Published in final edited form as:

J Proteome Res. 2018 January 05; 17(1): 234–242. doi:10.1021/acs.jproteome.7b00544.

Assessment of quantification precision of histone post-translational modifications by using an ion trap and down to 50,000 cells as starting material

Qi Guo¹, Simone Sidoli¹, Benjamin A. Garcia^{1,*}, and Xiaolu Zhao^{2,*}

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

²Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, P.R. China, 430072

Abstract

Histone post-translational modifications (PTMs) are fundamental players of chromatin regulation, as they contribute to editing histone chemical properties and recruiting proteins for gene transcription and DNA repair. Mass spectrometry (MS) based proteomics is currently the most widely adopted strategy for high throughput quantification of hundreds of histone PTMs. Samples such as primary tissues, complex model systems and biofluids are hard to retrieve in large quantities. Because of this, it is critical to know whether the amount of sample available would lead to an exhaustive analysis if subjected to MS. In this work, we assessed the reproducibility in quantification of histone PTMs using a wide range of starting material, i.e. from 5,000,000 to 50,000 cells. We performed the experiment using four different cell lines, i.e. HeLa, 293T, human embryonic stem cells (hESCs) and myoblasts, and we quantified a list of 205 histone peptides using ion trap MS and our in-house software. Results highlighted that the relative abundance of some histone PTMs deviated as little as just 4% when comparing high starting material with histone samples extracted from 50,000 cells, e.g. H3K9me2 (40% average abundance). Low abundance PTMs such as H3K4me2 (<3% average abundance) showed higher variability, but still around 34%. This indicates that most PTMs, and especially abundant ones, are quantified with high precision starting from low cell counts. This study will help scientists to decide whether specific experiments are feasible and to plan how much sample should be reserved for histone analysis using MS.

Graphical abstract

*Corresponding authors: 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. Xiaolu Zhao, Room 6016, College of Life Sciences, Wuhan University, Luojiashan Road, Wuhan, P.R. China, 430072. zhaoxiaolu@whu.edu.cn; Tel: + 86-27-68753797; Fax: + 86-27-68753797.

Keywords

data independent acquisition; histones; mass spectrometry; post-translational modifications; bottom-up

Introduction

DNA is organized by protein-DNA complexes called nucleosomes in eukaryotes. Nucleosomes are composed of 147 base pairs of DNA wrapped around a histone octamer containing two copies of each core histone protein H2A, H2B, H3, and H4¹. Histone proteins play significant roles in many nuclear processes because of their intimate association with DNA, including transcription, DNA damage repair, and heterochromatin formation. Histone proteins are extensively and dynamically post-translationally modified by nuclear proteins, and these post-translational modifications (PTMs) are thought to comprise a “histone code” where each specific combinatorial PTM profile of a histone dictates its specific function, such as activating transcription². Abnormal regulations of PTM may lead to developmental disorders and disease development such as cancer^{3, 4}. It is therefore critical to decipher the histone code to understand fundamental nuclear processes and how these are aberrantly regulated in disease.

Antibodies have been widely used to characterize histones and histone PTMs. However, antibody-based techniques have several limitations, such as: (i) they can only confirm the presence of a modification and cannot identify unknown PTMs; (ii) they are biased due to the presence of co-existing marks, which can influence binding affinity; (iii) they cannot identify combinatorial marks, as only very few antibodies are available for such purpose and (iv) they cross-react between highly similar histone variants or similar PTMs (e.g., di- and trimethylation of lysine residues). Mass spectrometry (MS) has therefore emerged as the most suitable analytical tool to quantify proteomes and protein PTMs⁵. The high speed (scan rate >10Hz), high resolution (>100,000) and high sensitivity of MS made it suitable for online chromatographic separation and detection of total modified histone peptides within an hour⁶. Histones can be analyzed with the traditional bottom-up MS strategy, and also via middle-down or top-down MS (reviewed in ⁷), in order to accurately identify and quantify not only individual PTMs, but also their co-frequency. The most commonly used strategy is still bottom-up MS, and the most widely adopted protocol includes derivatization of lysine residues in histones to allow trypsin to generate Arg-C like peptides (4–20 aa)^{8–10}. Data-independent acquisition (DIA) is currently the most suitable MS acquisition method, due to the large variety of isobarically modified peptides, which require MS/MS based quantification to discriminate their abundance if co-eluting during chromatography^{11–13}. Recently, Sidoli et al. assessed that the relatively low complexity of purified histone samples

combined with DIA allows for the use of low resolution MS instrumentation such as the ion trap¹², paving the way for a more affordable analysis of histone PTMs.

Since biological material is not always available in large quantities, it is important to assess the amount required for specific experiments. For example, primary cells are usually more biologically relevant tools than cell lines for biological studies; however, obtaining a pure population of primary cells can be a difficult and arduous process due to their requirement of additional nutrients not included in classical media. The amount of material required has been established for several biochemical techniques, including large-scale proteomics, i.e. 1–2 µg of peptides on column when running nano liquid chromatography (nanoLC). Histones is a peculiar example, as they are among the most abundant proteins in eukaryotic cells. Considering the length of the human genome (3 billion base pairs) and the average distribution of nucleosomes (about one each ~200 base pairs), it is safe to assume that histone proteins are present in millions and millions of copies in each individual cell. Because of this, a low cell number might be sufficient to perform exhaustive histone PTM analysis, but this still needed to be assessed.

In this work, we demonstrate that histone PTMs can be characterized using low resolution MS (ion trap) starting with as low as 50,000 cells. To put that in context, this cell number can be easily grown in a well of a 96-well plate when using average suspension cells. We assessed the reliability of quantification with such small amount of cells by using a variety of cell lines with remarkably different phenotypes: HeLa, 293T, human embryonic stem cells (hESC) and myoblasts. In the low cell number range, we established that histone marks with low abundance such as H3K4me had unfavorable coefficient of variation, and thus we suggest using more cells when characterizing those low abundance histone marks. However, abundant histone marks such as histone H4 acetylations were efficiently quantified with low cell counts. Collectively, our study addresses a simple, albeit critical, question about the amount of material required for MS analysis of histone PTMs. Moreover, the study was performed with low resolution MS, showing that this analysis is possible with cost efficient instrumentation.

Materials and Method

Cell culture

HeLa cell line (CCL-2) was cultured at 37°C with 5% CO₂ in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum, 2 mM L-glutamine and 50 µg/ml each of penicillin/streptomycin. Cultures at about 80% confluence were split 1:5 in 10 cm culture dishes. Briefly, the cells were washed in PBS. PBS containing 0.25% (w/v) trypsin was added to the dishes and placed at 37°C for 5–10 minutes. After the cells were detached from the dishes, prewarmed culture medium was added and the cells were pipetted up and down to render them single cells.

Primary myoblasts were cultured on dishes coated with 0.1% gelatin (Millipore) as described previously¹⁴. Briefly, cells were grown at 2000 to 3000 cells/cm² in proliferation medium composed of Dulbecco modified Eagle medium (DMEM), Medium 199 (Hyclone), 15% FBS, (Hyclone), 0.02 M HEPES buffer, 1.4 mg/l vitamin B12 (both Sigma-Aldrich, St.

Louis, MO, USA), 0.03 mg/l ZnSO₄ (Fisher Scientific, Fair Lawn, NJ, USA), 0.055 mg/l dexamethasone (Sigma-Aldrich), 2.5 µg/l hepatocyte growth factor and 10 µg/l basic fibroblast growth factor (both from Biovision Inc). At 80% confluency (usually every 5 to 7 days), cells were detached using 0.05% trypsin/EDTA (Gibco), quenched with DMEM/10% FBS dispersed into single cells and counted. H9 hESC from WiCell (Madison, WI) were maintained on Matrigel (Corning/Stem Cell Technologies) coated 6-well tissue culture plates in mTeSR[®]1 medium (Stem Cell Technologies, Vancouver, BC, Canada) as per manufacturer's recommendation. Medium was changed daily and cells were used when 60–80% confluent. Colonies were enzymatically detached using Accutase (eBioscience, San Diego, CA) and dispersed by pipetting for cell counting. 293T cells were grown adherent on plates with DMEM containing 10% FBS, 50 µg/ml each of penicillin and streptomycin. The cells were passaged every 3–4 days when at 80% confluency. Cells were harvested in 0.05% trypsin/EDTA and dispersed into single cells for counting.

Cell numbers were determined using a hemocytometer. To aliquot different cell numbers of 5e4, 1e5, 2e5, 5e5, 1e6, 2e6 and 5e6, about 10 million cells were set aside from each cell line in triplicate, washed in PBS thoroughly and collected by centrifugation at 150xg. One biological set was used to confirm extraction of histones using PAGE. Two other replicates were analyzed for MS.

Histone extraction

Histones were extracted as described previously with minor modification⁶. The cells were incubated in nuclear isolation buffer (NIB) (15 mM Tris-HCl, 15 mM NaCl, 60 mM KCl, 5 mM MgCl₂, 1 mM CaCl₂, 250 mM sucrose, pH 7.5, and 0.5 mM AEBSF, 10 mM sodium butyrate, 5 nM microcystin, 1 mM DTT added fresh) with 0.3% NP-40 on ice for 5 min. 300 µl, 200 µl, 100 µl, 100 µl, 50 µl, 50 µl and 50 µl of NIB with NP-40 were used for 5e6, 2e6, 1e6, 5e5, 2e5, 1e5 and 5e4 cells. The nuclei were collected by centrifuging at 700 × g at 4°C for 5 min. The resulting nuclear pellet was washed twice with the same volume of nuclear isolation buffer without NP-40. Histones were then acid-extracted with 0.2 M H₂SO₄ for 3 hours at 4°C with rotation. For all cells numbers, 100 µl of 0.2 M H₂SO₄ were used. The insoluble nuclear debris were pelleted at 3400 × g at 4°C for 5 min, and the supernatant was retained. Finally, histone proteins were precipitated overnight on ice after adding 100% trichloroacetic acid (TCA) in the ratio of 1:3 (v/v) to the acid extraction supernatant, in order to obtain a final TCA concentration of 33%. The precipitated proteins were washed once with 0.1% HCl in acetone (–20°C), twice with acetone (–20°C), and dried in speedvac. One biological set was redissolved in 10 µl of ddH₂O. Protein concentration was measured using the Bradford assay and histone proteins were subjected to 15% SDS-PAGE. Two other replicates were redissolved in 30 µl of 50 mM NH₄HCO₃ (pH 8.0) and subjected to histone propionylation and digestion.

Histone propionylation and digestion

The histone propionylation and digestion were performed as previously described with minor modification⁶. Propionic anhydride solution was freshly prepared by mixing propionic anhydride with acetonitrile in a ratio of 1:3 (v/v). 15 µl of such derivatization reagent was mixed with the histone sample in the ratio of 1:2 (v/v) and 7.5 µl of ammonium

hydroxide was added immediately to the mixture, followed by an incubation of 15 minutes at 37°C. After incubation, the samples were dried in a speed-vac to <10 µL and were filled up to 30 µL using 50 mM NH₄HCO₃. This derivatization reaction was repeated once. After drying down, histone proteins were then digested with trypsin (enzyme:sample ratio 1:20) in 50 mM NH₄HCO₃ overnight at room temperature. After digestion, the derivatization reaction was performed again twice to derivatize peptide N-termini. Samples were desalted using C₁₈ Stage-tips prior to LC-MS analysis.

NanoLC–MS/MS

Samples were analyzed by using a nanoLC-MS/MS setup. NanoLC was configured with a 75 µm ID x 15 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), packed in-house. The HPLC gradient was as follows: 2% to 28% solvent B (A = 0.1% formic acid; B = 95% MeCN, 0.1% formic acid) over 45 minutes, from 28% to 80% solvent B in 5 minutes, 80% B for 10 minutes at a flow-rate of 300 nL/min. nanoLC was coupled to an LTQ Velos mass spectrometer (Thermo Scientific, San Jose, CA, USA). For DIA, two full scan MS spectra (m/z 300–1100) were acquired in the ion trap within a DIA duty cycle, and 16 ms/ms were performed with an isolation window of 50 Da. Normalized collision energy (CE) was set to 35% with activation Q of 0.25¹².

Data Analysis

Raw MS data were analyzed adopting EpiProfile which enabled DIA analyses of histones¹⁵. Histone peptides with multiple PTMs were quantified in relative ratios retrieved by extracting retention time and peak area in mass spectra. Isobaric histone peptides were discriminated based on the distinguishing fragment ions in their tandem mass spectra and the chromatographic area were extracted under the curve using previous knowledge about peptide elution profiles. The relative abundance of PTMs is determined by dividing the area of the particular peptide in the summed total area of that histone peptide in all modified forms.

Results and Discussion

Histone PTMs have a wide range of relative abundance, spanning from common marks such as histone H4 acetylation (10–30% of the total chromatin), to low abundance, although biologically very important, marks such as H3K4me3 (<1%) to even rarer newly discovered PTMs such as derivatives of fatty acids and glycosylation^{16, 17}. Moreover, histone peptides can often be modified generating isobaric forms, i.e. same PTMs but in different amino acid residues. Because of this, a plethora of publications have focused on optimization of MS for accurate quantification of histone marks^{6, 12, 13, 18}. Currently, the most widely adopted protocol includes lysine and peptide N-terminal chemical derivatization prior to trypsin digestion⁶. This protocol reduces cleavage residues of trypsin in histone sequences, heavily rich in basic residues, and enhances the hydrophobicity of peptides for better HPLC retention, as they are rather short (4–20 aa) and hydrophilic. In this work, we used this protocol (Figure 1A) to analyze histone PTMs from four different cell lines, namely HeLa, 293T, human embryonic stem cells (hESCs) and myoblasts. These cells were selected for

their wide popularity as a model system, and also because their different phenotypes allowed us to assess our results as broadly applicable to multiple systems.

To test whether low-resolution ion trap is able to characterize histones properly with a small amount of starting material, we started from a range of cell numbers (5e4, 1e5, 2e5, 5e5, 1e6, 2e6 and 5e6) for all the seven conditions mentioned above. Protein bands visualized via SDS-PAGE of extracted histones indicated as expected that less histone was obtained from lower amounts of starting material (Figure 1A), although histones were still detectable at the low cell counts. Isolated histone proteins were propionylated on the protein level, tryptic digested, further N-term-propionylated on the peptide level and stage-tip desalted prior to nanoLC-MS/MS analysis using a data-independent acquisition (DIA) set-up on an ion trap (Thermo). In total, 205 different histone peptides with/without modifications were quantified by using EpiProfile¹⁵. Detailed information for all quantified histone peptides across all conditions and cell types were listed in Supplementary Table 1.

To evaluate the sensitivity of our MS analysis and the ability of discriminating signals from noise using low resolution MS, we performed linear regression between cell number and peak area of our peptides (two examples in Figure 1B). The correlation between the area of extracted ion chromatogram and cell count was above 0.5 for about 3/4 of all the peptides we quantified (N: 205) (Figure 1C). We then normalized by z-score both our peak areas and cell count values to see slope and intercept of our data (example in Supplemental Figure 1A). Poor sensitivity would be evidenced by the line intersecting the y axis below the origin of the axes, i.e. sample providing no signal. On the other hand, integration of background noise rather than peptide signals would be evidenced by the line intersecting the y axis above the origin of the axes, i.e. still signals with no cells. Results showed that our linear regression had an intercept close to 0 (Supplemental Figure 1B), indicating that the analysis is sensitive and unbiased. The correlation value had a median over 0.6 (Supplemental Figure 1C), which was in line with previous results we discussed in Figure 1C. The slope was close to 1 regardless of cell types, indicating the results were quite reproducible at different cell counts (Supplemental Figure 1D). These results supported a conclusion that low resolution ion trap could perform histone analysis with as low as 50,000 cells.

To further examine whether cell number would affect our final observation, clustering analysis of all quantified histone peptides from four different cell lines covering seven different starting cell amounts were applied using the calculated relative abundance (in %). A heat map representing all analyzed samples (z-score normalization) is illustrated in Figure 2. The clustering aggregates cell lines rather than cell counts even though the difference of relative abundance of histone PTMs among these cell lines is not remarkable (Table 1), indicating that cell counts is a minor factor for PTM relative abundance variation compared to cell types. In sum, myoblast, hESC and 293T cells exhibited similar histone modification states across all seven cell counts (orange refers to more abundant while green is less abundant). In comparison, histone peptides from Hela cells are also grouped together, however showing more differences in PTMs patterns. Collectively, we show that the quantification of histone PTMs with as low as 50,000 cells is sufficiently accurate to discriminate the cell lines we are currently investigating.

Accurate histone quantification is especially crucial in finding the changes of histone modifications under different conditions such as diseases and drug therapy. To evaluate the impact of starting cell amounts on the quantitation accuracy in our MS assay, we performed an experiment of sodium butyrate (NaB) treatment of HeLa cells, which inhibits HDACs and thus increases the overall histone acetylation states. We can observe an overall increased acetylation upon sodium butyrate treatment for acetylated peptides, in accordance with our expectation (detailed information for acetylation changes were calculated in Supplemental Table 4). We found a handful of acetylated peptides with lower abundance in sodium butyrate treatment, e.g. H3K9ac, H4K5acK8acK12acK16ac and H2AZK15ac. Considering that all peptides are normalized by their respective other modified forms, the reduced abundance of these peptides can be simply explained because their respective hyperacetylated forms (e.g. K9acK14ac) are increased. The majority of the increased acetylations (mostly H3 and H4 acetylations) is consistent across all seven starting cell amounts, confirming the ability and confidence of our MS analysis in accurately quantifying the histone PTMs changes upon treatment in as low as 50,000 cells.

Finally, we focused on the analysis of deconvoluted single PTMs. A given PTM, e.g. H3K9me3, is commonly present in different peptides due to potential combinatorial PTM patterns, e.g. H3K9me3K14ac. Because of this, we usually sum all peptides carrying a given PTM to roughly obtain its total relative abundance. We call this table “single PTMs” for simplicity. The MS signal of a histone peptide can vary due to their different modification state or sequence composition¹⁹. This is mostly because of the difference of the ionization efficiency, as well as the poor LC retention of short (<6 aa) and hydrophilic peptides. Therefore, the relative abundance and coefficient of variation (CV) of “single PTMs” we used in this study were just a rough estimation. They were not accurate for quantification purpose but were considered enough to show a trend of quantification reliability at different abundance level. Using histone H3 and H4 as examples, the average abundance, CV values and raw absolute intensities of quantified histone single PTMs in four cell lines are summarized in Table 1 (detailed information for quantified histone H3 and H4 peptides was summarized in Supple. Table 2). The coefficient of variations in Table 1 were calculated using seven different cell counts as replicates. The variations of single PTMs on histone H3 and H4 in four cell types were quantified and illustrated in Figure 3. Distinct histone marks exhibited different reproducibility across different amount of starting material. Those marks with higher abundances were quantified with on average higher reproducibility. For example, H3K9me2 and K14ac are commonly of high abundance (>30%); these marks were quantified with high reproducibility (CV as low as 4% in myoblast). In comparison, the histone H3 peptide TK₄QTAR (aa 3–8) is usually harder to detect, due to its low hydrophobicity and thus weak retention by C₁₈ chromatography¹². This peptide carries low abundance PTMs such as H3K4me2 (<3%), which showed a coefficient of variations up to 34%. Even lower abundance PTMs were found in the histone H3 peptide K₁₈QLATK₂₃AAR (aa 18–26), namely K18me1 and K23me1; these marks showed even higher CVs (>100%), due to their extremely low relative abundance (<1% in myoblast).

Taking all four cell lines into consideration (as illustrated in Figure 3), we can see that quite a number of histone marks (CV<25%) are reliable and quantifiable at all cell input levels, such as H3K14ac, H3K23ac, H3K27me1/me2 and modifications on H3K9ac as well as

H4K20me1 and acetylations on H4K5/K8/K12/K16 (refer to the orange highlighted marks in Figure 3). We also noticed that several marks were reliably quantified at the higher cell counts but became variable at lower cell counts. Therefore, in addition to a single CV value that covers data from all seven cell counts, we characterized each mark with average relative abundances (Ab) and different CVs which were calculated using different range of cell counts as replicates (as summarized in Table 2; detailed calculations as listed in Supple. Table 3). For example, “Ab 3” and “CV 3” were calculated using first three highest cell counts which were 5e6, 2e6 and 1e6 cells, whereas “Ab 7” and “CV 7” covered all seven cell counts. As highlighted in light orange in Figure 3, several histone H3 and H4 marks which are sensitive to lower cell counts (worse CV value covering seven starting cell counts as shown before slash), are generally consistent at higher numbers using more than 1,000,000 cells (much better CV value shown after slash). These marks are modifications on H3K4, H3K18ac, H3K18me1, H3K27me3, H3K36me2, H3K56ac/me2/me3, H3K79me1/me2/me3, H3K122ac and H4K20ac, among which H3K56me2 and H3K122ac need more than 2,000,000 cells as marked with “*” in hESC. Several marks which are variable at low cell counts but appear to be reasonably more reproducible ($25\% < CV < 35\%$) using more than 1,000,000 cells are highlighted in blue in Figure 3. The other marks like H3K27ac, H3K36me1/me3, H3K56me1, H3K79ac and H4K20me2/me3 are much less reproducible and variable across different cell counts, and thus require more careful inspection in studies with lower starting material. In general, as indicated in Figure 3, histone marks behave similarly among the four cells lines with minor exceptions that more cells might be needed for reliable quantification of some specific marks (H3.3K27/K36 marks as listed in Table 3) for human embryonic stem cells.

Based on our findings, researchers who are interested in those marks which are reliable at all cell counts could go down to 50,000 cells and might keep in mind that higher amount of starting material are required for those marks which are sensitive to cell counts. Collectively, our study provides reliable information for a reasonable estimation of the cell amounts used to perform a systematic analysis of any specific known histone marks.

Conclusion

In conclusion, our work demonstrated that an accurate quantification of abundant histone PTMs can be efficiently performed by using low resolution MS and as low as 50,000 cells as starting material. The use of DIA and EpiProfile compensates for the lack of high resolution, as the profile of the fragment ions generated by DIA and the intelligent peak picking of EpiProfile¹⁵ provided unambiguous quantification of histone peptides, including isobaric species. Low abundance histone marks showed more variability in quantification when comparing different amount of starting material, so a larger amount of starting material (at least 500,000 cells) is recommended. Nevertheless, we established that histone PTM analysis can be performed with μ l scale cell culture volume and with cost efficient MS instrumentation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

BAG acknowledges funding from NIH grants GM110174, AI118891 and CA196539; and a Robert Arceci Scholar award from the Leukemia and Lymphoma Society. XZ is supported by China Scholarship Council (CSC, 201506275204), the Natural Science Foundation of China (No. 31401087) and the 111 Project of China (B16036).

References

1. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997; 389(6648):251–60. [PubMed: 9305837]
2. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001; 293(5532):1074–80. [PubMed: 11498575]
3. Onder O, Sidoli S, Carroll M, Garcia BA. Progress in epigenetic histone modification analysis by mass spectrometry for clinical investigations. *Expert review of proteomics*. 2015; 12(5):499–517. [PubMed: 26400466]
4. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000; 403(6765):41–5. [PubMed: 10638745]
5. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003; 422(6928):198–207. [PubMed: 12634793]
6. Sidoli S, Bhanu NV, Karch KR, Wang X, Garcia BA. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *Journal of visualized experiments : JoVE*. 2016; (111)
7. Sidoli S, Cheng L, Jensen ON. Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *Journal of proteomics*. 2012; 75(12):3419–33. [PubMed: 22234360]
8. Garcia BA, Mollah S, Ueberheide BM, Busby SA, Muratore TL, Shabanowitz J, Hunt DF. Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols*. 2007; 2(4):933–8. [PubMed: 17446892]
9. Zhang C, Liu Y, Andrews PC. Quantification of histone modifications using (1)(5)N metabolic labeling. *Methods*. 2013; 61(3):236–43. [PubMed: 23454290]
10. Lin S, Garcia BA. Examining histone posttranslational modification patterns by high-resolution mass spectrometry. *Methods in enzymology*. 2012; 512:3–28. [PubMed: 22910200]
11. 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. *Molecular & cellular proteomics : MCP*. 2015; 14(9):2420–8. [PubMed: 25636311]
12. Sidoli S, Simithy J, Karch KR, Kulej K, Garcia BA. Low Resolution Data-Independent Acquisition in an LTQ-Orbitrap Allows for Simplified and Fully Untargeted Analysis of Histone Modifications. *Analytical chemistry*. 2015; 87(22):11448–54. [PubMed: 26505526]
13. 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]
14. Stadler G, Chen JC, Wagner K, Robin JD, Shay JW, Emerson CP Jr, Wright WE. Establishment of clonal myogenic cell lines from severely affected dystrophic muscles - CDK4 maintains the myogenic population. *Skeletal muscle*. 2011; 1(1):12. [PubMed: 21798090]
15. 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. *Molecular & cellular proteomics : MCP*. 2015; 14(6):1696–707. [PubMed: 25805797]
16. Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N, Lu Z, Ye Z, Zhu Q, Wysocka J, Ye Y, Khochbin S, Ren B, Zhao Y. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*. 2011; 146(6):1016–28. [PubMed: 21925322]

17. Sakabe K, Wang Z, Hart GW. Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(46): 19915–20. [PubMed: 21045127]
18. Krautkramer KA, Reiter L, Denu JM, Dowell JA. Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. *Journal of proteome research*. 2015; 14(8):3252–62. [PubMed: 26120868]
19. 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. *Molecular & cellular proteomics : MCP*. 2014; 13(9):2450–66. [PubMed: 25000943]

Figure 1. Workflow of the bottom-up strategy for histone analysis by MS and quality check
 (A) Workflow beginning from four types of cells (HeLa, 293T, hESC and myoblast) with a range of cell numbers (5×10^4 , 1×10^5 , 2×10^5 , 5×10^5 , 1×10^6 , 2×10^6 and 5×10^6), through nuclei extraction, acid extraction, propionylation, trypsin digestion, propionylation, stage-tipping and LC-MS/MS analysis.
 (B) Correlation between cell count and peptide raw signal intensity. Representative peptides showing correlation between Log10 transformed cell number vs Log10 transformed area of extracted ion chromatogram. Error bars represent standard deviation. (C) All correlation values for all peptides and cell lines analyzed, where peptides were sorted based on decreasing correlation. The figure shows that about 75% of the peptides have a correlation with the cell number of at least 0.5. Sample abbreviations are HeLa (HeLa, H (hESCs), M (Myoblasts) and T (293T).

Figure 2. Heatmap of all analyzed samples, z-score normalized across all conditions and cell lines
Clustering analysis of all peptides (x-axis) and conditions (y-axis) using the calculated relative abundance (in %) of the quantified peptides. Results indicate that the conditions group mostly based on cell lines rather than cell counts, implying that the number of cells used as starting material are a minor factor in the analysis.

Figure 3. PTM quantification precision on histone H3 and H4 in four cell lines using different amount of starting material

For each cell line, on the top, the average relative abundances of single PTMs on histone H3 and H4 (Detailed information as in Table 1). On the bottom, the coefficient of variations (using different cell counts as replicates) for the same histone marks. Those marks which are reliable across all different cell counts are highlighted in orange. The marks which are consistent at higher cell counts (more than 1,000,000 cells) are highlighted in light orange. For hESC, two marks which are reliable using more than 2,000,000 cells are marked with *.

A few marks which are less reliable however show better reliability using more than 1,000,000 cells are highlighted in blue.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

The average relative abundances, coefficient of variations (CV) and the raw absolute intensities of histone single PTMs on H3 and H4 in four cell lines.

Single mark	Average abundance				CV				Average Intensity			
	Hela	M	H	T	Hela	M	H	T	Hela	M	H	T
H3K4me1	9%	11%	8%	9%	21%	59%	33%	13%	1.7E+10	2.8E+10	3.4E+10	1.9E+10
H3K4me2	3%	2%	1%	1%	30%	25%	19%	34%	4.0E+09	3.1E+09	2.6E+09	1.8E+09
H3K4me3	0%	0%	0%	1%	80%	36%	29%	26%	4.2E+08	5.1E+08	1.2E+09	8.6E+08
H3K4ac	0%	5%	0%	0%	56%	69%	127%	86%	8.8E+08	9.2E+09	2.5E+09	1.2E+09
H3K9me1	12%	9%	9%	7%	17%	6%	27%	49%	3.8E+10	3.4E+10	6.3E+10	3.1E+10
H3K9me2	44%	46%	38%	53%	10%	4%	18%	12%	1.8E+11	1.5E+11	1.8E+11	2.4E+11
H3K9me3	19%	30%	20%	21%	9%	2%	20%	24%	7.9E+10	1.0E+11	1.2E+11	1.4E+11
H3K9ac	12%	5%	9%	5%	8%	12%	38%	17%	4.2E+10	1.9E+10	9.3E+10	2.5E+10
H3K14ac	34%	26%	22%	32%	8%	7%	24%	14%	1.4E+11	9.6E+10	1.5E+11	1.3E+11
H3K18me1	3%	1%	2%	5%	195%	41%	119%	122%	8.4E+08	6.1E+08	1.5E+09	8.0E+08
H3K18ac	10%	7%	11%	10%	29%	29%	26%	28%	2.6E+10	1.4E+10	3.2E+10	1.8E+10
H3K23me1	3%	1%	1%	3%	195%	21%	107%	125%	1.2E+09	1.6E+09	1.9E+09	8.3E+08
H3K23ac	22%	16%	20%	23%	16%	23%	23%	8%	5.0E+10	3.2E+10	5.4E+10	5.1E+10
H3K27me1	13%	19%	27%	20%	118%	50%	10%	82%	4.5E+10	1.8E+10	3.9E+10	8.6E+09
H3K27me2	15%	38%	18%	16%	24%	22%	64%	73%	1.9E+10	4.0E+10	3.2E+10	7.1E+10
H3K27me3	2%	18%	22%	34%	64%	68%	46%	87%	4.4E+09	1.1E+10	8.5E+09	6.1E+10
H3K27ac	6%	8%	5%	3%	71%	60%	90%	105%	3.0E+09	7.6E+09	2.5E+10	4.3E+09
H3K36me1	7%	19%	19%	35%	75%	68%	69%	87%	4.5E+09	1.4E+10	1.3E+10	6.4E+10
H3K36me2	15%	38%	23%	17%	40%	25%	49%	58%	2.1E+10	4.4E+10	5.4E+10	5.1E+10
H3K36me3	7%	12%	11%	8%	208%	90%	50%	82%	4.4E+10	1.0E+10	6.0E+09	7.7E+09
H3K27me1	20%	25%	24%	16%	26%	43%	38%	58%	1.1E+10	1.6E+10	2.3E+10	1.3E+10
H3K27me2	12%	30%	25%	39%	59%	33%	55%	75%	5.5E+09	1.9E+10	3.7E+10	6.6E+10
H3K27me3	21%	34%	20%	27%	36%	36%	53%	40%	2.3E+10	2.8E+10	2.9E+10	8.3E+10
H3K27ac	6%	13%	15%	16%	98%	97%	89%	108%	9.9E+09	3.6E+09	2.4E+10	3.1E+10
H3K36me1	16%	17%	18%	9%	38%	28%	43%	53%	9.7E+09	1.7E+10	5.3E+10	9.8E+09
H3K36me2	21%	32%	24%	30%	91%	37%	52%	23%	4.9E+10	1.8E+10	2.1E+10	6.2E+10

Single mark	Average abundance						CV						Average Intensity			
	Hela	M	H	T	Hela	M	H	T	Hela	M	H	T	M	H	T	T
H33K36me3	12%	23%	13%	8%	52%	45%	53%	48%	9.5E+09	1.7E+10	1.3E+10	1.3E+10	1.7E+10	1.3E+10	1.3E+10	1.5E+10
H3K56me1	26%	14%	19%	24%	51%	26%	47%	64%	1.2E+09	6.6E+08	1.3E+09	1.3E+09	6.6E+08	1.3E+09	1.3E+09	7.4E+08
H3K56me2	11%	8%	7%	11%	85%	32%	85%	70%	9.2E+08	3.9E+08	2.3E+08	2.3E+08	3.9E+08	2.3E+08	2.3E+08	2.0E+08
H3K56me3	20%	22%	21%	23%	52%	25%	31%	30%	1.2E+09	1.4E+09	1.4E+09	1.4E+09	1.4E+09	1.4E+09	1.4E+09	8.7E+08
H3K56ac	22%	18%	18%	12%	46%	33%	46%	34%	1.7E+09	1.1E+09	1.3E+09	1.3E+09	1.1E+09	1.3E+09	1.3E+09	6.8E+08
H3K79me1	7%	8%	20%	18%	110%	83%	24%	44%	4.9E+08	3.5E+08	1.4E+08	1.4E+08	3.5E+08	1.4E+08	1.4E+08	2.4E+09
H3K79me2	26%	57%	4%	1%	64%	30%	53%	55%	3.7E+09	6.0E+09	2.3E+09	2.3E+09	6.0E+09	2.3E+09	2.3E+09	5.2E+07
H3K79me3	42%	18%	36%	32%	62%	78%	39%	32%	3.3E+09	2.3E+09	3.3E+09	3.3E+09	2.3E+09	3.3E+09	3.3E+09	6.8E+09
H3K79ac	19%	6%	17%	15%	43%	68%	103%	52%	1.9E+09	5.0E+08	1.9E+09	1.9E+09	5.0E+08	1.9E+09	1.9E+09	8.2E+09
H3K122ac	23%	21%	27%	43%	96%	61%	95%	39%	4.5E+08	1.5E+09	2.3E+08	2.3E+08	1.5E+09	2.3E+08	2.3E+08	2.0E+09
H4K5ac	19%	20%	27%	24%	16%	37%	13%	19%	2.9E+10	2.7E+10	3.0E+10	3.0E+10	2.7E+10	3.0E+10	3.0E+10	1.6E+10
H4K8ac	16%	16%	25%	18%	7%	30%	13%	33%	2.2E+10	2.0E+10	2.3E+10	2.3E+10	2.0E+10	2.3E+10	2.3E+10	1.3E+10
H4K12ac	19%	18%	26%	19%	10%	29%	13%	29%	2.7E+10	2.3E+10	3.0E+10	3.0E+10	2.3E+10	3.0E+10	3.0E+10	1.4E+10
H4K16ac	26%	24%	32%	32%	26%	36%	14%	18%	4.3E+10	3.2E+10	4.5E+10	4.5E+10	3.2E+10	4.5E+10	4.5E+10	2.4E+10
H4K20me1	65%	52%	50%	55%	25%	12%	18%	21%	7.2E+08	8.6E+08	5.9E+08	5.9E+08	8.6E+08	5.9E+08	5.9E+08	5.5E+08
H4K20me2	3%	10%	1%	1%	123%	176%	141%	106%	1.0E+08	4.6E+08	2.5E+07	2.5E+07	4.6E+08	2.5E+07	2.5E+07	3.6E+06
H4K20me3	2%	2%	1%	2%	90%	81%	117%	74%	4.3E+07	2.2E+07	3.6E+07	3.6E+07	2.2E+07	3.6E+07	3.6E+07	9.7E+06
H4K20ac	8%	5%	8%	7%	121%	93%	42%	45%	2.6E+08	1.5E+08	9.1E+07	9.1E+07	1.5E+08	9.1E+07	9.1E+07	6.4E+07

Table 2

The average relative abundances and coefficient of variations (CV) of histone single PTMs on H3 and H4 in four cell lines. Based on relative abundances from seven cell counts, average relative abundances and CVs were calculated respectively using first two highest cell counts (namely Ab 2 and CV 2 as in Supplementary Table 3 for detailed information), first three highest (namely Ab 3 and CV 3), four, five, six and all seven cell counts). The relative abundances and CVs shown in this table are the average relative abundances and CVs of Ab 2-7 and CV 2-7. This was to characterize the influence of different cell counts regarding different single marks. Some easy marks are quite reliable at all seven cell counts, some marks are generally consistent at higher cell numbers (1,000,000 or 2,000,000 cells shown in orange), and some marks are much more variable.

Hela			Myoblast			hESC			293T				
Single Mark	Abundance	CV	CV 3	Abundance	CV	CV 3	Abundance	CV	CV 2	CV 3	Abundance	CV	CV 3
H3K4me1	8±0%	22±5%	23%	12±2%	72±11%	83%	10±1%	30±3%	27%	25%	9±0%	11±4%	7%
H3K4me2	3±0%	33±3%	30%	2±0%	23±7%	12%	1±0%	11±6%	8%	6%	1±0%	20±10%	13%
H3K4me3	0±0%	45±17%	41%	0±0%	50±17%	57%	0±0%	17±11%	6%	8%	1±0%	16±5%	15%
H3K4ac	0±0%	44±23%	22%	5±1%	47±23%	61%	0±0%	137±17%	117%	142%	0±0%	103±10%	112%
H3K9me1	11±1%	11±7%	3%	9±0%	7±0%	7%	10±1%	18±10%	5%	7%	7±1%	21±14%	14%
H3K9me2	46±2%	6±4%	1%	45±1%	6±1%	6%	38±2%	15±2%	17%	13%	54±1%	7±2%	5%
H3K9me3	20±1%	7±3%	3%	30±0%	3±0%	3%	22±1%	17±3%	15%	11%	24±2%	16±7%	8%
H3K9ac	11±0%	7±4%	3%	5±0%	13±1%	12%	10±1%	47±10%	59%	58%	5±0%	13±4%	7%
H3K14ac	35±1%	5±3%	1%	27±1%	5±2%	2%	21±1%	29±6%	36%	35%	29±2%	9±6%	2%
H3K18me1	1±1%	89±60%	34%	0±0%	29±17%	9%	1±1%	59±42%	11%	10%	2±2%	103±61%	33%
H3K18ac	12±1%	19±9%	10%	8±1%	21±8%	19%	11±1%	15±5%	12%	15%	10±0%	17±6%	15%
H3K23me1	1±1%	80±57%	50%	1±0%	7±7%	5%	1±0%	64±23%	55%	45%	1±1%	91±72%	19%
H3K23ac	24±2%	11±6%	3%	18±2%	17±6%	10%	19±1%	12±6%	11%	11%	23±0%	6±1%	6%
H3K27me1	15±4%	137±13%	144%	19±3%	56±11%	68%	27±1%	9±3%	6%	6%	13±6%	78±22%	112%
H3K27me2	14±1%	29±7%	31%	41±3%	17±10%	16%	24±5%	43±21%	19%	15%	18±3%	90±15%	99%
H3K27me3	2±1%	53±23%	23%	14±4%	64±11%	51%	19±3%	37±25%	2%	9%	47±13%	70±19%	42%
H3K27ac	4±1%	87±25%	100%	9±2%	72±8%	73%	6±1%	96±19%	112%	126%	2±1%	120±16%	141%
H3K36me1	5±2%	60±11%	69%	15±4%	62±15%	43%	17±4%	68±10%	73%	50%	47±13%	70±20%	41%
H3K36me2	15±1%	41±4%	36%	43±4%	21±7%	12%	27±5%	44±15%	26%	23%	16±1%	78±24%	89%
H3K36me3	12±5%	178±27%	156%	11±3%	96±13%	106%	13±1%	44±5%	48%	36%	6±2%	78±13%	96%
H3K27me1	20±1%	35±7%	45%	30±5%	32±10%	20%	22±2%	30±7%	24%	27%	11±3%	37±15%	29%

Single Mark	Hela			Myoblast			hESC			293T		
	Abundance	CV	CV 3	Abundance	CV	CV 3	Abundance	CV	CV 3	Abundance	CV	CV 3
H33K27me2	9±3%	66±18%	72%	29±2%	26±8%	18%	23±3%	68±10%	82%	50±12%	66±13%	47%
H33K27me3	25±3%	26±10%	13%	36±4%	27±11%	11%	21±2%	49±21%	10%	28±1%	41±7%	43%
H33K27ac	8±2%	94±16%	65%	10±4%	75±31%	40%	19±4%	70±18%	75%	22±5%	68±31%	55%
H33K36me1	14±1%	52±12%	64%	19±2%	19±9%	8%	18±2%	54±9%	51%	7±2%	63±8%	71%
H33K36me2	26±5%	102±12%	100%	32±2%	37±10%	36%	23±3%	47±18%	13%	28±2%	18±9%	14%
H33K36me3	15±2%	45±4%	45%	29±5%	35±13%	19%	12±2%	47±9%	40%	7±1%	43±9%	33%
H3K56me1	20±5%	57±11%	41%	13±1%	34±7%	42%	16±2%	46±7%	32%	15±6%	71±16%	69%
H3K56me2	14±3%	84±5%	84%	7±1%	41±11%	45%	6±1%	56±23%	23%	12±2%	74±14%	79%
H3K56me3	17±2%	46±8%	40%	23±2%	24±4%	24%	24±3%	22±7%	16%	25±1%	35±10%	39%
H3K56ac	26±3%	31±16%	19%	20±1%	21±12%	11%	21±3%	32±18%	7%	14±2%	31±6%	31%
H3K79me1	4±1%	66±23%	59%	4±2%	56±26%	38%	18±1%	24±5%	14%	15±2%	32±12%	23%
H3K79me2	35±8%	43±24%	11%	59±3%	34±9%	36%	3±1%	63±13%	45%	1±0%	41±13%	27%
H3K79me3	33±9%	62±22%	57%	18±3%	84±7%	89%	33±2%	22±9%	22%	36±4%	27±5%	24%
H3K79ac	19±2%	45±10%	48%	4±1%	86±12%	96%	18±3%	79±18%	50%	16±1%	60±13%	61%
H3K122ac	15±9%	126±25%	120%	18±3%	51±12%	33%	15±11%	85±55%	6%	42±5%	36±7%	26%
H4K5ac	21±1%	11±4%	6%	21±3%	40±19%	61%	27±1%	11±1%	10%	23±1%	18±3%	22%
H4K8ac	17±0%	7±1%	6%	17±2%	28±15%	43%	26±1%	17±3%	23%	18±2%	31±8%	39%
H4K12ac	20±1%	9±1%	8%	19±2%	30±14%	46%	27±1%	14±1%	15%	19±2%	26±7%	32%
H4K16ac	30±3%	18±6%	11%	26±4%	38±18%	58%	32±1%	12±5%	2%	33±3%	17±7%	22%
H4K20me1	56±7%	16±10%	5%	51±1%	12±2%	13%	53±3%	15±3%	14%	55±3%	17±3%	15%
H4K20me2	6±2%	79±42%	47%	18±9%	124±47%	93%	1±1%	104±29%	61%	1±0%	90±15%	99%
H4K20me3	3±1%	64±17%	50%	1±0%	55±27%	42%	2±1%	96±19%	96%	2±0%	58±28%	51%
H4K20ac	12±5%	94±37%	69%	6±1%	104±20%	112%	8±1%	26±8%	26%	6±1%	36±8%	27%