Supplemental Information for:

## **Separation and Characterization of Endogenous Nucleosomes by Native Capillary Zone Electrophoresis – Top-Down Mass Spectrometry (nCZE-TDMS)**

Kevin Jooß<sup>a</sup>, Luis F. Schachner<sup>a</sup>, Rachel Watson<sup>b</sup>, Zachary B. Gillespie<sup>b</sup>, Sarah A. Howard<sup>b</sup>, Marcus A. Cheek<sup>b</sup>, Matthew J. Meiners<sup>b</sup>, Amin Sobh<sup>c</sup>, Jonathan D. Licht<sup>c</sup>, Michael-Christopher Keogh<sup>b</sup> & Neil L. Kelleher<sup>a,\*</sup>

<sup>a</sup>Departments of Chemistry and Molecular Biosciences, the Chemistry of Life Processes Institute, and the Proteomics Center of Excellence, Northwestern University, Evanston, IL 60208, USA

<sup>b</sup>EpiCypher Inc., Durham, NC 27709, USA

<sup>c</sup>Division of Hematology/Oncology, University of Florida Health Cancer Center, Gainesville, FL 32610, USA

**\*Correspondence:** Prof. Dr. Neil Kelleher, 2070 Campus Dr, Evanston, Il 60208, USA, nkelleher@northwestern.edu



**Figure S1:** (**a**) Scheme of chromosome, chromatin, and nucleosome (Nuc). Chromosomes are macrostructures based on assemblages of chromatin, a chain repeating DNA : protein biopolymer. The Nuc, the smallest repeating unit, consists of ~147 bp DNA wrapped around two copies each of the core histones: H2A, H2B, H3, and H4. (**b**) Workflow of nCZE-TDMS system: Separation of different intact Nucs (e.g. proteoforms) in the nCZE dimension. TDMS analysis then follows in three steps: intact Nucs  $(MS<sup>1</sup>)$ , histone subunits  $(MS<sup>2</sup>)$ , and peptide fragments (pseudo-MS<sup>3</sup>).



**Figure S2.** Optimization of injection volume. Different volumes of H3G34V ( $c = 0.75 \mu M$ ,  $n = 3$ ), ranging from ~5.7 to 38.0 nL (2.5 psi, 15 to 100 sec) were injected, which corresponds to roughly 0.9 to 6.0% of the total capillary volume. (**a**) Full-width half maximum (FWHM) was plotted against injection volume. (**b**) Ratio between intensity and FWHM was plotted against injection volume.

Optimization of injection volume: As would be expected, CZE peak widths increase proportional with the applied injection volume (**Figure S2a**). One appropriate measure to evaluate peak shape is to look at the ratio between peak intensity and full width half maximum (FWHM). The highest ratio was observed at 11.4 nL injection volume (**Figure S2b**). In addition, there was no significant improvement of absolute intensity observed for higher injection volumes. For this reason, 11.4 nL (2.5 psi, 30 sec) was selected for further experiments.



**Figure S3.** Evaluation of influence of separation voltage on nucleosome migration. H2BK120ub (c = 500 nM) was analyzed at three different voltage settings: +12, +15, and +18 kV. The other method parameter settings were kept constant, including 40 mM AmAc (pH  $\approx$  6.8) as BGE and a supplemental pressure of 2.5 psi. The systematic increase in migration time by changing the HV from +12 to +18 kV, while maintaining other parameters, is a strong indication that net charge of the nucleosome complexes is negative under these conditions.



٦

**Figure S4.** Comparison of fragmentation performance between native CZE and native direct infusion (DI) for the analysis of histone H2A after ejection from H3K27me3 nucleosomes by pseudo MS<sup>3</sup> TD experiments. The same number of mass spectra were averaged for each technique for comparison. Fragmentation profiles and coverage were determined to be similar between these two approaches and most fragments were picked up by both techniques. This indicates no significant loss of fragmentation quality due to the front end CZE separation.



Figure S5. (a) Highlighted section of pseudo-MS<sup>3</sup> mass spectrum showing diagnostic trimethylated *b*-ions after ejection and fragmentation (HCD) of histone H3K27me3 from intact Nuc ( $c = 1 \mu M$ ). Predominantly, +14 and +15 charge states were observed for these fragment ions. (**b**) Graphical fragment maps of peptides observed from H3K27me3. Detected *b*-ions exhibiting a mass shift matching a tri-methylation are highlighted (red, dashed box).



**Figure S6:** Regression and determination of LOD: (**a**) Linear range was determined as 10 to 500 nM (110 amol to 5.7 fmol) with  $R^2 = 0.996$  and (**b**) dynamic range was observed between 10 to 3000 nM. (**c**) Mass spectrum of intact nucleosome at lowest analyzed concentration (10 nM) close to LOD with an average S/N of 12.3 ±1.0. (**d**) Deconvoluted isotopic distributions for histones ejected from unmodified recombinant Nucs measured at 62.5 nM.



**Figure S7.** Separation of endogenous nucleosomes (endoNucs; ~10 mg/mL) derived from TKO cells. Total Ion Electropherograms (**a**) divided into two major regions: (i) non-Nuc proteins and (ii) endoNucs. Raw and deconvoluted mass spectra (**b-d**) of three exemplary sections (75 – 80 min, 95 – 100 min, 115 -120 min). Several "nucleoforms" were observed in each section. In general, the intact mass of nucleoforms increased with migration time and an average mass shift of 620.2 ±18.5 Da was observed between neighboring peaks, which corresponds to a single base pair difference in length of the associated DNA ( $\Delta$ m<sub>GC</sub> = 618.4 Da,  $\Delta$ m<sub>AT</sub> = 617.4 Da).



**Figure S8.** Comparison of observed and theoretical isotopic distributions of H3 proteoforms. (**a**) Deconvoluted mass spectrum (blue) of NTKO in the range from 15,305 to 15,350 Da showing the three major H3 proteoforms. The theoretical isotopic distribution of the H3.2 proteoform exhibiting five methyl equivalents (ME) is overlayed (red), indicating a high match with the observed data. Calculated theoretical distributions of proteoforms (**b**) H3.2-ME5, (**c**) H3.1-ME4, and (**d**) H3.3-ME9.



**Figure S9.** Immunoblot analysis showing NSD2 (MMSET) and H3K36me2 levels in NSD2-high (NTKO) and NSD2 low (TKO) isogenic cells derived from the human multiple myeloma cell line KMS11. (A) Reduction of both fulllength (MMSET II) and the short isoform (MMSET I) of NSD2 in TKO compared to NTKO cells. Decrease in NSD2 levels is associated with a drop in global H3K36me2 levels. HDAC2 was used as a loading control for NSD2 blots. Total H3 was used as a loading control for H3K36me2 blots. (B, C) Relative quantification of NSD2 levels (normalized to HDAC2) and H3K36me2 (normalized to total H3) in NTKO and TKO cells. Data represent the mean  $\pm$  SD (n = 2).

Immunoblot Analysis: Nuclear extracts from NTKO and TKO cells were prepared using the Nuclear Extract Kit (Active Motif) following the manufacturer's protocol. Nuclear extracts (50  $\mu$ g proteins for NSD2/HDAC2 and 20 µg for H3K36me2/H3) were separated by SDS gel electrophoresis on a 4-12 % NuPAGE Bis-Tris gel (ThermoFisher Sceintific) and transferred using an iBlot system and iBlot Gel Nitrocellulose Transfer Stacks (ThermoFisher Scientific). Blots were blocked in Peirce Protein-Free Blocking Buffer (ThermoFisher Scientific) for 1 hour. Blots were then incubated with primary antibodies-NSD2 (PMID: [18156491\)](https://www.ncbi.nlm.nih.gov/pubmed/18156491), HDAC2 (Millipore # 05-814), H3K36me2 (Cell Signaling Technology #2901S), H3 (Cell Signaling Technology # 14269S) diluted in blocking buffer (1:1000), overnight at  $4^{\circ}$ C. Membranes were washed 3 times with PBS-Tween and incubated with the secondary antibody solutions - goat anti-mouse, Alexa Fluor Plus 800 (ThermoFisher Scientific # A32730) or goat anti-rabbit, Alexa Fluor Plus 680 (ThermoFisher Scientific # A32734) - diluted in PBS-tween  $(1:10,000)$  for 1 hour at room temperature. Membranes were then washed 3 times in PBS-Tween and scanned on an Odyssey CLx instrument (LICOR).



**Figure S10.** Relative quantification of H3.1/H3.2 proteoforms associated with K27 and K36 from NTKO (blue, n =3) and TKO (orange, n = 3) cells by bottom-up proteomics (LC-MS/MS). As expected, a decrease of K36me2 in TKO cells could be observed. Concurrently, an increase in K27me2, K27me3 and K27ac was detected. Samples were prepared and analyzed based on the protocol of Zheng *et al.* (Reference 51). Error bars represent standard deviations.



**Figure S11.** Relative quantification of H4K16 acetylation from NTKO (blue, n =3) and TKO (orange, n = 3) cells by bottom-up proteomics (LC-MS/MS). Samples were prepared and analyzed based on the protocol of Zheng *et al.* (Reference 51). Error bars represent standard deviations.

<b>Step</b>	Capillary	<b>Standard</b>	<b>High-resolution</b>	High-throughput	
Rinse: 0.1 M HCl	<b>SL</b> $5 \text{ min}$		$5 \text{ min}$	$3 \text{ min}$	
Rinse: BGE	CL.	$3 \text{ min}$	$3 \text{ min}$	$2 \text{ min}$	
Rinse: BGE	<b>SL</b>	$5 \text{ min}$	$5 \text{ min}$	$3 \text{ min}$	
Injection	<b>SL</b> $2.5$ psi, $30$ sec		$2.5$ psi, $30$ sec	$2.5$ psi, $30$ sec	
Ramp Up	SL.	$5$ psi, $1$ min	$2.2$ psi, 1 min	$15$ psi, $1$ min	
<b>SL</b> Separation		$+18$ kV, 5 psi, 30 min	$+18$ kV Synthetic Nucs: 2.2 psi, 115 min EndoNucs: 3.0 psi, 180 min	$+18$ kV, 15 psi, 6 min	
Ramp Down	SL.	$5$ psi, $5$ min	$5$ psi, $5$ min	$15$ psi, $4$ min	
<b>Total Run Time</b>		$\sim 50$ min	$\sim$ 135 min	$\sim$ 20 min	

**Table S1:** Overview of steps of the three developed native CE methods. Rinsing steps were performed at 100 psi.

 $SL =$  separation line,  $CL =$  conductive line



**Table S2:** Parameters of MS methods applied for Nuc characterization using CZE-TDMS.

**Table S3:** Determination of intra- and inter-day reproducibility for nCZE-TDMS method for nucleosome characterization. Designer nuc H3K27me3 (c = 1  $\mu$ M) was analyzed five consecutive times on two different days. Relative standard deviation (RSD) values were calculated for migration time (MT), peak area and peak intensity. Peak area and intensity were based on the extracted ion electropherograms (EIEs) of the five most intense charge states.

Day 1 (intra-day)				Day 2 (intra-day)				
Run	$MT$ [min]	Area [cts]	Intensity [cts]	Run	$MT$ [min]	Area [cts]	Intensity [cts]	
1	20.44	$2.60 \cdot 10^8$	$3.80 \cdot 10^6$	6	20.23	$2.92 \cdot 10^8$	$4.16 \cdot 10^6$	
$\overline{2}$	20.59	$2.74 \cdot 10^8$	$4.18 \cdot 10^6$	7	20.35	$2.59 \cdot 10^8$	$3.74 \cdot 10^6$	
3	20.59	$2.83 \cdot 10^8$	$4.29 \cdot 10^6$	8	20.20	$2.53 \cdot 10^8$	$3.82 \cdot 10^6$	
4	20.42	$2.63 \cdot 10^8$	$3.78 \cdot 10^6$	9	20.49	$2.57 \cdot 10^8$	$3.93 \cdot 10^{6}$	
5	20.61	$2.13 \cdot 10^8$	$3.73 \cdot 10^6$	10	20.51	$2.62 \cdot 10^8$	$3.91 \cdot 10^6$	
Mean	20.53	$2.58 \cdot 10^8$	$3.96 \cdot 10^6$	Mean	20.36	$2.62 \cdot 10^8$	$3.91 \cdot 10^6$	
Stdev	0.09	$2.70 \cdot 10^{7}$	$2.58 \cdot 10^5$	Stdev	0.14	$1.71 \cdot 10^{7}$	$1.58 \cdot 10^5$	
<b>RSD</b>	0.45%	10.46%	6.53%	<b>RSD</b>	0.70%	6.51%	4.04%	
				Day $1 \& 2$ (inter-day)				
				Mean	20.44	$2.60 \cdot 10^8$	$3.94 \cdot 10^{6}$	
				Stdev	0.15	$2.14 \cdot 10^{7}$	$2.03 \cdot 10^5$	
				<b>RSD</b>	0.71%	8.22%	5.16%	

**Table S4:** Relative quantification of NTKO/TKO histone proteoforms using nCZE-TDMS. Comparison between histone proteoforms from NTKO (IGH-NSD2; H3K36me2<sup>hi</sup>) and TKO (NSD2<sup>+/-</sup>; H3K36me2<sup>ho</sup>) cells using spectra averaged over the entire endoNuc migration range  $(n = 3)$ . Statistically significant increases (blue) and decreases (red) of histone proteoforms in TKO samples are highlighted.



PF = proteoform, met = methylation, Ac = acetylation, NtAc = N-terminal acetylation, ME = methyl equivalent

**Table S5:** Relative quantification of NTKO/TKO histones using nCZE-TDMS. Comparison between histones from NTKO and TKO cells using spectra averaged over the entire endoNuc migration range  $(n = 3)$ .



**Table S6:** Comparison between front, middle and end section of the endoNuc migration time window for NTKO/TKO endoNuc samples analyzed by nCZE-TDMS. Electropherograms were divided into three equally long sections and deconvoluted.



**Table S7:** Comparison between front, middle and end section of the endoNuc migration time window for **NTKO** (IGH-NSD2; H3K36me2hi) endoNuc samples analyzed by nCZE-TDMS. t-tests were performed between front and end sections. Statistically significant increases (blue) and decreases (red) of histone proteoforms over the course of the nCZE separation are highlighted.



PF = proteoform, met = methylation, Ac = acetylation, NtAc = N-terminal acetylation, ME = methyl equivalent

**Table S8:** Comparison between front, middle and end section of the endoNuc migration time window for **TKO**  $(NSD2<sup>+/-</sup>; H3K36me2<sup>lo</sup>)$  endoNuc samples analyzed by nCZE-TDMS. t-tests were performed between front and end sections. Statistically significant increases (blue) and decreases (red) of histone proteoforms over the course of the nCZE separation are highlighted.



PF = proteoforms, met = methylation, Ac = acetylation, NtAc = N-terminal acetylation, ME = methylation equivalent

## **Method S1: Preparation of TKO and NTKO mononucleosomes**

Cells were collected and resuspended in 2.5 pelleted cell volumes (PCV) of buffer A (10 mM HEPES, pH 7.9, 10 mM KCl, 340 mM sucrose, 0.5 mM PMSF, 0.5 mM benzamidine) supplemented with 5 mM β-mercaptoethanol and 1× cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). A similarly supplemented volume of buffer A (2.5 PCV) containing 0.2% (vol/vol) Triton X-100 detergent was also prepared simultaneously. The resuspended cell pellet was completely homogenized by pipetting and light vortexing, and the detergent-containing buffer A was added to the cell suspension, mixed by inversion, and allowed to incubate on ice for 10 min with occasional gentle mixing to thoroughly lyse the cells. The resultant nuclei were pelleted by centrifugation (1,300g, 4 °C, 5 min), and the supernatant was aspirated and discarded. Next, the nuclei pellet was resuspended in 6 PCV of supplemented buffer A and centrifuged through a 35 mL sucrose cushion (10 mM HEPES, pH 7.9, 30% (wt/vol) sucrose) in a 50-ml conical tube supplemented with 5 mM β-mercaptoethanol and  $1 \times$  cOmplete EDTA-free Protease Inhibitor Cocktail (1,300g, 4 °C, 10 min). 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 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 of 2 M NaCl. The suspension was diluted to a DNA concentration of 1.5 μg/μL, and CaCl<sub>2</sub> was 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 min. Digestion of the nuclei to nucleosomal species was initiated by adding 1 U MNase (Worthington, prepared in supplemented buffer A) for every 70 μg DNA, and the samples were incubated at 37 °C for 15 min. The digestion was then quenched by adding 0.5 M EGTA to a final concentration

of 10 mM, mixing by inversion and placing the samples on ice. The final volume of the digested material was noted. Approximately 10 μg DNA was purified (Qiagen DNA Cleanup kit) and resolved on a 2% agarose gel in  $0.5 \times$  TBE to confirm the extent of digestion by fragment size distribution in the population. NaCl (2 M) 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 by centrifugation at 12,000g and 4 °C to pellet any insoluble material before sizeexclusion chromatography. The material was fractionated with a HiPrep 26/60 Sephacryl S-300 HR column equilibrated with supplemented buffer A, as above, containing 650 mM NaCl using an ÄKTA Prime Plus FPLC (GE Life Sciences). Samples (10 μg) of the individual fractions containing nucleosomes were purified (Qiagen DNA Cleanup kit) and resolved on a 2% agarose gel in 0.5× TBE. Only those fractions that contained mononucleosomal-associated DNA fragments (~150 bp) were pooled for further analysis.

## **Discussion S1: Optimization of "higher-resolution" CZE method**

Though partial separation was achieved using the "standard" CZE method, there is still room for significant improvement. Therefore, several different method conditions were evaluated, including ammonium acetate concentration (20 – 40 mM), supplemental pressure (2.2 – 5.0 psi), and pH (6.0 – 8.0). The concentration of AmAc did not affect the overall separation performance significantly (e.g.  $R_{20mM} = 0.365$  vs.  $R_{40mM} = 0.366$ ). A change in the pH of the BGE did not lead to better separation performance. However, if the pH is too low  $(\leq 6.0)$ , no intact nucleosome signal was observed, possibly due to issues with complex stability. Moreover, appreciable metal adduct formation was observed at pH 8.0, decreasing overall signal quality. Given that the Nucs are counter-migrating in the system due to their negative net charge, balancing the supplemental pressure and mobility of the Nucs represents a key factor to improve separation performance. Although migration times and peak widths increase noticeably when decreasing the supplemental pressure from 5.0 psi to 2.2 psi, electrophoretic resolution is significantly improved from  $R = 0.37$ to 1.09 (see **Figure 3**).

## **Discussion S2: Optimization of "high-throughput" CZE method**

With potential for widespread deployment of the nCZE-TDMS platform, we developed a "highthroughput" CZE method tailored for the quality control environment during semi-synthetic Nuc manufacturing. For this purpose, we adapted the "standard" CZE method to decrease overall run time but still resolve Nuc-related peaks from salt/matrix compounds (**Figure S12**).



**Figure S12.** Comparison of Total Ion Electropherogram of "standard" (run time: 50 min) and "high-throughput" (run time: 20 min) CZE method analyzing unmodified recombinant Nuc ( $c = 1 \mu M$ ). SNR values were determined at FWHM. While the higher-throughput method can baseline-resolve Nuc from matrix, there is a two-fold drop in sensitivity compared to the standard CZE method.

The influence of supplemental pressure was tested in a range between 5.0 and 30 psi and either maintained constant over the course of analysis or varied in different sections of the method. A constant pressure above 15 psi lead to issues regarding resolution of Nuc and matrix compounds.

In addition, varying the pressure during analysis resulted in electrospray process instabilities and thus, potential negative effects on reproducibility. We determined that a constant supplemental pressure of 15 psi, in combination with some changes to the preconditioning steps, yielded stable and reproducible signal, and reduced total run time from 50 to < 20 min per analysis (**Figure S12**). The Nuc peak could be detected in < 5 min using the "high-throughput" CZE, which represents a major improvement versus ~20 min using the "standard" CZE method.