

# Limiting the impact of protein leakage in single-cell proteomics

Received: 5 November 2024

Accepted: 28 January 2025

Published online: 05 May 2025

 Check for updates

Andrew Leduc<sup>1</sup>✉, Yanxin Xu<sup>2</sup>, Gergana Shipkovenska<sup>2</sup>, Zhixun Dou<sup>2</sup> & Nikolai Slavov<sup>1,3</sup>✉

Limiting artifacts during sample preparation can significantly increase data quality in single-cell proteomics experiments. Towards this goal, we characterize the impact of protein leakage by analyzing thousands of primary single cells from mouse trachea. The cells were prepared either fresh immediately after dissociation or first cryopreserved and prepared at a later date. We directly identify permeabilized cells by imaging a cell permeable dye and use the data to define a signature for protein leakage. This signature is similar across diverse cell types and reflects increased leakage propensities for cytosolic and nuclear proteins compared to membrane and mitochondrial proteins. A classifier based on the signature allowed for the accurate identification of permeabilized cells across cell types and species. The classifier is integrated into QuantQC ([scp.slavovlab.net/QuantQC](http://scp.slavovlab.net/QuantQC)) to support its application to diverse samples and workflows.

Recent advances in throughput of single-cell proteomics by mass spectrometry have made it possible to quantify thousands of proteins across thousands of single cells and organelles<sup>1–5</sup>, including primary cells<sup>3,6–9</sup>. This will facilitate characterizing the influence of protein abundance, modification, synthesis and degradation on shaping the functions of single cells in heterogeneous tissue samples<sup>10</sup>. Realizing this potential requires minimizing changes in cells and proteins during the sample preparation, such as those that may be introduced by storage of dissociated tissue samples. However, freezing samples for short- or long-term storage may damage cells<sup>11,12</sup> and lead to potential bias in single-cell proteomics data.

Indeed, such effects have been observed in single-cell RNA sequencing, where transcripts have been found to leak out of cells with damaged membranes depending upon their localization in or outside of the mitochondria<sup>13</sup>. These cells are usually filtered out computationally based on a heuristic cut off which varies depending on the cell type<sup>13,14</sup>. However, an analogous characterization has not been performed in single-cell proteomics. Proteins vary significantly in subcellular localizations, physical properties and binding interactions, all of which may substantially affect their leakage propensities. Additionally, proteins are about tenfold smaller than the

mRNA that template them, making them more likely to leak upon membrane damage. For these reasons, we sought to characterize this effect.

Here, we show that protein leakage can affect protein quantification in mammalian cells frozen prior to sample preparation. We demonstrate that cells affected by protein leakage can be identified and excluded from analysis by using cell permeability staining or less directly by using a computational model that we develop and make available as part of an open source R package available at: <https://github.com/SlavovLab/QuantQC>.

## Results

We chose to work with primary tissue, mouse tracheal epithelium, to characterize the effect on diverse cell types. After dissociation using an enzyme cocktail as previously reported<sup>15</sup>, half of the cells were slowly frozen to  $-80^{\circ}\text{C}$  in 10% Dimethyl sulfoxide (DMSO) and 90% Fetal Bovine Serum (FBS) and prepared later, and the other half were immediately prepared fresh for single-cell proteomic analysis. Using nPOP with TMTpro 35-plex multiplexing, we prepared a total of 2784 single cells, 928 fresh and 1856 frozen. Samples were analyzed using the prioritized data acquisition (pSCOPE)<sup>6</sup> at a rate of 1018 cells/day

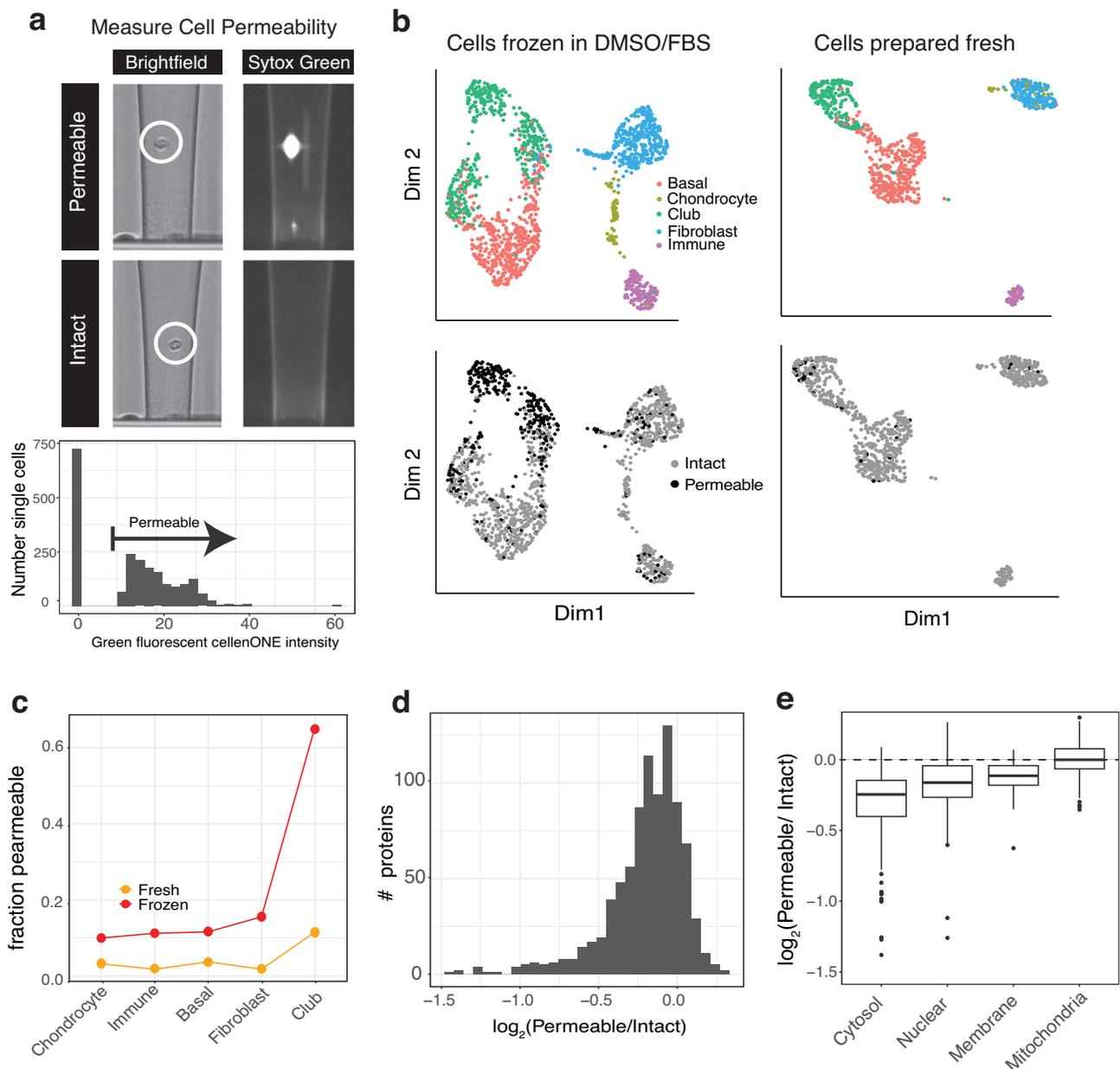
<sup>1</sup>Departments of Bioengineering, Biology, Chemistry and Chemical Biology, Single Cell Proteomics Center, Northeastern University, Boston, MA, USA. <sup>2</sup>Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>3</sup>Parallel Squared Technology Institute, Watertown, MA, USA.

✉ e-mail: [leduc.an@northeastern.edu](mailto:leduc.an@northeastern.edu); [nslavov@northeastern.edu](mailto:nslavov@northeastern.edu)

and average depth of 712 proteins quantified per cell. Summary reports of all data can be found in Supplementary Data 1 (frozen cells) and Supplementary Data 2 (fresh cells). Prior to cell isolation, the cells were stained with Sytox green to identify cells with compromised membrane permeability. We recorded the stain intensity of each cell and linked these measurements with downstream single-cell data using QuantQC (Fig. 1a). The distribution of Sytox green intensities across all single cells is bimodal, and the cells from the mode at 0 intensity were characterized as intact while the cells from the other mode as permeable (Fig. 1a). Further, the distribution of cell sizes from the two modes were indistinguishable within a cell type, suggesting that intensity difference is not due to cell size difference (Supplementary Fig. 1a). The Sytox green negative cells were 96% for fresh and 72% for frozen ones. The large increase in permeability for club cells may

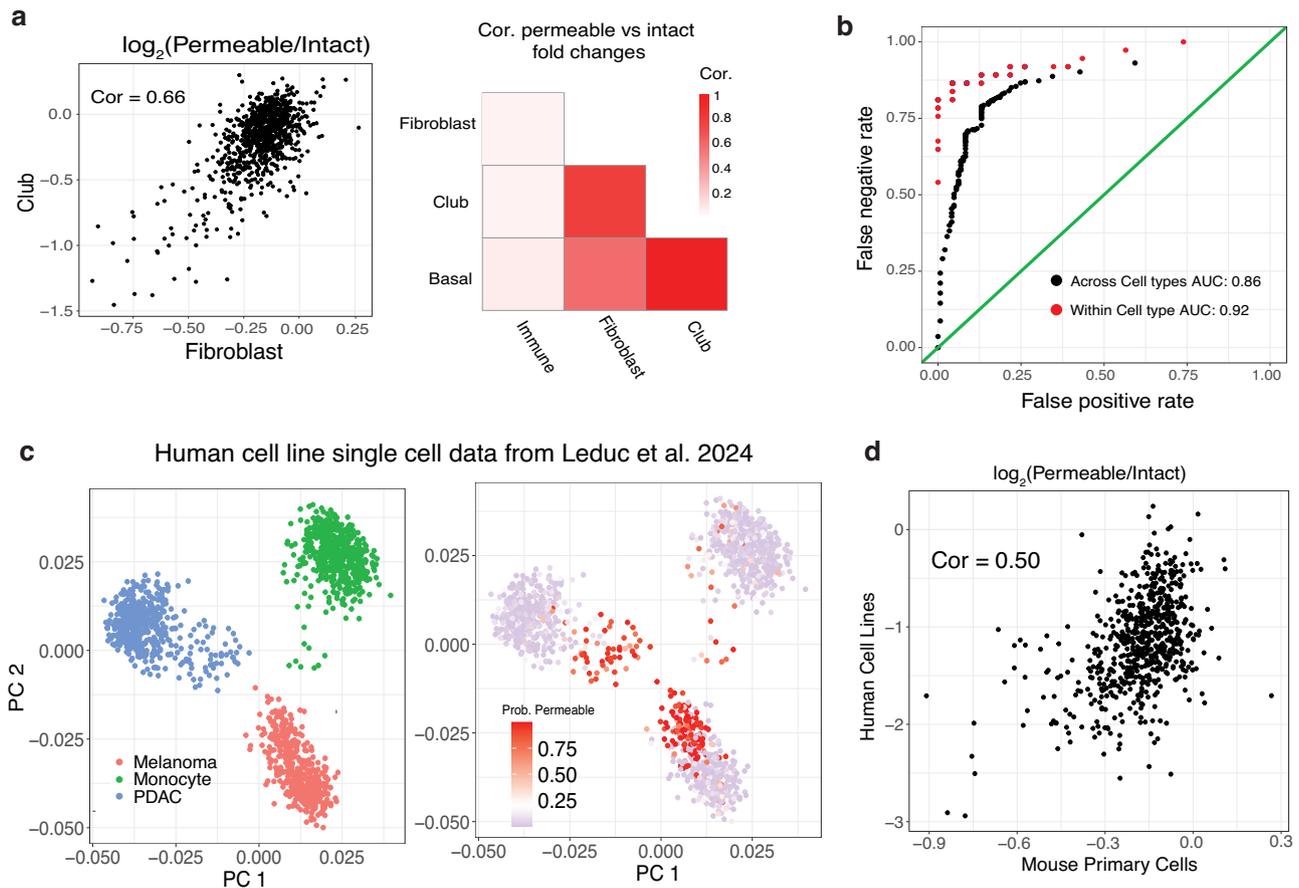
reflect their more fragile state, which has been noted in attempts to culture these cells<sup>16</sup>.

We then assigned cell types by using the LIGER<sup>17</sup> algorithm to perform label transfer from a previously annotated single-cell mRNA data set<sup>15,17</sup>. The correspondence to the single-cell mRNA data set was lower in frozen samples compared to the fresh. This was observed by reduced agreement in covariance patterns between shared genes<sup>18</sup> (Supplementary Fig. 1b). We also observed co-clustering of permeabilized cells in the frozen sample (Fig. 1b). We did not observe a separate permeable cell cluster in the fresh samples, which may have been due to the low number of permeable cells (Fig. 1b). Despite the lower confidence in integration score and co-clustering of permeabilized cells, we were still able to confidently assign cell types in the frozen condition based on the abundance of



**Fig. 1 | Quantifying protein leakage artifacts. a** Images of cells in the CellenONE nozzle taken with brightfield and with the green fluorescent channel. The two cells shown are not clearly distinguishable from the brightfield but one cell is permeable and thus positive for the Sytox green fluorescent dye. **b** The UMAP dimensionality reduction shows co-clustering of cells from fresh and frozen batches that were

recorded as permeable via Sytox green. **c** Percentages of permeable cells for the two sample handling conditions are shown in the dot plot. **d** Differences in protein fold changes between permeable and intact cells for all proteins. **e** Differences in protein fold changes between permeable and intact cells are categorized by sub-cellular compartment.



**Fig. 2 | Examining and predicting cell permeability in human cell lines.** **a** Log<sub>2</sub> average protein fold changes between permeable and intact Club cells and Fibroblasts are plotted against each other and show Pearson correlation of 0.66. The heatmap summarizes correlations between fold changes for all cell types is shown. **b** ROC curve for a classifier trained on permeability status of single cells using protein abundance profiles of the top 75 most significantly leaking proteins. Results for the model trained and tested on the same cell type are in red and trained and

tested on different cell types are in black. **c** PCA projection of single cells from Leduc et al., 2024. Cells are colored by cell type or their permeability score from the classifier. The cells towards the center of the two dimensional space of the first two principal components were enriched for high permeability scores. **d** When comparing the fold changes between cells with probability over 0.2 and under 0.2 to be permeable versus intact fold changes from the primary mouse tracheal cells, the fold changes strongly agree, Pearson correlation 0.50.

marker proteins for each cell type (Supplementary Fig. 1c–f). The proportion of permeable cells was unevenly distributed across cell types (Fig. 1c). This may reflect both the increased fragility of the more significantly affected cell types and the methods used for single-cell dissociation.

We next examined which proteins exhibited a significant change between permeable and intact cells (Fig. 1d). Most altered proteins were depleted from the permeabilized fraction, indicating that proteins leaked out from the cells characterized as permeable. The few proteins slightly more abundant in the non-permeabilized cells could suggest a cellular response pre-membrane rupture, but more likely reflects a small amount of measurement noise in estimating fold changes for proteins that do not leak. We also characterized the extent to which the reduction in protein levels was specific to different subcellular compartments. To this end, we plotted the difference in protein abundance between different subcellular compartments (Fig. 1e). Similar to the trends observed in mRNA sequencing, mitochondrial protein abundance is not significantly different between permeable and intact cells. In contrast, proteins localized to the cytosol have the most significant decrease in abundance between intact and permeable cells. Variation across the distribution of cytosolic proteins may reflect several factors. Certain proteins may differentially participate in large complexes that prevent easy escape from even a ruptured membrane. Additionally, this variation may reflect inexact localization knowledge

of these proteins. Examining specific proteins reveals metabolic enzymes such as peroxidases and enzymes involved in glycolysis such as Gapdh with approximately twofold reduced abundance. A full list of proteins with their fold change differences can be found in Supplementary Data 3.

To examine the generalizability of the protein leakage artifact, we estimated the average fold change for each protein between permeable and intact cells between cell types with greater than 5 permeable cells (Fig. 2a). Significant agreement of fold changes across cell types suggests that a similar mechanism of protein leakage is operating across cell types. However, weaker agreement between immune cells and other cell types may reflect a cell type specific component. The similarity in fold changes between cell types led us to explore the utility of a classifier for identifying cells with damaged membranes in data sets where permeability staining was not used. To do this, we trained an XGBoost model on cell permeability status using the abundances of the top 75 most significantly leaking proteins. XGBoost was chosen for its ability to handle missing data, as we preferred to not use imputed values in the classification task. To validate our model, we first trained and tested it on the same cell type leading to a high success rate of classification on the testing set with an AUC = 0.92 (Fig. 2b). We then sought to see how well the model could generalize by training it on fibroblasts, basal and immune cells and testing on basal cells. Performance decreased slightly but a significant portion of permeable

cells could be identified, AUC = 0.86, Fig. 2b. This may be partly the result of cell-type dependent leaking that may vary across cell types due to shifts in protein abundance and localization patterns. To facilitate easy usage of the classifier, it was incorporated into the QuantQC R package<sup>1</sup>.

To validate that this phenomenon was generalizable and not just specific to these samples, we next inspected a previously published data set of three human cell lines from our lab<sup>1</sup>. We trained our classifier for a final time on the entire mouse trachea data set and applied it to the human cells. Upon clustering and low dimensional projection in the principal component space, a population of cells closer to the center of the PCA space from each cell type was classified to have a high probability of permeabilization (Fig. 2c). We then compared the protein abundances between cells with probability greater or less than 0.2 of being permeable as assigned by the classifier. Fold changes between permeable and intact cells from our mouse trachea data set and found strong agreement with the human cell line fold changes (Fig. 2d). This strong agreement suggests these cells may suffer from the same artifact, albeit this does not provide conclusive validation as no Sytox Green intensity was measured for the human cells.

## Discussion

Our data demonstrate a substantial impact of protein leakage on single-cell proteomic measurements and a direct resolution based on excluding permeabilized cells from analysis. We also provide protein signatures of leakage and classification tools which may be used to detect and correct for this artifact. Protein leakage may also be mitigated by additional methods, such as cross-linking reagents, but we did not evaluate the effectiveness of such methods here.

While permeabilized cells may be excluded when performing sample preparation with methods that have fluorescent cell sorting capabilities, this capability is not ubiquitous. Indeed, recent methods have been developed using instruments that lack these capabilities<sup>19</sup>. In such cases, our classification model can help identify and correct for this artifact and thus improve data interpretation. Our results reinforce the importance of incorporating solutions for this problem in the community guidelines and best practices<sup>2</sup> and point to effective solutions. These solutions will reduce batch effects for and facilitate the analysis of frozen dissociated tissue samples, enabling experimenters greater flexibility in future studies.

## Methods

### Mouse model and handling

All mice experiments were performed in compliance with the Institutional Animal Care and Use Committee at Massachusetts General Hospital. 4-month-old C57BL/6 mice were ordered from the NIA. Mice were euthanized with CO<sub>2</sub> followed by cervical dislocation. The mouse used was male. Tissues were harvested post-euthanasia and perfusion with PBS.

### Tissue dissociation and cell suspension generation

Freshly dissected whole trachea preparations were submerged in 500  $\mu$ l of each enzyme dissociation cocktail for 30 min with gentle rocking at 37 °C. Papain (13.3  $\mu$ g/ml or 10  $\mu$ g/ml) was dissolved in EBSS buffer before mixing with activation buffer consisting of 0.067 mM  $\beta$ -mercaptoethanol, 1.1 mM EDTA and 5.5 mM cysteine-HCl in EBSS. Enzyme mix used for the dissociation protocol consisted of 25  $\mu$ l of 70 kU ml<sup>-1</sup> collagenase I, 25  $\mu$ l of 50 kU/ml hyaluronidase, 50  $\mu$ l of 7.5 kU/ml DNase, 120  $\mu$ l of 2.5  $\mu$ g/ml dispase and 400  $\mu$ l of 40  $\mu$ g/ml papain, to a final volume of 5 ml using DMEM. In all cases, single enzyme incubations were done for 30 min at 37 °C with gentle rocking, whereas the enzyme cocktail mix was incubated for 20 min at 37 °C

with gentle rocking. Cells were then either taken directly for single cell sample preparation, or frozen down in a cryopreservative buffer of 10% DMSO and 90% FBS.

### Proteomic sample preparation

Samples were prepared using the nPOP sample preparation method for multiplexed single cell proteomics<sup>1</sup>. Briefly, single cells were washed twice from either dissociation buffer or cryopreservation buffer with 1X PBS. Cells were then resuspended at a concentration of 1000 cells per  $\mu$ l and were incubated on ice and in the dark for 20 min with Sytox Green Dead Cell Stain (Thermo Fisher S34860). Cells were then washed one final time to remove dye and resuspended in 1X PBS at a concentration of 300 cells per  $\mu$ l for eventual cell sorting. Cells were then sorted in a volume of 300 pL into 9  $\mu$ l of 100% DMSO droplets on the surface of a fluorocarbon-coated glass slide for cell lysis using the CellenONE cell sorter and liquid handler. As cells were sorted, the fluorescent intensity of the Sytox Green stain was recorded using the CellenONE's green channel. The cells were isolated based on cell size and aspect ratios using threshold gates shown in Supplementary Fig. 2. The single cells were incubated for overnight digestion with the aid of the CellenONE's humidifier and slide cooling for evaporation prevention of droplets. The remaining single-cell suspension was used for making isobaric carriers and reference with the mPOP protocol<sup>20</sup>. Briefly, the cells were pelleted and resuspended in mass spectrometry grade water at a concentration of 1000 cells per  $\mu$ l. Then, the cells were lysed and the protein digested with 13.5  $\mu$ l of digestion buffer of 100 ng/ $\mu$ l Promega trypsin, 0.025% DDM, and 10 mM HEPES at pH 8.5.

The next day, the nPOP single-cell samples were labeled with 20  $\mu$ l of TMTpro 35-plex reagents dissolved in 100% DMSO at a concentration of 8.3  $\mu$ g/ $\mu$ l. Cells were labeled in sets of 29 as 126C and 127N are reserved for carrier and reference channels, 127C is excluded due to isotopic impurities from the carrier, and we did not have access to the full 35-plex set at the time of the sample preparation. The labeling reaction was then quenched with 20  $\mu$ l of 1% Hydroxylamine. Samples were pooled using the CellenONE in a 50%/50% solution of Acetonitrile and Water and dispensed into a 384 well PCR plate, dried down in a speed vac and stored at -20C for later injection for LC/MS analysis.

Cells frozen at -80 C in water were then lysed using heat at 90 C for 10 min following the mPOP sample preparation<sup>18</sup>. Cells were then digested overnight at 37 C in 100 mM TEAB buffer pH 8.5 and 10 ng/ $\mu$ l trypsin. Labels 126C and 127N were used to label carrier and reference samples, respectively. The samples were combined and diluted in 0.1% formic acid to a concentration of 10 ng/ $\mu$ l of peptide from the 126C labeled sample for carrier and 0.5 ng/ $\mu$ l from the reference. Samples from the plate were resuspended in 1  $\mu$ l of the carrier mix for injection.

### LC/MS analysis

Samples were run on an Exploris 480 mass spectrometer with a Vanquish Neo liquid chromatography and autosampler. A 25 cm 75  $\mu$ m ID IonOptics column was used for the chromatography column. The gradient ramped from 8% to 40% buffer B (80% Acetonitrile 20% 0.1% formic acid) over the course of 28 min with a 4 min wash at 90% buffer B at the end. Mass spectrometry data acquisition of single cells was performed using MaxQuant Live for Prioritized data acquisition with 60k MS1 and MS2 resolution, 118 ms maximum injection time and 1e6 maximum AGC<sup>5</sup>. Briefly, an original inclusion list was generated using a DIA run of the carrier sample which resulted in roughly 10,000 identified precursors. This DIA run utilized 60k MS1 resolving power and 45k MS2 resolving power and had 30 equally spaced MS2 windows from range 400 to 900 *m/z* with 1 DA overlap between MS2 frames. The max IT was 50 ms for MS2 scans and AGC was set to 300%. The LC gradient was identical to that used for the single-cell runs to maintain

accurate retention times. Priority tiers were assigned at three levels, which were equally sized and faceted based on precursor abundance, with highest abundance precursors placed on the highest priority tier. The inclusion list was then refined using two scout runs to identify and remove precursors which had low precursor ion fraction below a value of 0.7.

The DIA run for inclusion list generation used a single MS1 scan and 26 Th wide MS2 scans spanning 400–900 *M/Z* space with 1 Th overlap window to window. The chromatography gradient was the same as specified for the prioritized runs.

### Raw and processed data analysis

Raw data from data-dependent and prioritized data acquisition runs were searched by MaxQuant version 2.4.3.0 against a protein sequence database including entries from the appropriate murine SwissProt database (downloaded July 30, 2018) containing 20,386 proteins and known contaminants such as human keratins and common lab contaminants. The modifications for the TMTpro 35plex tags can be found in the supplemental data of the nPOP protocol<sup>1</sup>. Results were filtered at 1% FDR. The DIA run for inclusion list generation was searched using the DIA-NN software with TMT specified as a fixed modification.

Downstream data analysis was performed in R. Single-cell data was processed to obtain a protein X single-cell matrix of log<sub>2</sub> relative protein fold changes using the QuantQC package<sup>1</sup>. Cell types were then assigned by integrating of single-cell data with pre-annotated mRNA sequencing data from the same tissue type and dissociation procedure<sup>15</sup>. The LIGER algorithm was used to project the cells into the same high-dimensional space. Clustering was performed and protein single cells were assigned the identity of the predominant annotated mRNA single cells from the given cluster. Fold changes were then calculated between permeable and intact cells by taking the average fold change of permeable and intact cells within a cell type for each protein and subtracting the two vectors.

### Compartment GO term curation

To assign proteins to a given compartment, we used only proteins whose assignment to a specific compartment, i.e., if a protein was listed in both the cytoplasm and nucleus, it was not used. Thus not all proteins were used in the analysis of compartment specific protein leakage.

### Cell permeability classifier

The model used to classify permeable vs intact single cells XGboost was trained on the top 75 most significantly leaking proteins (Supplemental file 3) across single cells. Data was z-scored within each protein (across single cells) prior to training the model. Missing data was left as NA. For the within cell type comparison, the 420 club cells from the Frozen data set were randomly split into train and test sets of 80% and 20% of the data, respectively. For the across cell type comparison, the model was trained on all cells except for club cells and then tested on all club cells.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw and searched MS data have been deposited in accordance with community guidelines<sup>2</sup> and can be found at MassIVE repository MSV000094790. Additional processed data can be found at [https://scp.slavovlab.net/Leduc\\_et\\_al\\_2024](https://scp.slavovlab.net/Leduc_et_al_2024). Source data are provided with this paper.

### Code availability

The analysis can be reproduced by using resources from [https://scp.slavovlab.net/Leduc\\_et\\_al\\_2024](https://scp.slavovlab.net/Leduc_et_al_2024) and following instructions found at <https://github.com/SlavovLab/CellPermeability>. The classification tool has been incorporated into the QuantQC package, <https://github.com/SlavovLab/QuantQC> via the function FindPermeableCells.

### References

- Leduc, A., Koury, L., Cantlon, J. & Slavov, N. Massively parallel sample preparation for multiplexed single-cell proteomics using nPOP. *Nat. Protocols* <https://doi.org/10.1038/s41596-024-01033-8> (2024).
- Gatto, L. et al. Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments. *Nat. Methods* **20**, 375–386 (2023).
- Schoof, E. M. et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* **12**, 1–15 (2021).
- Kelly, R. T. Single-cell proteomics: progress and prospects. *Mol. Cell. Proteomics* **19**, 1739–1748 (2020).
- Derks, J. et al. Single-nucleus proteomics identifies regulators of protein transport. Preprint at *bioRxiv* <https://pubmed.ncbi.nlm.nih.gov/38948785/> (2024).
- Huffman, R. G. et al. Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. *Nat. Methods* **20**, 714–722 (2023).
- Rosenberger, F. A. et al. Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *Nat. Methods* **20**, 1530–1536 (2023).
- Karagach, S., Smollich, J., Atrakchi, O., Mohan, V. & Geiger, T. High throughput single-cell proteomics of in-vivo cells. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2024.11.01.621461v1.full> (2024).
- Khan, S. et al. Inferring post-transcriptional regulation within and across cell types in human testis. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.10.08.617313> (2024).
- Leduc, A., Harens, H. & Slavov, N. Modeling and interpretation of single-cell proteogenomic data. Preprint at <https://arxiv.org/abs/2308.07465> (2023).
- Massoni-Badosa, R. et al. Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol.* **21**, 112 (2020).
- Onat, B. et al. Cell storage conditions impact single-cell proteomic landscapes. *J. Proteome Res.* <https://doi.org/10.1021/acs.jproteome.4c00632> (2025).
- Ilicic, T. et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016).
- Osorio, D. & Cai, J. J. Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics* **37**, 963–967 (2021).
- Lin, B. et al. Airway hillocks are injury-resistant reservoirs of unique plastic stem cells. *Nature* <https://doi.org/10.1038/s41586-024-07377-1> (2024).
- Bukowy-Bieryłto, Z. et al. In vitro differentiation of ciliated cells in ALI-cultured human airway epithelium – The framework for functional studies on airway differentiation in ciliopathies. *Eur. J. Cell Biol.* **101**, 151189 (2022).
- Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
- Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
- Sanchez-Avila, X. et al. Easy and accessible workflow for label-free single-cell proteomics. *J. Am. Soc. Mass Spectrom.* **34**, 2374–2380 (2023).

20. Petelski, A. A. et al. Multiplexed single-cell proteomics using SCoPE2. *Nat. Protoc.* **16**, 5398–5425 (2021).

## Acknowledgements

We thank Sarah Sipe for help with MS data acquisition at PTI and members of the Slavov laboratory for thoughtful discussions. The work was funded by a Bits to Bytes award from MLSC to N.S., an NIGMS award R01GM144967 to N.S., and a MIRA award from the NIGMS of the NIH (R35GM148218) to N.S., and a UH3CA268117 award from NIH to N.S. and Z.D.

## Author contributions

Experimental design: A.L. and N.S. Sample preparation: A.L., Y.X., G.S., and Z.D. LC-MS/MS: A.L. Data analysis and writing: A.L. and N.S. All authors approved the final manuscript.

## Competing interests

N.S. is a founding director and CEO of Parallel Squared Technology Institute, which is a nonprofit research institute. The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56736-7>.

**Correspondence** and requests for materials should be addressed to Andrew Leduc or Nikolai Slavov.

**Peer review information** *Nature Communications* thanks Josh Elias and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025