

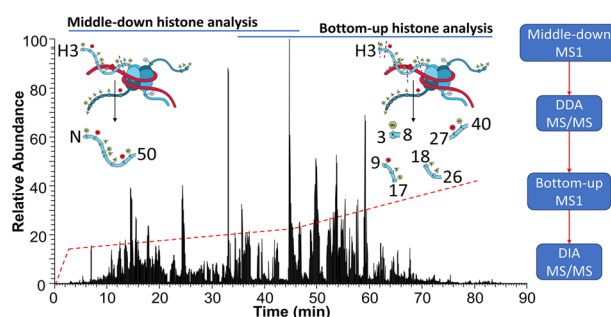
Quantitation of Single and Combinatorial Histone Modifications by Integrated Chromatography of Bottom-up Peptides and Middle-down Polypeptide Tails

Kevin A. Janssen,^{1,2} Mariel Coradin,^{1,2} Congcong Lu,² Simone Sidoli,^{2,3}
Benjamin A. Garcia^{1,2}

¹Biochemistry and Molecular Biophysics Graduate Group, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

²Epigenetics Institute, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

³Present Address: Department of Biochemistry, Albert Einstein College of Medicine, New York, NY, USA



Abstract. The analysis of histone post-translational modifications (PTMs) by mass spectrometry (MS) has been critical to the advancement of the field of epigenetics. The most sensitive and accurate workflow is similar to the canonical proteomics analysis workflow (bottom-up MS), where histones are digested into short peptides (4–20 aa) and quantitated in extracted ion chromatograms. However, this limits the ability to detect even very common co-occurrences of

modifications on histone proteins, preventing biological interpretation of PTM crosstalk. By digesting with GluC rather than trypsin, it is possible to produce long polypeptides corresponding to intact histone N-terminal tails (50–60 aa), where most modifications reside. This middle-down MS approach is used to study distant PTM co-existence. However, the most sensitive middle-down workflow uses weak cation exchange-hydrophilic interaction chromatography (WCX-HILIC), which is less robust than conventional reversed-phase chromatography. Additionally, since the buffer systems for middle-down and bottom-up proteomics differ substantially, it is cumbersome to toggle back and forth between both experimental setups on the same LC system. Here, we present a new workflow using porous graphitic carbon (PGC) as a stationary phase for histone analysis where bottom-up and middle-down sized histone peptides can be analyzed simultaneously using the same reversed-phase buffer setup. By using this protocol for middle-down sized peptides, we identified 406 uniquely modified intact histone tails and achieved a correlation of 0.85 between PGC and WCX-HILIC LC methods. Together, our method facilitates the analysis of single and combinatorial histone PTMs with much simpler applicability for conventional proteomics labs than the state-of-the-art middle-down MS.

Keywords: Histones, Epigenetics, Proteomics, Middle-down, PTMs, Chromatography

Received: 22 April 2019/Revised: 3 July 2019/Accepted: 8 July 2019/Published Online: 11 September 2019

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13361-019-02303-6>) contains supplementary material, which is available to authorized users.

Correspondence to: Benjamin Garcia;
e-mail: bgarci@penmedicine.upenn.edu

Introduction

The “histone code” is the hypothesis that post-translational modifications (PTMs) on histone proteins have specific combinatorial patterns associated with genetic readouts [1, 2]. The diverse array of modifications, most commonly acetylation and methylation, have indeed been demonstrated to play

important roles in the regulation of the genome [3, 4]. Histone PTMs are catalytically deposited and removed by enzymes termed *writers* and *erasers*. Critical to the function of these modifications are *reader* proteins, which are responsible for recognizing modifications and causing functional output. Because of this, the modification profile of histones at specific loci and even at global levels can provide important insight into cellular states and functions. Indeed, this has been demonstrated by discoveries showing that histone H3 phosphorylations are important for mitotic progression [5], histone H3 arginine methylation regulates pluripotency [6], and histone H4 acetylations regulate X-inactivation [7]. Mass spectrometry (MS) has played a critical role in the identification and quantitation of novel PTMs; to date, hundreds of different histone marks have been discovered [8]. In other words, almost every PTM discovered on other proteins has been identified on histones as well. However, most of the known modifications to date have still ambiguous biological function, mostly due to the fact that histone PTMs co-exist in combinatorial patterns affecting each other's roles [9–12]. This phenomenon is called “PTM crosstalk,” and we are still just scratching the surface of this concept due to the dearth of technology capable of accurately characterizing combinatorial marks. This highlights the demand for methods that can identify and quantitate single and combinatorial modifications.

Analysis of histone proteins by MS is most commonly performed by using a bottom-up workflow, where histones are digested into short peptides prior to analysis by liquid chromatography coupled with tandem MS (LC-MS/MS) [13–18]. This analysis provides the most reliable and accurate quantitation of PTMs, as short peptides are more easily resolved chromatographically and more accurately quantitated by MS as compared with intact proteins or long polypeptides. In the most commonly used protocols, histones are derivatized on unmodified and monomethylated lysine residues prior to trypsin digestion to prevent excessive cleavage [19], as histone sequences are highly enriched in lysine and arginine residues. This approach is highly efficient and highly utilized, especially to quantitate PTMs on histones H3 and H4 as it yields peptides that contain no more than 2 commonly modified lysine residues, with the sole exception of the histone H4 peptide 4-17, which contains 4 lysine residues commonly modified by acetylation (K5, K8, K12, and K16). Because tryptic peptides can have more than one potential modification site, this type of sample includes isobarically modified peptides, of which quantitation is discriminated at the fragment ion level. In order to perform accurate chromatographic quantitation of MS/MS spectra as well, samples are commonly acquired in data-independent acquisition (DIA) mode [20–23]. This workflow currently provides the highest sensitivity and reproducibility for the relative quantitation of histone marks at a global level. This not only has enabled the discoveries outlined above but has also led to work that identified correlates of multiple modifications on the same histone, such as the dynamics of H3K27 and H3K36 in H3G34L/W-bearing tumors [24] and the decrease in H3K27me3K36me2 in mouse embryonic stem cell

(mESC) differentiation [25]. This showcases the importance of bottom-up combinatorial modification analysis; however, analysis of tryptic peptides fails to deconvolute co-occurrence of distant modifications, meaning that it is intrinsically incapable of determining the quantity and identities of uniquely modified intact histone tails. This was exemplified by the ChIP-seq observation of overlapping populations of H3K4me3 and H3K27me3, which suggested that the two modifications could be on the same histone or on different histones in the same nucleosome [26]. However, bottom-up histone analysis was not sufficient to determine if the marks co-occur on the same histone, but middle-down, as it detects uniquely modified intact histone tails, was able to determine that the PTMs are found asymmetrically on nucleosomes [27].

To overcome the issue of analyzing long-distance combinatorial modifications, middle-down MS [28–30] and top-down MS [31–33] have been implemented. In particular, middle-down MS has been shown to provide quantitative accuracy of binary modifications similar to conventional bottom-up MS when both are compared with ionization efficiency-corrected bottom-up MS [34, 35]. Middle-down MS has been most commonly adopted using weak cation exchange-hydrophilic interaction chromatography (WCX-HILIC) [29, 30] combined with custom data processing and analysis as compared with bottom-up MS [36, 37]. This middle-down MS strategy yields hundreds of combinatorial PTM identifications on intact histone tails [29]. The use of this strategy requires a mass spectrometer capable of high mass resolution and ETD fragmentation due to the high mass and the high charge state of histone tails. In addition, due to the relatively young age of this strategy, there has been limited software development for analysis of middle-down datasets [38–40]. However, despite the achievements in accuracy and sensitivity, the method has not been widely applied for biological insights (discussed in [41]), partly because of the difficulty in setting up the chromatographic system. To overcome the challenges of WCX-HILIC technology, we have developed a porous graphitic carbon (PGC)-based method that is more convenient to use for proteomics labs because it uses the same buffer compositions as conventional C18 chromatography. PGC consists of monolayered sheets of carbon, similar to graphite but without regularly ordered successive layers. The carbon atoms are sp² hybridized and interact well with hydrophobic and aromatic molecules. Additionally, PGC has a polar retention effect that allows for binding and separation of both polar and non-polar molecules. Electrons in graphitic sheets can delocalize, inducing polarization of the sheet, although the exact mechanistic cause of this delocalization is not fully understood [42]. Due to the variety of interactions the stationary phase is capable of, PGC represents an attractive technology for separating highly complex solutions of peptides. We demonstrate that PGC is reliable and robust and provides accurate identification of combinatorially modified histone peptides despite lacking the chromatographic resolution and more extensive identification quantity of WCX-HILIC middle-down methods. Our work herein represents a step forward in the effort to improve the

convenience and reliability of middle-down analysis for mass spectrometrists and epigeneticists.

Methods

Histone Extraction

Histones were extracted from HeLa S3 cells as previously described [15, 43]. In brief, HeLa S3 cell pellets were resuspended in nuclear isolation buffer (NIB) with 0.2% NP-40 alternative, incubated for 5 min, and centrifuged at $1000\times g$ to collect the nuclear pellet. Two additional washes of NIB without NP-40 alternative were performed to reduce detergent concentration. Nuclei were resuspended in 0.2 M H_2SO_4 , shaken for 2 h, and histones were precipitated by adding trichloroacetic acid (TCA) to a final concentration of 25% w/v. Precipitates were washed with acetone and dried overnight. The Bradford assay was used to calculate protein concentrations. Histone extraction was performed in triplicate at the cell culture level, and aliquots from individual replicates were subjected to the subsequent sample processing and analysis.

Bottom-up Histone Preparation

Isolated histones were derivatized and digested as previously described [15, 43]. In brief, 40 μg of histones were resuspended in 10 μL of 50 mM NH_4HCO_3 (pH 8). To 15 μL of acetonitrile (ACN), 5 μL of propionic anhydride were added. The resulting solution was rapidly mixed, and 10 μL of it were added to the histone solution. To bring the solution back to pH 8, 5 μL of NH_4OH were immediately added. The solution was incubated at 37 °C for 15 min, causing unmodified and monomethylated lysines to be propionylated. This propionylation was performed twice, and then, histones were digested using trypsin at an enzyme-to-sample ratio of 1:20 at 37 °C overnight. The propionylation was again performed, twice, to ensure that the new N-termini of trypsin-generated peptides were propionylated. Histone peptides were desalted on in-house stage tips which were generated by wedging a 0.5-cm circular punch of a 3M Empore C18 paper disk into the bottom of a P200 pipette tip. The stage tips were conditioned with ACN, equilibrated with 0.1% trifluoroacetic acid (TFA), loaded with samples in 0.1% TFA, washed with 0.1% TFA, and eluted with 0.1% TFA in 70% ACN.

Middle-down Histone Preparation

One hundred microgram aliquots of isolated histones were fractionated on a 260×4.6 mm 5- μm Vydac C18 column using 0.2% TFA in 5% ACN as buffer A and 0.2% TFA in 95% ACN as buffer B. A flow rate of 0.8 mL/min and gradient of 47% B to 60% B over 50 min were used with UV detection at 220 nm, and 400 μL fractions were collected. Histone H4 separated well from H3 isoforms, but partially co-eluted with H2A. Histones H3.3 and H3.2 largely co-elute and were pooled into a single fraction (Figure S1). Histone H3.1 separated well from histones

H3.3 and H3.2 but was added to the histone H3 pool regardless. Histones were dried in a Savant SpeedVac SC100 and resuspended in 5 mM ammonium acetate at pH 4. All fractions were digested by adding 2 μg of GluC. Resulting peptides were desalted on in-house stage tips as described above.

Nano-liquid Chromatography

For all stationary phases except WCX-HILIC, 0.1% formic acid (FA) was used as buffer A, 0.1% FA in 80% ACN was used as buffer B, and the flowrate was set to 500 nL/min. Four stationary phases were used: C18 (Reprosil-Pur C18-AQ 3 μm ; Dr. Maisch), C30 (Develosil C30-UG 5 μm ; Phenomenex), PGC (Hypercarb 3 μm ; Thermo), and WCX-HILIC (PolyCAT A 3 μm ; PolyLC), each packed in 75- μm internal diameter in-house packed columns. Gradients were optimized on a Dionex UltiMate 3000 for each stationary phase. For bottom-up analyses, the analytical gradients were run over 45 min: C18, 3 to 36% B; C30, 5 to 40% B; and PGC, 15 to 50% B. For middle-down analyses, the analytical gradients were run over 75 min: C18, 5 to 40% B; C30, 7 to 43% B; and PGC, 14 to 23% B. For combined analyses, gradients were segmented to optimize both bottom-up and middle-down separations over 90 min. For C18, the gradient was 5 to 10% B over 28 min, 10 to 12% B over 1 min, 12 to 15% B over 14 min, and finally 15 to 40% B over 35 min. For PGC, the gradient was 14 to 23% B over 33 min and 23 to 50% B over 35 min. For WCX-HILIC, 20 mM propionic acid in 75% ACN, adjusted to pH 6.0 with ammonium hydroxide, was used as buffer A, 0.2% FA in 15% ACN (pH 2.5) was used as buffer B, and the flowrate was set to 300 nL/min. The gradient was 0% B over 5 min and then 70% B to 95% B over 90 min.

Mass Spectrometry

All analyses were performed on a Thermo Scientific Orbitrap Fusion. Three MS methods were used, one for bottom-up analyses, one for middle-down analyses, and a combination of both for middle-down and bottom-up mixtures. The bottom-up analyses were performed using 25% HCD collision energy DIA as previously described [15, 43]. First, a full scan from 300 to 1100 m/z at a resolution of 60,000 was obtained in the Orbitrap. Then, half of the full DIA range was analyzed in 50 m/z windows from 300 to 700 m/z with detection in the ion trap. A second full scan was performed, and then the DIA cycle was completed with the same scan parameters from 700 to 1100 m/z . The middle-down analyses were performed as previously described [44]. In brief, a full scan from 665 to 730 m/z , which encompasses all charge state 8 histone H3 and H4 peptides, was acquired at a resolution of 120,000 to ensure accurate charge state identification. Data-dependent MS2 was acquired by selecting peptides of charge state 8, subjecting the peptides to 20 ms of ETD at a reagent target of $1e5$, and detecting fragments at a resolution of 30,000 in the Orbitrap. The combined middle-down and bottom-up runs used similar scan parameters to the two independent runs. For this approach, three scans were used: the standard bottom-up full scan, a

single cycle of the entire bottom-up DIA range, 50 m/z windows from 300 to 1100 m/z , and the standard middle-down full scan with data-dependent MS2.

Data Analysis

Bottom-up data analysis was performed by processing raw files in EpiProfile 2.0 with retention time references disabled [45]. The retention time references native to EpiProfile 2.0 were generated from C18 runs, which would bias a comparison of methods to favor C18 chromatography, as it is the gold standard for bottom-up histone PTM analysis. Additionally, due to chromatographic differences, the combined middle-down and bottom-up runs would mismatch the expected retention times, even for the C18 analyses. Because of these potential biases, all runs, including those performed with C18, were processed with retention time references disabled. Middle-down data analysis was performed by processing raw files in Proteome Discoverer 2.2 using the Mascot search engine with a precursor mass tolerance of 2.1 Da and a fragment mass tolerance of 0.01 Da. Variable modifications were acetylation on N-termini and lysines, mono- and di-methylation on lysine and arginine, and trimethylation on lysine. Mascot output files were filtered through in-house software, HistoneCoderTool ProteoformQuant, which removes peptides without sufficient fragment ions to unambiguously localize PTMs [44, 46]. Quantitation was performed by summation of fragment ions and normalization to the total intensity of all peptides of identical sequence. Co-eluting peptides were assigned abundances by distributing the sum of common fragment ions based on the ratio of unique fragments. As dynamic exclusion was limited to 2 s, the same peptides can be quantitated multiple times, resulting in a histogram of intensity values that corresponds to the total abundance of each uniquely modified intact histone tail.

Results

Bottom-up Histone Analysis

The standard histone preparation, including lysine propionylation, was performed for all samples to directly compare novel chromatography to the established C18-based chromatography. With these studies, we optimized gradients for C30 and PGC stationary phases to match the C18 method runtime. As expected, PGC retained short hydrophilic peptides more strongly than C18 when using the chromatographic method optimized for C18 (Figure S2). Relative retention of amphipathic peptides was less predictable, and some peptides eluted slightly earlier from PGC than from C18. Because of this effect, the peptide elution order from PGC was inconsistent with the elution order from C18 (Figure S3). We analyzed data from each method using EpiProfile 2.0 with expected retention times ignored as they are based on prior C18-based chromatography. To ensure the comparison was balanced, the C18 runs were also analyzed without retention time references. The results showed similar quantitation of nearly all PTMs across stationary phases (Figure 1). Due to the fact that the majority of

modifications have relative abundances below 0.1, the heatmap scaling causes some relative abundances to appear to be more divergent than they are. The stacked bar plot extending from the heatmap shows this effect clearly. For example, K14ac appears to have dramatically different quantitation in C18 and PGC runs in the heatmap although the relative abundances differ by less than 0.05 and the stacked bars are more similar. The stacked bars also appear as a curve reflecting the consistency between all three stationary phases. Indeed, C18 and C30 runs appear almost identical while the relative abundances in PGC are more frequently different. This is likely due to the fundamental differences and similarities of the stationary phases; C18 and C30 are both alkyl chains bound to silica resin while PGC is monolithic and composed of layered graphene sheets. Regardless, these differences yield only minor discrepancies in quantitation and these results suggest that all three stationary phases can be used for accurate bottom-up histone analysis with conventional proteomics buffers.

Middle-down Histone Analysis

We evaluated middle-down histone analysis using WCX-HILIC and PGC stationary phases. To assess the quantitative accuracy of formic acid buffer-based middle-down analyses, we examined the single modification abundances across multiple runs. These results showed similar single PTM profiles between the two stationary phases (Figure 2a, b), with WCX-HILIC showing higher PTM abundances overall, especially for acetylations. Indeed, WCX-HILIC excels at separating differentially acetylated peptides due to the neutralization of positively charged lysine residues that interact with the WCX resin; however, unmodified peptides are poorly resolved [47]. Due to different buffer compositions during ionization, some abundance differences were expected, yet abundances of peptides detected in using both methods did not show dramatic differences (Figure 2c). Notably, middle-down peptide standards have not been generated, and quantitation accuracy cannot be compared between the two chromatographic approaches as each has biases of their own. Despite PGC yielding fewer than one-third of the peptide identifications of WCX-HILIC (Figure 2d), its 406 histone H3 peptide identifications are among the best alternatives to WCX-HILIC [48].

Next, we focused on the quality of the quantitation with middle-down MS. As this approach is used to examine combinatorial modifications rather than to optimally quantitate individual histone PTMs, reliability in histone PTM crosstalk is critical for method viability. To determine how well PGC quantitated combinatorial peptides, we examined binary modifications on histone H3 on two of the three lysine residues 9, 27, and 36 and compared the performance of PGC with that of WCX-HILIC (Figure 3a, b). This evaluation yields a circle plot where the thickness of a line indicates the relative abundance of peptides that contain only the two modifications. The two circle plots are nearly identical, though some minor differences can be observed. For example, PGC finds that K9unmodK27me1 and K9me2K36me3 are more common than WCX-HILIC.

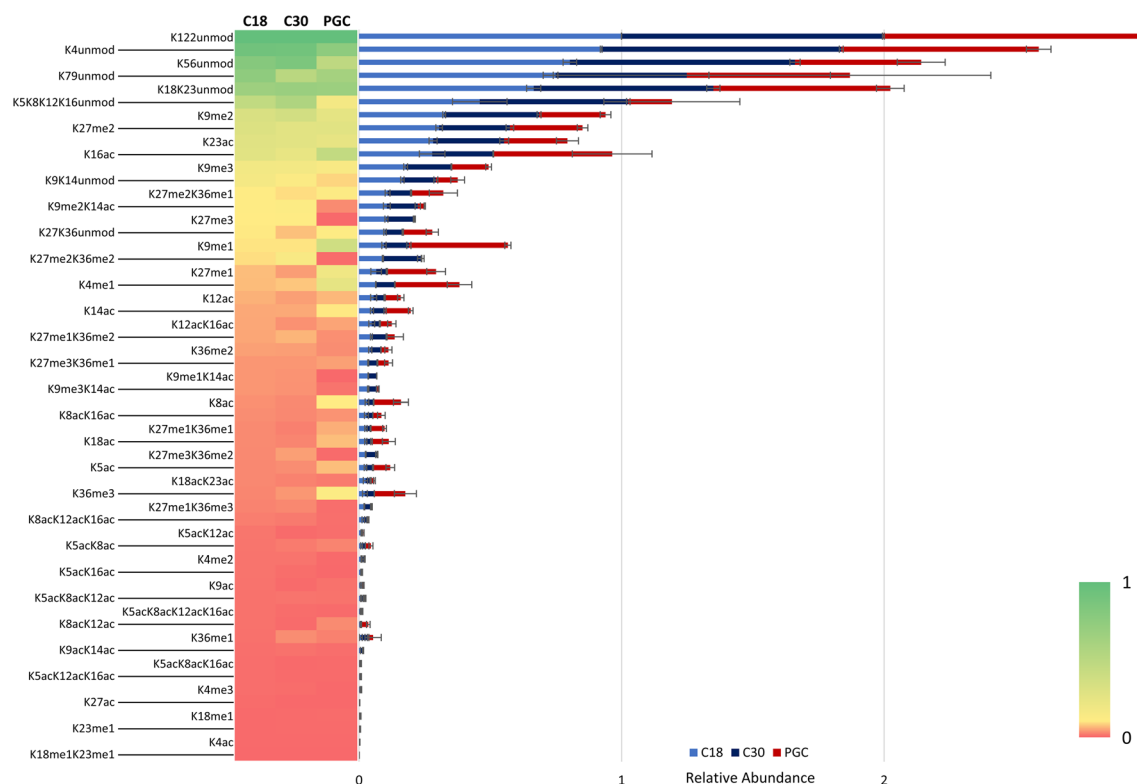


Figure 1. Bottom-up results of histone PTM quantitation from C18, C30, and PGC chromatographic methods are presented as a heatmap and corresponding stacked bar plot. Abundances of modifications were normalized to the total abundance of all forms of the modified peptide, giving very low values to the rare modifications such as K18me1K23me1 and K4ac and high values to common unmodified peptides. The histone corresponding to a modification is not specified in the plot; however, the lysine residues are informative: K5, K8, K12, K16, and K20 are exclusive to histone H4 while all other residues are exclusive to histone H3. There is a trend of agreement between all three stationary phases, with few exceptions. C18 and C30 showed greater agreement with each other than with PGC, likely due to the greater similarity in the structures of the stationary phases. All analyses were performed in triplicate at the cell culture level, and error bars represent standard deviations

Conversely, WCX-HILIC finds more K27me2K36unmod and K9me1K36me2 than PGC. Regardless, the abundances of combinatorially modified peptides are in greater agreement between WCX-HILIC and PGC (Figure 3c) than the abundances of all peptides (Figure 2c), confirming that PGC is a viable stationary phase for quantitating co-existing histone PTMs via middle-down MS.

Simultaneous Bottom-up and Middle-down Analyses

To optimize the analysis of both bottom-up and middle-down sized peptides, we tested a single, combined method to obtain accurate quantitation of individual histone PTMs (bottom-up) while examining changes in the combinatorial profiles of modifications across entire histone tails (middle-down). We analyzed samples of bottom-up tryptic histone peptides mixed with middle-down length peptides from all histone H3 fractions. Using PGC, this design scheme eluted most of the middle-down peptides in the first segment of the LC gradient and most of the bottom-up peptides in the second segment, though there was some overlap in elution and a small number of peptides

eluted outside of the indicated retention time ranges (Figure 4). The bottom-up scans were processed with EpiProfile 2.0, and, despite the longer duty cycle, the results closely matched those of the bottom-up-only runs (Figure 5). Bottom-up data analysis focused on histone H3 due to the importance and complexity of its PTM crosstalk. Bar plots of deconvoluted single PTMs show similar profiles between stationary phases for bottom-up-only data (Figure 5a–c) and bottom-up results from combined bottom-up and middle-down runs (Figure 5d–f). Although quantitation was not robustly reproducible in C18 and C30 combined runs, PGC combined runs yielded similar standard deviations to the bottom-up only runs. The high variability in C18 and C30 runs is likely due to poor separation of bottom-up and middle-down sized peptides. As the bottom-up DIA scan windows cover the +7 to +9 charge states of middle-down sized peptides, poor separation can lead to misidentification and misquantitation in EpiProfile 2.0. Additionally, systematic bottom-up overestimations of H3K23ac and H3K4me1 were present in combination runs, likely caused by reproducible peptide misidentification in EpiProfile 2.0. This is an important limitation to note; however, other modifications are not affected by this anomaly. Additionally, all PGC runs

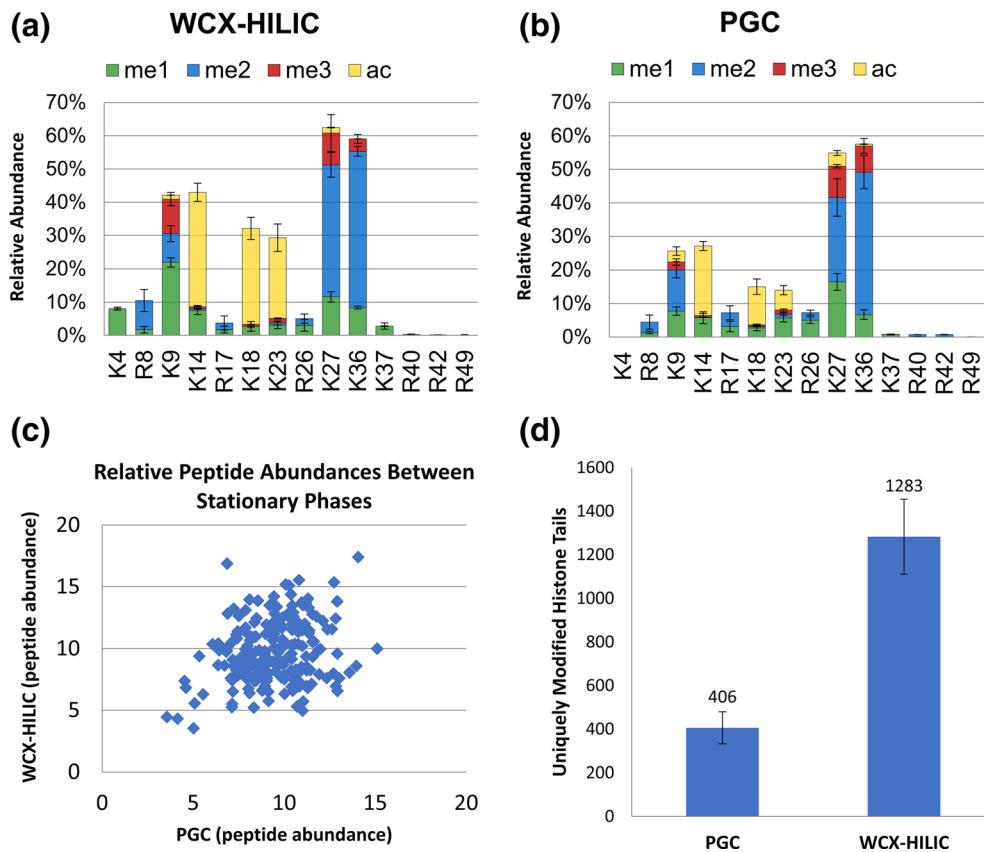


Figure 2. Histone H3 PTMs were deconvoluted from combinatorial marks detected by middle-down mass spectrometry to individual abundances on single residues. All analyses were performed in triplicate at the cell culture level, and error bars represent standard deviations. **(a)** The methylation and acetylation states of lysine and arginine residues as detected by using WCX-HILIC chromatography. **(b)** The methylation and acetylation states of lysine and arginine residues as detected by using formic acid buffer-based PGC chromatography. While some differences are evident between WCX-HILIC and PGC, the overall profile and trend of modifications are very similar. The most dramatic difference between the two stationary phases is the underestimation of acetylation marks by PGC chromatography. **(c)** The relative abundances of all peptides show similar trends between WCX-HILIC and PGC. Peptide abundance values were normalized by calculating $-\log_2$ (peptide intensity ratio). Although the correlation between the two stationary phases is not high, the plot demonstrates that there is neither an egregious outlier nor a substantial bias in peptide abundance. **(d)** The overall number of uniquely modified intact histone tails in WCX-HILIC runs is more than threefold higher than in PGC runs; however, this stark difference has not hampered the ability to obtain meaningful middle-down data through PGC chromatography

reproducibly show high abundance of K9me1, which can be adjusted by correcting for ionization efficiency using peptide standards [49]. Further, relative changes in the abundances of these modifications can still be observed between samples from two different conditions, meaning that PGC-based combined runs can provide accurate values of quantitative differences.

Correlation analysis of the combined run bottom-up results and bottom-up-only results yielded coefficients of at least 0.69, representing reasonable agreement between the methods (Figure 6). The bottom-up-only data showed the strongest correlations across stationary phases, likely due to the simplicity of the sample. These results show that retention of short hydrophilic peptides and accurate bottom-up quantitation of histone PTMs using EpiProfile 2.0 can be achieved through different hydrophobic stationary phases, even in the presence of a middle-down digest. Thus, the overestimations in PGC-based combined runs do not occlude its usage for histone

analysis, and advancements in data analysis software may address the noted limitations in the future.

The combined middle-down results closely resembled those of the middle-down-only runs as well (Figure S4). The closer correlations at the middle-down level compared with the bottom-up level are unsurprising as middle-down is performed by selecting for charge state 8 peptides while DIA-based bottom-up analysis can fragment middle-down sized peptides. We examined the correlation between combinatorial modifications identified by bottom-up and by middle-down (Figure 6), which showed that middle-down quantitation, regardless of the analytical method, correlates worse with bottom-up quantitation than nearly every other calculated correlation. Indeed, the PGC and WCX-HILIC middle-down runs yield a correlation of 0.85, despite the correlation coefficients between bottom-up and middle-down analyses being the lowest of all. This low correlation is unsurprising as middle-down is not expected to provide the most accurate

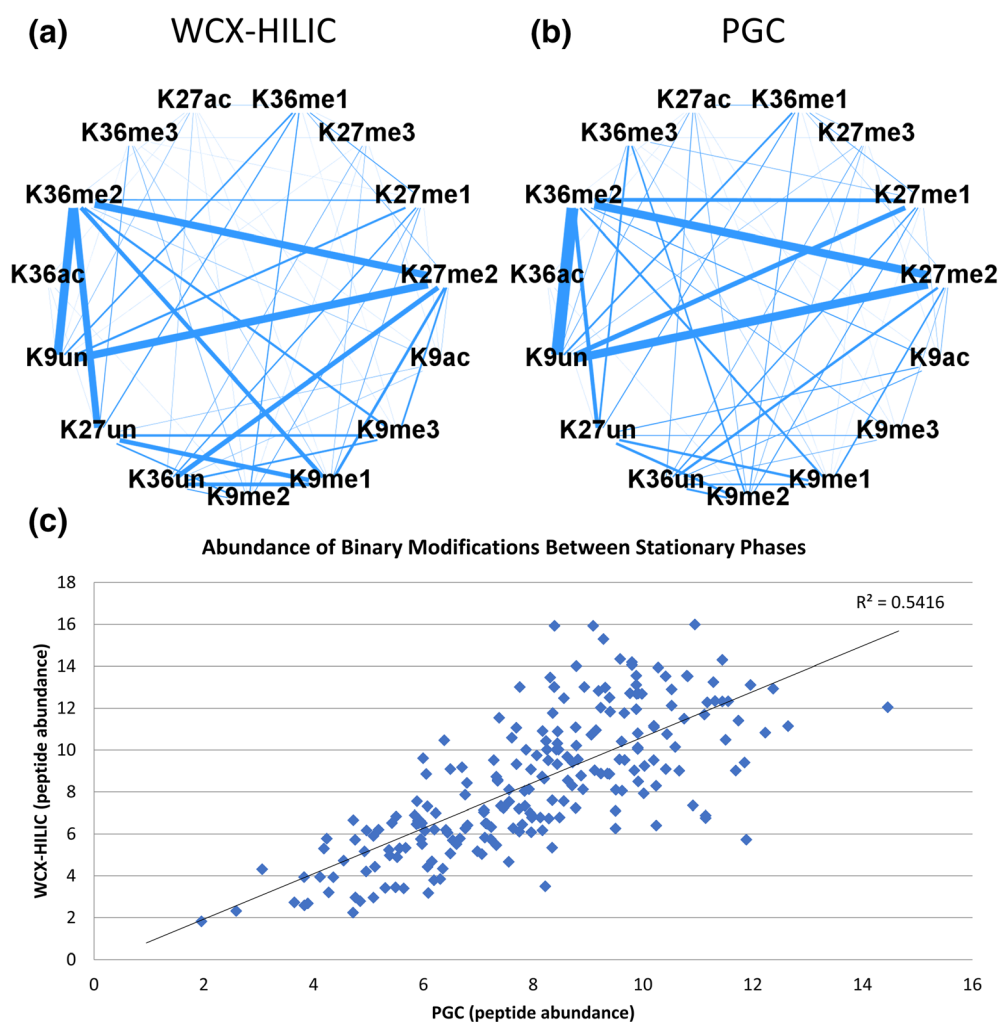


Figure 3. Binary modifications on histone tails were examined by filtering peptides to identify modification co-occurrence frequencies between two residues. **(a)** A circle plot showing the co-occurrence of binary modifications on K9, K27, and K36 from WCX-HILIC middle-down analysis. Line thickness indicates frequency of co-occurrence of the two connected PTMs. **(b)** A circle plot showing the co-occurrence of binary modifications on K9, K27, and K36 from PGC middle-down analysis. **(c)** A scatter plot demonstrating the close agreement between WCX-HILIC and PGC of abundances of binary histone H3 modifications. Peptide abundance values were normalized by calculating $-\log_2$ (peptide intensity ratio)

quantitation of individual PTMs; however, low quality middle-down runs provide substantially worse quantitative accuracy when compared with bottom-up. Thus, the bottom-up data are used to validate the middle-down data, which give context to the interplay of combinatorial modifications on a peptide and provide more insight into biological function than bottom-up. This demonstrates the utility of combined bottom-up and middle-down analyses: the approach reduces instrument time and hands-on time, reduces complexity in chromatographic setup, and provides internal validation for middle-down results without compromising the integrity of the data.

Discussion

We approached the challenge of middle-down analysis by prioritizing convenience and utility rather than quantity of

identifications. This allowed us to focus on creating a method that closely resembles the most common proteomic LC method. Using the common buffers of water and ACN with formic acid, we identified a stationary phase alternative to C18, PGC, which enables this histone analysis method to be easily adopted and used. While other methods have designed formic acid buffer-based approaches, they have been dependent on derivatization [50], shorter length middle-down peptides [48], or a specialized ion mobility mass spectrometer [51]. PGC-based chromatography presents the simplest formic acid buffer-based middle-down approach and yields the highest number of uniquely modified intact histone tail identifications.

Beyond the identification and quantitation of post-translational modifications on histone tails, the interconnectivity of such PTMs is the key to understanding how the histone code dictates the overall chromatin environment,

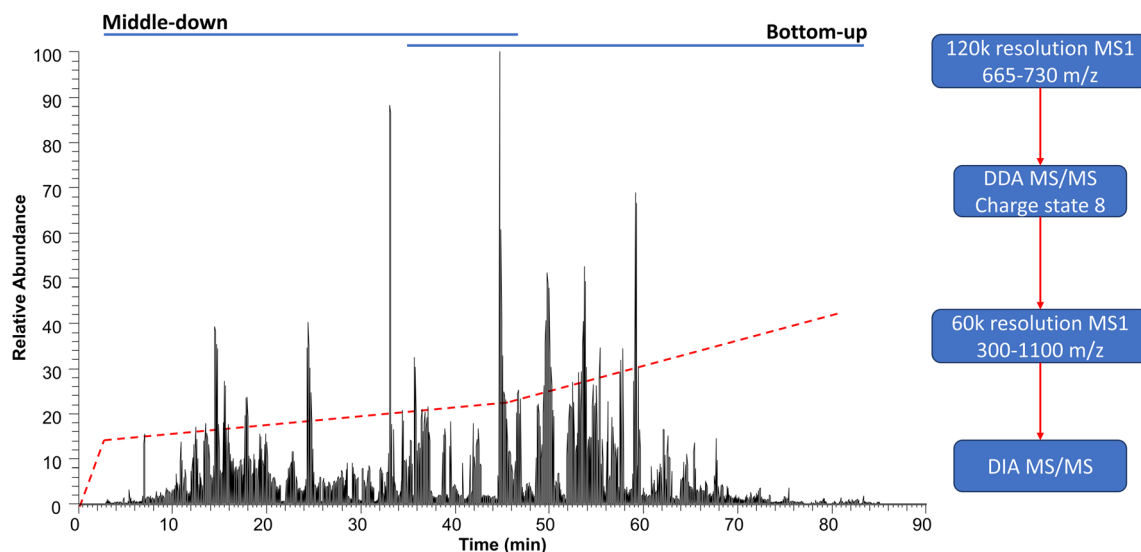


Figure 4. Design scheme for combined bottom-up and middle-down analyses using PGC chromatography. Blue lines (top) indicate where the majority of middle-down and bottom-up peptides elute; however, not all peptides follow this trend. MS scan parameters (right) were cycled throughout the entire run to ensure all peptides associated with middle-down and bottom-up analyses were detected. Sample chromatogram is from a PGC run and the analytical gradient is shown with y-axis values representing % B

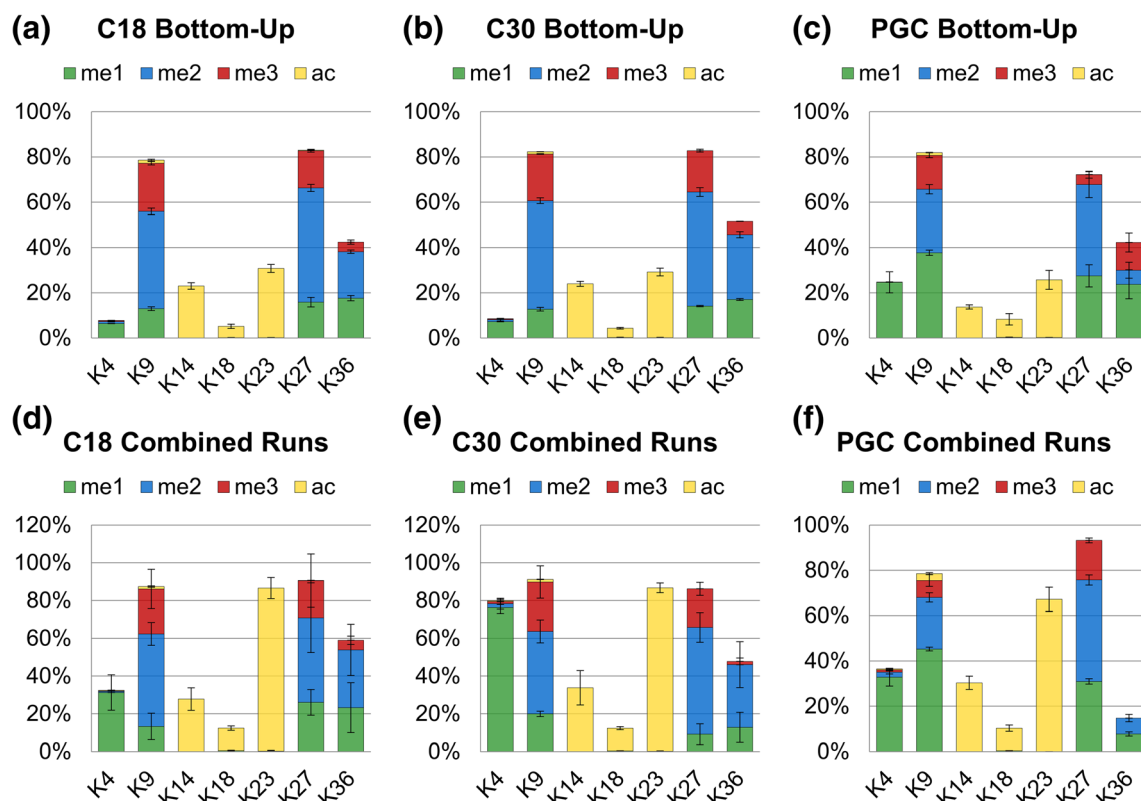


Figure 5. Histone H3 methylation and acetylation abundances across the entire HeLa S3 epigenome. All displayed data are from bottom-up analyses. Each lysine residue is shown with the relative occupation of single modifications, deconvoluted from combinatorial peptide analysis. The top row of bar plots shows results from analysis of tryptic peptides from bottom-up sample preparation, separated by (a) C18, (b) C30, and (c) PGC. The bottom row shows bottom-up results from samples containing a combination of both tryptic peptides from bottom-up sample preparation and middle-down polypeptides from GluC sample preparation, separated by (d) C18, (e) C30, and (f) PGC. All analyses were performed in triplicate at the cell culture level, and error bars represent standard deviations. Minor differences are present across stationary phases; however, the most prominent differences are observed on K4 and K23 when comparing bottom-up only with combined runs, regardless of the stationary phase

	C18 BU	C18 BU (Combined)	C30 BU	C30 BU (Combined)	PGC BU	PGC BU (Combined)	PGC MD	WCX-HILIC MD
C18 BU	1	0.81	0.99	0.68	0.74	0.67	0.55	0.62
C18 BU (Combined)	0.81	1	0.79	0.90	0.67	0.82	0.36	0.54
C30 BU	0.99	0.79	1	0.67	0.69	0.61	0.61	0.64
C30 BU (Combined)	0.68	0.90	0.67	1	0.64	0.80	0.26	0.52
PGC BU	0.74	0.67	0.69	0.64	1	0.80	0.28	0.45
PGC BU (Combined)	0.67	0.82	0.61	0.80	0.80	1	0.24	0.54
PGC MD	0.55	0.36	0.61	0.26	0.28	0.24	1	0.85
WCX-HILIC MD	0.62	0.54	0.64	0.52	0.45	0.54	0.85	1

1
0.8
0.6
0.4
0.2
0

Figure 6. A correlation matrix comparing single PTM bottom-up results (BU) from C18, C30, and PGC stationary phases and middle-down results (MD) from PGC and WCX-HILIC stationary phases. The combined runs indicate that tryptic bottom-up peptides were analyzed from a sample that also contained middle-down peptides. These data show strong agreement between BU runs regardless of the stationary phase and good agreement to the combined runs. The MD data did not correlate very well with the BU and combined data; however, this was expected as middle-down is used for identification of combinatorial PTMs rather than accurate quantitation of PTMs. Notably, the two MD analyses did correlate well with each other, indicating that using PGC for MD does not compromise the quality of the data

which thereby regulates biological processes. It has been well established that the presence of one mark on a histone tail can influence the addition or removal of other modifications on neighboring residues, deemed positive or negative interplay. While bottom-up is the method of choice for histone PTM identification and quantitation, the information provided is limited and can only interrogate interplay within 5 to 11 amino acids. In middle-down methods, where histone tails remained untouched after GluC digestion, histone PTM interplay can be assessed. For instance, without requiring perfect quantitation, the likelihood of an activating mark such as H3K4me3 and a silencing mark like H3K27me3 being present on the same tail can be easily deciphered. This type of information is crucial in the identification of novel poised genes and/or bivalent chromatin domains across the genome, even when peptide quantitation is not perfect. Additionally, middle-down can be used in combination with metabolic labeling [35], providing information about combinatorial code dynamics upon external stimuli.

Compared with bottom-up, middle-down-based approaches also help to remove biases caused by different ionization efficiencies and retention times often found with smaller peptides. One example of this effect is found when analyzing H3K4me3

by bottom-up [52]. After trypsin digestion, the resulting peptide is very hydrophilic (3-TKQTAR-8). Hydrophobicity is slightly increased by derivatization with propionic acid, but the poor charge density of the peptide decreases its ionization efficiency, especially compared with the peptide containing modifications such as H3K327ac or H3K36me1 (27-KSAPATGGVKKPHR-40). As a result, the H3K4me3-containing peptide is underestimated relative to the abundance of most other peptides, including the H3K27-containing peptides [49]. Middle-down methodologies mitigate these biases as the entire histone tail is analyzed, providing enough charge density for more consistent ionization efficiency (+7 to +9).

Previously, middle-down methods for histone analyses have been almost exclusively investigated utilizing WCX-HILIC chromatography. Our method showed similar performance to WCX-HILIC without compromising quantitation and identification of histone PTMs (Figure 2). Although the number of identifications with PGC is lower, the interplay score of modified tails is in high agreement with WCX-HILIC-based chromatography, showing that the goal of middle-down analysis is achieved regardless of the difference in quantity of uniquely modified intact histone tails detected. Additionally, as shown in Figure 3, H3K36me2 is more likely to be found in combination

with H3K27me2 rather than H3K27me3, a trend that is clearly described utilizing PCG chromatography. Indeed, the binary modification states of histone H3 observed using PGC chromatography are in strong agreement with WCX-HILIC (Figure 3b). To our knowledge, this is the first time histone PTM interplay has been demonstrated to be conserved across very different chromatographic methods. These results are critical not only because the data are reliable, but also because formic acid buffer-based methods require minimal LC-MS hardware configuration changes in proteomics labs. By utilizing the same buffers as in C18 chromatography, the PGC-based middle-down method can be rapidly configured as only the column needs to be changed.

In this novel method, we were also able to combine bottom-up and middle-down analyses in one run, decreasing instrument time and labor. Importantly, the combined runs achieved similar results to bottom-up-only analysis utilizing C18-based chromatography (Figure 5). This was also demonstrated by the correlation between bottom-up and middle-down runs, where bottom-up results are similar regardless of stationary phase and the presence middle-down polypeptides in the run (Figure 6). Of note is the higher abundance of H3K4me1 in PGC bottom-up runs (Figure 5C), as this peptide is not well retained by C18 columns and its abundance is commonly underestimated. This peptide is better retained on PGC columns, which leads to a more reliable peak that yields both more precise and more accurate quantitation. Additionally, C18 bottom-up and PGC bottom-up results are in agreement with a correlation coefficient of 0.74 (Figure 6). The agreement between C18's bottom-up-only and combined runs (0.81) nearly matches the agreement between PGC's bottom-up-only and combined runs (0.80). As noted, not all peptides are expected to yield identical quantitation values between the different stationary phases; thus, the agreement between the two analytical approaches employed is more important to the reliability of the data than the agreement between different stationary phases.

Conclusions

This work represents a step forward in the effort to improve, standardize, and increase the adoption of middle-down mass spectrometry for histone tails. While our formic acid buffer-based methods have disadvantages compared with classical WCX-HILIC, they dramatically improve the convenience, reliability, and ease of use of middle-down analyses. In addition, we have shown the ability to analyze histone tails at the bottom-up and middle-down levels simultaneously, allowing users to minimize instrument time and validate the quantitation of their middle-down results internally. These advantages provide the opportunity for more labs to explore the combinatorial histone code beyond the scope of bottom-up analysis. With this simplified method in hand, the most challenging aspect of the middle-down workflow is data analysis. Indeed, the next major advancement in middle-down analysis of histone tails will be

the development of simplified tools capable of processing data accurately and rapidly.

Acknowledgements

We gratefully acknowledge funding from NIH grants GM110174, AI118891, and CA196539.

References

1. Strahl, B.D., Allis, C.D.: The language of covalent histone modifications. *Nature*. **403**(6765), 41–45 (2000)
2. Jenuwein, T., Allis, C.D.: Translating the histone code. *Science*. **293**(5532), 1074–1080 (2001)
3. Kouzarides, T.: Chromatin modifications and their function. *Cell*. **128**(4), 693–705 (2007)
4. Bannister, A.J., Kouzarides, T.: Regulation of chromatin by histone modifications. *Cell Res*. **21**(3), 381–395 (2011)
5. Rossetto, D., Avvakumov, N., Côté, J.: Histone phosphorylation. *Epigenetics*. **7**(10), 1098–1108 (2012)
6. Torres-Padilla, M.-E., Parfitt, D.-E., Kouzarides, T., Zernicka-Goetz, M.: Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature*. **445**, 214 (2007)
7. Casas-Delucchi, C.S., Brero, A., Rahn, H.-P., Solovei, I., Wutz, A., Cremer, T., Leonhardt, H., Cardoso, M.C.: Histone acetylation controls the inactive X chromosome replication dynamics. *Nat. Commun.* **2**, 222–222 (2011)
8. Huang, H., Sabari, B.R., Garcia, B.A., Allis, C.D., Zhao, Y.: SnapShot: histone modifications. *Cell*. **159**(2), 458–458 e1 (2014)
9. Fischle, W., Wang, Y., Allis, C.D.: Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**(2), 172–183 (2003)
10. Lee, J.S., Smith, E., Shilatifard, A.: The language of histone crosstalk. *Cell*. **142**(5), 682–685 (2010)
11. Hunter, T.: The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell*. **28**(5), 730–738 (2007)
12. Atlasi, Y., Stunnenberg, H.G.: The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643 (2017)
13. Bonaldi, T., Imhof, A., Regula, J.T.: A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics*. **4**(5), 1382–1396 (2004)
14. Sidoli, S., Bhanu, N.V., Karch, K.R., Wang, X., Garcia, B.A.: Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: from histone extraction to data analysis. *J. Vis. Exp.* (111), e54112 (2016)
15. Janssen, K.A., Sidoli, S., Garcia, B.A.: Recent achievements in characterizing the histone code and approaches to integrating epigenomics and systems biology. *Methods Enzymol.* **586**, 359–378 (2017)
16. Sidoli, S., Garcia, B.A.: Characterization of individual histone posttranslational modifications and their combinatorial patterns by mass spectrometry-based proteomics strategies. *Methods Mol. Biol.* **1528**, 121–148 (2017)
17. Liao, R., Wu, H., Deng, H., Yu, Y., Hu, M., Zhai, H., Yang, P., Zhou, S., Yi, W.: Specific and efficient N-propionylation of histones with propionic acid N-hydroxysuccinimide ester for histone marks characterization by LC-MS. *Anal. Chem.* **85**(4), 2253–2259 (2013)
18. Maile, T.M., Izrael-Tomasevic, A., Cheung, T., Guler, G.D., Tindell, C., Masselot, A., Liang, J., Zhao, F., Trojer, P., Classon, M., Arnott, D.: Mass spectrometric quantification of histone post-translational modifications by a hybrid chemical labeling method. *Mol. Cell. Proteomics*. **14**(4), 1148–1158 (2015)
19. Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F.: Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* **2**(4), 933–938 (2007)
20. Sidoli, S., Lin, S., Xiong, L., Bhanu, N.V., Karch, K.R., Johansen, E., Hunter, C., Mollah, S., Garcia, B.A.: Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) analysis for characterization and

- quantification of histone post-translational modifications. *Mol. Cell. Proteomics*. **14**(9), 2420–2428 (2015)
21. Sidoli, S., Simithy, J., Karch, K.R., Kulej, K., Garcia, B.A.: Low resolution data-independent acquisition in an LTQ-Orbitrap allows for simplified and fully untargeted analysis of histone modifications. *Anal. Chem.* **87**(22), 11448–11454 (2015)
 22. Sidoli, S., Fujiwara, R., Garcia, B.A.: Multiplexed data independent acquisition (MSX-DIA) applied by high resolution mass spectrometry improves quantification quality for the analysis of histone peptides. *Proteomics*. **16**(15–16), 2095–2105 (2016)
 23. Krautkramer, K.A., Reiter, L., Denu, J.M., Dowell, J.A.: Quantification of SAHA-dependent changes in histone modifications using data-independent acquisition mass spectrometry. *J. Proteome Res.* **14**(8), 3252–3262 (2015)
 24. Shi, L., Shi, J., Shi, X., Li, W., Wen, H.: Histone H3.3 G34 mutations alter histone H3K36 and H3K27 methylation in cis. *J. Mol. Biol.* **430**(11), 1562–1565 (2018)
 25. Mao, H., Han, G., Xu, L., Zhu, D., Lin, H., Cao, X., Yu, Y., Chen, C.D.: Cis-existence of H3K27me3 and H3K36me2 in mouse embryonic stem cells revealed by specific ions of isobaric modification chromatogram. *Stem Cell Res Ther.* **6**(1), 132–132 (2015)
 26. Marchione, D.M., Garcia, B.A., Wojcik, J.: Proteomic approaches for cancer epigenetics research. *Expert Rev Proteomics*. **16**(1), 33–47 (2019)
 27. Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., Reinberg, D.: Asymmetrically modified nucleosomes. *Cell*. **151**(1), 181–193 (2012)
 28. Garcia, B.A.: What does the future hold for top down mass spectrometry? *J. Am. Soc. Mass Spectrom.* **21**(2), 193–202 (2010)
 29. Young, N.L., DiMaggio, P.A., Plazas-Mayorca, M.D., Baliban, R.C., Floudas, C.A., Garcia, B.A.: High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics*. **8**(10), 2266–2284 (2009)
 30. Shvartsburg, A.A., Zheng, Y., Smith, R.D., Kelleher, N.L.: Ion mobility separation of variant histone tails extending to the “middle-down” range. *Anal. Chem.* **84**(10), 4271–4276 (2012)
 31. Tran, J.C., Zamdborg, L., Ahlf, D.R., Lee, J.E., Catherman, A.D., Durbin, K.R., Tipton, J.D., Vellaichamy, A., Kellie, J.F., Li, M., Wu, C., Sweet, S.M.M., Early, B.P., Siuti, N., LeDuc, R.D., Compton, P.D., Thomas, P.M., Kelleher, N.L.: Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. **480**(7376), 254–U141 (2011)
 32. Shaw, J.B., Li, W., Holden, D.D., Zhang, Y., Griep-Raming, J., Fellers, R.T., Early, B.P., Thomas, P.M., Kelleher, N.L., Brodbelt, J.S.: Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J. Am. Chem. Soc.* **135**(34), 12646–12651 (2013)
 33. Ansong, C., Wu, S., Meng, D., Liu, X., Brewer, H.M., Kaiser, B.L.D., Nakayasu, E.S., Cort, J.R., Pevzner, P., Smith, R.D., Heffron, F., Adkins, J.N., Paša-Tolić, L.: Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella typhimurium* in response to infection-like conditions. *Proc. Natl. Acad. Sci. U. S. A.* **110**(35), 10153–10158 (2013)
 34. Sidoli, S., Lin, S., Karch, K.R., Garcia, B.A.: Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative abundance and stoichiometry. *Anal. Chem.* **87**(6), 3129–3133 (2015)
 35. Sidoli, S., Lu, C., Coradin, M., Wang, X., Karch, K.R., Ruminowicz, C., Garcia, B.A.: Metabolic labeling in middle-down proteomics allows for investigation of the dynamics of the histone code. *Epigenetics Chromatin*. **10**(1), 34 (2017)
 36. Zamdborg, L., LeDuc, R.D., Glowacz, K.J., Kim, Y.B., Viswanathan, V., Spaulding, I.T., Early, B.P., Bluhm, E.J., Babai, S., Kelleher, N.L.: ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **35**(Web Server issue), W701–W706 (2007)
 37. Baliban, R.C., DiMaggio, P.A., Plazas-Mayorca, M.D., Young, N.L., Garcia, B.A., Floudas, C.A.: A novel approach for untargeted post-translational modification identification using integer linear optimization and tandem mass spectrometry. *Mol. Cell. Proteomics*. **9**(5), 764–779 (2010)
 38. Pesavento, J.J., Mizzen, C.A., Kelleher, N.L.: Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal. Chem.* **78**(13), 4271–4280 (2006)
 39. DiMaggio, P.A., Young, N.L., Baliban, R.C., Garica, B.A., Floudas, C.A.: 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**(11), 2527–2543 (2009)
 40. Guan, S.H., Burlingame, A.L.: Data processing algorithms for analysis of high resolution MSMS spectra of peptides with complex patterns of posttranslational modifications. *Mol. Cell. Proteomics*. **9**(5), 804–810 (2010)
 41. Sidoli, S., Garcia, B.A.: Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev Proteomics*. **14**(7), 617–626 (2017)
 42. West, C., Elfakir, C., Lafosse, M.: Porous graphitic carbon: a versatile stationary phase for liquid chromatography. *J. Chromatogr. A*. **1217**(19), 3201–3216 (2010)
 43. Karch, K.R., Sidoli, S., Garcia, B.A.: Identification and quantification of histone PTMs using high-resolution mass spectrometry. In: Marmorstein, R. (ed.) *Methods in Enzymology*, pp. 3–29. Academic Press, Cambridge (2016)
 44. Jung, H.R., Sidoli, S., Haldbø, S., Sprenger, R.R., Schwämmle, V., Pasini, D., Helin, K., Jensen, O.N.: Precision mapping of coexisting modifications in histone H3 tails from embryonic stem cells by ETD-MS/MS. *Anal. Chem.* **85**(17), 8232–8239 (2013)
 45. Yuan, Z.-F., Sidoli, S., Marchione, D.M., Simithy, J., Janssen, K.A., Szurgot, M.R., Garcia, B.A.: EpiProfile 2.0: a computational platform for processing epi-proteomics mass spectrometry data. *J. Proteome Res.* **17**(7), 2533–2541 (2018)
 46. Greer, S.M., Sidoli, S., Coradin, M., Jespersen, M.S., Schwämmle, V., Jensen, O.N., Garcia, B.A., Brodbelt, J.S.: Extensive characterization of heavily modified histone tails by 193 nm ultraviolet photodissociation mass spectrometry via a middle-down strategy. *Anal. Chem.* **90**(17), 10425–10433 (2018)
 47. Sidoli, S., Schwämmle, V., Ruminowicz, C., Hansen, T.A., Wu, X., Helin, K., Jensen, O.N.: Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *Proteomics*. **14**(19), 2200–2211 (2014)
 48. Schröder, C.U., Ziemianowicz, D.S., Merx, K., Schriemer, D.C.: Simultaneous proteoform analysis of histones H3 and H4 with a simplified middle-down proteomics method. *Anal. Chem.* **90**(5), 3083–3090 (2018)
 49. Lin, S., Wein, S., Gonzales-Cope, M., Otte, G.L., Yuan, Z.-F., Afjehi-Sadat, L., Maile, T., Berger, S.L., Rush, J., Lill, J.R., Arnott, D., Garcia, B.A.: Stable-isotope-labeled histone peptide library for histone post-translational modification and variant quantification by mass spectrometry. *Mol. Cell. Proteomics*. **13**(9), 2450 (2014)
 50. Liao, R., Zheng, D., Nie, A., Zhou, S., Deng, H., Gao, Y., Yang, P., Yu, Y., Tan, L., Qi, W., Wu, J., Li, E., Yi, W.: Sensitive and precise characterization of combinatorial histone modifications by selective derivatization coupled with RPLC-ETD-MS/MS. *J. Proteome Res.* **16**(2), 780–787 (2017)
 51. Garabedian, A., Baird, M.A., Porter, J., Fouque, K.J.D., Shliaha, P.V., Jensen, O.N., Williams, T.D., Fernandez-Lima, F., Shvartsburg, A.A.: Linear and differential ion mobility separations of middle-down proteoforms. *Anal. Chem.* **90**(4), 2918–2925 (2018)
 52. Guo, Q., Sidoli, S., Garcia, B.A., Zhao, X.: Assessment of quantification precision of histone post-translational modifications by using an ion trap and down to 50 000 cells as starting material. *J. Proteome Res.* **17**(1), 234–242 (2018)