A simple method to quantify proteins from one thousand cells

Burcu Vitrinel[1,2], Dylan E. Iannitelli[2], Esteban O. Mazzoni[2,3], Lionel Christiaen*[2], and Christine Vogel*[1$]

1 Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY, USA

2 Center for Developmental Genetics, Department of Biology, New York University, New York, NY, USA

3 NYU Neuroscience Institute, NYU Langone Medical Center, New York, NY, USA

*equally contributing authors

[$]corresponding author: cvogel@nyu.edu

Supporting Information Placeholder

ABSTRACT: Recent developments in single-cell transcriptomics have created an urgent need for similar approaches to map the proteome in samples from a minimal number of cells. We optimized multiple steps in the mass spectrometry protocol to develop such a method, MinPut, with improved sensitivity to quantify proteins from as few as 1,000 mammalian cells. MinPut uses chemical peptide labeling and does not require specific equipment, antibodies, or other materials. MinPut quantifies >2,500 proteins with high reproducibility. We established and validated the method by comparing mouse embryonic stem cells and *in vitro* differentiated motor neurons. MinPut correctly identifies differentially expressed proteins with small fold-changes, and a dynamic range in abundance similar to that of standard methods. Protein abundance measurements obtained with MinPut compare well to corresponding transcript abundance and to measurements using standard inputs. Therefore, MinPut offers a robust and accurate method to acquire proteomics data from minimal input samples.

Introduction

Typical proteomics experiments require a minimum of ~50,000 mammalian cells (ideally 500,000) to derive enough sample to identify 2,000 to 5,000 proteins in a single-shot tandem mass spectrometry experiment, depending on instrumentation and setup (**Figure 1a**). To increase the number of identified proteins, more material is required to conduct extensive chromatographic separation. Moreover, numerous biological systems, including biopsies and other rare and precious samples, yield much lower numbers of cells, in the order of hundreds to thousands. Therefore, there remains a need for proteomic methods for minimal sample input with optimized coverage and reproducibility[1].

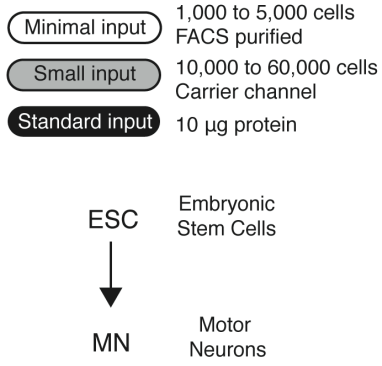
Despite these challenges, recent years have seen several advances in processing small samples. Some of these methods have made use of specialized systems such as custom designed platforms[2–4] that enabled high resolution in tens of cells or even single cells. Other systems have employed specific antibodies, such as CyTOF[5], CITE-seq[6] and proximity ligation assays[7], which create sensitive but limited measurements. Further, some methods have relied on specific material such as paramagnetic beads[8] or collection microreactors[9] to maximize yield from little starting material.

A recently developed protocol, SCoPE-MS, analyzes proteins from single, hand-picked mammalian cells via Tandem Mass Tagging (TMT) coupled to conventional mass spectrometry[10](**Figure 1a**). In a standard 10plex TMT experiment, peptides from 10 different samples are labeled with sample-specific mass tags and then pooled. During the subsequent tandem-mass spectrometry experiment, the tags are indistinguishable by mass at the first 'precursor' level and therefore isolated together. In the subsequent second level of analysis, the peptides and tags are fragmented: the peptides can be sequenced and each tag's channel is quantified through ion intensity measurements (**Figure 1a**). In SCoPE-MS, one channel in the setup is dedicated to a 'carrier' with peptides at high abundance that produce enough signal for reliable peptide identification (**Figure 1a**). The remaining channels contain the experimental samples. Their peptides' abundance is too low for identification, but the intensities of mass tags are available for quantitation. The fundamental idea of the approach is therefore to separate peptide identification and quantitation. However, single cell proteomics still struggles with proteome coverage and reproducibility (**Figure 1a**).

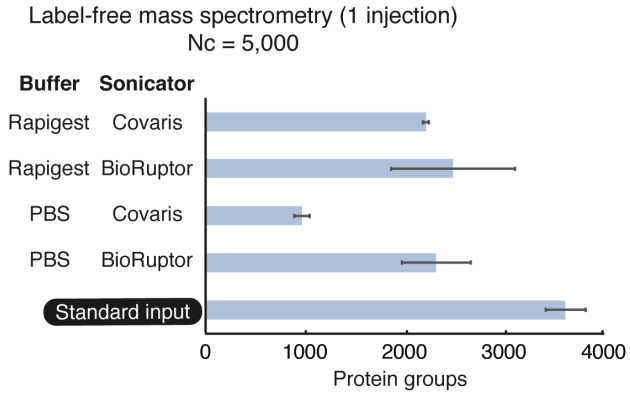
a.

	Number of cells needed for input	Proteins identified	Correlation with standard input sample (Rs)	Correlation between replicates (Rs)	Correlation with RNA abundance (Rs)
Standard	>500,000	~2,700	-	~0.99	~0.43
SCoPE-MS	1	~900	~0.62	~0.5	~0.25
MinPut	**1,000**	**~2,500**	**~0.81**	**~0.99**	**~0.39**

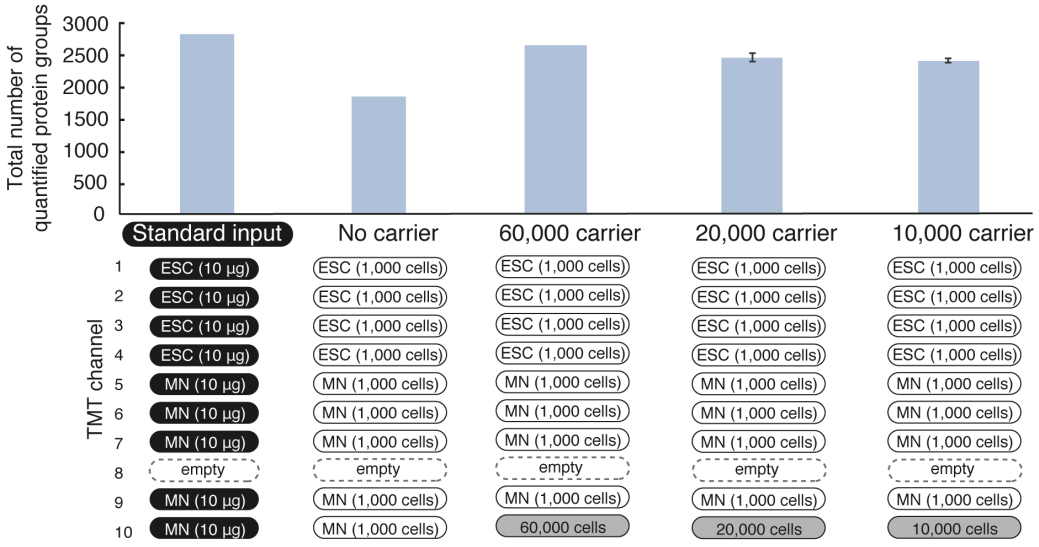
b.



c.



d.



e.

1	126	**10,000 ES cells**
2	127N	empty
3	127C	empty
4	128N	empty
5	128C	empty
6	129N	empty
7	129C	empty
8	130N	empty
9	130C	empty
10	**131**	**10,000 ES cells**

TMT channel #

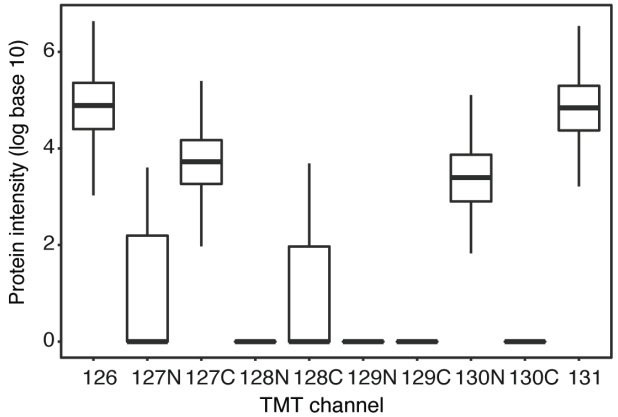


Figure 1. Optimization of the minimal sample protocol - requirements, expectations, and areas of optimization. The panels describe testing of the need for carrier channel, carrier channel input size and position. **a.** The table provides an overview of requirements and expected results for standard proteomics protocols, a recently developed single-cell method, and MinPut. **b.** We optimized the buffer composition and cell lysis method; the graph shows the numbers of protein groups detected by label-free mass spectrometry using FACS-purified 5,000 ES cells. The error bars show standard deviation. **c.** The panel shows an overview of the cell numbers used for different purposes in the protocol and comparisons. **d.** We compared TMT setups for using 1,000 cells with and without carrier channel and using 60,000,

20,000, and 10,000 cells for the carrier channel. The graph shows the numbers of protein groups detected using the experimental setup below. The error bars show standard deviation. **e.** We show that the carrier channel affects neighboring channels at 1 Thompson distance using a 10plex TMT experiment with empty channels 2-8 and sample from 10,000 cells in channels 1 and 10. ESC - embryonic stem cells, MN - motor neurons, FACS - fluorescence-activated cell sorting, TMT - tandem mass tag

To address these challenges, we developed a straightforward and robust method for mass spectrometry analysis of samples from a small number of cells. The minimal input method, called MinPut, uses 1,000 cells, does not require special tools or reagents and quantifies >2,500 proteins per sample (**Figure 1a**). Quantification accuracy and reproducibility as well as correlation with transcript abundances are comparable with those of standard proteomics. We established and validated the method in mouse embryonic stem cells and *in vitro* differentiated motor neurons.

Results

Simplified protein extraction to maximize sample retention

To develop the MinPut protocol and assess its performance, we optimized the mass spectrometry workflow at several steps and employed an established *in vitro* differentiation paradigm to compare between mouse embryonic stem cells (ESC) and motor neurons (MN) (**Figure 1b**)[11]. First, we tested different sonicators and buffers for cell lysis using 5,000 mouse embryonic stem cells purified by fluorescence-activated cell sorting (FACS) and label-free mass spectrometry (**Figure 1c**). We compared the results to those from a standard input sample, which was prepared from ~10 µg of protein. This amount of protein can typically be derived from ~500,000 mammalian cells. The standard sample preparation included cell lysis with a phosphate buffered saline-based buffer without detergent, using the Bioruptor sonicator and clean-up of peptides with reserve phase filters. For mass spectrometry analysis, we injected all of the peptides derived from the minimal input samples and 600 ng from the standard samples.

In contrast with other methods for small-sample analysis[4,10], the Bioruptor sonicator outperformed the Covaris sonicator with respect to protein identification (**Figure 1b**). Further, we omitted detergent from the lysis buffer, as it did not improve protein identification. We also omitted peptide reduction and alkylation which is used in standard sample preparations, as the peptide loss outweighed the procedures' benefits (*not shown)*. The results are reproducible using peptide instead of protein identifications (**Suppl. Figure 1**). For all cell lysis methods, protein abundances correlated well between technical replicates and more so than across methods (average $R^2 = 0.93$, **Suppl. Figure 2**).

Experimental design for protein quantitation with peptide mass tags

Next, we optimized the use of 10plex TMT labeling with a carrier channel for peptide quantitation. As a carrier, we used a pool of equal proportions of both cell types analyzed in the samples. First, we showed that the use of a carrier channel was essential to improve protein identification compared to a setup with minimal input in all channels ("No carrier", **Figure 1d,** first 3 bars). Second, we showed that using samples as small as 1,000 cells per channel yielded comparable protein identifications and reproducibility to those using 5,000 cells (**Suppl. Figure 3**).

Third, we showed that the carrier should ideally be placed in channel 10 with channel 8 empty as it produced erroneous signals in specific neighboring channels. To test this "leakiness", we used a TMT setup with channels empty except for positions 1 and 10 which both contained peptides prepared from 10,000 cells (**Figure 1e**). The carrier in channel 1 produced signals in channels 2, 3 and 5. The carrier in channel 10 did not affect channel 9 immediately adjacent to it. However, we observed a substantial signal at the -1 Thompson distance to the carrier in channel 10, i.e. in channel 8, indicating contamination of the mass tag with the light isotope. Indeed, the observed intensity in channel 8 was about 3-4% of the total intensity in channel 10, consistent with the contamination with the light isotope as reported by the company. For this reason, we placed the carrier into channel 10 and left channel 8 empty in all subsequent TMT experiments.

Finally, we minimized the size of the carrier channel, as a low carrier-to-sample ratio is advantageous with respect to signal strength (**Suppl. Figure 4**). In addition, smaller carriers require less cell sorting. We tested carrier channels with peptides derived from 60,000, 20,000, and 10,000 cells. We found that the carrier from 10,000 cells provided similar protein identification and reproducibility similar to the larger carrier channels. We thus used this carrier size in subsequent experiments, resulting in a final carrier-to-sample ratio of 10,000:1,000 = 10:1 (**Figure 1d,** last 3 bars).

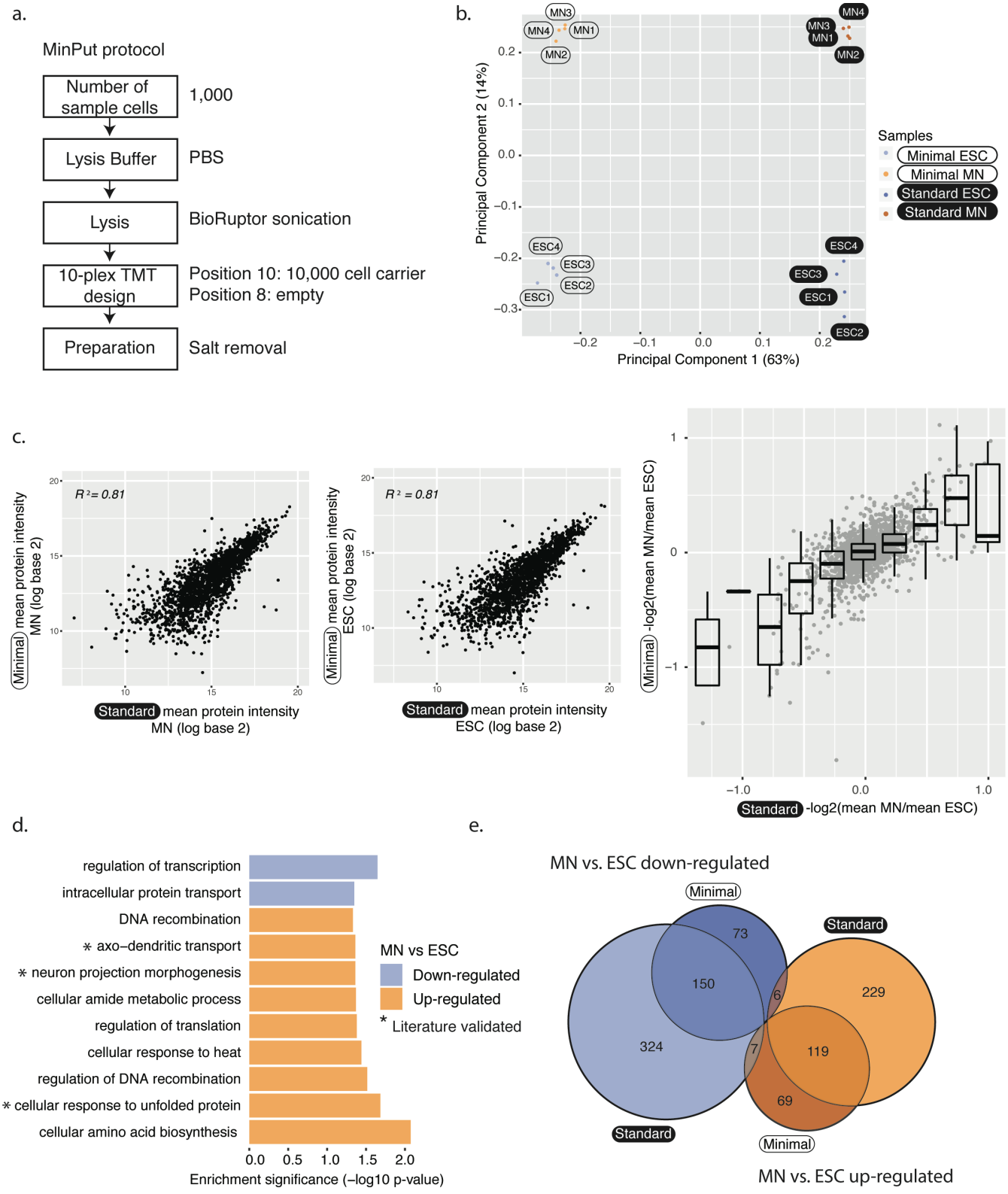


Figure 2. MinPut's protein abundance measurements are accurate and reproducible **a.** The flowchart illustrates the MinPut protocol with points of optimization. **b.** The first principal components of minimal and standard sample input data separate based on the protocol used as well as the two different cell types. The analysis was done using 1763 protein groups that were identified in both minimal and standard preparations. **c.** Protein abundances fold changes from minimal and standard sample input preparations correlate well (R^2 = 0.78 and 0.77, respectively). The analysis was done using 1763 protein groups that were identified in both minimal and standard preparations. **d.** Differentially expressed proteins have similar function enrichments between the MinPut and standard protocol (p-value<0.05). **e.** The MinPut and standard input protocols overlap in their results with respect to identification of proteins that are significantly differentially up- or down-regulated (p-value<0.05). There is significant overlap between the up- and down-regulated proteins from either preparation (p=2.070e^{-192} and

$p=7.639e^{-164}$ for MN vs. ESC down-regulated and up-regulated, respectively, hypergeometric test). The proteins found in common in the opposing groups are not significant (p=0.221 and p=0.191, hypergeometric test).

We evaluated MinPut's ability to reveal differential protein expression using an established *in vitro* motor neuron differentiation protocol from mouse embryonic stem cells[11]. The system is well-established with known markers of successful differentiation. While requiring only 1/50 of the number of cells, MinPut's proteome coverage reached almost 90% of what we detected in single-shot standard proteome analysis, i.e. 2,483 compared to 2,828 proteins, **Figure 1d**). Replicate experiments correlated with an average of R^2=0.99 in log-log abundance plots (**Suppl. Figure 5**), indicating high reproducibility.

Validating accuracy of measured protein abundance

Figure 2a shows an overview of the final MinPut protocol which uses 1,000 cells in experimental sample channels and 10,000 cells in the carrier channel. We validated MinPut's protein quantitation accuracy in samples from mouse embryonic stem cells and differentiated motor neurons in several ways. First, we compared the protein abundances analyzing samples from the standard and minimal input preparations. The first two principal components of the respective experiments explained a total of 77% of the variation in the data **(Figure 2b)**. The first component separated the two experiments according to the number of cells used as input. The second component separated the two cell types providing a first indication of MinPut's ability to produce biologically meaningful protein quantitation.

Second, we confirmed the consistency between the MinPut and standard protocol by direct correlation of the measured protein abundances in the two cell types: the correlation coefficient ranged between 0.81 and 0.77 between the two protocols (**Figure 2c**). Third, both minimal and standard input preparations showed similar correlation with corresponding transcript abundances as taken from bulk RNA sequencing samples from the same differentiation paradigm[12], with coefficients ranging from 0.39 to 0.43 (**Suppl. Figure 6**).

Finally, we validated MinPut's ability to identify differentially expressed proteins. To do so, we selected 229 and 195 proteins from the MinPut experiments with significant up- and down-regulation in motor neurons compared to stem cells, respectively (q-value<0.01, Student's t-test, **Figure 2d**). The proteins up-regulated in motor neurons were significantly enriched in several biological functions such as axo-dendritic transport and neuron projection morphogenesis (p-value<0.05, Fisher's exact test, **Figure 2d, Suppl. Figure 7**). For example, the neuron marker genes AINX[13], MAP1B[14], RABP1[14,15], STMN2[16] and TBB3[17] are all up-regulated in motor neurons (**Suppl. Figure 9**). In addition, the differentially expressed proteins are significantly enriched in plasma membrane proteins (p-value<0.01), some of which are the known neuron markers MAP1B and STMN2.

We also compared significantly differentially expressed proteins between the MinPut and standard sample preparations (**Figure 2e**). While more proteins were identified using the standard preparation and more proteins were differentially expressed between the two cell types (**Suppl. Figure 8**), the up- and down-regulated proteins from either preparation overlapped significantly (p<0.0001 for MN vs. ESC down-regulated and up-regulated, hypergeometric test). There are virtually no proteins in the opposing groups (p=0.22 and p=0.19, hypergeometric test). Combined, these results support MinPut's ability to quantify proteins correctly and to identify significantly differentially expressed proteins.

Discussion

We present a straightforward and robust protocol, MinPut, for use of minimal input (1,000 cells) for proteomics analysis. The method quantifies ~2,500 proteins in mammalian cells and sensitively identifies differential expression. Further, we show that protein abundances measured by MinPut correlate well with transcript abundances of similar samples, and have high reproducibility across replicates. The correlation coefficients are similar to those reported for other cell systems[18].

In sum, we provide a protocol that uses sample amounts achievable by, for example, dissection of specific cell types or tissues *in vivo* or FACS purification of rare cell populations, enabling analysis of highly specific cell populations. MinPut achieves quantitation accuracy and reproducibility similar to that of standard proteomics. In addition, MinPut does not require specific equipment or reagents. Therefore, it enables analysis of systems in which it is very difficult to obtain large numbers of cells.

Data availability

The mass spectrometry data including the MaxQuant output files have been deposited to the ProteomeXchange Consortium via the PRIDE[19] partner repository with the dataset identifier PXD015123.

AUTHOR INFORMATION

Corresponding Author

[$]cvogel@nyu.edu

Author Contributions

[*]C.V. and L.C. contributed equally to the completion of the study.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

B.V. acknowledges funding by American Heart Association grant #18PRE33990254. D.E.I. acknowledges funding by the NIH/NINDS (1F31NS103447-01A1). E.O.M acknowledges funding by Project ALS (A13-0416). L.C. acknowledges funding by NIH/NHLBI R01 HL108643 and by the Leducq Foundation Trans-Atlantic network of excellence award 15CVD01. C.V. acknowledges funding by the NIH/NIGMS (R35GM127089), and the Zegar Family Foundation Fund for Genomics Research at New York University.

REFERENCES

(1) Marx, V. A Dream of Single-Cell Proteomics. *Nat. Methods* **2019**. https://doi.org/10.1038/s41592-019-0540-6.

(2) Lombard-Banek, C.; Moody, S. A.; Nemes, P. Single-Cell Mass Spectrometry for Discovery Proteomics: Quantifying Translational Cell Heterogeneity in the 16-Cell Frog (Xenopus) Embryo. *Angew. Chem. Int. Ed Engl.* **2016**, *55* (7), 2454–2458.

(3) Zhu, Y.; Piehowski, P. D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R. J.; Shukla, A. K.; Petyuk, V. A.; Campbell-Thompson, M.; Mathews, C. E.; et al. Nanodroplet Processing Platform for Deep and Quantitative Proteome Profiling of 10–100 Mammalian Cells. *Nature Communications*. 2018. https://doi.org/10.1038/s41467-018-03367-w.

(4) Li, S.; Plouffe, B. D.; Belov, A. M.; Ray, S.; Wang, X.; Murthy, S. K.; Karger, B. L.; Ivanov, A. R. An Integrated Platform for Isolation, Processing, and Mass Spectrometry-Based Proteomic Profiling of Rare Cells in Whole Blood. *Mol. Cell. Proteomics* **2015**, *14* (6), 1672–1683.

(5) Orecchioni, M.; Bedognetti, D.; Newman, L.; Fuoco, C.; Spada, F.; Hendrickx, W.; Marincola, F. M.; Sgarrella, F.; Rodrigues, A. F.; Ménard-Moyon, C.; et al. Single-Cell Mass Cytometry and Transcriptome Profiling Reveal the Impact of Graphene on Human Immune Cells. *Nat. Commun.* **2017**, *8* (1), 1109.

(6) Stoeckius, M.; Hafemeister, C.; Stephenson, W.; Houck-Loomis, B.; Chattopadhyay, P. K.; Swerdlow, H.; Satija, R.; Smibert, P. Simultaneous Epitope and Transcriptome Measurement in Single Cells. *Nat. Methods* **2017**, *14* (9), 865–868.

(7) Genshaft, A. S.; Li, S.; Gallant, C. J.; Darmanis, S.; Prakadan, S. M.; Ziegler, C. G. K.; Lundberg, M.; Fredriksson, S.; Hong, J.; Regev, A.; et al. Multiplexed, Targeted Profiling of Single-Cell Proteomes and Transcriptomes in a Single Reaction. *Genome Biology*. 2016. https://doi.org/10.1186/s13059-016-1045-6.

(8) Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J. Ultrasensitive Proteome Analysis Using Paramagnetic Bead Technology. *Mol. Syst. Biol.* **2014**, *10*, 757.

(9) Myers, S. A.; Rhoads, A.; Cocco, A. R.; Peckner, R.; Haber, A. L.; Schweitzer, L. D.; Krug, K.; Mani, D. R.; Clauser, K. R.; Rozenblatt-Rosen, O.; et al. Streamlined Protocol for Deep Proteomic Profiling of FAC-Sorted Cells and Its Application to Freshly Isolated Murine Immune Cells. *Mol. Cell. Proteomics* **2019**, *18* (5), 995–1009.

(10) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: Mass Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity during Cell Differentiation. *Genome Biol.* **2018**, *19* (1), 161.

(11) Mazzoni, E. O.; Mahony, S.; Closser, M.; Morrison, C. A.; Nedelec, S.; Williams, D. J.; An, D.; Gifford, D. K.; Wichterle, H. Synergistic Binding of Transcription Factors to Cell-Specific Enhancers Programs Motor Neuron Identity. *Nat. Neurosci.* **2013**, *16* (9), 1219–1227.

(12) Velasco, S.; Ibrahim, M. M.; Kakumanu, A.; Garipler, G.; Aydin, B.; Al-Sayegh, M. A.; Hirsekorn, A.; Abdul-Rahman, F.; Satija, R.; Ohler, U.; et al. A Multi-Step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. *Cell Stem Cell* **2017**, *20* (2), 205–217.e8.

(13) Chien, C. L.; Liem, R. K. Characterization of the Mouse Gene Encoding the Neuronal Intermediate Filament Protein Alpha-Internexin. *Gene* **1994**, *149* (2), 289–292.

(14) Bouquet, C.; Soares, S.; von Boxberg, Y.; Ravaille-Veron, M.; Propst, F.; Nothias, F. Microtubule-Associated Protein 1B Controls Directionality of Growth Cone Migration and Axonal Branching in Regeneration of Adult Dorsal Root Ganglia Neurons. *J. Neurosci.* **2004**, *24* (32), 7204–7213.

(15) Maden, M. Retinoic Acid in the Development, Regeneration and Maintenance of the Nervous System. *Nat. Rev. Neurosci.* **2007**, *8* (10), 755–765.

(16) Stein, R.; Mori, N.; Matthews, K.; Lo, L. C.; Anderson, D. J. The NGF-Inducible SCG10 mRNA Encodes a Novel Membrane-Bound Protein Present in Growth Cones and Abundant in Developing Neurons. *Neuron* **1988**, *1* (6), 463–476.

(17) Katsetos, C. D.; Legido, A.; Perentes, E.; Mörk, S. J. Class III Beta-Tubulin Isotype: A Key Cytoskeletal Protein at the Crossroads of Developmental Neurobiology and Tumor Neuropathology. *J. Child Neurol.* **2003**, *18* (12), 851–866; discussion 867.

(18) McManus, J.; Cheng, Z.; Vogel, C. Next-Generation Analysis of Gene Expression Regulation – Comparing the Roles of Synthesis and Degradation. *Molecular BioSystems*. 2015, pp 2680–2689. https://doi.org/10.1039/c5mb00310e.

(19) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450.