Supporting Information

On-capillary cell lysis enables top-down proteomic analysis of single mammalian cells by CE-MS/MS

Kendall R. Johnson[†], *Yunfan Gao*[†], *Michal Greguš, Alexander R. Ivanov**

Barnett Institute of Chemical and Biological Analysis, Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Ave., Boston, Massachusetts 02115, United States

*Email: a.ivanov@northeastern.edu

Table of Contents

1. Experimental

Materials and Reagents

Optima LC-MS grade water was purchased from Fisher Scientific (Pittsburgh, PA). Ammonium acetate, acetic acid (HAc), sodium hydroxide (NaOH), methanol (MeOH), HPLC-MS grade formic acid (FA), and acridine orange were obtained from Sigma-Aldrich (St. Louis, MO). Aquapel was purchased from PGW Auto Glass (Cranberry Twp, PA). Premium fetal bovine serum (FBS) was obtained from R&D Systems (Minneapolis, MN). Gibco penicillin−streptomycin (P/S) was purchased from Thermo Fisher Scientific (Waltham, MA). F-12K Medium (Kaighn's Modification of Ham's F-12 Medium) was purchased from ATCC (Manassas, VA).

Cell Culture

HeLa-S3 cells were converted to grow in suspension culture following a published protocol,[1] after which they were cultured in F-12K medium supplemented with 10% FBS, 100 I.U./mL penicillin, and 100 µg/mL streptomycin. The cell density was maintained between 4.0E5 and 1.5E6 cells/mL. THP-1 cells were cultured in RPMI 1640 medium supplemented with 10% FBS, 0.05 mM 2-mercaptoethanol, 100 I.U./mL penicillin, and 100 µg/mL streptomycin. The cell density of THP-1 cells was maintained between 2E5 and 1E5 cells/mL. Cell viability and density were determined by trypan blue staining and hemocytometer cell counting upon each cell passage.

Additional mass spectrometry settings

The ion transfer tube was set to 110°C. For MS1 scan, the scan range was set to 600 to 2000 m/z with a maximum injection time of 100 ms and an AGC target setting of 1×10^6. The RF Lens was set to 50%. The top two intensity precursor ions were selected with an intensity threshold of 2×10^6. The dynamic exclusion window was set to 90 s, and only precursors with charge states 5-24 were selected for fragmentation. Precursor ions were isolated with a 3.0 m/z window and fragmented with an HCD collision energy of 23%. A 500 ms ion injection time and an AGC target of 1×10^6 was used in MS2 scans.

Additional details on data analysis

Gene ontology (GO) term enrichment analysis was performed in FunRich (v3.1.3)[2] using the UniProt human database (downloaded July 2022).

Open-source web tools ClustVis[3] and TBtools[4] were used for principal component analysis and unsupervised hierarchical heat map clustering, respectively. Further data analysis was performed in the R environment.[5] Graphical outputs were generated using the package ggplot2.[6]

2. Figure S1. Representative ion density maps of HeLa cell TDP

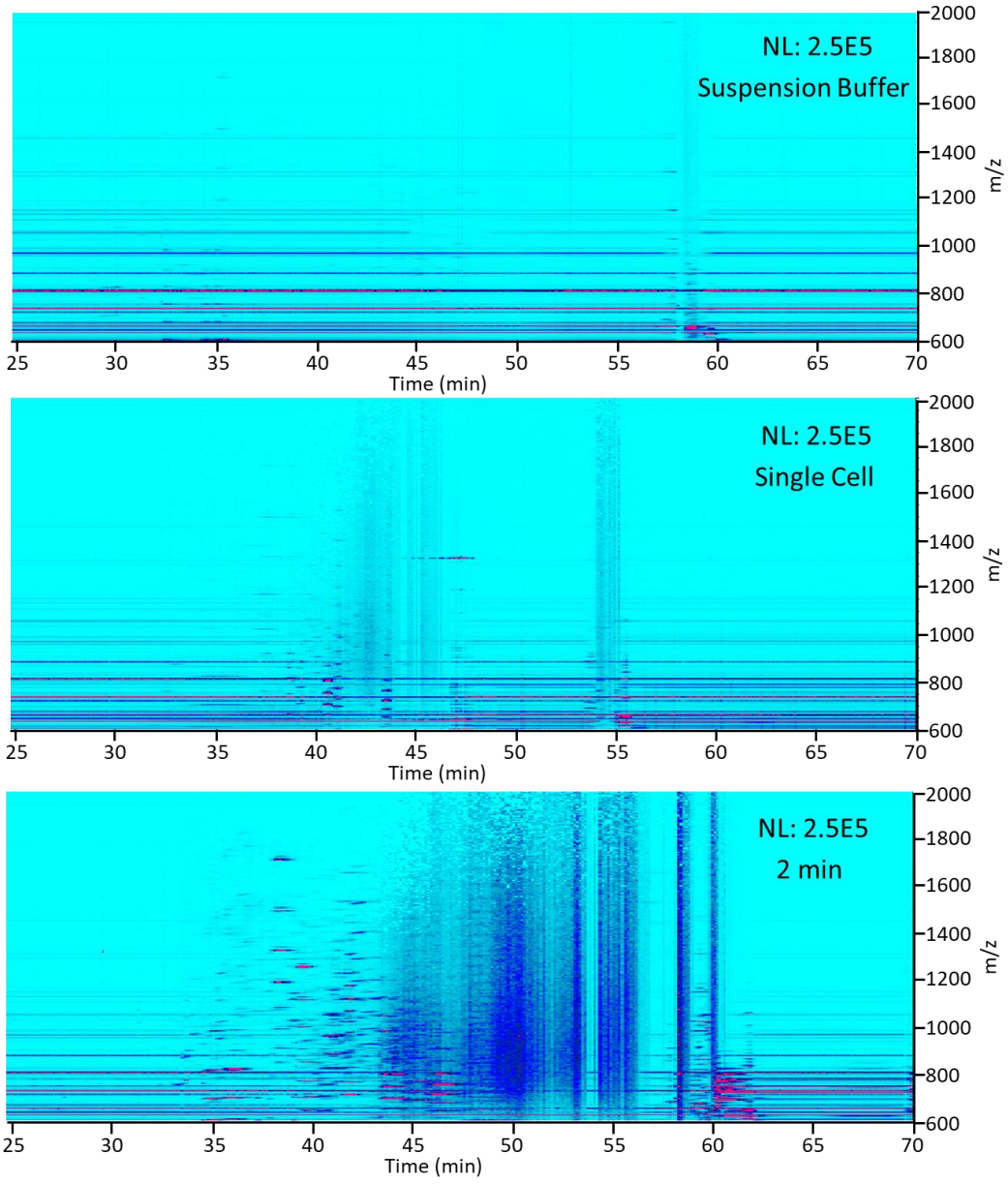

Figure S1. Representative ion density maps of HeLa cell TDP. Cell suspension buffer (top), a single HeLa cell (middle) and 2-min spray voltage injection of cells (bottom). Normalized to 2.5E5 intensity.

3. Figure S2. Measuring single HeLa cell diameters from microscope images

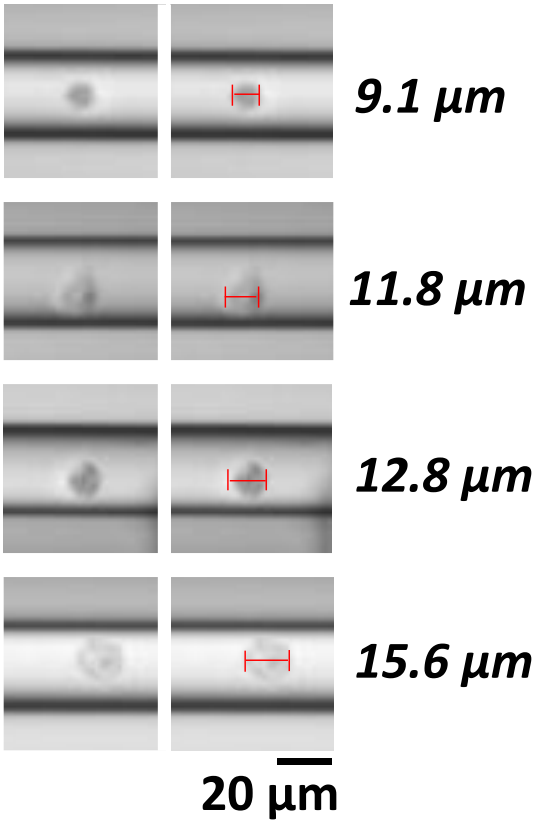

Figure S2. Measuring single HeLa cell diameters from microscope images. Zoomed in images of single HeLa cells inside the capillary prior to cell lysis and subsequent TDP analysis by CE-MS/MS. Red lines superimposed over the cells were used to calculate cell diameters.

4. Figure S3. Match-between-runs after migration time alignment

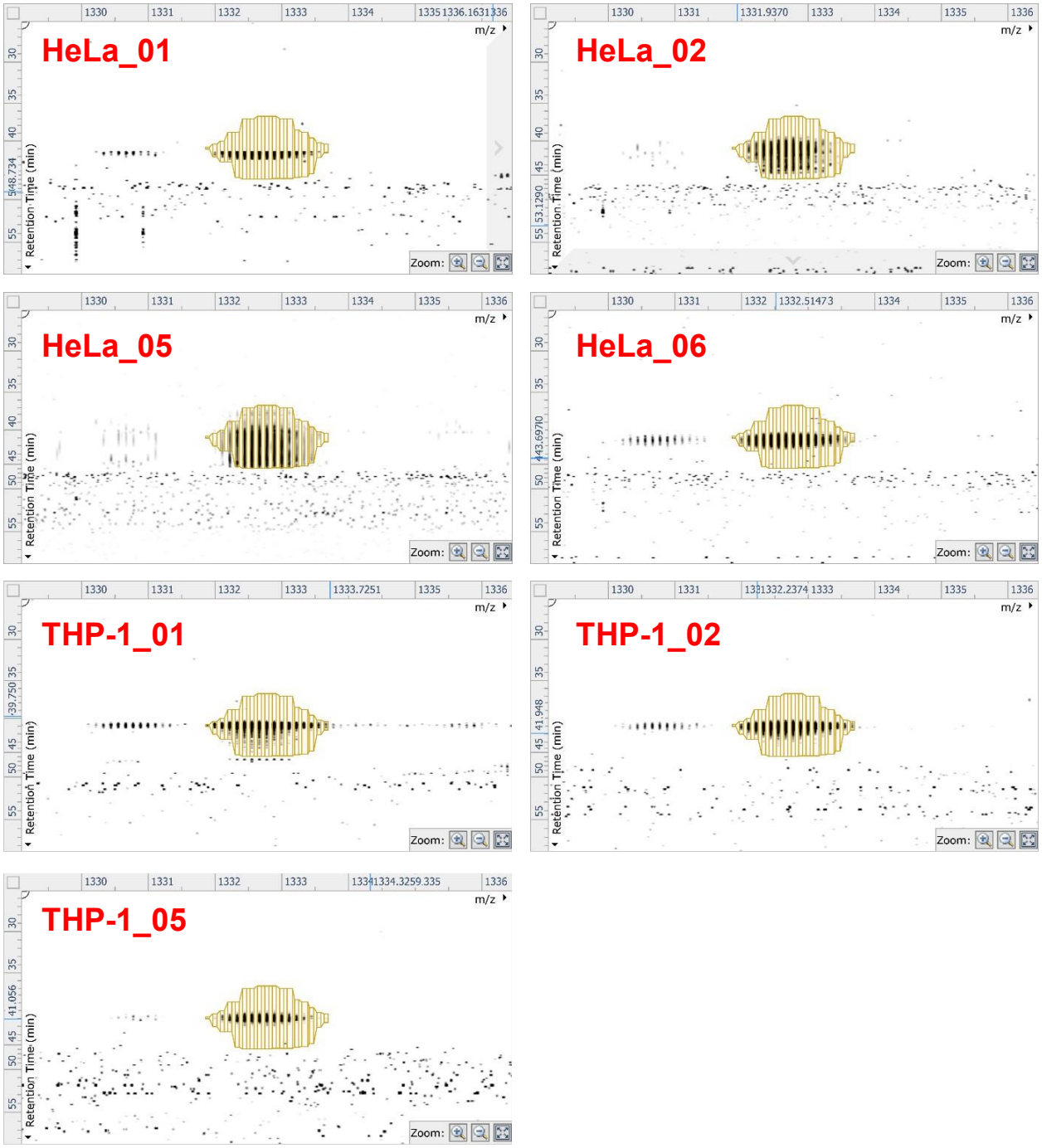

Figure S3. Match-between-runs after migration time alignment. Representative images of peak picking in each HeLa and THP-1 single-cell data file for MBR using Progensis QI software. The proteoform shown here corresponds to a variant of prothymosin alpha isoform 2, with the initiator methionine removed and N-terminal acetylation.

5. Figure S4. THP-1 single cell protein and proteoform identifications with and without match-between-runs

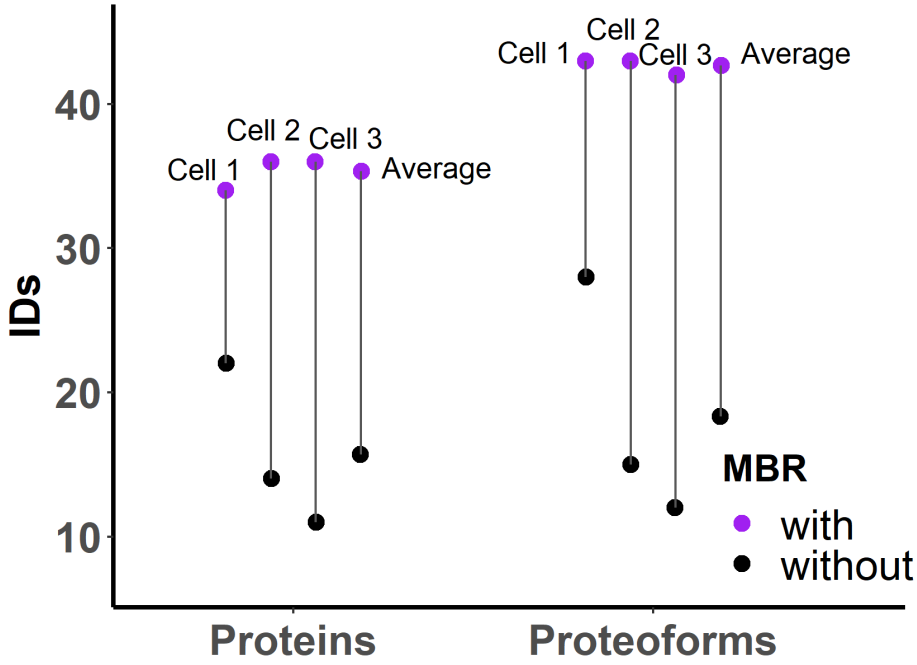

Figure S4. THP-1 single cell protein and proteoform identifications with and without match-between-runs. Protein and proteoform identifications for each THP-1 single cell replicate shown with and without match-between-runs. Average number of identifications with and without match-between-runs is also shown. The improvements in identifications of proteins and proteoforms observed for THP-1 cells are similar to the gains reported in Figure 6A for HeLa cells using the same match-between-runs approach.

6. Table S1. Protein and proteoform identification numbers with two-level filtering (1% FDR and additional e-value cutoff)

Sample ID	Protein		Proteoform	
	1% FDR only	1% FDR and e-value ≤ 0.01	1% FDR only	1% FDR and e-value ≤ 0.01
blank1	1	0	1	0
blank2	2	0	2	0
blank3	0	0	0	0
blank4	0	0	0	0
suspension_buffer_1	2	2	4	4
suspension_buffer_2	3	2	5	4
suspension_buffer_3	4	2	5	4
suspension_buffer_4	2	2	2	2
2min_sprayvoltage_1	36	33	46	40
2min_sprayvoltage_2	61	54	83	74
2min_sprayvoltage_3	46	44	74	69
2min_sprayvoltage_4	62	57	93	88
1min_sprayvoltage_1	15	17	21	25
1min_sprayvoltage_2	36	32	54	49
1min_sprayvoltage_3	38	33	55	50
1min_sprayvoltage_4	28	24	43	40
1cell_HeLa_1	18	17	24	24
1cell_HeLa_2	22	21	25	24
1cell_HeLa_3	44	40	56	50
1cell_HeLa_4	26	25	36	33
1cell_THP1_1	22	22	30	28
1cell_THP1_2	15	14	17	15
1cell_THP1_3	12	11	13	12

7. Table S2. TopFD Search Parameters

Max. charge	**30**
Max. mass (Da)	**100000**
MS1 S/N ratio	**3.0**
MS2 S/N ratio	**1.0**
Precursor window size (m/z)	**3.0**
M/z error (m/z)	**0.02**

8. Table S3. TopPIC Search Parameters

Decoy Database	Checked
Spectrum level FDR	0.01
Proteoform level FDR	0.01
Mass error tolerance (ppm)	15
PrSM cluster error tolerance (Da)	1.2
Maximum number of mass shifts	1
N-terminal forms	None, NME, NME_Acetylation, M_acetylation
Maximum mass shift (Da)	500
Minimum mass shift (Da)	-500
Number of combined spectra	1

* NME: N-terminal methionine excision, NME_Acetylation: N-terminal methionine excision and acetylation, M_acetylation: N-terminal methionine acetylation.

9. References

1. Abdeen, S. H.; Abdeen, A. M.; El-Enshasy, H. A.; Shereef, A. A., HeLa-S3 Cell Growth Conditions in Serum-Free Medium and Adaptability for Proliferation in Suspension Culture. *J Biol Sci* **2011,** *11* (2), 124-134.
2. Pathan, M.; Keerthikumar, S.; Ang, C. S.; Gangoda, L.; Quek, C. Y.; Williamson, N. A.; Mouradov, D.; Sieber, O. M.; Simpson, R. J.; Salim, A.; Bacic, A.; Hill, A. F.; Stroud, D. A.; Ryan, M. T.; Agbinya, J. I.; Mariadason, J. M.; Burgess, A. W.; Mathivanan, S., FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **2015,** *15* (15), 2597-601.
3. Metsalu, T.; Vilo, J., ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res* **2015,** *43* (W1), W566-70.
4. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H. R.; Frank, M. H.; He, Y.; Xia, R., TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* **2020,** *13* (8), 1194-1202.
5. R Core Team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2019.
6. Hadley, W., *Ggplot2 : elegant graphics for data analysis*. Second edition. ed.; Springer-Verlag New York: 2016; p 1 online resource.