

# Middle-Down Characterization of the Cell Cycle Dependence of Histone H4 Posttranslational Modifications and Proteoforms

Tingting Jiang, Michael E. Hoover, Matthew V. Holt, Michael A. Freitas, Alan G. Marshall, and Nicolas L. Young\*

Post-translational modifications (PTMs) of histones are important epigenetic regulatory mechanisms that are often dysregulated in cancer. We employ middle-down proteomics to investigate the PTMs and proteoforms of histone H4 during cell cycle progression. We use pH gradient weak cation exchange-hydrophilic interaction liquid chromatography (WCX-HILIC) for on-line liquid chromatography-mass spectrometry analysis to separate and analyze the proteoforms of histone H4. This procedure provides enhanced separation of proteoforms, including positional isomers, and simplifies downstream data analysis. We use ultrahigh mass accuracy and resolution Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer to unambiguously distinguish between acetylation and tri-methylation ( $\Delta m = 0.036$  Da). In total, we identify and quantify 233 proteoforms of histone H4 in two breast cancer cell lines. We observe significant increases in S1 phosphorylation during mitosis, implicating an important role in mitotic chromatin condensation. A decrease of K20 unmodified proteoforms is observed as the cell cycle progresses, corresponding to an increase of K20 mono- and di-methylation. Acetylation at K5, K8, K12, and K16 declines as cells traverse from S phase to mitosis, suggesting cell cycle-dependence and an important role during chromatin replication and condensation. These new insights into the epigenetics of the cell cycle may provide new diagnostic and prognostic biomarkers.

## 1. Introduction

Histone posttranslational modifications (PTMs) are crucial to the regulation and maintenance of eukaryotic genomes. The basic template of our heredity originates from genomic DNA. In the nuclei of all eukaryotic cells, genomic DNA is organized and compacted into chromatin that consists of repeating units of nucleosomes. Each nucleosome is comprised of an octamer (an H3-H4 tetramer and two H2A-H2B dimers) of core histone proteins wrapped around by  $\sim 146$  base pairs of DNA.<sup>[1,2]</sup> Core histone proteins consist of a globular domain and a structurally dynamic N-terminal tail that is prone to a variety of PTMs. Commonly observed modifications include acetylation, methylation, and phosphorylation, and to a lesser extent ubiquitination, sumoylation, arginine deimination, proline isomerization, and ADP ribosylation.<sup>[3,4]</sup> These PTMs function in conjunction with histone variants, DNA methylation, and ATP-dependent chromatin

T. Jiang, Dr. A. G. Marshall  
Department of Chemistry and Biochemistry  
Florida State University  
95 Chieftain Way, Tallahassee, FL, 32306, USA

Dr. M. E. Hoover, Dr. M. A. Freitas  
Department of Cancer Biology and Genetics  
Ohio State University  
460 West 12th Avenue, Columbus, OH, 43210, USA

M. V. Holt, Dr. N. L. Young  
Verna and Marrs McLean Department of Biochemistry and Molecular Biology  
Baylor College of Medicine  
One Baylor Plaza, Houston, TX, 77030, USA  
E-mail: Nicolas.Young@bcm.edu

Dr. A. G. Marshall  
Ion Cyclotron Resonance Program  
National High Magnetic Field Laboratory  
1800 East Paul Dirac Drive, Tallahassee, FL, 32310, USA

Dr. N. L. Young  
Department of Molecular and Cellular Biology  
Baylor College of Medicine  
One Baylor Plaza, Houston, TX, 77030, USA

See the associated Commentary by Simone Sidoli,  
<https://doi.org/10.1002/pmic.201800131>

DOI: 10.1002/pmic.201700442

## Significance Statement

In the present study, a middle-down approach is employed for the identification, characterization, and quantitation of the posttranslational modifications (PTMs) and proteoforms of histone H4 in two breast cancer cell lines. Our experimental strategy not only preserves the combinatorial attributes of most existing PTMs, but also enhances the separation ability and reduces the complexity of downstream data analysis. The fluctuations in the relative abundances of common modifications throughout the cell cycle, including phosphorylation, acetylation, and methylation, suggest that they play significant roles at different stages of the cell cycle in breast cancer and could participate in carcinogenesis. In summary, our analysis advances the understanding and elucidation of cancer epigenetics with preservation of single molecule connectivity between the major N-terminal PTMs and lays the foundation for the further study of histones or other proteins.

remodeling.<sup>[5]</sup> These interrelated mechanisms generate a dynamic epigenetic machinery to modulate and regulate the chromatin structure and biological processes, such as transcription, DNA replication and repair, and chromatin condensation.<sup>[6]</sup>

Histone PTMs function in concert at the single molecule level. The histone code hypothesis proposes that histone PTMs may function in concert to orchestrate the regulation of gene expression.<sup>[6]</sup> This theory has been examined in numerous biological, biochemical, and biophysical studies. ChIP-chip and ChIP-seq techniques have been employed to demonstrate the genomic localization of histone modifications and suggest strong evidence of combinatorial effects of histone PTM marks.<sup>[3,7–9]</sup> However, these efforts often lack quantitative capacity and are not unbiased. The discovery of histone code “writers” and “erasers,” histone-modifying enzymes such as histone acetyltransferases (HATs), histone deacetylases (HDACs), and histone lysine methyltransferases (HKMTs), has revealed some basic elements of homeostasis of combinatorial patterns of histone modifications.<sup>[10]</sup> Previous studies reveal direct biophysical interactions that are mediated by multiple site-specific PTM recognition domains in proteins or protein complexes that function in tandem to recognize specific combinations of PTMs on single histone molecules.<sup>[11–15]</sup> These findings suggest that histone combinatorial PTMs serve as a signal platform to regulate the recruitment of downstream effectors or “readers,” often containing tandem PTM recognition domains, to determine the ultimate biological outcome.<sup>[6,16]</sup> Thus, development of a robust and effective analytical method to decipher the histone codes is highly desirable.

A proteoform is a protein product with all possible variations thereof (PTMs, sequence variants, etc.) unambiguously defined in combination at the single molecule level.<sup>[17]</sup> This terminology has recently been rapidly adopted in the scientific lexicon of the proteomics community because it concisely describes a complex concept that often requires continual repetition. For example, the histone H4 K20me2 PTM is distinct from the H4 K20me2 proteoform. The H4 K20me2 PTM simply indicates occupancy of a single residue with a PTM. The H4 K20me2 proteoform

indicates that only that site is occupied and all other sites of variable modification are unambiguously unmodified. The concept of binary (or higher order) combinations expresses only how two (or more) PTMs exist or function in concert, without consideration of other PTMs. These concepts are exceptionally useful in describing the nuances of hyper-modified proteins such as histone H4. Note that we use proteoform in slight violation of this definition because we do not capture a few minor PTMs deep in the sequence; however, the utility of this concept remains essential to understanding the hyper-modified histone H4 N-terminal tail.

Liquid chromatography-mass spectrometry (LC-MS) methods, when properly designed, can enable the unbiased quantitation of single-molecule-level combinations of PTMs (proteoforms). Mass spectrometry (MS)-based methods are rapidly evolving and have become an important analysis platform for the characterization and analysis of histones and other proteins with PTMs due to its high mass accuracy, resolution, and selectivity.<sup>[18–21]</sup> Traditional bottom-up MS and antibody-based methods are incapable of maintaining the connectivity between modifications, making it impossible to observe single molecule combination of histone PTMs.<sup>[22]</sup> In addition, antibody-based techniques are susceptible to cross-reactivity and epitope occlusion.<sup>[3]</sup> Thus, the current most effective methods to connect two distal PTMs are via LC-MS approaches that achieve gas-phase sequencing of molecules containing both sites of variable modification.

Middle-down LC-MS provides an effective alternative to top-down LC-MS methods and maintains most of the single-molecule-level combinatorial information. Great efforts have been made with top-down MS studies of intact histone H4. However, they either suffer from an inability to couple on-line LC separation to high quality MS analysis or lack reliable quantification metrics to interpret the complicated spectra with reproducible quantitation.<sup>[23–28]</sup> Middle-down mass spectrometric analysis relies on digestion of the protein into large peptides (usually between 3 and 10 kDa) by use of proteases such as AspN and GluC, to provide a promising alternative approach to investigate histone proteins with high accuracy, sensitivity, and specificity.<sup>[29–32]</sup> Middle-down analysis maintains molecular connectivity over large stretches of sequence, and circumvents many of the challenges of top-down methods.

A quantitative understanding of histone H4 proteoforms in the cell cycle of breast cancer cells will contribute to an expanded understanding of breast cancer and fundamental biology. Breast cancer is the most commonly diagnosed cancer and the second leading cause of death for women throughout the world. Identification of new biomarkers of breast cancer, which can be evaluated as indicators of normal or pathologic processes, is challenging but exceptionally appealing. Crosstalk networks among histone modifications have been implicated in breast cancer.<sup>[33–36]</sup> Moreover, the current proteoform-level understanding of the fundamental biology of the cell cycle is very limited. Previously, our group has reported a significant increase in phosphorylation of two histone H1 variants in two breast cell lines as the cell cycle advances from S phase to mitosis,<sup>[36]</sup> suggesting that this modification is cell cycle-dependent and could serve as a marker for proliferation. Additionally, histone H4 PTMs, such as methylation, acetylation, and phosphorylation are known to change during cell cycle progression.<sup>[37–39]</sup> Further, increasing evidence

indicates that these modifications function in concert rather than independently.<sup>[3,6-9]</sup> Studies have also shown that cell survival, cell death, and cell cycle progression are interconnected.<sup>[40]</sup> Thus, an investigation of the role of combinatorial histone PTMs in cell proliferation pathways during cell cycle progression will enhance our understanding of carcinogenesis. This may ultimately offer the potential to manipulate cancer cells into apoptosis or other cellular states prone to treatment and effectively control cancer progression.

Here, we present a robust, accurate, effective, and automated middle-down workflow to investigate the proteoforms of histone H4. This method uses chromatography with high selectivity to distinguish isomeric proteoforms, electron transfer dissociation (ETD) for extensive fragmentation of the highly basic H4 tail peptide, and automated quantitative data analysis. This approach enables us to make relative abundance comparisons with high quantitative accuracy. We apply this method to study the fluctuations of the combinatorial PTM patterns of histone H4 in two breast cancer cell lines during cell cycle progression. Our analysis reveals new insights on how histone H4 PTMs function in concert and how histone H4 proteoforms vary throughout the cell cycle in cultured breast cancer cells.

## 2. Experimental Section

**Sample Preparation:** Breast cancer cells from MDA-MB-231 and MCF-10A cell lines were grown as previously described.<sup>[36]</sup> Two rounds of 2 mM thymidine (Sigma-Aldrich, St. Louis, MO, USA) were applied to the cell growth medium of each cell line to attain an S phase block and one round of 100 nM nocodazole (Sigma-Aldrich, St. Louis, MO, USA) to achieve a mitotic phase block. Cell nuclei from 50 million cells at each phase were isolated by standard NP-40 alternative-based methods. After nuclei isolation, the resulting pellets were acid-extracted with 0.4 N H<sub>2</sub>SO<sub>4</sub>, precipitated with 20% trichloroacetic acid (TCA), washed with acetone and air-dried overnight. Extracts from the previous two steps were purified and separated by use of reversed-phase high-performance liquid chromatography (RP-HPLC) with a C<sub>18</sub> column (4.6 × 250 mm, 5 μm, 300 Å pore size, 218TP54, Grace, Columbia, Maryland) with a Dionex UltiMate 3000 × 2 Dual Analytical System (Thermo Scientific, Waltham, MA, USA). Solvent A consists of 5% CH<sub>3</sub>CN (LC grade) and 0.2% trifluoroacetic acid (TFA; LC-MS grade) and solvent B contains 95% CH<sub>3</sub>CN and 0.188% TFA. Fractions were eluted at a flow rate of 0.8 mL min<sup>-1</sup> and collected with a linear gradient of solvent B as follows: 5% to 30% B in 5 min, 35% to 60% B in 105 min. Each fraction was collected with an autosampler and vacuum-dried. Histone H4 was subsequently digested with AspN endoprotease (Roche Applied Science) in 100 mM ammonium bicarbonate (pH = 8) buffer with a protein-to-enzyme ratio of 10:1 for 6 h at 37 °C. Digestion was quenched by freezing at -80 °C. The target 23-amino acid peptide of the N-terminal tail of histone H4 was further purified on a second C<sub>18</sub> column (2.1 × 230 mm, 5 μm, 300 Å pore size, Grace, Columbia, Maryland) implemented on the second channel of the above HPLC system with a gradient of 1% B per min at a flow rate of 0.2 mL min<sup>-1</sup>.

**WCX-HILIC-MS/MS:** The N-terminal tails of histone H4 were separated in the second dimension by pH gradient-driven weak

cation exchange-hydrophilic interaction LC (WCX-HILIC) with a nanoLC system (Eksigent 2D-NanoLC Ultra), coupled on-line to a custom-built hybrid Velos Pro-14.5 or 21 tesla Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer.<sup>[20,21]</sup> Even though the two instruments are highly similar in design except for magnet field strength, it could possibly account for part of the analytical differences we observed from two cell lines. The Velos Pro is equipped with front-end electron transfer dissociation (FETD), for which fluoranthene is used to generate reagent ions.<sup>[41]</sup> Approximately 1.0–2.0 μg of purified histone H4 1-23aa peptide sample was loaded onto a custom-packed HILIC column (~20 cm, 100 μm ID) with PolyCAT A resin (3 μm particles, 1500 Å pore size, PolyLC, Columbia, MD, USA). Solvent A consists of 95% methanol (LC-MS grade), 5 mM propionic acid (>99.5%, Fluka Puriss, Sigma), and ammonium hydroxide (>99.99%) to bring the pH to 6.0. Solvent B is comprised of 4.5% methanol (LC-MS grade) with 0.5% formic acid (LC-MS grade) added to adjust the pH to 2.0. The gradient starts at 25% solvent B, followed by a linear gradient of 36–99.8% B in 100 min. A wash step of 99.8% B is maintained for 30 min before returning back to 25% B for 60 min for column reconditioning at a flow rate of 250 nL min<sup>-1</sup>. Data acquisition was performed by use of Xcalibur software (Thermo Fisher Scientific) with a mass range of *m/z* 210–2000 at a mass resolving power of 152 500 at *m/z* 406. Each full-range mass spectrum was followed by five data-dependent tandem ETD product ion spectra for the five most abundant precursor ions at a charge state of 6+ with 50 ms reaction period. The automatic gain control (AGC) target value was set at 1e6 for MS<sup>1</sup> and 2e5 for MS<sup>2</sup>. For a large ion population, ten fills of the external quadrupole trap were employed to transfer more ions to the ICR cell for MS<sup>2</sup> analysis. The isolation width was set at *m/z* 2.2 and maximum ion injection period of 500 ms. Precursor charge states 1+, 2+, and 3+ were excluded and dynamic exclusion was disabled throughout the analysis to maximize the number of MS<sup>2</sup> spectra for the analytes of interest. This abundant and consistent sampling enables higher confidence and consistent partitioning of signal to proteoforms for quantitative analysis. All time-domain data were Hanning apodized, zero-filled once, fast Fourier transformed to magnitude mode, and then converted to mass-to-charge ratio spectra by a two-term calibration equation.<sup>[42]</sup>

**Data Processing and Analysis:** Data processing is performed by a custom analysis suite adapted from DiMaggio, Jr. et al.<sup>[43]</sup> Briefly, theoretical MS<sup>2</sup> fragment spectra are generated for all possible chosen PTM combinations. MS<sup>2</sup> spectra are matched to theoretical spectra with 10 ppm error, and are within 6.8 Da window of the precursor mass (a di-methylated equivalent precursor will not be compared with a mono-methylated theoretical spectra). A mixed integer linear optimization solution is used to determine which combinations best account for the most intensity in a given spectrum. The matched MS<sup>2</sup> peak magnitude specific to a given combination (only the magnitude of MS<sup>2</sup> peaks which match a given assigned spectra) is multiplied by the MS<sup>1</sup> precursor intensity and integrated with all identically matching MS<sup>2</sup> spectra to determine a given proteoform's abundance. Values are normalized to total percent of histone H4 and discrete and combinatorial values are obtained through marginalization of proteoform quantities. For example, the relative abundance of K16ac is simply the sum of the relative

abundances of all proteoforms containing K16ac. The relative abundance of the single molecule binary combination of K16ac and K20me2 is simply the sum of the relative abundance of all proteoforms containing both K16ac and K20me2. Statistical analysis and figure generation is done by use of R and Python programming languages and packages (matplotlib). Statistical significance was defined as  $p < 0.05$  derived from two-tail Student's *t*-test and Pearson correlations were used to determine reproducibility.

### 3. Results and Discussion

#### 3.1. Experimental Workflow

Off-line RP-HPLC is used to separate the crude histone acid extract into different families and the histone H4 fraction is subsequently digested and further purified. In **Figure 1A**, the core histones H2B, H2A, H4, and H3 and linker histone H1 are baseline-separated. After collection of each fraction, endoprotease AspN is used to digest the histone H4 fraction to produce the first 23-amino acid peptide of the N-terminal tail where the majority of PTMs are located. To reduce the sample complexity and facilitate data analysis, the digested peptides are subjected to a second offline RP-HPLC to further purify the target N-terminal peptide of 23 amino acids of histone H4 from the internal peptides. As demonstrated in **Figure 1B**, the N-terminal tail of histone H4 elutes starting from approximately 23 to 25 min. Proteoforms with different acetylations or other modifications are not well separated, thereby exposing the limitations of  $C_{18}$ -based RP-HPLC for histone proteoform separation.

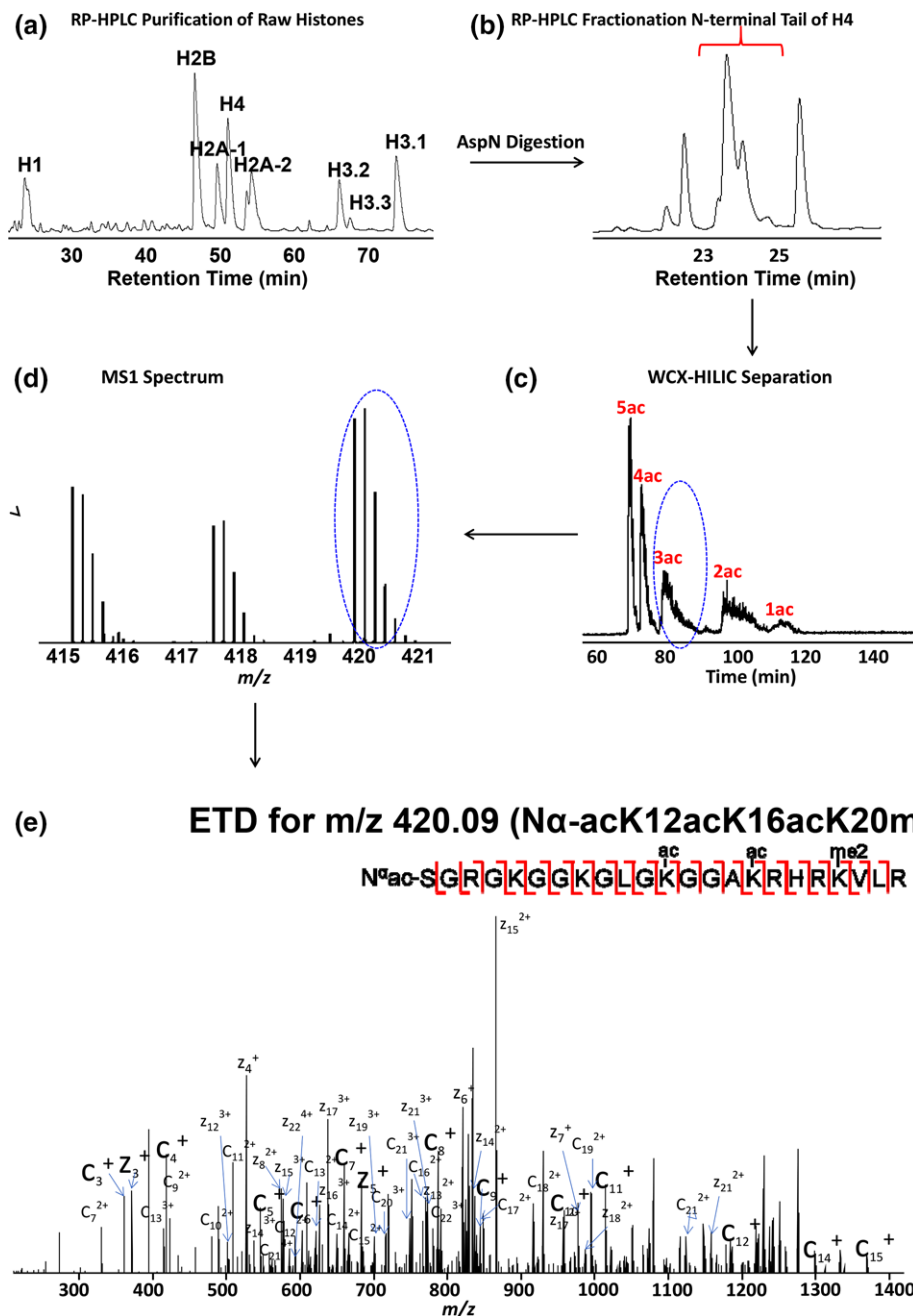
To effectively resolve histone H4 proteoforms, on-line pH gradient weak cation exchange-hydrophilic interaction liquid chromatography (WCX-HILIC) is used to separate the proteoforms of histone H4 based on net charge state and hydrophilicity. Due to the charge neutralization of the acetyl groups, proteoforms are primarily separated by the extent of acetylation. They are further separated by methylation state and location of both acetylations and methylations (**Figure 1C**). The eluent is introduced by electrospray ionization (ESI) into a front-end electron transfer dissociation (ETD) equipped FT-ICR mass spectrometer<sup>[41]</sup>. The precursor masses are determined by high resolution  $MS^1$  (**Figure 1D**). After isolation, ETD is used for tandem mass spectrometry and localization of PTMs (**Figure 1E**). As an example, we present the annotated tandem mass spectrum of the N-terminal histone H4 tail proteoform  $N\alpha$ -acK12acK16acK20me2 PTM (**Figure 1E**). The fragmentation map suggests that our ETD induces sufficiently extensive fragmentation to confidently detect and quantify the PTMs and proteoforms, especially when further constrained by high resolution  $MS^1$  precursor mass.<sup>[30,43]</sup> One of the biggest challenges for histone investigations is to identify and characterize the positional isomers, which exhibit identical masses and very similar chromatography. **Figure 1** shows separation based mainly on the degree of acetylation. However, both  $MS^1$  and  $MS^2$  results reveal that some of the isomerically modified peptides are eluted and isolated at different times. Thus, our separation technique does in fact increase the separation capability. For example, **Figure 1E** shows only the proteoform,

$N\alpha$ -acK12acK16acK20me2, from that ETD  $MS^2$  product ion mass spectrum; other positional isomers exist but are isolated and fragmented at different elution times and  $MS^2$  product ion mass spectra. The implemented 2D three-step separation technique greatly enhances our capability to effectively separate these positional isomers and reduces the complexity of data analysis. Notably, the ultrahigh mass accuracy and resolution FT-ICR MS permits us to distinguish unambiguously between acetylation and tri-methylation at the precursor mass level. These PTMs exhibit the same nominal mass and a mass difference ( $\Delta m = 0.036$  Da) that is challenging to distinguish for most mass spectrometers for larger peptides and higher charge states observed here.

#### 3.2. Bioinformatics and the Application to Explore the Dynamics of PTMs of Histone H4 during Breast Cancer Invasion

We employ this effective middle-down workflow to investigate changes in PTMs and proteoforms of histone H4 in two breast cancer cell lines during cell cycle progression. The cell lines include: MCF10A, a precancerous breast epithelial cell line that is often used as a surrogate of normal breast epithelial cells; and a highly invasive breast cancer cell line MDA-MB-231. For each cell line, thymidine and nocodazole are used to achieve an S phase and G2/M phase block. Asynchronous cells are also harvested, and represent a time-average of the cell cycle, but are substantially enriched for the G1 phase of the cell cycle. Two biological replicates of asynchronous cells and cells in S and G2/M phase for each cell line are subjected to 1–3 technical replicate runs depending on each sample amount. To examine the reproducibility, we plot a Pearson correlation matrix heat map for all sample runs. Color schemes are employed to show quantitative differences between data sets. As shown in **Figure 2**, cell lines and treatments mostly cluster together. Pearson correlations between technical replicates are generally about 0.9, indicating excellent technical reproducibility. One of the sample runs from nocodazole treatment in MDA-MB-231 cell line stands out as a mild outlier with a lower correlation to the other technical replicates, likely resulting from insufficient sample loading for the third technical replicate. We include it here as part of the complete data set. The ETD tandem mass spectra reveal that the acetylations occur on the N-terminus and at K5, K8, K12, and K16; phosphorylation at S1; monomethylation at R3; and un-, mono-, di-, and tri-methylation at K20, consistent with previous literature.

We performed hierarchical clustering of all proteoforms of histone H4 commonly observed in both cell lines (**Figure 3**). The quantitative data is expressed in the form of  $\log_{10}$  ratios. Our statistical analysis reveals distinct proteoform abundance differences between the MCF10A and MDA-MB-231 cell lines. Treatments with thymidine and nocodazole to achieve an S phase and G2/M phase blocks appear to not induce global proteoform disruption, but rather greatly influence a few proteoforms, such as proteoforms containing S1 phosphorylation. The differences between the two cell lines may result from differences in cellular and cancer phenotype; however, two cell lines are analyzed with two different FT-ICR mass spectrometers (14.5 T vs 21 T) on

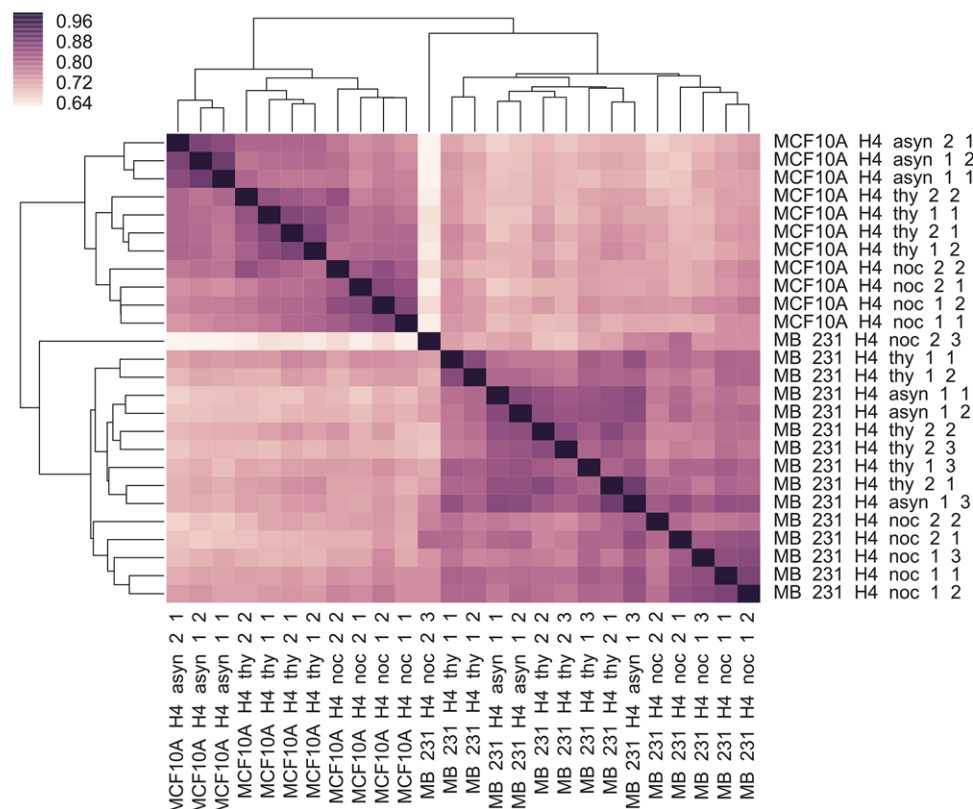


**Figure 1.** Overall experimental workflow for the H4  $N\alpha$ -acK12acK16acK20me2 proteoform. a) Reversed-phase HPLC separates histones into different families (H2A, H2B, H3.1, H3.2, H3.3, H4, and H1). b) The target 23 amino acid peptide of the N-terminal tail is further purified by RP-HPLC. c) Isoforms are primarily separated by extent of acetylation, and then by methylation state and location by use of WCX-HILIC. d) Representative broadband mass spectrum of the desired 23 amino acid peptide of histone H4 at a retention time of 82.36 min. e) Representative ETD product ion spectrum for precursor ion  $m/z$  420.09 with matched fragments marked.

different days and may simply reflect analytical differences, even though the two instruments are highly similar in design except for magnet field strength.

The proteoform data may be analyzed at many different levels, including the simplest which is discrete PTMs, without consideration of the presence of other PTMs (Table 1). Binary

combinations of PTMs may be queried in a similar manner. Such reduced representations dispose of valuable information but are extremely helpful in identifying patterns and relationships that may be explored with more detailed questions regarding how these PTMs function in concert.



**Figure 2.** Pearson correlation matrix heat map for cell lines MDA-MB-231 and MCF-10A at different cell cycle stages.

## 4. H4 S1 Phosphorylation

Phosphorylation at serine 1 of histone H4 has been found in a wide range of organisms, including worms, flies, and mammals, and is associated with chromatin condensation during mitosis.<sup>[44]</sup> As shown in **Figure 4A**, the abundance of S1 phosphorylation in both cell lines dramatically increases as the cell cycle traverses from S phase to G2/M phase, in agreement with previous studies,<sup>[44,45]</sup> and consistent with an important role for S1ph in mitotic chromatin condensation. In **Figure 4B**, all proteoform changes observed are represented as a volcano plot of  $-\log_{10}(p \text{ value})$  versus  $\log_{10}(\text{fold change})$ . Individual proteoforms that are above the statistical threshold of  $p < 0.05$  and change by more than fivefold are considered meaningful. Each point on this graph corresponds to a row in Table S1, Supporting Information.

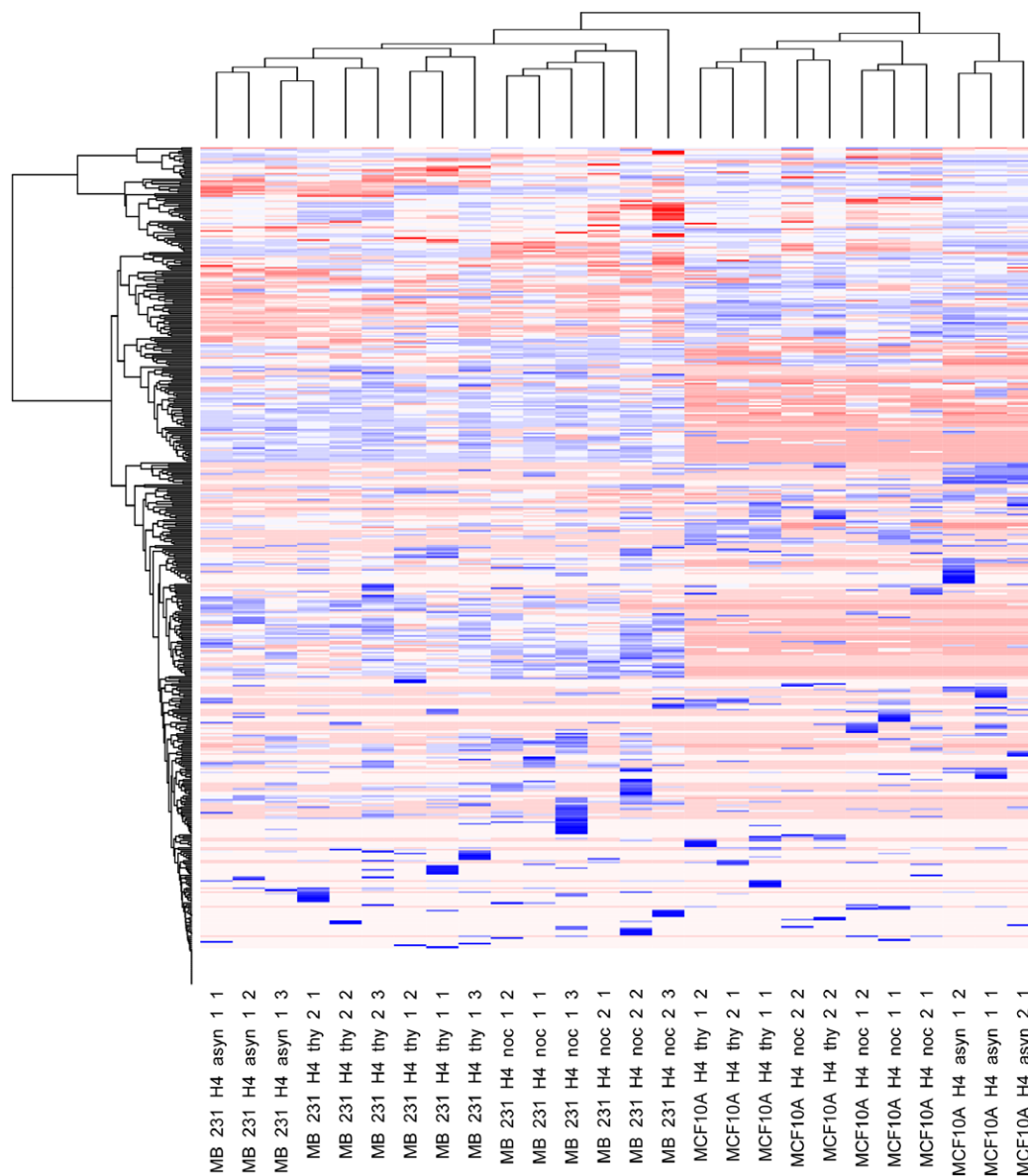
### 4.1. H4S1ph ± ac

With our middle-down analysis, we are also able to explore the correlation between S1ph and other PTMs occurring on the N-terminus. From the fold change bar graph in binary combinations of PTMs shown in **Figure 4C**, we observe that the relative abundances of S1ph-containing binary combinations mostly dramatically increase during cell cycle progression in both cell lines. However, the specific deviations from this increase are the most interesting. The moderate increases in S1ph are mostly

specific to lysine acetylation-containing proteoforms, with the exception of K16 in MDA-MB-231 cells. This behavior is most obvious in MCF10A cells, for which there is a dramatic shift from S1ph co-occurring with acetylation in S phase to co-occurring with unacetylation in G2/M. The increase in S1ph reaches an apex in G2/M and this increase is highly anticorrelated with Nε-lysine acetylation at all sites at the single molecule level, with the exception of K16ac only in MDA-MB-231 cells. MCF10A cells show strikingly opposite behavior for the binary combination of S1phK16ac. This single molecule combination is instead depleted relative to asynchronous cells (surrogate for G1) for both S and G2/M.

### 4.2. H4S1ph ± me

The relationship between S1ph and site of methylation is also interesting. Our results exhibit increases in S1phR3un, S1phK20me1, and S1phK20me2 and huge decrease for S1phK20un in both cell lines. Phosphorylation of S1 and the methylation of R3 site may be mutually impeded due to steric hindrance. The correlation between mono- and di-methylation of K20 and S1ph distinct from the unmodified state at K20 is interesting and may be due to localization of the kinase to regions enriched in these marks. K20me3 is characterized as heterochromatic and less accessible; however, the structural characteristics of K20un are not well understood.



**Figure 3.** Cross-correlation analysis between two breast cancer cell lines (MDA-MB-231 and MCF-10A) and the histone H4 PTM  $\log_{10}$ (abundance). The y-axis represents individual proteoforms, which are too numerous to list explicitly in the figure and are included in Table S1, Supporting Information.

## 5. H4 Acetylation

Acetylation is the most commonly observed modification in histone H4. We investigate the levels of universally conserved acetylation sites at each individual lysine residue of the N-terminal tail of histone H4 through the cell cycle. K16 is the most abundant  $N\epsilon$ -lysine acetylation site, followed in order of decreasing abundance by K12, K8, and K5. We observe a pronounced decline for acetylation at K16 and slight decrease for acetylation at K5 and K8 whereas K12ac remains almost the same (Figure S1, Supporting Information), consequently resulting in a steady increase in the abundances of K5un, K8un, and K16un during the cell cycle progression. Previous reports have

indicated a role for acetylation at K5 of histone H4 in histone deposition during chromatin replication and assembly in S phase.<sup>[46]</sup> The deacetylated form of K16 is pivotal to the formation of condensed chromatin during M phase.<sup>[47]</sup> Thus, our observed decrease of acetylation at these sites is consistent with previous findings.

### 5.1. H4 $N\alpha$ -ac

We find that more than 99.5% of cells in MCF10A cell line, and 95.9% in MDA-MB-231 cell line are acetylated on the N-terminus. This result contradicts the assumption in previous

**Table 1.** The percent abundance relative to the total histone H4 extracted from the cells for all discrete PTMs considered in our analysis with error expressed as standard deviation. Measurements are made on combinatorially modified peptides and these “discrete” PTM numbers are derived by summing those peptides containing each PTM while ignoring, or marginalizing, other PTMs. Note the large range of basal abundances of PTMs and that cell cycle dependence occurs for both abundant and scarce PTMs. Thus, fold change does not accurately represent the significance of all changes and absolute change is sometimes critical to understanding the function of PTMs.

PTM	MCF10A Asyn/G1		MCF10A Thy/S		MCF10A Noc/M		MB231 Asyn/G1		MB231 Thy/S		MB231 Noc/M	
	Average [%]	STDEV	Average [%]	STDEV	Average [%]	STDEV	Average [%]	STDEV	Average [%]	STDEV	Average [%]	STDEV
N $\alpha$ -ac	99.57	$\pm 0.02$	99.69	$\pm 0.08$	99.48	$\pm 0.36$	99.89	$\pm 0.03$	95.87	$\pm 5.35$	98.44	$\pm 1.05$
N $\alpha$ -un	0.43	$\pm 0.02$	0.31	$\pm 0.08$	0.52	$\pm 0.36$	0.11	$\pm 0.03$	4.14	$\pm 5.35$	1.56	$\pm 1.05$
S1ph	0.27	$\pm 0.02$	0.34	$\pm 0.24$	0.85	$\pm 0.36$	0.09	$\pm 0.08$	0.14	$\pm 0.09$	0.50	$\pm 0.22$
S1un	99.73	$\pm 0.02$	99.66	$\pm 0.24$	99.15	$\pm 0.36$	99.91	$\pm 0.08$	99.86	$\pm 0.09$	99.50	$\pm 0.23$
R3me1	0.86	$\pm 0.10$	0.98	$\pm 0.21$	1.08	$\pm 0.16$	0.57	$\pm 0.05$	0.59	$\pm 0.14$	0.85	$\pm 0.23$
R3un	99.14	$\pm 0.10$	99.02	$\pm 0.21$	98.92	$\pm 0.16$	99.44	$\pm 0.05$	99.41	$\pm 0.14$	99.15	$\pm 0.23$
K5ac	4.38	$\pm 0.27$	3.50	$\pm 0.44$	2.74	$\pm 0.34$	2.05	$\pm 0.12$	2.25	$\pm 0.87$	1.90	$\pm 0.53$
K5un	95.62	$\pm 0.27$	96.50	$\pm 0.44$	97.26	$\pm 0.34$	97.95	$\pm 0.12$	97.75	$\pm 0.87$	98.10	$\pm 0.53$
K8ac	4.34	$\pm 0.35$	4.20	$\pm 0.63$	2.80	$\pm 0.19$	3.38	$\pm 0.42$	4.09	$\pm 1.10$	2.20	$\pm 0.54$
K8un	95.66	$\pm 0.35$	95.80	$\pm 0.63$	97.20	$\pm 0.19$	96.62	$\pm 0.42$	95.91	$\pm 1.10$	97.80	$\pm 0.54$
K12ac	9.67	$\pm 1.02$	9.70	$\pm 1.02$	9.55	$\pm 0.74$	6.30	$\pm 0.32$	6.63	$\pm 2.87$	5.19	$\pm 1.94$
K12un	90.33	$\pm 1.02$	90.30	$\pm 1.02$	90.45	$\pm 0.74$	93.70	$\pm 0.32$	93.37	$\pm 2.87$	94.81	$\pm 1.94$
K16ac	33.94	$\pm 2.30$	29.17	$\pm 0.59$	19.89	$\pm 1.90$	41.03	$\pm 2.81$	37.53	$\pm 5.16$	33.77	$\pm 6.81$
K16un	66.06	$\pm 2.30$	70.83	$\pm 0.59$	80.11	$\pm 1.90$	58.97	$\pm 2.81$	62.47	$\pm 5.16$	66.23	$\pm 6.81$
K20me1	16.42	$\pm 0.48$	4.53	$\pm 0.48$	5.90	$\pm 0.27$	13.87	$\pm 0.80$	6.65	$\pm 0.57$	11.48	$\pm 0.91$
K20me2	68.40	$\pm 0.89$	84.17	$\pm 0.61$	92.58	$\pm 0.53$	72.33	$\pm 1.95$	82.37	$\pm 1.50$	86.67	$\pm 1.27$
K20me3	0.48	$\pm 0.08$	1.64	$\pm 0.37$	0.58	$\pm 0.17$	0.06	$\pm 0.02$	0.21	$\pm 0.09$	0.17	$\pm 0.10$
K20un	14.71	$\pm 1.23$	9.66	$\pm 0.72$	0.95	$\pm 0.12$	13.74	$\pm 1.40$	10.77	$\pm 1.72$	1.67	$\pm 0.65$

studies that histone H4 is constitutively N-terminally acetylated as a co-translational event<sup>[45]</sup> with 100% of histone H4 being N-terminally acetylated. Instead, we reproducibly observe a small but measurable amount of not N-terminally acetylated histone H4 and this effect varies across the cell cycle. This observation further demonstrates the precision and effectiveness of our implemented middle-down workflow.

## 5.2. H4ac $\pm$ ac

We also investigated the change in abundance of binary combinations of PTMs containing two acetylations. Figure S2, Supporting Information exhibits a universal decrease for all diacetylated marks for both cell lines. A similar trend is observed for proteoforms with three acetylations and four acetylations.

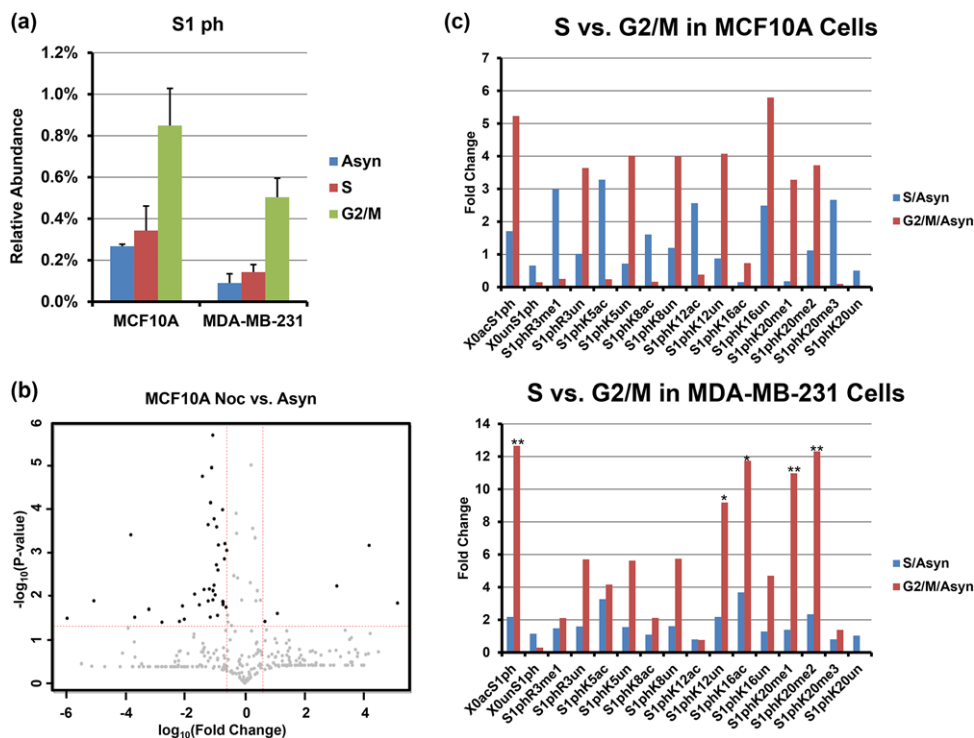
## 6. H4 Methylation

Both K20 methylation and R3 methylation are investigated here. Both of these events draw the methylation from the same physiological source, S-adenosylmethionine; however, the biological function of lysine acetylation and arginine methylation are not strongly similar.

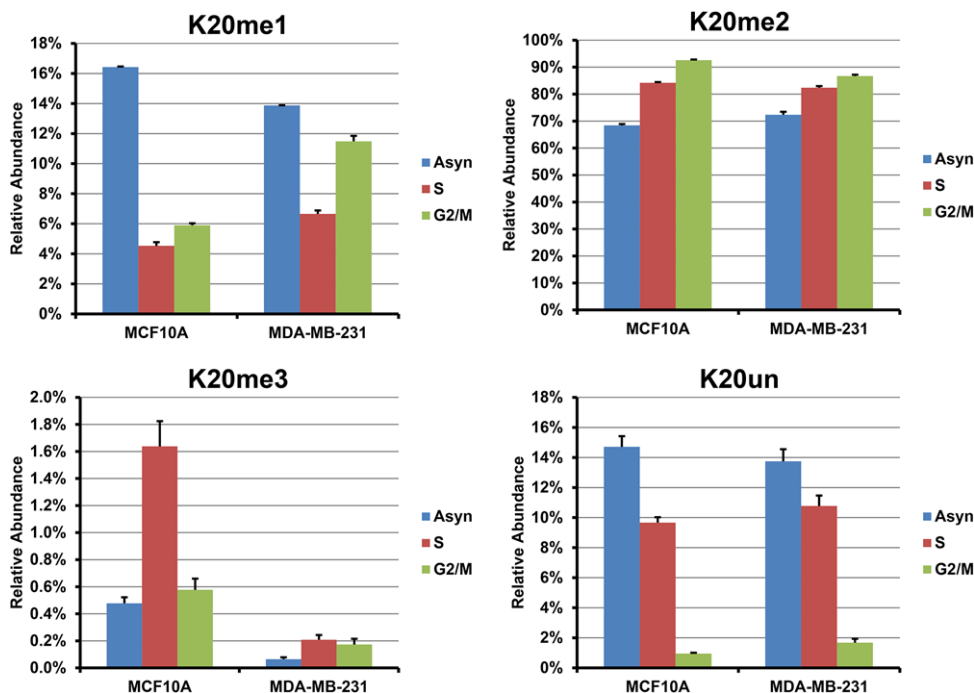
### 6.1. K20 Methylation

K20 di-methylation is the predominant methylation state at K20, accounting for over 70% of all histone H4 in both cell lines, followed by K20 mono-methylated, unmodified, and K20 tri-methylated proteoforms. Histone H4 K20 methylation is involved in mitosis and other chromatin-regulated processes, including transcriptional activation, gene silencing, and DNA repair.<sup>[48–50]</sup> As shown in **Figure 5**, the relative abundances of both K20me2 and K20me1 in both cell lines increase between S phase to mitosis, resulting in a substantial decrease of K20 unmodified proteoforms and not an increase in K20me3. Also, we observe that N $\alpha$ -acK20un, K5acK20un, K8acK20un, K12acK20un, and K16acK20un all decrease substantially, consistent with the overall decrease of K20 unmodified proteoforms. This observation suggests that K20un proteoforms play a role in histone replication and deposition processes in S phase and then they are extensively converted to K20me1 and K20me2 as the cell cycle progresses to G2/M phase. In contrast to the upregulation in K20me2 and K20me1, we notice a decrease for K20 tri-methylation as cells advance through the cell cycle, known to be mostly located in heterochromatic region and participating in modulating transcriptional silencing. And this decrease is more obvious in the MCF10A cell line and we find that the decline of K20me3 is mainly associated with other unmodified acetylation sites, such as K5unK20me3, K8unK20me3, K12unK20me3, and





**Figure 4.** a) Relative abundances of total phosphorylation on serine 1 observed from cell lines MCF-10A and MDA-MB-231 at different cell cycle stages. Error bars represent standard error (S.E.) from two biological replicates in S and G2/M phase for each cell line, two biological replicates of asynchronous cells for cell line MDA-MB-231, and one biological replicate for cell line MCF-10A. Each biological replicate contains one to three technical replicates depending on sample amount. b) Volcano plot of proteoforms containing S1ph comparison between mitosis and asynchronous cells from cell line MCF-10A. c) Fold change of binary combinations of PTMs containing S1ph and another PTM relative to asynchronous cells. Top: MCF10A; Bottom: MB-231. Asterisks denote  $p$  values  $< 0.05$  as determined by Student's  $t$ -test.



**Figure 5.** Relative abundances of mono-, di-, and tri-methylation and unmodified proteoforms on lysine 20 observed from cell lines MCF-10A and MDA-MB-231 at different cell cycle stages. Error bars represent S.E. from two biological replicates in S and G2/M phase for each cell line, two biological replicates of asynchronous cells for cell line MDA-MB-231, and one biological replicate for cell line MCF-10A. Each biological replicate contains one to three technical replicates depending on sample amount.

K16unK20me3 (Figure S3, Supporting Information). This result is consistent with the changes in the heterochromatic regions, because reduced acetylation is also associated with heterochromatin. Note that we do observe K20me3 in acetylated histone H4; however, such proteoforms do not decrease during the cell cycle.

## 6.2. R3 Methylation

In addition to K20, methylation at R3 is also interrogated. The overall relative abundances of R3 mono-methylation and unmodified forms remain at a similar level during the cell cycle for both cell lines, suggesting that the regulation of methylation at R3 is cell cycle independent. We do not observe any di-methylation at this site. To explore the relationship between R3 and K20 methylation, we observe the presence of R3me1K20me1, R3me1K20me2, R3me1K20me3, and R3me1K20un, in contrast to a previous report that methylation of H4R3 is observed only in the presence of H4K20 di-methylation.<sup>[29]</sup> The discrepancy further suggests that our method is more advantageous and sensitive in detecting low-abundance proteoforms and combinations, given that the overall relative abundance of R3me1 is less than 1% and the relative abundances of R3me1K20me1, R3me1K20un, and R3me1K20me3 are orders of magnitude less and can be easily overlooked by other methods. Relative to the fluctuations in abundances of methylation at K20 through the cell cycle, we postulate that there is no direct interdependence between methylations at these two different sites.

## 7. Concluding Remarks

Here we present a robust and effective three-step 2D RP/WCX-HILIC-MS/MS workflow to investigate histone H4 at both the discrete PTMs and proteoform levels. We demonstrate the advantages of this method in separating isobaric and isomeric proteoforms, thereby greatly alleviating the complexity of spectra and the downstream data analysis to increase our quantitation confidence. Moreover, this optimized method requires approximately 1  $\mu$ g sample, which, given the relative abundance of histone H4, makes many biological studies feasible. The study of two breast cancer cell lines through the cell cycle presented here reveals distinct epigenetic patterns in models of different progression states of breast cancer. At the PTM level, our results show dramatic increase for S1 phosphorylation, K20 mono- and di-methylation and downregulation for K20 unmodification and acetylation at K5, K8, K12, and K16 for both cell lines between S phase and G2/M phase, consistent with previous studies. N-acetylation and R3 mono-methylation remain almost the same. At the proteoform level, we analyze the fluctuation of combinatorial PTMs for both cell lines. In total, 233 proteoforms are reproducibly observed and quantitated for both cell lines. An additional 202 proteoforms were observed for a total of 435; however, these additional 202 proteoforms were not reproducibly detected and are considered below our limit of quantitation. Proteoforms that contain the “binary combinations” (note that other sites may or may not be

modified): S1phK5un, S1phK8un, S1phK12un, and S1phK16un show clear synergistic effect during cell cycle progression, whereas R3 methylation and K20 methylation status appear uncoordinated. To further understand the mechanism of breast carcinogenesis, an extension of the research to combinatorial histone codes across different histone family members in each nucleosome or across different nucleosomes would be necessary. Additional follow-up with other molecular biology techniques, with these quantitative and single molecule specific results in mind, could provide complementary information to advance our understanding of the complexity of epigenetic regulations during the cell cycle of breast cancer cells.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgments

This work was supported by NSF Division of Materials Research through DMR 11–57490, and the State of Florida. This work was supported by NIH awards CA016058, OD018056 and support from the Ohio State University. Raw mass spectrometry data was deposited into ProteomeXchange via the PRIDE partner repository with the dataset identifier PXD008296.

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

epigenetics, ETD, FT-ICR, mass spectrometry, PTM, WCX-HILIC

Received: November 27, 2017

Revised: February 28, 2018

Published online:

- [1] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251.
- [2] R. D. Kornberg, Y. L. Lorch, *Cell* **1999**, *98*, 285.
- [3] N. L. Young, P. A. Dimaggio, B. A. Garcia, *Cell. Mol. Life Sci.* **2010**, *67*, 3983.
- [4] A. J. Bannister, T. Kouzarides, *Cell Res.* **2011**, *21*, 381.
- [5] C. Bonisch, S. M. Nieratschker, N. K. Orfanos, S. B. Hake, *Expert Rev. Proteomics* **2008**, *5*, 105.
- [6] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41.
- [7] M. F. Dion, S. J. Altschuler, L. F. Wu, O. J. Rando, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5501.
- [8] C. L. Liu, T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, N. Friedman, O. J. Rando, *PLOS Biol.* **2005**, *3*, 1753.
- [9] S. K. Kurdastani, S. Tavazoie, M. Grunstein, *Cell* **2004**, *117*, 721.
- [10] P. Chi, C. D. Allis, G. G. Wang, *Nat. Rev. Cancer* **2010**, *10*, 457.
- [11] B. Li, M. Gogol, M. Carey, D. Lee, C. Seidel, J. L. Workman, *Science* **2007**, *316*, 1050.

- [12] A. J. Ruthenburg, H. Li, D. J. Patel, C. D. Allis, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 983.
- [13] P. Trojer, G. Li, R. J. Sims, A. Vaquero, N. Kalakonda, P. Bocconi, D. Lee, H. Erdjument-Bromage, P. Tempst, S. D. Nimer, Y. H. Wang, D. Reinberg, *Cell* **2007**, *129*, 915.
- [14] J. Moriniere, S. Rousseaux, U. Steuerwald, M. Soler-Lopez, S. Curtet, A. L. Vitte, J. Govin, J. Gaucher, K. Sadoul, D. J. Hart, J. Krijgsveld, S. Khochbin, C. W. Muller, C. Petosa, *Nature* **2009**, *461*, 664.
- [15] N. Saksouk, N. Avvakumov, K. S. Champagne, T. Hung, Y. Doyon, C. Cayrou, E. Paquet, M. Ullah, A. J. Landry, V. Cote, X. J. Yang, O. Gozani, T. G. Kutateladze, J. Cote, *Mol. Cell* **2009**, *33*, 257.
- [16] M. Yun, J. Wu, J. L. Workman, B. Li, *Cell Res.* **2011**, *21*, 564.
- [17] L. M. Smith, N. L. Kelleher, C. T. D. Proteomics, *Nat. Methods* **2013**, *10*, 186.
- [18] A. G. Marshall, C. L. Hendrickson, G. S. Jackson, *Mass Spectrom. Rev.* **1998**, *17*, 1.
- [19] N. K. Kaiser, J. P. Quinn, G. T. Blakney, C. L. Hendrickson, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1343.
- [20] T. M. Schaub, C. L. Hendrickson, S. Horning, J. P. Quinn, M. W. Senko, A. G. Marshall, *Anal. Chem.* **2008**, *80*, 3985.
- [21] C. L. Hendrickson, J. P. Quinn, N. K. Kaiser, D. F. Smith, G. T. Blakney, T. Chen, A. G. Marshall, C. R. Weisbrod, S. C. Beu, *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1626.
- [22] S. M. Fuchs, K. Krajewski, R. W. Baker, V. L. Miller, B. D. Strahl, *Curr. Biol.* **2011**, *21*, 53.
- [23] N. L. Kelleher, *Anal. Chem.* **2004**, *76*, 196a.
- [24] J. J. Pesavento, C. A. Mizzen, N. L. Kelleher, *Anal. Chem.* **2006**, *78*, 4271.
- [25] J. J. Pesavento, C. R. Bullock, R. D. LeDuc, C. A. Mizzen, N. L. Kelleher, *J. Biol. Chem.* **2008**, *283*, 14927.
- [26] Z. Tian, N. Tolic, R. Zhao, R. J. Moore, S. M. Hengel, E. W. Robinson, D. L. Stenoien, S. Wu, R. D. Smith, L. Pasa-Tolic, *Genome Biol.* **2012**, *13*, R86.
- [27] X. Dang, J. Scotcher, S. Wu, R. K. Chu, N. Tolic, I. Ntai, P. M. Thomas, R. T. Fellers, B. P. Early, Y. Zheng, K. R. Durbin, R. D. Leduc, J. J. Wolff, C. J. Thompson, J. Pan, J. Han, J. B. Shaw, J. P. Salisbury, M. East-erling, C. H. Borchers, J. S. Brodbelt, J. N. Agar, L. Pasa-Tolic, N. L. Kelleher, N. L. Young, *Proteomics* **2014**, *14*, 1130.
- [28] M. Rea, T. Jiang, R. Eleazer, M. Eckstein, A. G. Marshall, Y. N. Fondufe-Mittendorf, *Mol. Cell. Proteomics* **2016**, *15*, 2411.
- [29] D. Phanstiel, J. Brumbaugh, W. T. Berggren, K. Conard, X. Feng, M. E. Levenstein, G. C. McAlister, J. A. Thomson, J. J. Coon, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4093.
- [30] N. L. Young, P. A. DiMaggio, M. D. Plazas-Mayorca, R. C. Baliban, C. A. Floudas, B. A. Garcia, *Mol. Cell. Proteomics* **2009**, *8*, 2266.
- [31] A. Moradian, A. Kalli, M. J. Sweredoski, S. Hess, *Proteomics* **2014**, *14*, 489.
- [32] S. Sidoli, V. Schwammle, C. Ruminowicz, T. A. Hansen, X. Wu, K. Helin, O. N. Jensen, *Proteomics* **2014**, *14*, 2200.
- [33] P. K. Lo, S. Sukumar, *Pharmacogenomics* **2008**, *9*, 1879.
- [34] S. Sharma, T. K. Kelly, P. A. Jones, *Carcinogenesis* **2010**, *31*, 27.
- [35] S. W. Harshman, M. E. Hoover, C. S. Huang, O. E. Branson, S. B. Chaney, C. M. Cheney, T. J. Rosol, C. L. Shapiro, V. H. Wysocki, K. Huebner, M. A. Freitas, *J. Proteome Res.* **2014**, *13*, 2453.
- [36] Y. Chen, M. E. Hoover, X. Dang, A. A. Shomo, X. Guan, A. G. Marshall, M. A. Freitas, N. L. Young, *Mol. Cell. Proteomics* **2016**, *15*, 818.
- [37] J. Brustel, M. Tardat, O. Kirsh, C. Grimaud, E. Julien, *Trends Cell Biol.* **2011**, *21*, 452.
- [38] M. J. Kruhlak, M. J. Hendzel, W. Fischle, N. R. Bertos, S. Hameed, X. J. Yang, E. Verdin, D. P. Bazett-Jones, *J. Biol. Chem.* **2001**, *276*, 38307.
- [39] T. Banerjee, D. Chakravarti, *Mol. Cell. Biol.* **2011**, *31*, 4858.
- [40] S. Maddika, S. R. Ande, S. Panigrahi, T. Paranjothy, K. Weglarczyk, A. Zuse, M. Eshraghi, K. D. Manda, E. Wiechec, M. Los, *Drug Resist. Updat.* **2007**, *10*, 13.
- [41] C. R. Weisbrod, N. K. Kaiser, J. E. P. Syka, L. Early, C. Mullen, J. J. Dunny-ach, A. M. English, L. C. Anderson, G. T. Blakney, J. Shabanowitz, C. L. Hendrickson, A. G. Marshall, D. F. Hunt, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1787.
- [42] E. B. Ledford, D. L. Rempel, M. L. Gross, *Anal. Chem.* **1984**, *56*, 2744.
- [43] P. A. DiMaggio, Jr., N. L. Young, R. C. Baliban, B. A. Garcia, C. A. Floudas, *Mol. Cell. Proteomics* **2009**, *8*, 2527.
- [44] C. M. Barber, F. B. Turner, Y. Wang, K. Hagstrom, S. D. Taverna, S. Mollah, B. Ueberheide, B. J. Meyer, D. F. Hunt, P. Cheung, C. D. Allis, *Chromosoma* **2004**, *112*, 360.
- [45] K. Yamamoto, Y. Chikaoka, G. Hayashi, R. Sakamoto, R. Yamamoto, A. Sugiyama, T. Kodama, A. Okamoto, T. Kawamura, *Mass Spectrom. (Tokyo)* **2015**, *4*, A0039.
- [46] M. A. Koprinarova, G. C. Russev, *Cell Cycle* **2008**, *7*, 414.
- [47] A. Vaquero, M. B. Scher, D. H. Lee, A. Sutton, H. L. Cheng, F. W. Alt, L. Serrano, R. Sternglanz, D. Reinberg, *Genes Dev.* **2006**, *20*, 1256.
- [48] T. Doerks, R. R. Copley, J. Schultz, C. P. Ponting, P. Bork, *Genome Res.* **2002**, *12*, 47.
- [49] J. J. Pesavento, H. Yang, N. L. Kelleher, C. A. Mizzen, *Mol. Cell. Biol.* **2008**, *28*, 468.
- [50] S. Jorgensen, G. Schotta, C. S. Sorensen, *Nucleic Acids Res.* **2013**, *41*, 2797.