HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2021 August 15.

Published in final edited form as:

Nat Methods. 2021 March ; 18(3): 303–308. doi:10.1038/s41592-020-01052-9.

Decoding the Protein Composition of Whole Nucleosomes with Nuc-MS

Luis F. Schachner[1,2], **Kevin Jooß**[1,3], **Marc A. Morgan**[4,5], **Andrea Piunti**[4,5], **Matthew J. Meiners**[6], **Jared O. Kafader**[1,3], **Alexander S. Lee**[2,3,5], **Marta Iwanaszko**[4,5], **Marcus A. Cheek**[6], **Jonathan M. Burg**[6], **Sarah A. Howard**[6], **Michael-Christopher Keogh**[6], **Ali Shilatifard**[4,5], **Neil L. Kelleher**[1,2,3,4,5,*]

[1]Department of Chemistry, Northwestern University, Evanston, Illinois, USA

[2]The Chemistry of Life Processes Institute, Northwestern University, Evanston, Illinois, USA

[3]The Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, USA

[4]Simpson Querrey Institute for Epigenetics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[5]Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[6]EpiCypher, Inc., Research Triangle Park, North Carolina, USA

Abstract

Current proteomics approaches disassemble and digest nucleosome particles, blurring the readout of the “histone code”. To preserve nucleosome-level information, we developed Nuc-MS which displays the landscape of histone variants and their PTMs in a single mass spectrum. Combined with immunoprecipitation, Nuc-MS quantified nucleosome co-occupancy of histone H3.3 with variant H2A.Z (6-fold over bulk) and the co-occurrence of oncogenic H3.3K27M with euchromatic marks (e.g., a >15-fold enrichment of H3K79me2). Nuc-MS is highly concordant with *ChIP-seq* and offers a new readout of nucleosome-level biology.

*Address reprint requests and communications to: Neil L. Kelleher, 2170 Tech Dr., Silverman Hall, Northwestern University, Evanston, Illinois, 60208, United States, Ph. 847-491-3731, Fax. 847-467-1566, n-kelleher@northwestern.edu.

Author Contributions

LFS performed Nuc-MS data acquisition and analysis. KJ assisted with proteoform quantitation and MS analysis. AP, MM, ASL and AS prepared and made available H3.3K27M-FLAG and H3.3WT-FLAG mononucleosomes. MM and MI conducted and analyzed *ChIP-seq* experiments. JOK assisted with acquisition of multiplexed individual ion MS data on endogenous nucleosomes. MJM, MAC, JMB and SAH assisted with synthesis, purification, and verification of modified nucleosomes. MCK coordinated modified nucleosome synthesis and provided insightful feedback on the manuscript. LFS and NLK conceived of the project and wrote the manuscript.

Competing Financial Interests

NLK serves as a consultant to Thermo Fisher Scientific. EpiCypher is a commercial developer and supplier of reagents, including the recombinant semi-synthetic modified nucleosomes (dNucs) used in this study. The other authors declare no competing interests.

Introduction

Chromatin is a highly structured fiber of nucleosome core particles, octameric complexes of histone proteins that bind, compact and rearrange DNA inside the nucleus.[1] Access to chromatin is regulated in part by post-translational modifications to the four core histones in a nucleosome, H2A, H2B, H3, and H4, and by their substitution with isoforms and variants.[2] These nucleosome-level changes encode structural information for nuclear effectors that trigger defined cellular events critical to health and disease.[3-6] Many studies also indicate that the precise arrangement of histone marks presented on nucleosomes is necessary for the recruitment of multi-valent readers (e.g. BPTF binding to both H3K4me3 and H4K16ac for transcription regulation).[7-9] Thus, new methods are needed that can provide nucleosome-level compositional information underlying epigenetic phenomena (*i.e.*, a nucleosome code).[1]

For decades, affinity reagents, mass spectrometry (MS) and proteomics have played key roles in characterizing the histone isoforms and PTMs involved in epigenetic regulation.[10, 11] However, existing protocols in mass spectrometry remove the linkage between modifications and their nucleosomes of origin by denaturing[10] and digesting[12] the constituent histones prior to MS (Fig. 1a, **at left).** Specifically, digesting histone mixtures into small peptides forfeits clear correlations among PTMs, blurring assertions about the original composition of the intact histones in the mixture. Similarly, denaturing a nucleosome population destroys information regarding co-localization of histone isoforms and PTMs within the intact particle. The inference problem resulting from such procedures diminishes our understanding of the “nucleosome code”: namely, the organization and impact of co-occurring modifications, isoforms and mutations in epigenetics and disease pathogenesis.[6, 9]

To close these knowledge gaps and shift from inference to direct measurement, we developed Nuc-MS, a method based on electrospray operated in ‘native’ mode that neither denatures nor digests nucleosomes,[10] and consists of three stages of tandem MS[13] to capture the protein composition of synthetic and endogenous nucleosomes. Here, we apply Nuc-MS to several systems to demonstrate three central points: (1) Nuc-MS provides a wide-lens, quantitative view of the histone modification landscape of nucleosomes in a single spectrum; (2) correlations among histone PTMs and variants are preserved; and (3) the method is highly concordant and complimentary with *ChIP-seq*.

Results

We begin with synthetic nucleosomes[14] to establish proof-of-concept and benchmark Nuc-MS on highly defined samples with known structural stabilities,[15] providing the first example of whole nucleosome fragmentation by tandem MS (Fig. 1a, **far right**). For unmodified nucleosomes containing 108.6 kDa of protein and 91.3 kDa of DNA, the Nuc-MS process is shown in Fig. 1b. In brief, intact nucleosomes were measured with a 3.1 Da error. Isolation and activation of the 35+ charge state of the nucleosome using high-energy collisional fragmentation (HCD) resulted in the ejection of all core histones, whose intact

masses were measured with isotopic resolution (Fig. 1b, **MS^2**). Additional details of the three-stage process of Nuc-MS are provided in Supplemental Figs. 1, 2, and 3.

With this first report of controlled disassembly of nucleosome core particles, we next asserted the *quantitative* nature of Nuc-MS by mixing two synthetic nucleosomes bearing H3K36me1 and H3K36me2 in a 1:1 ratio. Accordingly, the Nuc-MS readout of this sample showed a ratio of 49.2 ±2.5% and 50.8 ±3.3% (n = 3) for the integrated areas of ejected H3K36me1 and H3K36me2 proteoforms, respectively. This result demonstrates that relative quantitation – measuring proteoform abundances for the same histone – is possible and accurate by Nuc-MS (Supplemental Fig. 4). To further benchmark the assay, we used the three-stages of Nuc-MS to decode the composition of nucleosomes treated *in vitro by* acetyl-transferase p300/CBP associated Factor (PCAF) and methyl-transferase Polycomb repressive complex 2 (PRC2) (Supplemental Figs. 5-6).

Given the low bias of Nuc-MS to capture the proteoform landscape of histones in a population of nucleosomes, we challenged the approach with the variants and PTM patterns present on endogenous nucleosomes from human HEK 293T and HeLa cells (Fig. 2 and Supplemental Fig. 7). The MS^1 measurement of intact mononucleosomes profiles the mass distribution of these mononucleosomes, centered at ~200 kDa, and is consistent with the predominant source of nucleosome heterogeneity arising from differences in DNA-strand length after MNase digestion (Supplemental Fig. 8). Upon activation of the *entire* population of mononucleosomes for the tandem MS experiment, it proved straightforward to eject and detect >115 histone proteoforms (histone variants and PTMs) at high resolution in the same spectrum (Fig. 2a; Supplemental Table 1).

Examination of spectral regions for each core histone reveals a clear snapshot of the proteoform distributions present in bulk chromatin at >1% relative abundance. Of note is the quantitative readout of H2A and H2B distributions of gene family members, a challenging measurement to make by any other technique, and unveils cell-specific isoform distributions (insets in Fig. 2a-b, Supplemental Fig. 7).[10] Additionally, the major proteoform of histone H4 was characterized by tandem MS^3 as H4K20me2-$N_t$ac in both cell lines (Supplemental Fig. 7c,d). Thus, a Nuc-MS experiment yields a single MS^2 spectrum displaying the dominant histone proteoform landscape in cells without upfront separation or data ‘recombineering’.

To further probe the utility of this new data type in chromatin research, we tested the agreement between Nuc-MS and *ChIP-seq* in quantifying the co-enrichment of marks and variants within H3.3-containing nucleosomes. We chose H3.3 given that previous reports showed H2A.Z co-localizing with this variant, thereby creating “unstable” nucleosomes that facilitate access for transcription factors.[16] To this end, we transfected HEK cells with a plasmid containing the *H3.3F3A* gene and a C-terminal FLAG-HA epitope-tag extension. Next, we immuno-precipitated H3.3-FLAG-HA nucleosomes via anti-FLAG. The Nuc-MS readout for these H3.3-enriched nucleosomes showed six-fold co-enrichment of H2A.Z with H3.3 relative to bulk chromatin ($p = 2.7\text{x}10^{-7}$, Fig. 2b, see H2A.Z peaks at left). In parallel, we also found that H3.3 coincides with a higher abundance of H4 un- and mono-methylation (H4 proteoforms with “+3” and “+4” methyl equivalents in Supplemental Fig. 9,

Supplemental Table 2). Overall, the co-occurrence of H2A.Z, H3.3, and hypo-methylated H4 proteoforms suggests that these nucleosomes are in regions of active transcription[17, 18] and high H3-H4 turnover.[19]

Parallel *ChIP-seq* analysis revealed highly similar peak tracks for H3.3 and H2A.Z (Fig. 2c), summarized by a heatmap that aggregates ~73,000 reads into four clusters (Fig. 2d). The elevated signal intensity in clusters 1-3 estimates up to 20% co-occurrence for H3.3 and H2A.Z in introns and promoters (Supplemental Fig. 10), consistent with the Nuc-MS measurement of 13.7 ±0.2% co-occurrence (n = 3).

Finally, we used Nuc-MS to profile the composition of nucleosomes harboring H3.3 K27M. This 'toxic' oncohistone arises in up to 80% of diffuse intrinsic pontine gliomas (DIPG), an aggressive tumor of the pediatric brain stem.[20] This mutation is a model system for investigating the profound cellular effects of changing a single histone residue.[21, 22] Previous work found that H3.3K27M associates with K27ac on H3WT, and that the mutation and K27ac co-localize with RNA pol II, indicating these marks are present in sites of active transcription.[23]

As above, we immuno-precipitated mononucleosomes from HEK 293T cells expressing the inducible transgenes H3.3 K27M-FLAG-HA or H3.3 WT-FLAG-HA as a control (Supplemental Fig. 11). Readout by Nuc-MS displayed the dominant histone proteoform landscapes for H3.3K27M and H3.3WT nucleosomes (Fig. 3a, 3b). Importantly, H3.3K27M nucleosomes contain a 33.7 ±1.4% increase in H4K16ac, a proteoform correlated with active transcription[24] and accessible chromatin ($p = 1.01\text{x}10^{-4}$, Supplemental Fig. 12). The data also reveal that H3.3K27M has a lower methylation state than H3.3WT upon loss of the K27 methylation site (Fig. 3d,h, Supplemental Table 3). In terms of nucleosome symmetry, we detected far lower levels of the wild-type H3 tail relative to H3.3K27M-FLAG and H3.3WT-FLAG, indicating that these nucleosomes are >80% homotypic for H3.3-FLAG tails (Supplemental Fig. 13).[9] Methylated proteoforms of endogenous wild-type H3.2 and H3.1 were detected in H3.3WT-FLAG nucleosomes, also indicating that <20% of nucleosomes were heterotypic and affirming that Nuc-MS can quantitatively interrogate the homotypic vs. heterotypic nature of nucleosomes.

Fragmentation of the H3.3K27M-FLAG proteoforms detected the high co-occurrence of K27M with H3K79me2, a mark linked to active transcription (Supplemental Fig. 14).[25] This finding is notable as H3.3K79me2 has been reported present in <2.5% of H3.3K79 proteoforms,[26] and is therefore enriched in H3.3K27M nucleosomes by >15-fold relative to bulk. Examination of H2A proteoforms revealed elevated acetylation and a three-fold decrease in H2A.Z abundance in H3.3K27M-FLAG nucleosomes compared to H3.3WT-FLAG (Fig. 3e, Supplemental Fig. 15), suggesting that K27M may be correlated with changes in H2A.Z occupancy.[27] Additional data and discussion on histone ubiquitination is provided in Supplemental Fig. 16.[28] Authenticating the differential enrichment of marks detected by Nuc-MS, western blot analyses confirmed co-enrichment of H3.3K27M with H3.3K79me2 and H4K16ac (Supplemental Fig. 17). Moreover, orthogonal *ChIP-seq* on the co-localization of H3.3 variant with H4K16ac and H3K79me2 showed strong co-association

of these marks (Fig. 3f-h; Supplemental Figs. 10, 18-19), informing an integrated model for average nucleosome composition present at promoters (Fig. 3b).

Discussion

The results above establish a platform for the direct compositional analysis of intact nucleosome particles by mass spectrometry, capturing the status of dozens of histone proteoforms and their PTMs in a single, quantitative experiment (e.g., Figs. 2a and 3a). Nuc-MS detected >115 histone proteoforms in the same spectrum without chromatography or optimization for specific histone marks (Supplemental Table 1). This new data-type offers a low-bias survey of the modification landscape in a population of nucleosomes down to ~1% proteoform abundance. Interestingly, we detected a variety of unidentified histone proteoforms in HeLa nucleosomes, many of which have masses consistent with 'short H2A' variants[29] (Supplemental Table 1, bottom). The assay can thus quantify the co-enrichment of multiple variants and PTMs simultaneously without *a priori* knowledge. As such, it is complementary to targeted methods like *ChIP-seq* or CUT&RUN[30] that provide positional information on specific histone marks.

Notably, coupling Nuc-MS with immuno-affinity techniques enables a profile of the nucleosome modification landscape in sub-regions of chromatin. The nucleosome particle lies at the intersection of atomic regulatory elements acting upon histones/DNA and the diverse phenomena of epigenomic regulation operative at higher-orders of chromatin structure.[1] Thus, an integrated readout of nucleosome composition, either from bulk or after immuno-enrichment, connects specific arrangements of PTMs and variants to chromatin-level effects. For example, Nuc-MS quantified the co-enrichment of H2A.Z with H3.3,[16] as well as the linked hypo-methylation of histone H4, painting a picture of the nucleosome composition in regions of active transcription and high H3-H4 turnover.[31] Our findings on the nucleosomal characteristics that follow the loss of the H3K27 methylation site (Fig. 3b) provide observations that inform the model of active transcription and chromatin de-condensation correlated with the K27M mutation, complementing more established data streams from many studies using single-PTM measurement technologies like *ChIP-seq*.[28, 32]

Future development of Nuc-MS will incorporate alternative fragmentation techniques to improve sequence coverage and identifications of ejected proteoforms, even on samples containing <1000 nucleosomes. We envisage coupling this approach with immuno-enrichment of rare marks to reveal the extent and function of a 'nucleosome code', where co-occurring histone proteoforms combine to potentiate regulation of gene expression and progression of disease.

Online Methods

Nucleosome Assembly

Nucleosome particles were assembled by salt dialysis.[33] Specifically, 50 μg of 601 nucleosome positioning sequence DNA (*EpiCypher*) was resuspended in 50 μL nuclease-free 2M NaCl, heated to 37°C and mixed thoroughly. We then mixed the following reagents in order: 54.2 μL 2M NaCl, 50 pmol 601 DNA (45.8 μL), 50 μL of 20 μM H2A/H2B dimer

(*New England BioLabs*), and 50 μL of 10 μM H3/H4 tetramer (*New England BioLabs*). The NaCl concentration was gradually lowered by adding increasing volumes of 10 mM Tris HCl (pH 8) every thirty minutes (final concentrations at each step: 2M; 1.48M; 1.0M; 0.6M; 0.25M). Sample was then added to a 20 kDa MWCO Slide-A-Lyzer MINI dialysis device (*Thermo Fisher Scientific*) and dialyzed against three buffer changes (the second being overnight) of 100 mM ammonium acetate (pH 6.8). After dialysis, samples were exchanged into 10 mM Tris HCl, pH 8 and concentrated to 100 μL using 30 kDa MWCO spin filters (*Millipore-Sigma*).

For MNase digestion of free DNA, samples were brought to 200 μL with 10 mM Tris buffer (pH 8), and supplemented with 4 μL of 100 mM $CaCl_2$ (2 mM final), and 20 μL of 20 U/μL MNase (*New England BioLabs*, 1/1000 dilution of commercial stock). Reactions were digested for 1 minute; quenched with 2 μL 500 mM EDTA and thorough pipetting; then desalted 4-5 times into 150 mM ammonium acetate (pH 6.8) using 100 kDa MWCO spin filter. Effective nucleosome assembly was confirmed by native TBE gels (*BioRad*). The nucleosome ubiquitinated at H2AK119 (*EpiCypher* #16-0363) and mono- and di-methylated at K36 (*EpiCypher* #16-0322 and #16-0319) were desalted 10 times into 150 mM ammonium acetate using 30 kDa MWCO spin filters prior to MS analysis.

Nucleosome Enzymatic Modification (PCAF and PRC2)

PCAF: For nucleosome acetylation, the following reaction was set up: 50 μL 5x Histone Acetyltransferase (HAT) buffer (250 mM Tris HCl, 50% glycerol, 0.5 mM EDTA, 5 mM DTT), 8 μL Acetyl CoA at 10 mM, 40 μL of 2.8 μM nucleosome, 2 μL recombinant PCAF (aa 492-658) at 197 μM, 150 μL 150 mM ammonium acetate. Mixture incubated at RT for 5 minutes and desalted using 30 kDa MWCO spin filters (*Millipore-Sigma*) into 150 mM ammonium acetate for MS analysis.

PRC2: for nucleosome methylation, the following reaction was set up: 50 μL nucleosome at 2.8 μM, 1 μL 100 mM TCEP, 0.2 μL 1M $MgCl_2$, 3.33 μL 1.5 M Tris (pH 8.8), 1 μL 10 mM SAM, 2 μL PRC2 (2.8 μM, Active Motif), 42.47 μL water. Mixture was incubated at RT for 18 hours and desalted into 150 mM ammonium acetate for MS analysis using 100 kDa MWCO spin filters (*Millipore-Sigma*).

Preparation of HEK and HeLa Mononucleosomes

Harvested cells were resuspended in 2.5 pelleted cell volumes (PCV) of buffer A (10mM HEPES, pH 7.9, 10mM KCl, 340mM Sucrose, 0.5mM PMSF, 0.5mM Benzamidine) supplemented with 5 mM βME and 1x cOmplete™ EDTA-free protease inhibitor cocktail (Roche). A similarly supplemented buffer A volume (2.5 PCV) containing 0.2% (v/v) Triton X-100 detergent was also prepared during this time. The resuspended cell pellet was completely homogenized by pipetting and light vortexing and the detergent containing buffer A added to the cell suspension, mixed by inversion, and allowed to incubate on ice for 10 minutes with occasional gentle mixing to thoroughly lyse the cells. The resultant nuclei were pelleted via centrifugation (1300 x g, 4°C, 5 minutes) and the supernatant aspirated and discarded.

The nuclei pellet was next resuspended in 6 PCV of supplemented buffer A and spun through a 35 mL sucrose cushion (10 mM HEPES pH 7.9, 30% (w/v) sucrose) in a 50 mL conical tube supplemented with 5 mM βME and 1x cOmplete™ EDTA-free protease inhibitor cocktail (1300xg, 4°C, 10 minutes). The sucrose cushion purification of the nuclei was repeated until the cushion remained free of cell debris after centrifugation.

The pelleted nuclei were then gently resuspended in 2 PCV of buffer A (supplemented as above). The DNA concentration in the nuclei resuspension was calculated (A260) by hypertonic lysing of the nuclei by mixing 5 μL of the nuclei resuspension with 495 μL 2M NaCl. The suspension was diluted to a DNA concentration of 1.5 μg/μL and $CaCl_2$ added to the nuclei to a final concentration of 1 mM. The suspension was aliquoted into 2 mL microcentrifuge tubes and equilibrated in a 37°C water bath for 5 minutes. Digestion of the nuclei to nucleosomal species was initiated by adding 1U MNase (*Worthington*; prepared in supplemented buffer A) for every 70 μg of DNA and the samples were allowed to incubate at 37°C for 15 minutes. The digestion was then quenched by adding 0.5 M EGTA to a final concentration of 10 mM, mixing via inversion, and placing the samples on ice. The final volume of the digested material was noted. Approximately 10 μg of DNA was purified (*QIAGEN* DNA cleanup kit) and resolved on a 2% agarose gel in 0.5x TBE to confirm the extent of digestion by fragment size distribution in the population.

2M NaCl was added dropwise to the digested chromatin while mixing on a magnetic stir plate at 4°C to a final concentration of 650 mM. The resulting material was cleared via centrifugation at 12,000xg at 4 °C to pellet any insoluble material before size exclusion chromatography. The material was fractionated via a HiPrep™ 26/60 Sephacryl® S-300 HR column equilibrated with supplemented buffer A, as above, containing 650mM NaCl using an ÄKTA Prime Plus FPLC (*GE Lifescience*). 10 μg samples of the individual fractions containing nucleosomes were purified (*QIAGEN* DNA cleanup kit) and resolved on a 2% agarose gel in 0.5x TBE. Only those fractions that contained mononucleosomal-associated DNA fragments (~150bp) were pooled for further analysis.

Mononucleosome FLAG-immunoprecipitation

Immunoprecipitation of FLAG-tagged H3.3WT and H3.3K27M histones was performed with 120 uL of anti-FLAG beads (*Sigma Aldrich*, #F1804) in 4mL of lysate, as previously.[21, 23]

Chromatin Immunoprecipitation for ChIP-seq

HEK293T cells ($1x10^8$) were crosslinked with 1% formaldehyde for 15 minutes and quenched with 0.125 M glycine. Cell lysis and chromatin preparation were performed as previously (Lee et al., 2006; Vo et al., 2017). For sonication, the sample was resuspended in 2 mL of ChIP buffer #3 and sonicated on a *Covaris* E220 sonicator for 5 minutes, 200 cycles per burst, 140W peak intensity pulse and 20% duty factor. Following sonication, samples were cleared by centrifugation at 20,000xg for 15 minutes. Chromatin was then diluted to a concentration of 0.4 mg/mL and Triton X-100 added to a final concentration of 1%. For each ChIP binding reaction, 0.4 mg of chromatin was incubated with 10 μL of antibody overnight at 4°C. The following day, 20 uL of Protein-A/G plus beads (*Santa Cruz* sc-2003) were

added to each ChIP and incubated for 2 hours. Beads were washed 5 times in ChIP wash buffer and 2 times in TEN buffer. ChIP'ed DNA was purified using phenol: chloroform as previously (Lee et al., 2006; Vo et al., 2017). Full references available in Supplemental Information. Antibodies used: H2A.Z (*Cell Signaling Technologies*, *CST*, #2718S) Lot #2; H2A.Z (*Abcam*, #ab4174) Lot# GR3198864-2; H3.3 (*EMD/Millipore*, #09-838) Lot #3310680; H4K16ac (*CST*, #13534S) Lot #3; H3K79me2 (*CST*, #5427S) Lot#4. Western blot analysis was performed using 1:1000 dilution of these antibodies.

ChIP-sequencing and data processing

ChIP-seq libraries were synthesized using the Illumina TruSeq kit, size selected (200-400bp) with SPRI select beads and sequenced on an *Illumina* Novaseq instrument. Base-calling was performed using bcl2fastq v2.17.1.14 and read quality was assessed with FastQC v0.11.9 (Andrews, 2010). *ChIP-seq* reads were aligned to the human genome (hg38) using Bowtie v0.12.9 (Langmead et al., 2009) allowing for 2 mismatches and retaining only uniquely mapped reads. MACS v2.1.0 (Zhang et al., 2008) was used to call peaks using a false discovery rate filter of 0.01. Annotation of *ChIP-seq* data was performed using HOMER v4.10 (Heinz, 2010) and Pearson correlation for samples calculated using R package DiffBind v3.0.6 (Ross-Innes et al. 2012). Functional gene enrichment analysis was performed using R package clusterProfiler v3.18.0 (Yu et al., 2012). The following R packages were also used in the analysis: samtools v1.6, deepTools v3.1.1 and bedtools v2.29.1. Resulting *P*-values shown in Supplemental Figure 19 are adjusted for multiple comparison using the Benjamini-Hochberg (BH) Procedure and are denoted in the color scale of the figure. Genome wide occupancy heatmaps were generated using deepTools v3.1.1 (Ramirez et al., 2016) and centered on H3.3 peaks. Full references available in the Supplemental Information. The ChIP-seq data presented in this manuscript have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE149484 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149484).

Native Mass Spectrometry

Nucleosomes samples at a concentration of 2 μM and desalted into 150 mM ammonium acetate using 30 kDa MWCO 0.5 mL spin filters (*Millipore-Sigma*). Samples were analyzed using a Q Exactive HF mass spectrometer with Extended Mass Range (QE-EMR) and a Q Exactive HF Ultra-High Mass Range (QE-UHMR), both by *Thermo Fisher Scientific*. Data were collected using XCalibur QualBrowser 4.0.27.10 (*Thermo Fisher Scientific*). The native electrospray platform is coupled to a three-tiered tandem MS process. First, the analysis of the intact nucleosome (MS^1) provides the total complex mass (reported as a deconvoluted neutral average mass value).[13] In stage two, the nucleosome is activated by collisions with nitrogen gas to eject histones (MS^2). In stage three, further vibrational activation of the ejected histones via collisions with nitrogen gas yields backbone fragmentation products from each monomer (MS^3) that are recorded at isotopic resolution (120,000 resolving power at *m/z* 400). These fragments can be mapped onto the primary sequence of the histones in order to localize posttranslational modifications. A step-by-step protocol for Nuc-MS data acquisition and analysis is available on the Nature Protocol

Exchange repository (DOI 10.21203/rs.3.pex-1288/v1) and in the Supplemental Information.[34]

QE-EMR parameters: The Nuc-MS workflow utilizes native electrospray ionization (nESI) source held at +2 kV, C-trap entrance lens voltage setting between 1.8 - 4 V, HCD gas pressure setting between 2-4 V, and CID voltage set at 15-25 V for desalting and 75-100 V for histone ejection. HCD energy set to 100-120 V for histone fragmentation with a pressure of 2. Microscans set to 20 and max injection time to 2000 ms for collection of fragmentation data.

QE-UHMR parameters: HCD gas pressure between 0.5-1 for detection of histones at isotopic resolution; in-source trapping voltage of −100 to −150 V for histone ejection; CE of 49-65 eV, microscans set to 20 and max inject time to 1000 ms for fragmentation of proteoforms.

Prior to using the QE-UHMR for nucleosome analysis, quantitative ejection of the PCAF- and PRC2-treated histones for fragmentation was achieved using front-end infrared activation coupled to the QE-EMR, which used a 20 W continuous-wave CO_2 laser (*Synrad Firestar V20*) at an average power of 1.2 W. The laser was attenuated with a 1.0 optical density (O.D.) nickel-coated zinc selenide neutral density filter and aligned unfocused to the inlet capillary with protected gold mirrors. Once we transitioned the Nuc-MS platform to the QE-UHMR, the in-source trapping capability proved to be a reliable method for ejecting histones at sufficient intensities for adequate fragmentation.

MS Data Analysis

Intact mass values for nucleosome complexes and ejected histones, the MS^1 and MS^2 measurements, were determined by deconvolution to convert data from the *m/z* to the mass domain using MagTran 1.03[35] (mass range: 15,000-300,000 Da; max no. of species: 10-15; S/N threshold: 1; mass accuracy: 0.05 Da; charge determined by: charge envelop only). Intact mass measurements are reported as neutral average masses; errors represent 1σ deviation from the mean of the masses calculated for all sampled charge states.

UniDec 3.2.0[36] was used to create isotopically resolved deconvoluted mass spectra, depicted in the butterfly diagrams. Data processing: Range 500 – 2500 Th, Bin every: 0; UniDec parameters: Charge Range: 5 – 15, Mass Range: 10 – 20 kDa, Sample mass every (Da): 0.05. Peak area values were calculated by integrating the assigned mass range of proteoforms in the deconvoluted mass spectrum.

High-resolution fragmentation data were processed using Xtract (Signal-to-Noise threshold ranging from 1-30, *Thermo Fisher Scientific*), mMass 5.5.0 (www.mmass.org), ProSight Lite 1.4[37] (precursor mass type: average; fragmentation method: HCD; fragmentation tolerance: 10-15 ppm), and TDValidator 1.0[38] (max ppm tolerance: 25 ppm; cluster tolerance: 0.35; charge range: 1-10; minimum score: 0.5; S/N cutoff: 3; Mercury7 Limit: 0.0001; minimum size: 2) to assign recorded fragment ions to the primary sequence of the subunits. Specifically, ProSight Lite and TDValidator were used to analyze fragmentation spectra in medium throughput to assign and validate *b* and *y* fragment ions to the histone

sequences, and for generating a p-score. mMass was used to interrogate individual fragment ions within a spectrum not identified by TDValidator or ProSight Lite. The histones H2A, H2B, H3, and H4 were identified by mapping backbone fragment ions to their amino acid sequence using ProSight Lite.[37] Unexplained mass shifts (Δm) observed at the MS^1, MS^2, and MS^3 levels for the intact complex and subunits, respectively, were manually interrogated using the UNIMOD database (http://www.unimod.org/modifications_list.php) as a reference for candidate modifications.

Statistics and Reproducibility

Plots for the relative quantitation of proteoforms in Figure 3 and Supplemental Figures 4, 9 and 13 were based on direct infusion experiments (n = 3). The scatter plots are centered on the mean relative peak area values, which are denoted in the bar chart. Individual data points are ordered left to right according to data file of origin. Statistical significance was evaluated using two-sided, two-sample t-tests. The α-values (based on $\alpha = 0.05$) were adjusted according to the Bonferroni correction and both α- and *p*-values are reported in Supplemental Tables 2-3.[39] The DNA/protein gel results shown in Supplemental Figures 1a, b, 11a-b, 13a-b, and 15a-b were each reproduced at least once with comparable results. Additional information on statistics and reproducibility is provided in the Life Sciences Reporting Summary.

Ion Collection and Data Acquisition for Individual Ion Mass Spectrometry (I^2MS)

Detailed methods for this new technique are reported in Kafader, *et al.* (2019).[40] Briefly, this new method uses direct assignment of charge states on individual ions inside an Orbitrap-style mass spectrometer with a harmonic potential. To provide populations of single ions of endogenous mononucleosomes, the transmission of the instrument was detuned to lower the number of ions entering the Orbitrap analyzer and achieve detection of a single ion per *m/z* value for each acquisition event (disabled Automatic Gain Control and maximum injection time between 0.03 - 1 ms). Both the *m/z* and charge (*z*) were necessary to determine the mass of the ion. In the Orbitrap portion of a Q Exactive instrument (*Thermo Fisher Scientific*), the *m/z* of ions were determined from the frequency of ion rotation around the central electrode; the charge, *z*, was given by the rate of the induced charge on the outer electrode, also known as Selective Temporal Overview of Resonant Ions (STORI), described in detail elsewhere.[40] Plotting of the I^2MS spectrum from this multiplexed, I^2MS procedure was achieved by binning ~1M acquired individual ions of mononucleosomes in 0.2 Da increments. In parallel, to validate the charge assignment, the calculated charge of the ions used for the individual ion MS spectrum were binned in quantized domains, as reported previously.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of General Medical Sciences P41 GM108569 for the National Resource for Translational and Developmental Proteomics at Northwestern University and NIH grants

S10OD025194 and RF1AG063903 (Kelleher lab) and R44GM116584, R44CA212733 and R44CA214076 (EpiCypher). LFS is a Gilliam Fellow of the Howard Hughes Medical Institute. Research in this publication is also supported by Thermo Fisher Scientific and a fellowship associated with the Chemistry of Life Processes Predoctoral Training Grant T32GM105538 at Northwestern University. A.P. is supported by the transition to independence grant K99CA234434-01. We also thank M. Senko, P. Compton, C. Koo, L. Szymczak, and M. McAnnally for technical assistance, and S. Judge and A. Rosenzweig for providing thoughtful suggestions to the manuscript.

Data Availability

MS^1, MS^2, and MS^3 spectra presented in the manuscript are available online in the MassIVE database under accession code MSV000085238 and can be visualized with Thermo Qual Browser.

References

1. Moller J; de Pablo JJ, Bottom-Up Meets Top-Down: The Crossroads of Multiscale Chromatin Modeling. Biophysical Journal 2020, 118 (9), 2057–2065. [PubMed: 32320675]
2. Campos EI; Reinberg D, Histones: annotating chromatin. Annu Rev Genet 2009, 43, 559–99. [PubMed: 19886812]
3. Bannister AJ; Kouzarides T, Regulation of chromatin by histone modifications. Cell Res 2011, 21 (3), 381–95. [PubMed: 21321607]
4. Portela A; Esteller M, Epigenetic modifications and human disease. Nat Biotechnol 2010, 28, 1057. [PubMed: 20944598]
5. Patel DJ; Wang Z, Readout of epigenetic modifications. Annual review of biochemistry 2013, 82, 81–118.
6. Zink LM; Hake SB, Histone variants: nuclear function and disease. Curr Opin Genet Dev 2016, 37, 82–9. [PubMed: 26826795]
7. Ruthenburg AJ; Li H; Patel DJ; Allis CD, Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol 2007, 8 (12), 983–94. [PubMed: 18037899]
8. Ichikawa Y; Connelly CF; Appleboim A; Miller TCR; Jacobi H; Abshiru NA; Chou H-J; Chen Y; Sharma U; Zheng Y; Thomas PM; Chen HV; Bajaj V; Müller CW; Kelleher NL; Friedman N; Bolon DNA; Rando OJ; Kaufman PD, A synthetic biology approach to probing nucleosome symmetry. eLife 2017, 6, e28836. [PubMed: 28895528]
9. Voigt P; LeRoy G; Drury WJ 3rd; Zee BM; Son J; Beck DB; Young NL; Garcia BA; Reinberg D, Asymmetrically modified nucleosomes. Cell 2012, 151 (1), 181–93. [PubMed: 23021224]
10. 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]
11. Shah RN; Grzybowski AT; Cornett EM; Johnstone AL; Dickson BM; Boone BA; Cheek MA; Cowles MW; Maryanski D; Meiners MJ; Tiedemann RL; Vaughan RM; Arora N; Sun Z-W; Rothbart SB; Keogh M-C; Ruthenburg AJ, Examining the Roles of H3K4 Methylation States with Systematically Characterized Antibodies. Molecular Cell 2018, 72 (1), 162–177.e7. [PubMed: 30244833]
12. Compton PD; Kelleher NL; Gunawardena J, Estimating the Distribution of Protein Post-Translational Modification States by Mass Spectrometry. Journal of Proteome Research 2018, 17 (8), 2727–2734. [PubMed: 29945451]
13. Belov ME; Damoc E; Denisov E; Compton PD; Horning S; Makarov AA; Kelleher NL, From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. Anal Chem 2013, 85 (23), 11163–73. [PubMed: 24237199]
14. Azegami N; Saikusa K; Todokoro Y; Nagadoi A; Kurumizaka H; Nishimura Y; Akashi S, Conclusive evidence of the reconstituted hexasome proven by native mass spectrometry. Biochemistry 2013, 52 (31), 5155–7. [PubMed: 23879667]

15. Lercher L; Raj R; Patel NA; Price J; Mohammed S; Robinson CV; Schofield CJ; Davis BG, Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. Nat Commun 2015, 6, 7978. [PubMed: 26305776]

16. Jin C; Zang C; Wei G; Cui K; Peng W; Zhao K; Felsenfeld G, H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet 2009, 41 (8), 941–5. [PubMed: 19633671]

17. Cui K; Zang C; Roh T-Y; Schones DE; Childs RW; Peng W; Zhao K, Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell 2009, 4 (1), 80–93. [PubMed: 19128795]

18. Talasz H; Lindner HH; Sarg B; Helliger W, Histone H4-Lysine 20 Monomethylation Is Increased in Promoter and Coding Regions of Active Genes and Correlates with Hyperacetylation. Journal of Biological Chemistry 2005, 280 (46), 38814–38822.

19. Svensson JP; Shukla M; Menendez-Benito V; Norman-Axelsson U; Audergon P; Sinha I; Tanny JC; Allshire RC; Ekwall K, A nucleosome turnover map reveals that the stability of histone H4 Lys20 methylation depends on histone recycling in transcribed chromatin. Genome Res 2015, 25 (6), 872–883. [PubMed: 25778913]

20. Lulla RR; Saratsis AM; Hashizume R, Mutations in chromatin machinery and pediatric high-grade glioma. Science Advances 2016, 2 (3), e1501354. [PubMed: 27034984]

21. Herz H-M; Morgan M; Gao X; Jackson J; Rickels R; Swanson SK; Florens L; Washburn MP; Eissenberg JC; Shilatifard A, Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. Science 2014, 345 (6200), 1065–1070. [PubMed: 25170156]

22. Brumbaugh J; Kim IS; Ji F; Huebner AJ; Di Stefano B; Schwarz BA; Charlton J; Coffey A; Choi J; Walsh RM; Schindler JW; Anselmo A; Meissner A; Sadreyev RI; Bernstein BE; Hock H; Hochedlinger K, Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. Nat Cell Biol 2019, 21 (11), 1449–1461. [PubMed: 31659274]

23. Piunti A; Hashizume R; Morgan MA; Bartom ET; Horbinski CM; Marshall SA; Rendleman EJ; Ma Q; Takahashi Y.-h.; Woodfin AR; Misharin AV; Abshiru NA; Lulla RR; Saratsis AM; Kelleher NL; James CD; Shilatifard A, Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. Nature Medicine 2017, 23 (4), 493–500.

24. Taylor GCA; Eskeland R; Hekimoglu-Balkan B; Pradeepa MM; Bickmore WA, H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res 2013, 23 (12), 2053–2065. [PubMed: 23990607]

25. Nguyen AT; Zhang Y, The diverse functions of Dot1 and H3K79 methylation. Genes Dev 2011, 25 (13), 1345–58. [PubMed: 21724828]

26. Sweet SMM; Li M; Thomas PM; Durbin KR; Kelleher NL, Kinetics of Re-establishing H3K79 Methylation Marks in Global Human Chromatin. Journal of Biological Chemistry 2010, 285 (43), 32778–32786.

27. Wang Y; Long H; Yu J; Dong L; Wassef M; Zhuo B; Li X; Zhao J; Wang M; Liu C; Wen Z; Chang L; Chen P; Wang Q.-f.; Xu X; Margueron R; Li G, Histone variants H2A.Z and H3.3 coordinately regulate PRC2-dependent H3K27me3 deposition and gene expression regulation in mES cells. BMC Biology 2018, 16 (1), 107. [PubMed: 30249243]

28. Lowe BR; Maxham LA; Hamey JJ; Wilkins MR; Partridge JF, Histone H3 Mutations: An Updated View of Their Role in Chromatin Deregulation and Cancer. Cancers (Basel) 2019, 11 (5), 660.

29. Talbert PB; Henikoff S, Histone variants — ancient wrap artists of the epigenome. Nature Reviews Molecular Cell Biology 2010, 11 (4), 264–275. [PubMed: 20197778]

30. Skene PJ; Henikoff S, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 2017, 6.

31. Huang C; Zhang Z; Xu M; Li Y; Li Z; Ma Y; Cai T; Zhu B, H3.3-H4 tetramer splitting events feature cell-type specific enhancers. PLoS Genet 2013, 9 (6), e1003558–e1003558. [PubMed: 23754967]

32. Zhao Z; Shilatifard A, Epigenetic modifications of histones in cancer. Genome Biology 2019, 20 (1), 245. [PubMed: 31747960]

References for Methods

33. Luger K; Mader AW; Richmond RK; Sargent DF; Richmond TJ, Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 1997, 389 (6648), 251–60. [PubMed: 9305837]
34. Schachner LF; Lee A; Kelleher NL, Protocol for Decoding the Protein Composition of Whole Nucleosomes with Nuc-MS: Sample Preparation, Data Acquisition and Analysis. Protocol Exchange 2020. DOI: 10.21203/rs.3.pex-1288/v1.
35. Zhang Z; Marshall AG, A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. J. Am. Soc. Mass Spectrom 1998, 9 (3), 225–33. [PubMed: 9879360]
36. Marty MT; Baldwin AJ; Marklund EG; Hochberg GKA; Benesch JLP; Robinson CV, Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. Analytical Chemistry 2015, 87 (8), 4370–4376. [PubMed: 25799115]
37. Fellers RT; Greer JB; Early BP; Yu X; LeDuc RD; Kelleher NL; Thomas PM, ProSight Lite: graphical software to analyze top-down mass spectrometry data. Proteomics 2015, 15 (7), 1235–8. [PubMed: 25828799]
38. Fornelli L; Srzentić K; Huguet R; Mullen C; Sharma S; Zabrouskov V; Fellers RT; Durbin KR; Compton PD; Kelleher NL, Accurate sequence analysis of a monoclonal antibody by top-down and middle-down orbitrap mass spectrometry applying multiple Ion activation techniques. Anal. Chem 2018, 90 (14), 8421–8429. [PubMed: 29894161]
39. Bland JM; Altman DG, Multiple significance tests: the Bonferroni method. BMJ 1995, 310 (6973), 170. [PubMed: 7833759]
40. Kafader JO; Melani RD; Durbin KR; Ikwuagwu B; Early BP; Fellers RT; Beu SC; Zabrouskov V; Makarov AA; Maze JT; Shinholt DL; Yip PF; Tullman-Ercek D; Senko MW; Compton PD; Kelleher NL, Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. Nature Methods 2020, 17 (4), 391–394. [PubMed: 32123391]

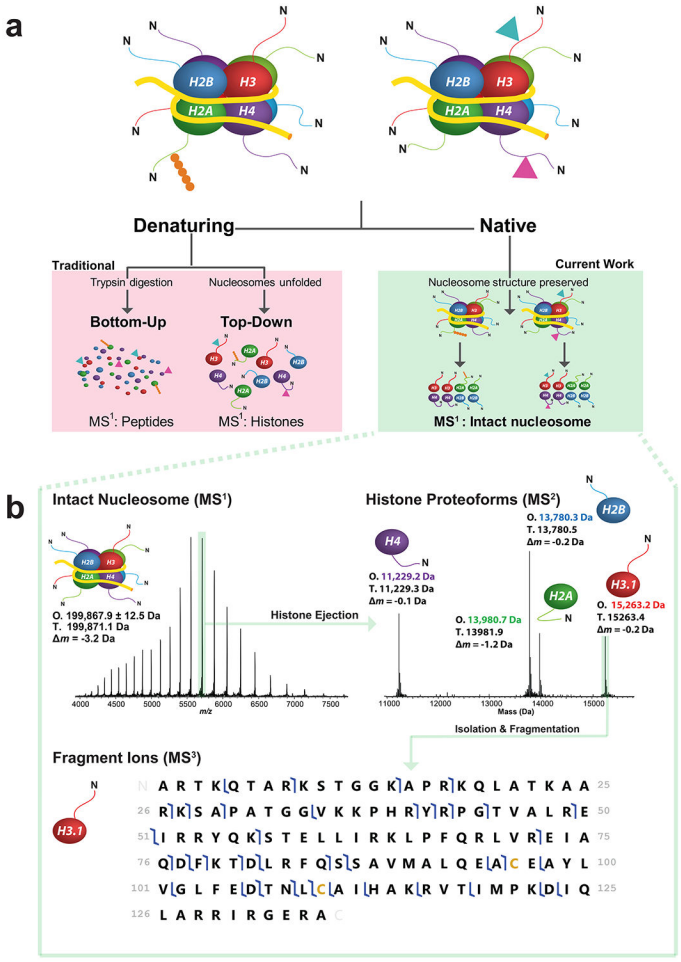

Figure 1. Three strategies of histone analysis, including Nuc-MS for the direct interrogation of intact nucleosomes.

(a) In the figure's hypothetical nucleosome mixture **(a, top)**, the Nuc-MS workflow **(a, right side)** detects the co-localization of H3 and H4 methylation (blue and pink triangles) and determines that the H2A ubiquitination (in orange) is present on a separate nucleosome. In contrast, traditional methods **(a, left side)** use either protease-derived peptides or whole histones under denaturing conditions to detect histone PTMs, blurring the modification states of intact nucleosomes in a mixture. Nuc-MS detects proteoforms and their PTMs present in intact nucleosomes by employing top-down MS in native mode **(a, right side). (b)** Data from the three steps of Nuc-MS on an intact unmodified nucleosome. **First**, the mass of intact nucleosomes is measured (MS^1: O, observed average mass; T, theoretical mass; Δ*m*, error). **Second**, a single nucleosome charge state (e.g. 35+ ions highlighted in green) is isolated and activated by collisions with nitrogen to eject all intact histones and detect them simultaneously at isotopic resolution (MS^2, reporting monoisotopic masses). **Third**, each histone is isolated and further activated to create backbone fragmentation products that characterize the proteoforms, revealing PTMs or sequence events (MS^3; blue flags indicate fragment ions matching uniquely to human histone H3.1, depicted as a graphical fragment map at bottom; green rectangle in the upper right of the panel highlights the intact precursor).

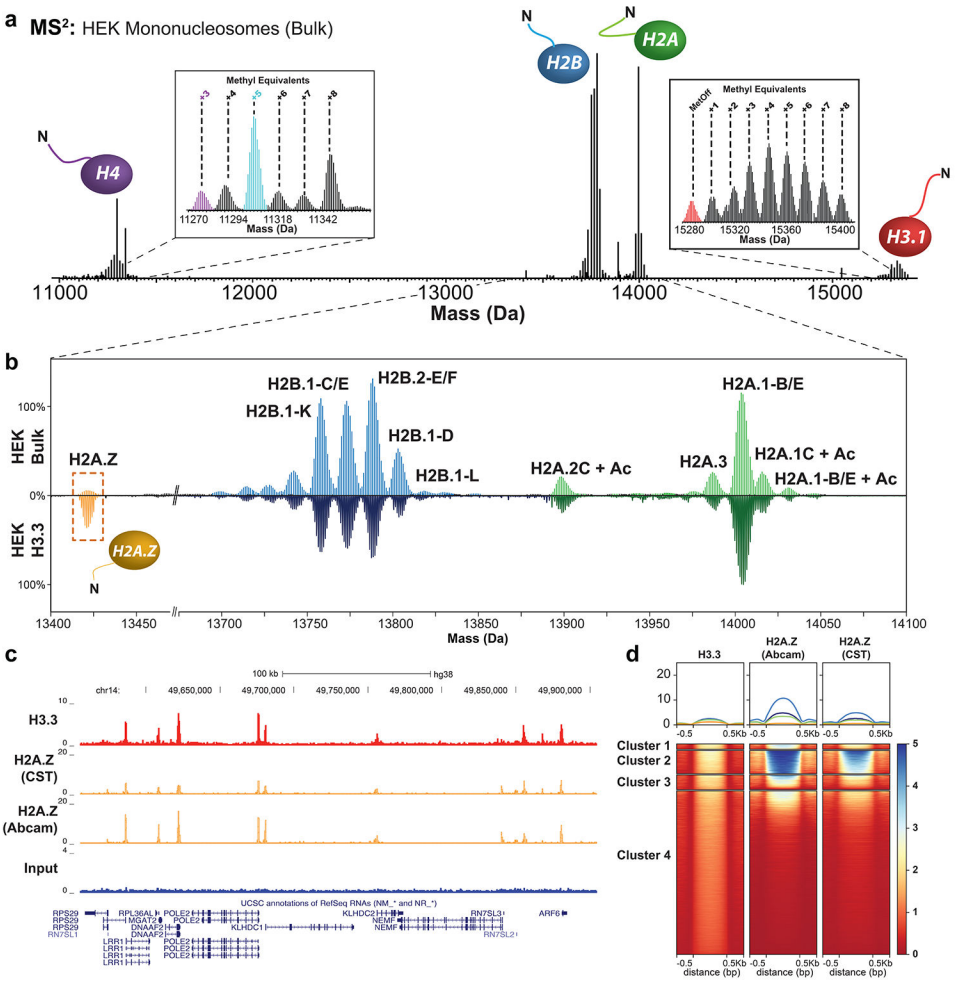

Figure 2. Nuc-MS analysis of endogenous mononucleosomes from HEK cells.

(a) MS^2 spectrum of ejected histones from fragmented nucleosomes in the range of 6000-9000 *m/z* (average of three measurement replicates), demonstrating detection of all core histones and their proteoform distributions. Insets show spectral regions in the mass domain containing the isotopic distributions for H3.1 and H4 proteoforms. The H4 proteoform with five methyl equivalents (highlighted in cyan) was isolated and fragmented to produce the fragmentation map in Supplemental Fig. 7, thus characterizing the proteoform as N-terminally acetylated H4K20me2. **(b)** Comparison of H2A and H2B proteoform profiles in HEK bulk chromatin *(top panel)* vs. H3.3-containing nucleosomes *(bottom panel),* shows latter to be enriched for H2A.Z (yellow box; 14% abundance. All proteoform peaks are normalized to the intensity of the peak of H2A.1-B/E (n = 3). **(c)** Example tracks from *ChIP-seq* reads in HEK cells showing input, H3.3 and H2A.Z targets (antibody details in Methods) supporting co-localization of these two variants. **(d)** Heatmap centered on H3.3 peaks ±0.5 kb showing the correlation of *ChIP-seq* signal between H3.3 and H2A.Z. Clusters of loci are compositionally defined in Supplemental Fig. 10 and described with gene ontology terms in Supplemental Fig. 19.

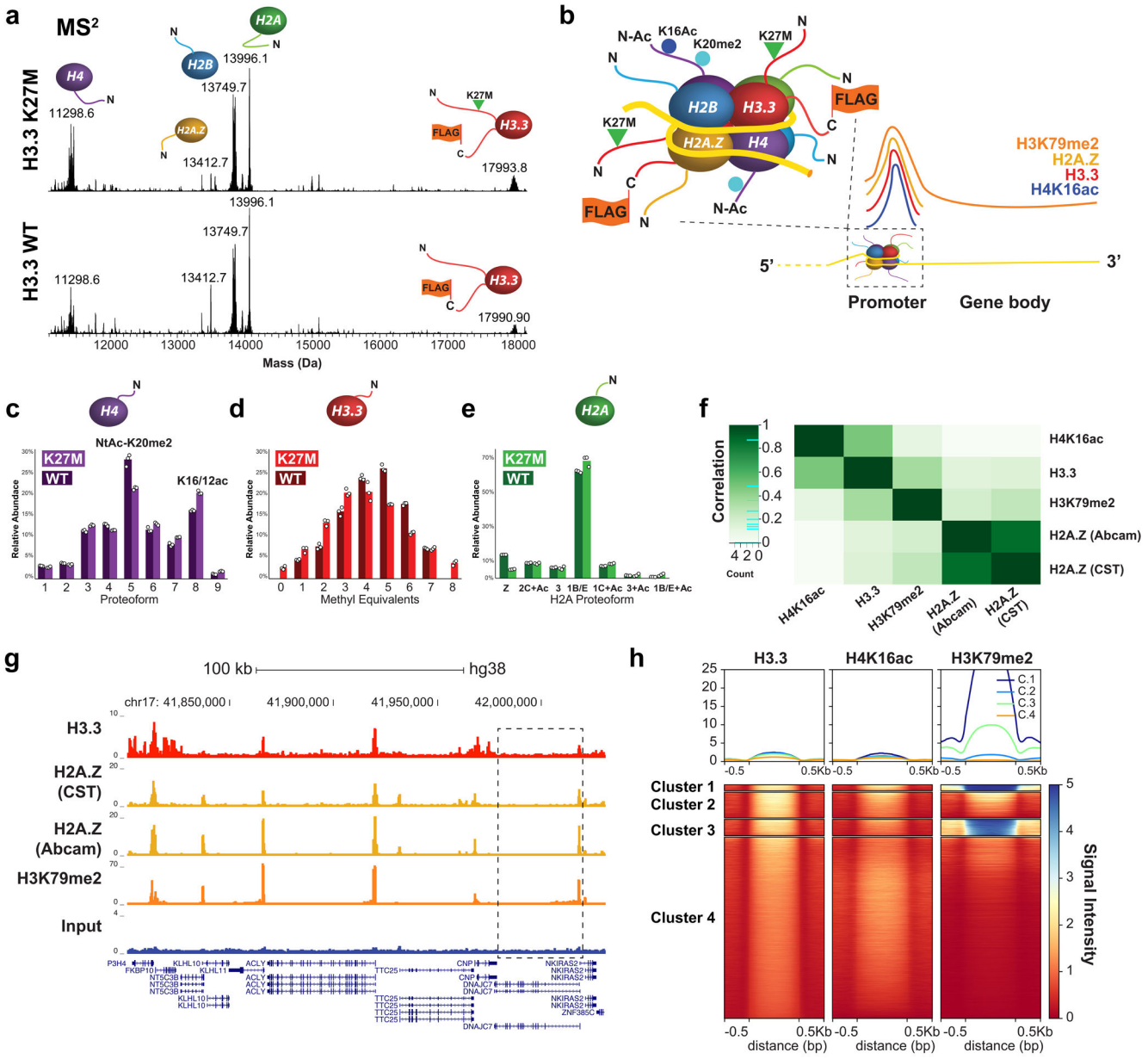

Figure 3. Nuc-MS of endogenous nucleosomes prepared from cells with H3.3-FLAG-HA WT or K27M.

(a) MS^2: measurement of the histone proteoforms ejected from mononucleosomes isolated from two cell lines (all measured at isotopic resolution; two biological replicates and three measurement replicates). Mononucleosomes were isolated from 6000-9000 *m/z* to eject histone proteoforms for MS^2 measurement. Note the ~2.5 kDa shift in H3 due to the addition of the FLAG-HA-tag in comparing the MS^2 spectra of H3.3-enriched HEK vs HEK bulk mononucleosomes in Fig. 2. **(b)** Depiction of the composition for the most abundant nucleosomes determined by Nuc-MS, reflecting high enrichment for histone proteoforms and variants present at promoters and highly expressed genes. **(c-e)** Quantitative analysis of proteoform abundances for ejected histones by MS^2. Data points from three replicates are

displayed as a scatter plot and ordered left to right according to data file of origin. The mean integrated peak area for each histone proteoform is represented with the histogram. **(f)** Pearson correlation plot showing association among histone PTMs and variants characterized by Nuc-MS and targeted by *ChIP-seq* (H3.3, H3K79me2 and H2A.Z, and H4K16ac). **(g)** Example tracks showing *ChIP-seq* reads in HEK cells for Input, H3.3, H3K79me2 and H2A.Z support co-localization of the latter three. A zoom-in of the gene highlighted with a black box is in Supplemental Fig. 18. **(h)** Heatmap centered on H3.3 peaks ±0.5 kb showing the correlation of *ChIP-seq* signal between H3.3, H3K79me2 and H4K16ac.