could not be reliably reproduced (Sidoli et al., 2015). In the next section we discuss crucial differences between the three major strategies for histone analysis, including very recent methodological improvements.

3. Mass Spectrometry-based Strategies for Histone Analysis

3.1 Bottom-up mass spectrometry

Bottom-up is the conventional method used for analysis of PTMs because it provides the highest sensitivity and throughput. The technique allows for interrogation of a broad variety of histone PTMs. The basic principle of bottom-up mass spectrometry analysis involves the digestion of proteins down to peptides of 5–20 amino acid residues in length. This approach is the most similar to conventional shotgun proteomics, though the presence of several isobaric peptides demands particular attention when choosing the correct MS acquisition method.

3.1a DIA vs. DDA—Data-dependent acquisition (DDA) remains the most common MS data acquisition method in proteomics in general. DDA uses the most abundant ions in an MS1 spectrum to select for fragmentation. This relies on sample purity and abundance, as any contaminants can be selected for fragmentation, especially if the desired analyte does not provide adequate signals. Dynamic exclusion is used to combat this issue and enhance sample coverage; after an ion is selected for fragmentation, it will not be selected again for a set period of time, allowing the mass spectrometer to select less abundant ions. One of the challenges of this approach is that it remains possible for not all co-eluting peptides to be selected for fragmentation. If the MS is not fast enough, the lower abundance peptides will not be selected for fragmentation before the peptide has completely eluted. Thus, DDA produces very clean, high quality MS2 spectra, but it cannot select for all eluting ions within a small limit of time. While improvements in mass spectrometers have increased data acquisition rate greatly, very low abundance ions commonly remain unidentified. Analysis of these data is typically performed by matching MS2 spectra with known peptide sequences, resulting in output of a listed of identified peptides.

Histone peptides have peculiar issues. For instance, they often have isobaric forms that generate mixed MS2 spectra because they do not all separate well by chromatography and can have the same precursor mass (e.g. histone H3 aa 9–17, K_{ac}STGGKAPR and KSTGGK_{ac}APR). To solve this issue, we and others recently proposed to adopt data-independent acquisition (DIA) in order to plot fragment ion chromatographically and discriminate isobaric forms by using unique fragment ions (Krautkramer, Reiter, Denu, & Dowell, 2015; Sidoli, Fujiwara, & Garcia, 2016; Sidoli, Simithy, Karch, Kulej, & Garcia, 2015; Sidoli, Lin, et al., 2015). DIA for histones commonly uses large isolation windows (50 *m/z*) to select precursor ions for fragmentation, as noted in Figure 1A. This leads to the generation of a different chromatogram for every fragmentation window, as the repetitive cycle of the MS allows for the same amount of data from a fragmentation event as the precursor scan (Figure 1B). However, because many ions are selected for fragmentation, this method creates high complexity MS2 spectra. DIA is thus intuitively not ideal to identify new peptides, but it is very effective for accurate quantitation if the retention time and the mass of the peptide of interest is known. Fortunately, most histone modifications are known and analyses do not need to identify new peptides, thus, data analysis it is only reliant on peptide lists that are used to generate selected extracted ion chromatographs. We recently developed the in-house software EpiProfile, which performs peak area extraction using empirically designed retention time prediction and automatically performs differential quantitation of isobaric species by using the profile of the fragment ions (Yuan et al., 2015).

Another challenge of histone analysis is that the mass difference between highly abundant modifications such as acetylation (42.011 Da) and trimethylation (42.047 Da) is very small. Luckily, such modifications generate a stark difference in HPLC retention time, such that the elution times can be used to easily discriminate the two species. This difference is caused by hydrophobicity; acetylation is a more hydrophobic modification, and thus in reversed phase chromatography it will elute later.

3.1b Isotopic Labeling Methods—Isotopic labeling can also be applied in histone analysis to enhance the accuracy of relative quantitation between two samples. Labeling of

histones is done by addition of ^{13}C and ^{15}N isotopes of lysine and arginine. These amino acids are selected because they are the cleavage sites of trypsin, allowing confidence that all peptides will have at least one isotopically labeled amino acid residue. Additionally, it generally provides more confident quantitation because samples are mixed at early stages of sample preparation. Ideally, the advantage of this labeling is negligible in case of accurate sample preparation in label-free mode. However, unique and interesting applications of SILAC are the so-called “pulsed” experiments, which involves exposing cultured cells to the isotopic amino acids for a given period of time. This allows for monitoring of their turnover by calculating the ratio of heavily vs light labeled proteins. Cells can also be treated with a stimulus while plated with a heavy labeled media in order to verify which histone proteins are produced after the stimulus.

While labeling of lysine and arginine has clear advantages in relative quantitation, this approach not designed to optimally analyze the dynamic nature of histone PTMs. The principle behind the study of PTM dynamics is to provide cells with an isotopically labeled substrate that will be metabolized into a PTM, such as a methylation or acetylation. Histones can then be analyzed at set time points, and isotopic labels will reveal which PTMs are new and which are old. In the case of acetylations, $^{13}\text{C}_6$ glucose will be metabolized into $^{13}\text{C}_2$ acetyl groups on histone tails, and peptides can be observed to change from entirely unlabeled acetyl groups to multiple labeled acetyl groups, showing which acetylations are catalyzed first at what turnover rate. This can be used to determine which modifications have roles in the cell cycle (Alabert et al., 2015). To study methylation dynamics, ^{13}C methionine is used in cell culture. Methionine is then converted to s-adenosylmethionine SAM and used by histone methyltransferases to directly methylate histones. The effect is similar to the use of glucose to study acetylation, however, multiple methylations can be added to the same lysine residue. The presence of an unlabeled and a labeled methyl group on a lysine indicates a new methyl group has joined an old methyl group to create a new effect on environment of the histone. Further, a method has recently been developed to label O-linked β -N-acetylglucosamine (O-GlcNAc), a modification that can be added to many proteins, including histones H2A, H2B, and H4 (Wang et al., 2016). This technique is based on the combination of multiple labeling methods: heavy $^{13}\text{C}_6$ glucose, lysine($^{13}\text{C}_6$, $^{15}\text{N}_2$), and arginine($^{13}\text{C}_6$, $^{15}\text{N}_4$) are all used simultaneously. The amino acids are used to easily identify newly synthesized peptides, allowing for kinetic studies, and the glucose is rapidly metabolized to UDP-GlcNAc, causing nearly complete heavy labeling of O-GlcNAc modifications within 24 hours. The method depends heavily on an O-GlcNAc peptide enrichment strategy, which is outlined in 3.1e. Notably, although isotopic labeling of phosphorylation modifications has not been accomplished in cell culture, it has been accomplished *in nucleio* (Molden, Goya, Khan, & Garcia, 2014).

3.1c Combining Genomics with MS Interrogation (ChIP-MS)—Chromatin immunoprecipitation (ChIP) has recently had a great increase in use due to the advancement of genomic techniques such as next-generation sequencing. The standard of the technique uses formaldehyde to create crosslinks between protein-protein and protein-nucleic acid complexes, sonication to shear DNA into ~500 base pairs in length, and antibodies to precipitate analytes of interest. This is most commonly performed using antibodies against

transcription factors, histone PTMs (Soldi et al., 2016), histone readers (LeRoy et al., 2012), or histone variants (Won et al., 2015), followed by sequencing or qPCR to determine the genomic loci of the analyte. Many aspects of transcriptional regulation have been discovered in this manner, and correlations can be identified between an antibody's target and transcriptional activation or inactivation with inclusion of an additional analysis such as RNA-seq or proteomics. The ChIP protocol can be adapted to prepare samples for MS instead of genomic analysis, allowing for a broader interrogation of the state of the chromatin (Mohammed, Taylor, Brown, Papachristou, Carroll, & D'Santos, 2016), and presents the opportunity for reverse-ChIP, an approach that targets genomic sequences for pull-down rather than proteins (Byrum, Raman, Taverna, & Tackett, 2012; Waldrip et al., 2014). Using antibodies, antisense oligonucleic acids, or known high affinity proteins such as LexA or TetR, specific molecules can be targeted for pull-down followed by bottom-up MS analysis. The MS analysis can be applied to provide information of the histone variants and PTMs, and also to identify proteins associated with the target of the immunoprecipitation. Indeed, ChIP-MS based methods have been successfully used to identify interaction partners of H2A in yeast, interactors of the male-specific lethal (MSL) complex in *Drosophila*, and proteins associated with the lncRNA Xist (Wierer & Mann, 2016). Further development and use of these technologies will inevitably yield a wealth of epigenetic data describing mechanistic functions of proteins, complexes, and nucleic acids in chromatin.

3.1d Nucleosome Symmetry—While the analysis of histones and their modifications has advanced quite rapidly in the last decade, the analysis of macromolecular complexes like intact nucleosomes remains a challenge for *in vivo* and cell culture experiments. This is primarily due to the necessity to disassemble the nucleosome to perform analyses sensitive enough to detect histone modifications, preventing the possibility of specifically determining all modifications on both copies of an individual histone within a single nucleosome. However, larger-scope studies have been performed that utilize ChIP-MS to detect some (a)symmetry of modifications within a nucleosome, but it is not yet possible to perform in a completely unbiased approach. By use of a high quality antibody, a nucleosome containing a specific modification such as H4K20me1 can be isolated, then bottom-up MS will reveal the occupancy of the modification in the precipitated nucleosomes. If only one copy of the histone contains the modification, only 50% of the H4K20-containing peptides will be monomethylated; if the modification is symmetric, 100% of the peptides will show the H4K20me1 modification (Voigt et al., 2012). Because of the dependence on antibodies, this approach is low-throughput and relies heavily on antibody quality to produce meaningful data. Additionally, the detection of asymmetric marks suggests that there are more nucleosome isoforms than histone isoforms, meaning that separation of nucleosome isoforms for analysis is impractical. Though unaddressed, of particular interest would be determination of whether or not symmetric modifications are always added simultaneously though the use of isotopic labeling, but this would be best performed in an unbiased analysis. Thus, a creative approach to identifying the symmetry of all histone modifications in a sample would be high value advancement, but antibody-targeted methods can still provide meaningful information for directed hypotheses.

While histone modification symmetry analysis is a great challenge, it is simpler to determine of the symmetry of histone variants. Because the histone variants have different sequences, their genetic sequences can easily be modified to contain amino acid tags, such as FLAG, to enable simple and robust nucleosome purification. Thus, ChIP-MS can be used on FLAG-tagged histone variants in intact nucleosomes and subsequent analysis of the sample will show the nucleosome occupancy of the variant. For instance, Won et al. observed that H3.3 and H2A.Z are present mostly, if not completely, asymmetrically as they represent only ~50% of their respective histone populations (Won et al., 2015). Interestingly, a very small portion of nucleosomes were observed to contain both H3.3 and macroH2A, suggesting a highly specialized role for these nucleosomes. Because these data have been integrated with sequencing information, some of the regulatory roles of histone variants can be hypothesized based on genomic loci, but due to the overwhelming diversity of nucleosome isoforms, it is currently impossible to determine all of roles of an individual histone variant or modification.

3.1e Low Abundance Modification Enrichment—Conventional bottom-up is very effective for most acetylations and methylations, however, some modifications must be enriched for in order to be detected at quantitative levels. Presently, phosphorylation is the most commonly studied low abundance histone modification. The modification on histones has been strongly associated with the cell cycles, but much of this information could not have been obtained without enrichment. Enrichment is most commonly used to study epigenetic effects of protein kinase activity, both directly and indirectly. Phosphate enrichment improves detection and quantitation of phosphate modifications, though the inability to isotopically label phosphate groups *in vitro* remains a challenge to optimally quantitate the modifications and study kinetics. Phosphorylation enrichment is performed by passing a trypsin digested sample through a titanium dioxide column. The phosphopeptides are bound in 80% acetonitrile (ACN), 1% trifluoroacetic acid (TFA) and eluted in 40% ACN, 15% NH₄OH, thus requiring evaporation before resuspension and analysis by conventional RP-nanoLC-MS/MS. Enrichment has proven to be very useful by demonstrating the ability to detect far more phosphopeptides by MS than methods that exclude TiO₂ enrichment (Humphrey, Azimifar, & Mann, 2015; Thingholm & Larsen, 2016).

Other modifications like ubiquitination are usually enriched using antibodies. Enrichment and separation of glycosylated peptides remains a challenge, but enrichment of O-GlcNAc modifications have improved significantly in recent years. Most approaches required chemical biology techniques to alter GlcNAc modifications for specific affinity based enrichments, but very recently, the affinity of phenylboronic acid (PBA) for GlcNAc modifications has been exploited to enable efficient enrichment. Peptides are bound to PBA resin in DMSO, unmodified peptides are removed by washing with ACN, and O-GlcNAcylated peptides are removed in 0.1% TFA, which can be used to load the sample on the LC. The procedure resulted in a 39-fold enrichment of O-GlcNAc peptides, which has the capability of enabling detection of far more modified peptides than conventional analyses (Wang et al., 2016).

3.2 Middle-down mass spectrometry

Middle-down is a practical intermediate between bottom-up and top-down MS. The approach of middle-down is difficult to generalize to heterogeneous samples of protein due to the diversity of amino acid content due to inconsistency in uncommon amino acid content of different proteins. However, histone samples are particularly suitable for the purpose, as by cleaving acidic amino acid residues the intact N-terminal tail is obtained as large and analyzable polypeptide (50–60 aa length). Thus, middle-down MS has been recently presented as a major step in the deciphering of the histone code. In the early 2000s, the Allis lab postulated the existence of a histone code, which by definition includes the presence of combinatorial marks (Jenuwein & Allis, 2001). Essentially, the postulate suggests that multiple modifications on histones collectively contribute to determining which proteins and nucleic acids interact with or bind to the nucleosome. In the ensuing years, this theory proved to be correct, as cross-talk was identified and characterized between multiple PTMs, and enzymes that recognize combinatorial patterns were discovered. Middle-down analyzes modifications in relatively close proximity, increasing the ability to determine which modifications co-occur and which are mutually exclusive (Schwämmle et al., 2016; Schwämmle, Aspalter, Sidoli, & Jensen, 2014; Sidoli et al., 2014), however, the increased size of the peptides produces additional challenges. Electrospray ionization produces more charge states in larger analytes, decreasing signal from each charge state, and thus decreasing the amount of analyte that can be selected for fragmentation. The MS2 must be high resolution due to challenges of analyzing poorly separated peptides with isobaric forms, and electron transfer dissociate (ETD) is a more appropriate fragmentation method compared to collision induced dissociation (CID). Large peptides are usually separated by LC with lower efficiency than bottom-up sized peptides, so the ability to determine the location of each modification highly relies on the resolution of the MS2. Finally, computational analysis dramatically increases in complexity due to the presence of co-fragmented isobaric species in the MS2. Software like ProSightPC (Thermo), PILOT_PTMs (DiMaggio, Young, Baliban, Garcia, & Floudas, 2009) and isoScale slim (Schwämmle et al., 2014) have been introduced to overcome this issue. The advancement of data analysis tools is critical for enabling middle-down MS to become widespread in use for both histones and general proteomics.

3.3 Top-down mass spectrometry

Top-down MS is used to detect intact histone proteins in order to analyze the maximum number of PTMs on a single protein. The technique determines the entire mass of the protein, thus accounting for all PTMs, and fragmentation of the protein can determine the location of some of the modifications. The Kelleher lab pioneered the use of top-down MS analysis, which has been used to characterize most histone variants without any proteolytic digestion (Boyne, Pesavento, Mizzen, & Kelleher, 2006; Pesavento, Kim, Taylor, & Kelleher, 2004; Siuti, Roth, Mizzen, Kelleher, & Pesavento, 2006; Thomas, Kelleher, & Mizzen, 2006). However, difficulties arise due to the complexity of analyzing high numbers of isoforms; multiple isoforms can co-elute and co-fragment, decreasing sensitivity of detection, and making identification of the MS2 spectra highly challenging. Because of this, histone analysis via top-down MS remains sparsely used for large scale analyses, but some improvements have increased the capabilities of top-down MS and are encouraging for the

future of the technique. Previously, direct infusion of proteins was most commonly used on histones due to the challenge of separating multiple isoforms, however, improvements to both in-line and off-line separation methods have allowed for better analyses to be performed. Tran et al. demonstrated use of isoelectric focusing for separation by the isoelectric point (pI) of the protein, which is altered by the modification states of lysine residues and the presence of phosphate modifications (Tran et al., 2011). Further fractionation by mass using gel electrophoresis followed by on-line reverse phase chromatography has allowed for the detection of over 400 unique isoforms of core histones. It is estimated that there are more than 1,000 isoforms of total core histone proteins in a single sample, thus, without better methods of offline fractionation, identification of all histone isoforms is not feasible. While top-down MS can retrieve data on many modifications on a single histone, sensitivity disadvantages often prevent a complete analysis of one a single isoform of an individual histone. Additionally, high-throughput methods for analysis of two of the same histones types in a single nucleosome do not yet exist, preventing detailed characterization of the symmetry of histone PTMs throughout the genome.

4. Future Perspective/Outlook

Accurate quantitation of histone PTMs and their combinatorial patterns are already playing a major role in determining molecular changes in specific diseases and in general model systems (reviewed in (Bonaldi & Nuberini, 2016; Önder, Sidoli, Carroll, & Garcia, 2015)). In part due to the rising importance of epigenetic therapeutics to treat cancers and other diseases, MS histone analysis has become a fundamental tool for cell biology. However, solely the abundance of a histone modification is often an insufficient parameter to determine its biological function. As mentioned, techniques like ChIP-seq have gained massive importance, as they allow the genome-wide mapping of histone PTMs. However, being based on antibodies, it is a low throughput strategy and often inapplicable for rare or combinatorial marks. Additionally, even though some histone PTMs have had their roles in chromatin decoded, such as histone H3K4me3 being highly enriched on active promoters, it is not sufficient to map the localization of a modification to determine whether the gene downstream is expressed. Transcriptomics techniques like RNA-seq, are required to characterize gene expression levels.

Genomics, transcriptomics and proteomics all have limitations that prevent a comprehensive understanding of our biological model systems, implying that each individual strategy does not suffice for comprehensive systems biology. Integration of those methods is an attractive frontier of the “post –omics” era. Overall, there is overwhelming consent that integrating –omics dataset is a proper path to follow. Numerous attempts have been already made, several with success. However, while some data seem simpler to integrate, such as transcriptomics and proteomics, others like ChIP-seq and proteomics are more challenging to integrate, or sometimes even defined as irrelevant (Gomez-Cabrero et al., 2014). In other cases, –omics dataset are not integrated in a comprehensive manner, where data are combined without a predefined hypothesis, but rather used in a “targeted mode,” where only specific and incomplete information is extracted. This last example is the case of a proteomics dataset used only to confirm the abundance of proteins of interest, or a ChIP-seq dataset used only

to confirm the presence of a transcription factor on specific genes. Obviously, all approaches are valid if they help understanding a biological mechanism, however, a narrow view of a fraction of such rich data can lead to biased conclusions and overlook critical information that most accurately defines the model.

Histone analysis has a great potential to become the missing bridge between genomics and proteomics, as changes in chromatin conformation can be linked to changes in gene expression and thus, phenotype. For instance, an external stimulus like an inhibitor of a receptor triggers a cascade of predominantly phosphorylation events which can easily be monitored by proteomics. Such phosphorylations ultimately affect histone writers, readers, and erasers by changing their activities; e.g. methyltransferase EZH2 cannot catalyze demethylation or trimethylation on histone H3 residue K27 if the enzyme phosphorylated at threonine 487 (Wei et al., 2011). A recent collaboration between the Garcia and Zhao labs generated a comprehensive table of all known histone PTMs and their writers as shown in figure 2 (Zhao & Garcia, 2015). By knowing the changes in global abundance (by MS) and localization (ChIP-seq) of histone PTMs, it would be possible to predict in an unbiased manner which genes are likely to have their expressions modulated, especially if the observed phosphorylation events are on enzymes catalyzing the histone PTMs that change their abundance. Ultimately, RNA-seq and proteomics analyses would confirm those changes in gene expression and protein abundance. By using knowledge of protein-protein interactions from public databases like STRING, BioGRID or IntAct, it could become possible to reconstruct entire pathways from external stimulus to protein abundance changes, and histone PTM quantitation would be a fundamental link in the path. Such analyses could pave the way to a fully unbiased approach to systems biology.

Acknowledgments

BAG gratefully acknowledges support from NIH grants GM110174 and AI118891; and a Leukemia and Lymphoma Society Dr. Robert Arceci Scholar award

References

- Alabert C, Barth TK, Reverón-Gómez N, Sidoli S, Schmidt A, Jensen ON, Groth A. Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes & Development*. 2015; 29(6): 585–590. <http://doi.org/10.1101/gad.256354.114>. [PubMed: 25792596]
- Bonaldi T, Nuberini R. Recent advances in mass spectrometry analysis of histone post-translational modifications: potential clinical impact of the PAT-H-MS approach. *Expert Review of Proteomics*. 2016; 13(3):245–250. <http://doi.org/10.1586/14789450.2016.1147960>. [PubMed: 26822344]
- Boyne MT, Pesavento JJ, Mizzen CA, Kelleher NL. Precise Characterization of Human Histones in the H2A Gene Family by Top Down Mass Spectrometry. *Journal of Proteome Research*. 2006; 5(2): 248–253. <http://doi.org/10.1021/pr050269n>. [PubMed: 16457589]
- Britton LMP, Newhart A, Bhanu NV, Sridharan R, Gonzales-Cope M, Plath K, Garcia BA. Initial characterization of histone H3 serine 10 O-acetylation. *Epigenetics*. 2013; 8(10):1101–1113. <http://doi.org/10.4161/epi.26025>. [PubMed: 23949383]
- Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: A Method for Identification of Proteins and Histone Posttranslational Modifications at a Single Genomic Locus. *Cell Reports*. 2012; 2(1):198–205. <http://doi.org/10.1016/j.celrep.2012.06.019>. [PubMed: 22840409]
- Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, Kouzarides T. Histone Deimination Antagonizes Arginine Methylation. *Cell*. 2004; 118(5):545–553. <http://dx.doi.org/10.1016/j.cell.2004.08.020>. [PubMed: 15339660]

- Dehennaut V, Leprince D, Lefebvre T. O-GlcNAcylation, an Epigenetic Mark. Focus on the Histone Code, TET Family Proteins, and Polycomb Group Proteins. *Frontiers in Endocrinology*. 2014; 5:155. <http://doi.org/10.3389/fendo.2014.00155>. [PubMed: 25309514]
- DiMaggio PA, Young NL, Baliban RC, Garcia BA, Floudas CA. 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. *Molecular & Cellular Proteomics : MCP*. 2009; 8(11):2527–2543. <http://doi.org/10.1074/mcp.M900144-MCP200>. [PubMed: 19666874]
- Egelhofer TA, Minoda A, Klugman S, Lee K, Alekseyenko AA, Cheung M, Lieb JD. An assessment of histone-modification antibody quality. *Nature Structural & Molecular Biology*. 2011; 18(1):72–90. <http://doi.org/10.1038/nsmb.1972.An>.
- Garcia BA, Mollah S, Ueberheide BM, Busby SA, Muratore TL, Shabanowitz J, Hunt DF. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protocols*. 2007; 2(4): 933–938. Retrieved from <http://dx.doi.org/10.1038/nprot.2007.106>. [PubMed: 17446892]
- Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merckenschlager M, Gisel A, Tegnér J. Data integration in the era of omics: current and future challenges. *BMC Systems Biology*. 2014; 8(Suppl 2):I1–I1. <http://doi.org/10.1186/1752-0509-8-S2-I1>.
- Hebbes TR, Thorne AW, Crane-Robinson C. A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO Journal*. 1988; 7(5):1395–1402. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC458389/>. [PubMed: 3409869]
- Humphrey SJ, Azimifar SB, Mann M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat Biotech*. 2015; 33(9):990–995. Retrieved from <http://dx.doi.org/10.1038/nbt.3327>.
- Jenuwein T, Allis CD. Translating the Histone Code. *Science*. 2001; 293(5532):1074–1080. Retrieved from <http://science.sciencemag.org/content/293/5532/1074.abstract>. [PubMed: 11498575]
- Krautkramer KA, Reiter L, Denu JM, Dowell JA. Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. *Journal of Proteome Research*. 2015; 14(8):3252–3262. <http://doi.org/10.1021/acs.jproteome.5b00245>. [PubMed: 26120868]
- LeRoy G, Chepelev I, DiMaggio PA, Blanco MA, Zee BM, Zhao K. Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. *Genome Biol*. 2012; 13 <http://doi.org/10.1186/gb-2012-13-8-r68>.
- Li W, Nagaraja S, Delcuve GP, Hendzel MJ, Davie JR. Effects of histone acetylation, ubiquitination and variants on nucleosome stability. *Biochemical Journal*. 1993; 296(Pt 3):737–744. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1137757/>. [PubMed: 8280071]
- Liao R, Wu H, Deng H, Yu Y, Hu M, Zhai H, Yi W. Specific and Efficient N-Propionylation of Histones with Propionic Acid N-Hydroxysuccinimide Ester for Histone Marks Characterization by LC-MS. *Analytical Chemistry*. 2013; 85(4):2253–2259. <http://doi.org/10.1021/ac303171h>. [PubMed: 23339652]
- Lin, S., Garcia, BA. Chapter One – Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. In: C. W. and C. D. A. B. T.-M. in *Enzymology*, editor. *Nucleosomes, Histones & Chromatin Part A*. Vol. 512. Academic Press; 2012. p. 3-28. <http://doi.org/http://dx.doi.org/10.1016/B978-0-12-391940-3.00001-9>
- Maile TM, Izrael-Tomasevic A, Cheung T, Guler GD, Tindell C, Masselot A, Arnott D. Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method. *Molecular & Cellular Proteomics*. 2015; 14(4):1148–1158. <http://doi.org/10.1074/mcp.O114.046573>. [PubMed: 25680960]
- Meert P, Dierickx S, Govaert E, De Clerck L, Willems S, Dhaenens M, Deforce D. Tackling aspecific side reactions during histone propionylation: The promise of reversing overpropionylation. *Proteomics*. 2016; 16(14):1970–1974. <http://doi.org/10.1002/pmic.201600045>. [PubMed: 27139031]
- Meert P, Govaert E, Scheerlinck E, Dhaenens M, Deforce D. Pitfalls in histone propionylation during bottom-up mass spectrometry analysis. *Proteomics*. 2015; 15(17):2966–2971. <http://doi.org/10.1002/pmic.201400569>. [PubMed: 26010583]

- Mohammed H, Taylor C, Brown GD, Papachristou EK, Carroll JS, D'Santos CS. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protocols*. 2016; 11(2):316–326. <http://doi.org/10.1038/nprot.2016.020>. [PubMed: 26797456]
- Molden RC, Goya J, Khan Z, Garcia Ba. Stable Isotope Labeling of Phosphoproteins for Large-scale Phosphorylation Rate Determination. *Molecular & Cellular Proteomics*. 2014; 13:1106–18. <http://doi.org/10.1074/mcp.O113.036145>. [PubMed: 24532841]
- Oliver SS, Denu JM. Dynamic Interplay between Histone H3 Modifications and Protein Interpreters: Emerging Evidence for a “Histone Language. *ChemBioChem*. 2011; 12(2):299–307. <http://doi.org/10.1002/cbic.201000474>. [PubMed: 21243717]
- Önder Ö, Sidoli S, Carroll M, Garcia BA. Progress in epigenetic histone modification analysis by mass spectrometry for clinical investigations. *Expert Review of Proteomics*. 2015; 12(5):499–517. <http://doi.org/10.1586/14789450.2015.1084231>. [PubMed: 26400466]
- Paternoster V, Edhager AV, Sibbersen C, Nielsen AL, Børglum AD, Christensen JH, Palmfeldt J. Quantitative assessment of methyl-esterification and other side reactions in a standard propionylation protocol for detection of histone modifications. *Proteomics*. 2016; 16(14):2059–2063. <http://doi.org/10.1002/pmic.201500425>. [PubMed: 27080621]
- Pesavento JJ, Kim YB, Taylor GK, Kelleher NL. Shotgun Annotation of Histone Modifications: A New Approach for Streamlined Characterization of Proteins by Top Down Mass Spectrometry. *Journal of the American Chemical Society*. 2004; 126(11):3386–3387. <http://doi.org/10.1021/ja039748i>. [PubMed: 15025441]
- Price BD, D'Andrea AD. Chromatin Remodeling at DNA Double Strand Breaks. *Cell*. 2013; 152(6):1344–1354. <http://doi.org/10.1016/j.cell.2013.02.011>. [PubMed: 23498941]
- Rousseaux S, Khochbin S. Histone Acylation beyond Acetylation: Terra Incognita in Chromatin Biology. *Cell Journal (Yakhteh)*. 2015; 17(1):1–6. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393657/>.
- Sadakierska-Chudy A, Filip M. A Comprehensive View of the Epigenetic Landscape. Part II: Histone Post-translational Modification, Nucleosome Level, and Chromatin Regulation by ncRNAs. *Neurotoxicity Research*. 2015; 27:172–197. <http://doi.org/10.1007/s12640-014-9508-6>. [PubMed: 25516120]
- Schwämmle V, Aspalter CM, Sidoli S, Jensen ON. Large Scale Analysis of Co-existing Post-translational Modifications in Histone Tails Reveals Global Fine Structure of Cross-talk. *Molecular & Cellular Proteomics*. 2014; 13(7):1855–1865. <http://doi.org/10.1074/mcp.O113.036335>. [PubMed: 24741113]
- Schwämmle V, Sidoli S, Ruminowicz C, Wu X, Lee CF, Helin K, Jensen ON. Systems Level Analysis of Histone H3 Post-translational Modifications (PTMs) Reveals Features of PTM Crosstalk in Chromatin Regulation. *Molecular & Cellular Proteomics*. 2016; 15(8):2715–2729. <http://doi.org/10.1074/mcp.M115.054460>. [PubMed: 27302890]
- Shiio Y, Eisenman RN. Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(23):13225–13230. <http://doi.org/10.1073/pnas.1735528100>. [PubMed: 14578449]
- Sidoli S, Bhanu, NV., Karch, KR., Wang, X., Garcia, BA. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis; *Journal of Visualized Experiments: JoVE*. 2016. p. 54112 <http://doi.org/10.3791/54112>
- Sidoli S, Cheng L, Jensen ON. Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *Journal of Proteomics*. 2012; 75(12):3419–3433. <http://doi.org/http://dx.doi.org/10.1016/j.jprot.2011.12.029>. [PubMed: 22234360]
- Sidoli S, Fujiwara, R., Garcia, BA. Multiplexed data independent acquisition (MSX-DIA) applied by high resolution mass spectrometry improves quantification quality for the analysis of histone peptides. *Proteomics*. 2016. Retrieved from <http://dx.doi.org/10.1002/pmic.201500527>
- Sidoli S, Garcia BA. Properly reading the histone code by MS-based proteomics. *Proteomics*. 2015; 15(17):2901–2902. <http://doi.org/10.1002/pmic.201500298>. [PubMed: 26223514]

- Sidoli S, Lin S, Xiong L, Bhanu NV, Karch KR, Johansen E, Garcia BA. Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications. *Molecular & Cellular Proteomics*. 2015; 14(9):2420–2428. <http://doi.org/10.1074/mcp.O114.046102>. [PubMed: 25636311]
- Sidoli S, Schwämmle V, Ruminowicz C, Hansen TA, Wu X, Helin K, Jensen ON. Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *Proteomics*. 2014; 14(19):2200–2211. <http://doi.org/10.1002/pmic.201400084>. [PubMed: 25073878]
- Sidoli S, Simithy J, Karch KR, Kulej K, Garcia BA. Low Resolution Data-Independent Acquisition in an LTQ-Orbitrap Allows for Simplified and Fully Untargeted Analysis of Histone Modifications. *Analytical Chemistry*. 2015; 87(22):11448–11454. <http://doi.org/10.1021/acs.analchem.5b03009>. [PubMed: 26505526]
- Sidoli S, Yuan ZF, Lin S, Karch K, Wang X, Bhanu N, Garcia BA. Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis. *Proteomics*. 2015; 15(9):1459–1469. <http://doi.org/10.1002/pmic.201400483>. [PubMed: 25641854]
- Siuti N, Roth MJ, Mizzen CA, Kelleher NL, Pesavento JJ. Gene-Specific Characterization of Human Histone H2B by Electron Capture Dissociation. *Journal of Proteome Research*. 2006; 5(2):233–239. <http://doi.org/10.1021/pr050268v>. [PubMed: 16457587]
- Soldi M, Cuomo A, Bonaldi T. Quantitative assessment of chemical artefacts produced by propionylation of histones prior to mass spectrometry analysis. *Proteomics*. 2016; 16(14):1952–1954. <http://doi.org/10.1002/pmic.201600173>. [PubMed: 27373704]
- Thingholm, TE., Larsen, MR. The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides. In: von Stechow, L., editor. *Phospho-Proteomics: Methods and Protocols*. New York, NY: Springer New York; 2016. p. 135-146. http://doi.org/10.1007/978-1-4939-3049-4_9
- Thomas CE, Kelleher NL, Mizzen CA. Mass Spectrometric Characterization of Human Histone H3: A Bird's Eye View. *Journal of Proteome Research*. 2006; 5(2):240–247. <http://doi.org/10.1021/pr050266a>. [PubMed: 16457588]
- Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Letters to Nature*. 2011; 480:254–261. <http://doi.org/10.1038/nature10575>.
- Voigt P, LeRoy G, Drury WJ, Zee BM, Son J, Beck D, Reinberg D. Asymmetrically Modified Nucleosomes. *Cell*. 2012; 151(1):181–193. <http://doi.org/10.1016/j.cell.2012.09.002>. [PubMed: 23021224]
- Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, Tackett AJ. A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics*. 2014; 9(9):1207–1211. <http://doi.org/10.4161/epi.29919>. [PubMed: 25147920]
- Wang X, Yuan ZF, Fan J, Karch KR, Ball LE, Denu JM, Garcia BA. A Novel Quantitative Mass Spectrometry Platform for Determining Protein O-GlcNAcylation Dynamics. *Molecular & Cellular Proteomics*. 2016; 15(7):2462–2475. <http://doi.org/10.1074/mcp.O115.049627>. [PubMed: 27114449]
- Wei Y, Chen YH, Li LY, Lang J, Yeh SP, Shi B, Hung M-C. CDK1-dependent phosphorylation of EZH2 suppresses methylation of H3K27 and promotes osteogenic differentiation of human mesenchymal stem cells. *Nature Cell Biology*. 2011; 13(1):87–94. <http://doi.org/10.1038/ncb2139>. [PubMed: 21131960]
- Wierer M, Mann M. Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Human Molecular Genetics*. 2016; 00(00):1–9. <http://doi.org/10.1093/hmg/ddw208>.
- Won KJ, Choi I, LeRoy G, Zee BM, Sidoli S, Gonzales-Cope M, Garcia BA. Proteogenomics analysis reveals specific genomic orientations of distal regulatory regions composed by non-canonical histone variants. *Epigenetics & Chromatin*. 2015; 8(1):1–11. <http://doi.org/10.1186/s13072-015-0005-9>. [PubMed: 25621012]
- Yuan ZF, Lin S, Molden RC, Cao XJ, Bhanu NV, Wang X, Garcia BA. EpiProfile Quantifies Histone Peptides With Modifications by Extracting Retention Time and Intensity in High-resolution Mass

Spectra. Molecular & Cellular Proteomics : MCP. 2015; 14(6):1696–1707. <http://doi.org/10.1074/mcp.M114.046011>. [PubMed: 25805797]

Zhao Y, Garcia BA. Comprehensive Catalog of Currently Documented Histone Modifications. Cold Spring Harbor Perspectives in Biology. 2015; 7:a025064. [PubMed: 26330523]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Data independent acquisition (DIA)

(A) DIA is an acquisition method where the MS is programmed to select parent ions for fragmentation in the same mass windows at every cycle. By doing this, it generates several layers of chromatograms (B) that can be used to perform extracted ion chromatography at the MS1 and MS2 level (C).