HHS Public Access

Author manuscript

J Proteome Res. Author manuscript; available in PMC 2019 February 24.

Published in final edited form as:

J Proteome Res. 2018 July 06; 17(7): 2533–2541. doi:10.1021/acs.jproteome.8b00133.

EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data

Zuo-Fei Yuan[1], **Simone Sidoli**[1], **Dylan M. Marchione**[2], **Johayra Simithy**[1], **Kevin A. Janssen**[1], **Mary R. Szurgot**[1], and **Benjamin A. Garcia**[1,*]

[1]Epigenetics Institute, Department of Biochemistry and Biophysics, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA 19104, USA.

[2]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA 19104, USA.

Abstract

Epigenetics has become a fundamental scientific discipline with various implications for biology and medicine. Epigenetic marks, mostly DNA methylation and histone post-translational modifications (PTMs), play important roles in chromatin structure and function. Accurate quantification of these marks is an ongoing challenge due to the variety of modifications and their wide dynamic range of abundance. Here, we present EpiProfile 2.0, an extended version of our 2015 software (v1.0) for accurate quantification of histone peptides based on liquid chromatography – tandem mass spectrometry (LC-MS/MS) analysis. EpiProfile 2.0 is now optimized for data-independent acquisition through the use of precursor and fragment extracted ion chromatography to accurately determine the chromatographic profile and to discriminate isobaric forms of peptides. The software uses an intelligent retention time prediction trained on the analyzed samples to enable accurate peak detection. EpiProfile 2.0 supports label-free and isotopic labeling, different organisms, known sequence mutations in diseases, different derivatization strategies, and unusual PTMs (such as acyl-derived modifications). In summary, EpiProfile 2.0 is a universal and accurate platform for the quantification of histone marks via LC-MS/MS. Being the first software of its kind we anticipate that EpiProfile 2.0 will play a fundamental role in epigenetic studies relevant to biology and translational medicine. EpiProfile is freely available at https://github.com/zfyuan/EpiProfile2.0_Family.

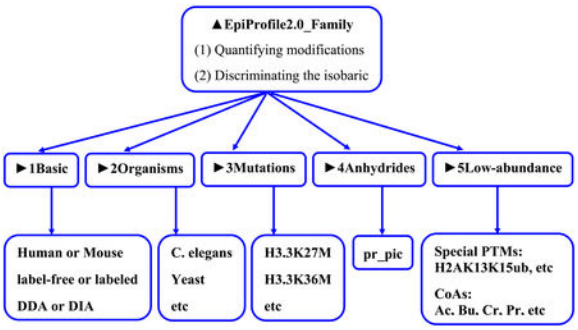
Graphical Abstract

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

Supporting Information

The following supporting information is available free of charge at ACS website http://pubs.acs.org.

MS files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD004166. EpiProfile 2.0 (the Matlab code, user manual, and export annotation) is freely available via GitHub at https://github.com/zfyuan/EpiProfile2.0_Family. For operating it requires Matlab and Xcalibur.



Keywords

epigenetics; histone; post-translational modification; mutations; acylation; quantification

Introduction

Histones are proteins wrapped by DNA to form the structure of chromatin. Histone post-translational modifications (PTMs) play important roles in chromatin function, such as regulating transcription, DNA repair, DNA replication, and chromosome condensation[1]. Analysis of histone PTMs is an ongoing challenge as antibody based methods are usually biased towards the complexity of epitope recognition due to the large variety and density of PTMs on histone sequences[2, 3]. Mass spectrometry (MS) has become the method of choice for large-scale analysis of global PTMs[4], especially on histones (i.e. Epi-Proteomics[5]). The bottom-up MS approach, where proteins are digested into short peptides, is the most widely used method, as it allows for more efficient chromatographic separation and MS detection. Methods like top-down (intact protein analysis) and middle-down (analysis of intact N-terminal tails) MS are also used for histone analysis, but more for specific applications such as analysis of combinatorial PTM patterns[6, 7]. Bottom-up MS poses a particular challenge for histone analysis because of the numerous isobaric forms of histone peptides, i.e. differently modified sequences that have the same precursor mass. One example is mono-acetylation on either K9 or K14 of the histone H3 peptide $K_9STGGK_{14}APR$ (aa 9−17). As these mono-acetylated peptides often co-elute, tandem mass spectrometry (MS2) is needed to discriminate their relative abundance.

Traditionally, data-dependent acquisition (DDA) is the most used MS acquisition method used to gather MS2, which selects peptides by their intensity acquired during the full scan (MS1)[8]. The drawback of this method is that the selection of peptides depends on the instrument duty cycle, so that low abundant species might be excluded for MS2 fragmentation in slower scanning instruments. Moreover, when dynamic exclusion is enabled isobaric peptides are selected for one or few scans, making it impossible to draw the entire fragment profiles to discriminate isobaric forms. In a DDA experiment, it is possible to include targeted scans that repeat at every duty cycle, which would enable more accurate MS2 extracted ion chromatograms (XICs). However, this precludes the possibility to re-mine datasets for a specific isobaric form if this was not considered during the acquisition. Data-independent acquisition (DIA) is an effective alternative wherein setting MS2 targets is

unnecessary, as the entire mass range is fragmented at every instrument duty cycle[9–11]. Although specific software can deconvolve peptide identifications from DIA runs[12, 13], this acquisition method is currently mostly suitable for quantification of previously identified peptides. Quantification is supposedly more accurate than by using DDA datasets, as the proper chromatographic peak is selected with more specificity due to the possibility to integrate both precursor and fragment ion profiles. DIA is thus gaining interest in histone peptide analysis[14, 15], as most histone PTM studies aim to accurately quantify PTMs rather than identifying new ones. Histone DIA has reached impressive levels of accuracy, so that it can be currently performed just by acquiring runs in a simple low resolution ion trap[16].

Accurate quantification of histone PTMs has been conducted by many research groups for over a decade[17–26]. Each group has faced the same challenge of determining correct peptide retention time and discriminating isobaric peptides. These problems have been partially solved in different ways. For example, one study[23] discriminated isobaric peptides on H3 and H4 by manual calculation of fragment intensities exported from Skyline[27] or Spectronaut. Another study[24] discriminated isobaric peptides only on H3 by developing a new software named Fishtones. Finally, one more study[25] discriminated isobaric peptides on H3 and H4 by deconvolution of isomer mass spectra from synthetic peptides using an in-house software named Iso-PeptidAce. We present EpiProfile 2.0 as a more robust and universal software tool for a wide variety of experimental designs. EpiProfile 2.0 can quantify histone data obtained from different sample preparations (e.g. label-free or isotopic labeling for both protein sequences and PTMs), different organisms, different histone mutations[28, 29], different derivatization strategies, or histones modified by low-abundance PTMs such as acylations[30]. This software is an upgrade from our previous version (EpiProfile 1.0[31]), which could only process high resolution DDA datasets with a very limited number of experimental options. To our knowledge, EpiProfile 2.0 is the most flexible software for histone analysis. It differentiates from widespread software for DIA data analysis due to its ability to accurately discriminate highly modified histone peptides.

Materials and methods

Materials and methods include the description of preparation of samples subjected to LC-MS/MS and EpiProfile analysis. Later, we describe the principles and workflow of EpiProfile 2.0.

Cell culture

As previously described[32], HeLa cells were cultured in suspension with Minimum Essential Medium Eagle (MEM) Joklik modification for Suspension Cultures with 10% Newborn Calf Serum (Thermo Scientific SH30118.03), 1% Glutamax (Gibco, 35050–079), and 1% Penicillin/Streptomycin (Gibco, 15140–122). The cell density was maintained within the range of $1–10\times10^5$ cells/mL.

To obtain histones with mutated sequences (namely “oncohistones”), C8-D1A primary murine astrocytes transduced with a lentivirus encoding H3.3 WT-FLAG-HA or H3.3 K27M-FLAG-HA were cultured in Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% Newborn Calf Serum (Thermo Scientific SH30118.03) and 1

μg/mL puromycin. Adherent cells were maintained on 100 × 20 mm tissue culture dishes (Sarstedt 83.3902) and serially passaged 1:5 by gentle trypsinization every 3–4 days. Cells were harvested by gentle scraping and washed twice with PBS before histone extraction.

Histone extraction and digestion

Histone purification and analysis was performed as previously described[18, 33]. Briefly, histones were acid-extracted from nuclei with 0.2 M H_2SO_4 for 2 hours and precipitated with 33% trichloroacetic acid (TCA) for 1 hour. Purified histones were dissolved in 30 μL of 50 mM NH_4HCO_3 (pH8.0). Derivatization reagent was prepared by mixing propionic anhydride with acetonitrile in a ratio of 1:4 (v/v), and was added to the histone sample in 1:2 (v/v) ratio for 15 minutes at 37°C. This reaction was performed twice. Histones were then digested with trypsin (enzyme to sample ratio of 1:20) in 50 mM NH_4HCO_3 overnight at room temperature. After digestion, the derivatization reaction was performed again twice to cap peptide N-termini. Samples were desalted prior LC-MS/MS analysis by using C_{18} Stage-tips.

NanoLC-MS/MS

Samples were analyzed using a nanoLC-MS/MS setup. 1 μg of sample was loaded into an in-house packed 75 μm ID x 20 cm Reprosil-Pur C18-AQ (3 μm; Dr. Maisch GmbH, Germany) nano-column using an EASY-nLC nano-HPLC (Thermo Scientific, San Jose, CA, USA). The HPLC gradient was as follows: 0% to 26% solvent B (A = 0.1% formic acid; B = 95% acetonitrile, 0.1% formic acid) over 45 minutes, from 26% to 80% solvent B in 5 minutes, 80% B for 10 minutes at a flow-rate of 300 nL/min. nLC was coupled online with an Orbitrap Elite MS (Thermo Scientific, San Jose, CA, USA). Runs were acquired using DIA as previously described[16]. Briefly, one full scan MS spectrum (m/z 300−1100) was acquired in the Orbitrap at a resolution of 120,000 (at 200 m/z FWHM) followed by 16 MS/MS events spanning through the mass range, each acquired in the ion trap with an isolation window of 50 m/z. Fragmentation was performed using collision induced dissociation (CID) set at 35%. For the DDA mode, dynamic exclusion was enabled (repeat count: 1, exclusion duration: 0.5 min). One full MS scan (m/z 300 to 1100) was collected every cycle followed by 10 MS/MS scans using CID. From 15 to 41 min, five masses were set as targets (528.30, 570.84, 754.93, 761.94, and 768.95). The isolation window was set at 3.0 m/z. Ions with a charge state of one and a rejection list of common contaminant ions (including keratin, trypsin and BSA) (exclusion width=10 ppm) were excluded from MS/MS spectra.

In vitro histone acylation assay

For the analysis of histone acylation (e.g. crotonylation), we incubated histone H3 with histone acetyltransferase (HAT) p300 and crotonylation[30]. *In vitro* enzymatic assays were carried out by incubating 0.5 μg of p300 with 10 μg of recombinant histone H3 in the presence of 0.5 mM crotonyl-CoA (Sigma-Aldrich) in 1X HAT buffer (25 mM pH 8 Tris-HCl, 25 mM KCl, 1mM DTT, 0.1mM AEBSF and 5 mM sodium butyrate) for 60 min at 30°C; final volume was 50 μl. Histones were derivatized with propionic anhydride, desalted and injected into an online nanoLC-MS/MS as described above.

Principle of operation of EpiProfile 2.0

Currently, EpiProfile 2.0 supports only Thermo Scientific RAW data as they are converted into open formats using the conversion tool pXtract (pfind.ict.ac.cn). However, all vendor formats are potentially suitable. In the future, we will try ProteoWizard (proteowizard.sourceforge.net) to transform different vendor raw files into MS1 and MS2 files. EpiProfile automatically determines the MS acquisition method (DIA or DDA), fragmentation (CID or HCD), and resolution of the full MS scan. Before the analysis, EpiProfile extracts the retention time (RT) of all unmodified peptides from the analyzed runs. The assumption is that all runs analyzed together will have consistent retention times for specific peptides, as the column performance is usually stable over several days. Using the unmodified peptides of histone H3 3−8, 9−17, 18−26, and 27−40, EpiProfile 2.0 makes serial decisions: 1) which run has the greatest summed intensity of these four peptides, and in this selected run which of these four peptides is the most abundant (i.e. selecting a confident reference peptide), 2) which runs are within a default RT window of ± 2 minutes from the RT of this reference peptide (i.e. setting the RT window to exclude outliers), 3) among the runs in the RT window, which run has the median RT for the reference peptide, 4) among the runs in the RT window in which the reference peptide eluted earlier than the median RT, which run has the most abundant reference peptide, and 5) among the runs in the RT window in which the reference peptide eluted later than the median RT, which run has the most abundant reference peptide. Based on the three selected runs, all the retention times are extracted for 77 unmodified peptides from H3, H4, H1, H2A, and H2B, and the median of retention times for each unmodified peptide is set as the reference in the future peak selection.

Next, EpiProfile performs XICs for each peptide group (i.e. the same peptide sequence with different modification forms). For example, histone H3 peptide TKQTAR (aa 3−8) has five forms (i.e. unmodified (un), me1, me2, me3, and ac). The kernel workflow is as follows. (1) Peptide information is obtained (e.g. peptide sequence, modification, m/z of each charge state). (2) The XICs of the unmodified form with one charge state (the most abundant one between all charge states) are obtained, the correct XIC for the unmodified is determined by the reference retention time for the unmodified as described above, and the area under curve (AUC) for all charge states is calculated and summed. (3) The retention time of other modified peptides is determined by their retention time relationship between the modifications and unmodified (e.g. the pattern of me3 < me2 < ac < unmodified < me1, in which < means the peptide on the left elutes earlier). If it is DDA, the most abundant XIC is selected. If it is DIA, the fragment XIC is also used, and the XIC with the best score (i.e. cosine) to measure the similarity of precursor XIC and fragment XIC is selected. (4) AUCs of the modified peptides with all charge states are calculated as described in the second step. The isobaric peptides are discriminated using fragment XICs by DIA or targets in DDA. (5) The relative abundance of PTMs is calculated by dividing the area of a given form by the summed total area of all forms for a given peptide. After EpiProfile quantifies each peptide group, it exports tables and figures. In case of experiments adopting isotopic labeling, the labeled peptides are quantified assuming they have the same retention time of the respective unlabeled peptides. EpiProfile 2.0 is freely available for download at https://github.com/

zfyuan/EpiProfile2.0_Family. In the Results section, the details of specific applications will be described.

Results

First, we briefly review the upgrades of EpiProfile 2.0 as compared to our previous version (v1.0). Then, we present the performance of EpiProfile 2.0 in analyzing HeLa cell extracts and more applications such as “oncohistones” (histones with sequence mutations) and histone crotonylation.

Upgrades of EpiProfile 2.0 as compared to 1.0 for histone analysis

Compared with the previous version of our software EpiProfile 1.0[31], EpiProfile 2.0 has evolved according to the multiple requests we received in the last 3 years by collaborators and colleagues. (1) EpiProfile 2.0 supports all histones (i.e. H3, H4, H2A, H2B, and H1) and 13 different organisms, while EpiProfile 1.0 only supports histones H3 and H4 and 2 organisms (i.e. human and mouse, which have identical H3 and H4 peptides). (2) EpiProfile 2.0 supports both DIA and DDA, while EpiProfile 1.0 only supports DDA. In EpiProfile 1.0, only 5 isobaric acetyl peptides from H3 and H4 are discriminated, thanks to targeted scans during the acquisition. In EpiProfile 2.0, 16 isobaric peptides belonging to H3, H4 and H2A are discriminated, thanks to the fully untargeted MS acquisition method. (3) EpiProfile 2.0 supports both high and low resolution MS1, while EpiProfile 1.0 only supports high resolution MS1. (4) EpiProfile 2.0 supports multiple experimental designs. EpiProfile 1.0 only supports SILAC (Arg10) and ^{13}C glucose by pre-calculated peptide masses. In contrast, EpiProfile 2.0 automatically supports more SILAC options (Arg10, Lys8Arg10, Arg6, Lys6Arg6)[34, 35], ^{13}C glucose[36, 37], ^{15}N labeling[38, 39], and $^{13}CD_3$ labeled heavy methylation and methionine[40, 41]. (5) EpiProfile 2.0 exports figures for quality control and data interpretation (e.g. peptide numbers, peptide intensities, principal component analysis, and PTM relative abundance). The software outputs two tables for canonical analysis, one including the relative abundances of peptides and the other the relative abundance of individual PTMs (calculated by summing the relative abundance of all peptides carrying a given modifications), or three tables in case of labeling, where the extra table contains the calculated heavy/light ratios for all histone peptides. In summary, EpiProfile 2.0 was designed to greatly enhance flexibility in terms of supported data and ideas for original experimental designs.

Performance of EpiProfile 2.0 on HeLa cells

By using a HeLa cell histone extract acquired using DIA on an Orbitrap Elite (OT-IT), we tested the performance of EpiProfile 2.0 by aligning 217 quantified peptides from all histone variants H3, H4, H1, H2A, and H2B with a spiked-in library of 93 synthetic histone peptides.

Proper extraction of chromatographic profiles can be monitored by using the output figures named “profile layouts”, which show the retention time relationship between histone variants or different PTMs on the same peptide sequence. An example for the peptide of histone H3 KSTGGKAPR (aa 9−17) is displayed in Figure 1. The first layout of five

peptides is K9me3 < K9me2 < K9ac < unmodified < K9me1. K9me3 and K9me2 are close at about 0.5 min. Similarly, the second layout of five peptides is K9me3K14ac < K9me2K14ac < K9acK14ac < K14ac < K9me1K14ac. Other peptides elute according to the same principle (Figure S1A-J). Importantly, the retention time helps distinguish K9me3 and K9ac though they have a close peptide mass (i.e. 0.036385 Da mass difference).

Next, EpiProfile 2.0 is designed to accurately discriminate the relative abundance of isobaric forms using the fragment ion profile acquired by DIA. Figure 2 shows the procedure to discriminate H3 K9ac and K14ac. The whole chromatographic peak is split to two components using the ratios at each acquired MS2 spectrum by summing the intensities of unique fragments for each species. The overall ratio of K9ac to K14ac is calculated as the area under the curve (Figure S2A). Other isobaric peptides that currently considered and discriminated are H3 aa 18−26 K18ac/K23ac, H4 aa 4−17 1ac, 2ac, 3ac, H3 aa 9−17 K9acS10ph/S10phK14ac, H3.1/3 aa 27−40 K36me3/K27me2K36me1, H2AV/Z aa 1−19 1ac, 2ac, 3ac, H2A aa 4−11 K5ac/K9ac, H2A aa 12−17 K13ac/K15ac (Figure S2B-P). In summary, EpiProfile 2.0 adopts both the order of elution of modifications and unique fragment ions to determine the confident XIC and discriminate isobaric forms.

To validate the correctness of EpiProfile operation, we used synthetic histone peptides labeled to discriminate them from their respective endogenous forms. The library consists of 93 synthetic heavy labeled histone peptides, including the most abundant modified and unmodified forms[42]. Heavy amino acids (e.g., 6.0138 Da for P, and 3.0037 Da for G) were used to discriminate the endogenous isobaric peptides, which have the same mass. For example, the peptide H3K9ac (KacSTGGKAPR) was labeled with one heavy P and H3K14ac (KSTGGKacAPR) with one heavy P and one heavy G (Figure S3). The analysis of synthetic peptides was performed using both manual peak integration (using Thermo Xcalibur) and automatically with EpiProfile; results showed a very high similarity between the two extraction methods (Table S1 and Figure 3A).

Furthermore, we compared the performance of EpiProfile using different types of MS acquisition methods, including both DDA vs DIA, and high vs low resolution acquisition. For that, we used histones extracted from HeLa cells. Table S1 and Figure 3B show that EpiProfile 2.0 and manual calculation via Thermo Xcalibur have similar results for the endogenous histone peptides, demonstrating that the quantification of histone DIA data by EpiProfile 2.0 is as accurate as the manual peak picking. Then, we assessed the reproducibility of EpiProfile in performing XIC using three technical replicates. All four experiments showed reproducible technical replicates indicated by the low coefficient of variation, similarity in calculated PTM relative abundance, and correlation significance p-value for all data (Table S2 and Figure 3C-E). Together, we proved that EpiProfile is a robust software with accuracy comparable to manual peak integration.

EpiProfile 2.0 quantifies histone peptides in various applications

EpiProfile 2.0 has been adapted to quantify histones from different organisms, histones with point mutations known or predicted to have disease relevance, histones derivatized by different chemical anhydrides, and histones with unusual or low-abundance PTMs. Collectively we refer to this set of programs as the EpiProfile 2.0 family, which has 74

different versions of EpiProfile 2.0 (as shown in Figure S4). The 73 variants of EpiProfile 2.0 are described below.

Different organisms ——We have adapted EpiProfile 2.0 to include 11 additional species used as model systems for epigenetics studies, i.e. *Bos taurus, Caenorhabditis elegans, Harpegnathos saltator, Heterocephalus glaber, Neurospora crassa, Oxytricha trifallax, Theileria annulata, Plasmodium falciparum, Saccharomyces cerevisiae, Saccharum officinarum*, and *Xenopus laevis*. The quantification of some peptides are different because of the sequence variance. For example, in *Neurospora crassa* the peptide H2A aa 1−19 TGGGKSGGKASGSKNAQSR has three lysines, but in human/mouse the peptide H2AV aa 1−19 AGGKAGKDSGKAKAKAVSR has five lysines. Thus, the components of isobaric acetyl peptides are different.

Different mutations ——We have adapted EpiProfile 2.0 to include 30 known or predicted missense histone mutations (as shown in Table 1). For example, the endogenous peptide H3.3 aa 27−40 KSAPSTGGVKKPHR has 15 unmodified and modified forms (e.g. K27me1, K36me1, K27me2K36me1, K27me1K36me2, K27ac, etc. as shown in Figure S1E), while the mutant peptide H3.3 K27M MSAPSTGGVKKPHR has 8 different forms (e.g. K36me1, K36me2, K36me3, M27oxK36me1, etc. as shown in Figure S5 obtained by DIA). The layouts of corresponding forms (i.e. the retention time relationship) for these two peptide groups are different.

Different anhydrides ——The software now supports peptide N-termini derivatization using phenyl isocyanate – PIC[24]. This protocol has emerged as helpful to enhance the hydrobophicity of short histone peptides, and it is therefore a valuable alternative to the propionylation protocol. EpiProfile 2.0 does not currently support derivatization protocols using other anhydrides, as we showed that they lead to less accurate analyses[43].

Low-abundance PTMs ——As instruments enhance their sensitivity, more PTMs have emerged as suitable for quantification in standard LC-MS/MS analyses. We have included H2AK13K15ub, H3K27acK36me1/2/3, H3R17me1/2R42me1/2, and H3T3ph. Moreover, we included 27 acyl-CoA species or combinations for eight H3 peptides (i.e. 3−8, 9−17, 18−26, 27−40, 54−63, 64−69, 73−83, and 117−128) and seven H4 peptides (i.e. 4−17, 20−23, 24−35, 41−45, 56−67, 68−78, and 79−92) (as shown in Table 2). First, there are 10 single acyl-CoA species (i.e. ac, bu, cr, glu, hex, bhb, hmg, mal, pr, and suc, see the abbreviation list). In Figure S6, H3 9−17 unmodified, K9cr, K14cr, K9crK14cr can be observed, and K9cr and K14cr are discriminated by DIA. Second, there are 6 labeled acyl-CoA species (i.e. bu, cr, glu, mal, pr, and suc). Third, there are 7 combinations of acyl-CoA species (i.e. acbu, accr, acglu, acbhb, acmal, acpr, and acsuc). Fourth, there are 2 D5 of acyl-CoA species (i.e. bu and pr). At last, there are 2 multi acyl-CoA species (i.e. D0 and D5). From the acylation data, we know that in terms of retention time, bu, cr, and hex elute later than pr, while ac, glu, bhb, hmg, mal, and suc elute earlier than pr. In conclusion, EpiProfile 2.0 has now become a highly flexible software suitable for a wide variety of experiments, including more organisms in study, sequence mutations, derivatization techniques, and low-abundance modifications.

Discussion

We developed EpiProfile 2.0 by significantly improving EpiProfile 1.0. Because of a large demand for histone PTM quantification, we developed an entire family of EpiProfile 2.0 tools, containing 74 versions that include all the current requests we have received from users around the world. We show that layout (i.e. retention time relationship of peptides) and discrimination of isobaric peptides are the basis for quantification of histone PTMs, which are the unique features of EpiProfile (though Skyline can deal with DIA data, it is not designed for issues related to histone analysis as EpiProfile[31]). Therefore, to do quantitative epigenetics correctly in various applications, the usage of EpiProfile 2.0 family can simplify significantly the data extraction. Currently, potential errors in peak extraction because of low-abundance peptides are manually corrected (e.g. checking SILAC or synthetic labeled peptides) through changing the source code. For future versions, we speculate that a graphic user interface will be easy to correct peak extraction errors. Now EpiProfile only supports Thermo raw files due to the available conversion tools. We intend to edit the software to support other file formats such as mzML. The quantification approach is also being extended to non-histone PTMs. For example, this approach can be applied to phosphorylation in non-histone proteins[44–46]. To the best of our knowledge, we are the first to quantify histone PTMs using the retention time relationship of PTMs (i.e. layouts), which is valuable to validate histone PTM identifications and evaluate database search engines[47]. In summary, we developed a universal platform to quantify histone PTMs by bottom-up MS, including all histones and different types of MS. It is our hope that EpiProfile 2.0 will allow proteomics researchers in the epigenetics field to utilize DIA experiments for enhanced characterization of histone modifications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We thank other members of the Garcia lab for testing the program and for helpful feedback. BAG was supported by NIH grants GM110174, AI118891, and CA196539; DOD grant W81XWH-113–1-0426; and an UPenn Epigenetics Institute pilot grant. DM was supported by NIH grants TL1TR001880 and T32GM008275.

Abbreviations

PTM	post-translational modification
DIA	data-independent acquisition
DDA	data-dependent acquisition
AUC	area under curve
un	unmodified
me1/2/3	mono-/di-/tri-methylation
ac	acetylation

ph	phosphorylation
bu	butyryl
cr	crotonyl
glu	glutaryl
hex	hexanoyl
bhb	β-hydroxybutyryl
hmg	3-hydroxy-3-methylglutaryl
mal	malonyl
pr	propionyl
suc	succinyl
va	valeryl
hbu	heavy labeled butyryl
hpr	heavy labeled propionyl
D0	unlabeled
D5	5 deuterium labeled

Reference

1. Kouzarides T, Chromatin modifications and their function. Cell 2007, 128, (4), 693–705. [PubMed: 17320507]
2. Cheung P, Generation and characterization of antibodies directed against di-modified histones, and comments on antibody and epitope recognition. Methods Enzymol 2004, 376, 221–34. [PubMed: 14975309]
3. Egelhofer TA; Minoda A; Klugman S; Lee K; Kolasinska-Zwierz P; Alekseyenko AA; Cheung MS; Day DS; Gadel S; Gorchakov AA; Gu T; Kharchenko PV; Kuan S; Latorre I; Linder-Basso D; Luu Y; Ngo Q; Perry M; Rechtsteiner A; Riddle NC; Schwartz YB; Shanower GA; Vielle A; Ahringer J; Elgin SC; Kuroda MI; Pirrotta V; Ren B; Strome S; Park PJ; Karpen GH; Hawkins RD; Lieb JD, An assessment of histone-modification antibody quality. Nat Struct Mol Biol 2011, 18, (1), 91–3. [PubMed: 21131980]
4. Aebersold R; Mann M, Mass spectrometry-based proteomics. Nature 2003, 422, (6928), 198–207. [PubMed: 12634793]
5. Zheng Y; Huang X; Kelleher NL, Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. Curr Opin Chem Biol 2016, 33, 142–50. [PubMed: 27371874]
6. Moradian A; Kalli A; Sweredoski MJ; Hess S, The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. Proteomics 2014, 14, (4–5), 489–97. [PubMed: 24339419]
7. Tvardovskiy A; Wrzesinski K; Sidoli S; Fey SJ; Rogowska-Wrzesinska A; Jensen ON, Top-down and Middle-down Protein Analysis Reveals that Intact and Clipped Human Histones Differ in Post-translational Modification Patterns. Mol Cell Proteomics 2015, 14, (12), 3142–53. [PubMed: 26424599]

8. Bantscheff M; Lemeer S; Savitski MM; Kuster B, Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 2012, 404, (4), 939–65. [PubMed: 22772140]

9. Venable JD; Dong MQ; Wohlschlegel J; Dillin A; Yates JR, Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat Methods 2004, 1, (1), 39–45. [PubMed: 15782151]

10. Silva JC; Gorenstein MV; Li GZ; Vissers JP; Geromanos SJ, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 2006, 5, (1), 144–56. [PubMed: 16219938]

11. Gillet LC; Navarro P; Tate S; Rost H; Selevsek N; Reiter L; Bonner R; Aebersold R, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 2012, 11, (6), O111 016717.

12. He L; Diedrich J; Chu YY; Yates JR, 3rd, Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. Anal Chem 2015, 87, (22), 11361–7. [PubMed: 26499134]

13. Tsou CC; Avtonomov D; Larsen B; Tucholska M; Choi H; Gingras AC; Nesvizhskii AI, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods 2015, 12, (3), 258–64. [PubMed: 25599550]

14. Sidoli S; Lin S; Xiong L; Bhanu NV; Karch KR; Johansen E; Hunter C; Mollah S; Garcia BA, Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications. Mol Cell Proteomics 2015, 14, (9), 2420–8. [PubMed: 25636311]

15. Sidoli S; Fujiwara R; Garcia BA, Multiplexed data independent acquisition (MSX-DIA) applied by high resolution mass spectrometry improves quantification quality for the analysis of histone peptides. Proteomics 2016, 16, (15–16), 2095–105. [PubMed: 27193262]

16. Sidoli S; Simithy J; Karch KR; Kulej K; Garcia BA, Low Resolution Data-Independent Acquisition in an LTQ-Orbitrap Allows for Simplified and Fully Untargeted Analysis of Histone Modifications. Anal Chem 2015, 87, (22), 11448–54. [PubMed: 26505526]

17. Garcia BA; Mollah S; Ueberheide BM; Busby SA; Muratore TL; Shabanowitz J; Hunt DF, Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat Protoc 2007, 2, (4), 933–8. [PubMed: 17446892]

18. Karch KR; Sidoli S; Garcia BA, Identification and Quantification of Histone PTMs Using High-Resolution Mass Spectrometry. Methods Enzymol 2016, 574, 3–29. [PubMed: 27423855]

19. Freitas MA; Sklenar AR; Parthun MR, Application of mass spectrometry to the identification and quantification of histone post-translational modifications. J Cell Biochem 2004, 92, (4), 691–700. [PubMed: 15211567]

20. Beck HC; Nielsen EC; Matthiesen R; Jensen LH; Sehested M; Finn P; Grauslund M; Hansen AM; Jensen ON, Quantitative proteomic analysis of post-translational modifications of human histones. Mol Cell Proteomics 2006, 5, (7), 1314–25. [PubMed: 16627869]

21. Xu M; Chen S; Zhu B, Investigating the cell cycle-associated dynamics of histone modifications using quantitative mass spectrometry. Methods Enzymol 2012, 512, 29–55. [PubMed: 22910201]

22. Zhang C; Liu Y; Andrews PC, Quantification of histone modifications using (1)(5)N metabolic labeling. Methods 2013, 61, (3), 236–43. [PubMed: 23454290]

23. Krautkramer KA; Reiter L; Denu JM; Dowell JA, Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. J Proteome Res 2015, 14, (8), 3252–62. [PubMed: 26120868]

24. Maile TM; Izrael-Tomasevic A; Cheung T; Guler GD; 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 2015, 14, (4), 1148–58. [PubMed: 25680960]

25. Abshiru N; Caron-Lizotte O; Rajan RE; Jamai A; Pomies C; Verreault A; Thibault P, Discovery of protein acetylation patterns by deconvolution of peptide isomer mass spectra. Nat Commun 2015, 6, 8648. [PubMed: 26468920]

26. Yang L; Tu S; Ren C; Bulloch EM; Liao CL; Tsai MD; Freitas MA, Unambiguous determination of isobaric histone modifications by reversed-phase retention time and high-mass accuracy. Anal Biochem 2010, 396, (1), 13–22. [PubMed: 19699711]
27. Schilling B; Rardin MJ; MacLean BX; Zawadzka AM; Frewen BE; Cusack MP; Sorensen DJ; Bereman MS; Jing E; Wu CC; Verdin E; Kahn CR; Maccoss MJ; Gibson BW, Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol Cell Proteomics 2012, 11, (5), 202–14. [PubMed: 22454539]
28. Lewis PW; Muller MM; Koletsky MS; Cordero F; Lin S; Banaszynski LA; Garcia BA; Muir TW; Becher OJ; Allis CD, Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 2013, 340, (6134), 857–61. [PubMed: 23539183]
29. Mohammad F; Helin K, Oncohistones: drivers of pediatric cancers. Genes Dev 2017, 31, (23–24), 2313–2324. [PubMed: 29352018]
30. Simithy J; Sidoli S; Yuan ZF; Coradin M; Bhanu NV; Marchione DM; Klein BJ; Bazilevsky GA; McCullough CE; Magin RS; Kutateladze TG; Snyder NW; Marmorstein R; Garcia BA, Characterization of histone acylations links chromatin modifications with metabolism. Nat Commun 2017, 8, (1), 1141. [PubMed: 29070843]
31. Yuan ZF; Lin S; Molden RC; Cao XJ; Bhanu NV; Wang X; Sidoli S; Liu S; Garcia BA, EpiProfile Quantifies Histone Peptides With Modifications by Extracting Retention Time and Intensity in High-resolution Mass Spectra. Mol Cell Proteomics 2015, 14, (6), 1696–707. [PubMed: 25805797]
32. Zee BM; Britton LM; Wolle D; Haberman DM; Garcia BA, Origins and formation of histone methylation across the human cell cycle. Mol Cell Biol 2012, 32, (13), 2503–14. [PubMed: 22547680]
33. Sidoli S; Bhanu NV; Karch KR; Wang X; Garcia BA, Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. J Vis Exp 2016, (111), doi:10.3791/54112.
34. Lin S; Yuan ZF; Han Y; Marchione DM; Garcia BA, Preferential Phosphorylation on Old Histones during Early Mitosis in Human Cells. J Biol Chem 2016, 291, (29), 15342–57. [PubMed: 27226594]
35. Ong SE; Blagoev B; Kratchmarova I; Kristensen DB; Steen H; Pandey A; Mann M, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002, 1, (5), 376–86. [PubMed: 12118079]
36. Evertts AG; Zee BM; Dimaggio PA; Gonzales-Cope M; Coller HA; Garcia BA, Quantitative dynamics of the link between cellular metabolism and histone acetylation. J Biol Chem 2013, 288, (17), 12142–51. [PubMed: 23482559]
37. Cargile BJ; Bundy JL; Grunden AM; Stephenson JL, Jr., Synthesis/degradation ratio mass spectrometry for measuring relative dynamic protein turnover. Anal Chem 2004, 76, (1), 86–97. [PubMed: 14697036]
38. Conrads TP; Alving K; Veenstra TD; Belov ME; Anderson GA; Anderson DJ; Lipton MS; Pasa-Tolic L; Udseth HR; Chrisler WB; Thrall BD; Smith RD, Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and 15N-metabolic labeling. Anal Chem 2001, 73, (9), 2132–9. [PubMed: 11354501]
39. Wu CC; MacCoss MJ; Howell KE; Matthews DE; Yates JR, 3rd, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem 2004, 76, (17), 4951–9. [PubMed: 15373428]
40. Zee BM; Levin RS; Xu B; LeRoy G; Wingreen NS; Garcia BA, In vivo residue-specific histone methylation dynamics. J Biol Chem 2010, 285, (5), 3341–50. [PubMed: 19940157]
41. Ong SE; Mittler G; Mann M, Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods 2004, 1, (2), 119–26. [PubMed: 15782174]
42. Lin S; Wein S; Gonzales-Cope M; Otte GL; Yuan ZF; Afjehi-Sadat L; Maile T; Berger SL; Rush J; Lill JR; Arnott D; Garcia BA, Stable-isotope-labeled histone peptide library for histone post-translational modification and variant quantification by mass spectrometry. Mol Cell Proteomics 2014, 13, (9), 2450–66. [PubMed: 25000943]

43. Sidoli S; Yuan ZF; Lin S; Karch K; Wang X; Bhanu N; Arnaudo AM; Britton LM; Cao XJ; Gonzales-Cope M; Han Y; Liu S; Molden RC; Wein S; Afjehi-Sadat L; Garcia BA, Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis. Proteomics 2015, 15, (9), 1459–69. [PubMed: 25641854]

44. Sidoli S; Fujiwara R; Kulej K; Garcia BA, Differential quantification of isobaric phosphopeptides using data-independent acquisition mass spectrometry. Mol Biosyst 2016, doi: 10.1039.

45. Tsai CF; Wang YT; Yen HY; Tsou CC; Ku WC; Lin PY; Chen HY; Nesvizhskii AI; Ishihama Y; Chen YJ, Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. Nat Commun 2015, 6, 6622. [PubMed: 25814448]

46. McLachlin DT; Chait BT, Analysis of phosphorylated proteins and peptides by mass spectrometry. Curr Opin Chem Biol 2001, 5, (5), 591–602. [PubMed: 11578935]

47. Yuan ZF; Lin S; Molden RC; Garcia BA, Evaluation of proteomic search engines for the analysis of histone modifications. J Proteome Res 2014, 13, (10), 4470–8. [PubMed: 25167464]

Synopsis

To quantify histone PTMs in various applications we present EpiProfile 2.0, which consists of discriminating the mixture of isobaric peptides and determining the retention time of modified peptides for all histones. These are challenges of histone analysis, while data-independent acquisition is the key technique to resolve the problems. EpiProfile 2.0 is a universal and accurate platform for histone PTM quantification by bottom-up MS. Thus we believe it will serve as a valuable software tool for the Epi-proteomics field.

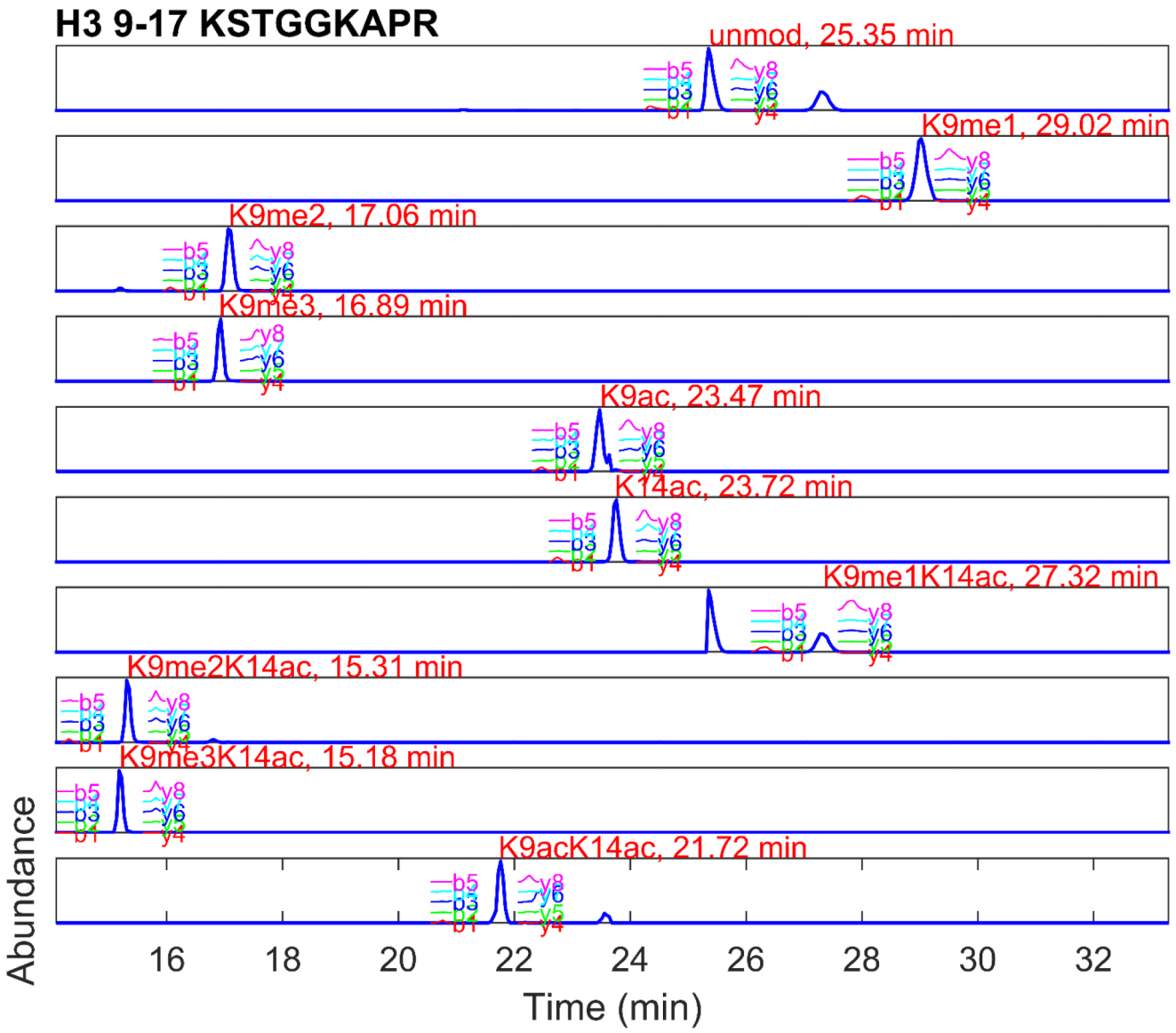


Figure 1. Layouts on H3 9−17 KSTGGKAPR in the endogenous samples.

Two layouts are: K9me3 < K9me2 < K9ac < unmodified < K9me1, and K9me3K14ac < K9me2K14ac < K9acK14ac < K14ac < K9me1K14ac, where the later layout is earlier than the former layout. Subplots show the extracted ion chromatography (XIC) for each peptide based on the precursor m/z (i.e. x-axis is the retention time, y-axis is the intensity). PTM type and retention time of each peptide are above the corresponding chromatographic peak. Fragment ions have the same retention time as the precursor ion. Because fragment ions are usually much lower than the precursor ion, they are aside the precursor ion rather than under the precursor ion profile.

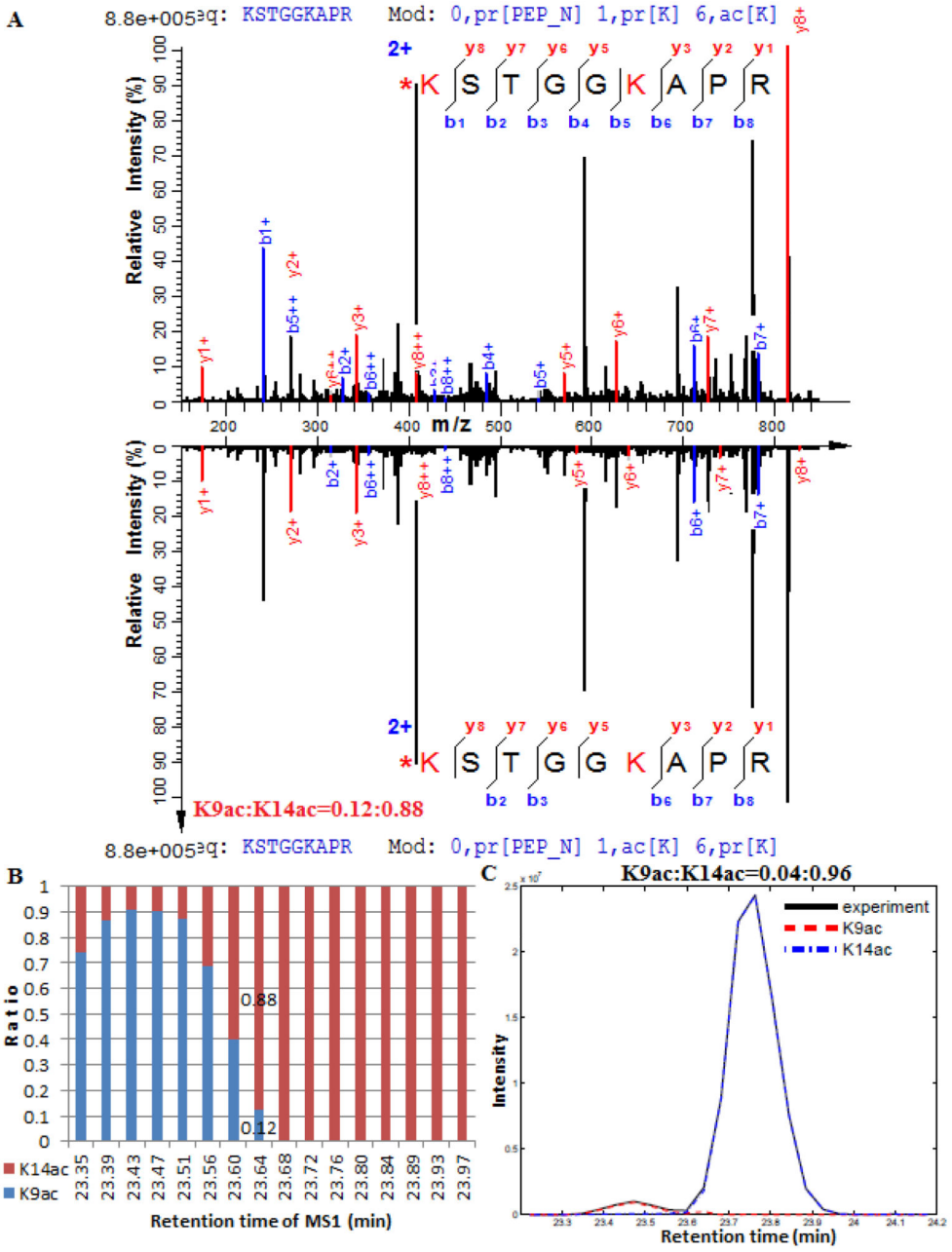
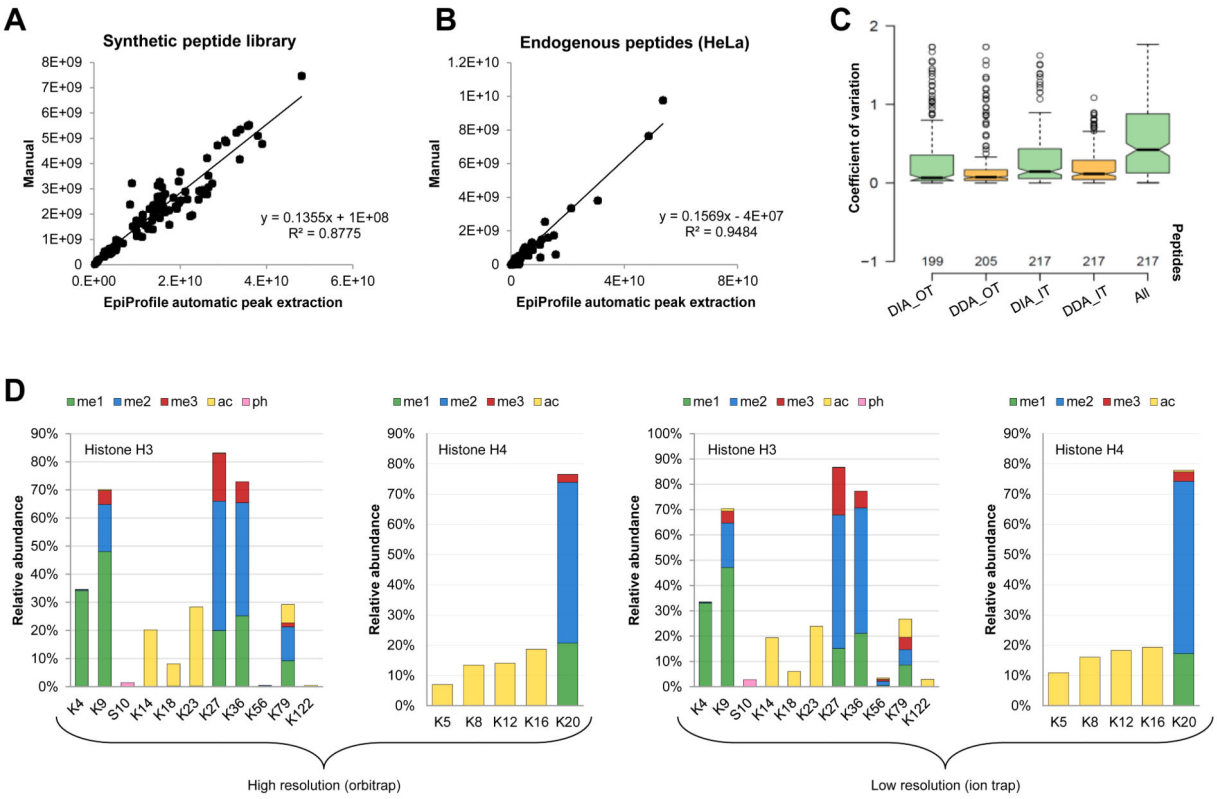


Figure 2. Discrimination of H3 K9ac and K14ac by unique fragment ions.

(A) The same MS2 from DIA is matched to both H3 K9ac and K14ac. Unique fragment ions between K9 and K14 are extracted to calculate the ratio of K9ac to K14ac on this MS2. (B) During the retention time of the chromatographic peak there are 16 MS2. On each MS2 the ratio of K9ac to K14ac is calculated. (C) The whole chromatographic peak is split to two components by the ratios at each time point. The overall ratio of K9ac to K14ac is calculated by the area under curve.



E

-Log10 correlation p-value

R correlation	DIA_OT_1	DIA_OT_2	DIA_OT_3	DDA_OT_1	DDA_OT_2	DDA_OT_3	DIA_IT_1	DIA_IT_2	DIA_IT_3	DDA_IT_1	DDA_IT_2	DDA_IT_3
DIA_OT_1		86.13	92.52	85.75	85.90	87.86	46.94	57.02	67.93	47.87	51.20	50.47
DIA_OT_2	0.92		96.21	90.60	92.41	92.89	43.51	52.35	66.23	44.93	47.78	49.10
DIA_OT_3	0.93	0.93		70.64	71.37	72.25	46.89	64.06	81.13	47.79	50.46	52.17
DDA_OT_1	0.91	0.92	0.88		207.27	203.17	43.10	56.34	61.73	45.21	53.27	48.81
DDA_OT_2	0.91	0.93	0.88	0.99		203.11	42.33	54.86	61.01	44.16	50.59	46.34
DDA_OT_3	0.92	0.93	0.88	0.99	0.99		41.93	55.24	61.47	44.40	50.54	46.89
DIA_IT_1	0.79	0.77	0.79	0.77	0.77	0.76		97.50	87.73	124.45	104.28	145.50
DIA_IT_2	0.84	0.82	0.86	0.83	0.83	0.83	0.93		126.28	91.60	116.25	96.44
DIA_IT_3	0.87	0.87	0.91	0.85	0.85	0.85	0.92	0.97		82.46	90.99	92.25
DDA_IT_1	0.80	0.78	0.79	0.78	0.78	0.78	0.96	0.93	0.91		96.78	148.39
DDA_IT_2	0.81	0.79	0.81	0.82	0.81	0.81	0.94	0.96	0.92	0.93		108.95
DDA_IT_3	0.81	0.80	0.82	0.80	0.79	0.79	0.98	0.93	0.93	0.98	0.95	

Figure 3. High fidelity quantification using EpiProfile 2.0 for high and low resolution data and for DDA and DIA acquisition methods.

(A) Correlation of areas of ion chromatograms for synthetic peptides extracted using EpiProfile 2.0 vs manually. (B) Same plot using signals of histone peptides from HeLa cells. (C) Coefficient of variation of three technical replicates for histone peptide relative abundance from HeLa cells acquired using DIA or DDA in high resolution MS (OT) or low resolution MS (IT). Below, number of peptides quantified. On the right, coefficient of variation estimated across all results from all acquisitions (using the average of technical replicates). (D) Relative abundance of single histone H3 and H4 PTMs using high resolution (left) and low resolution (right) acquisition from HeLa cells acquired by DIA. (E)

Correlation (in green) and correlation significance p-value (in blue) for all replicates and acquisition methods, calculated using HeLa cells peptide relative abundance.

Table 1.

Histone mutations supported by EpiProfile 2.0

H33A29V_T32I	H33A15G	H33R17G	H33A29P	H33P121R	H33K27M	H33G34R
H33G34V	H33G34W	H33K36M	H31K27M	H33T45I	H33G90R	H33G33E
H33G34A	H33V35L	H33K36A	H33K36I	H33K36R	H33K36T	H33K36Q
H33K36E	H33K36Nle	H33K37E	H33K37Q	H33K37T	H33K37N	H33K37R
H31G34W	H33K27R_G34R					

Table 2.

Histone PTMs modified by various acyl-CoA species

Types	Examples (H3 3–8 TKQTAR)
10 single acyl-CoA species (i.e. ac, bu, cr, glu, hex, bhb, hmg, mal, pr, and suc)	In ac-CoA, unmodified (0,pr;2,pr;) and K4ac (0,pr;2,ac;)
	In pr-CoA, unmodified (0,va;2,va;) and K4pr (0,va;2,pr;)
6 labeled acyl-CoA species (i.e. bu, cr, glu, mal, pr, and suc)	In bu-CoA, unmodified (0,pr;2,pr;) and K4bu (0,pr;2,hbu;)
	In pr-CoA, unmodified (0,pr;2,pr;) and K4pr (0,pr;2,hpr;)
7 combinations of acyl-CoA species (i.e. acbu, accr, acglu, acbhb, acmal, acpr, and acsuc)	In acbu-CoA, unmodified (0,pr;2,pr;) K4ac (0,pr;2,ac;), and K4bu (0,pr;2,bu;)
	In acpr-CoA, unmodified (0,prD5;2,prD5;), K4ac (0,prD5;2,ac;) and K4pr (0,prD5;2,pr;)
2 D5 of acyl-CoA species (i.e. bu and pr)	In bu-CoA, unmodified (0,prD5;2,prD5;) and K4bu (0,prD5;2,bu;)
	In pr-CoA, unmodified (0,prD5;2,prD5;) and K4pr (0,prD5;2,pr;)
2 multi acyl-CoA species (i.e. D0 and D5)	In D0, unmodified (0,pr;2,pr;), K4me1, K4me2, K4me3, K4ac, K4glu, K4bhb, K4mal, K4suc, K4cr (0,pr;2,cr;)
	In D5, K4bu (0,prD5;2,bu;) and K4pr (0,prD5;2,pr;) as well as from unmodified (0,prD5;2,prD5;) to K4cr (0,prD5;2,cr;)