

¹ scpviz: A Python bioinformatics toolkit for Single-cell Proteomics and multi-omics analysis

³ Marion Pang  ^{1¶}, Baiyi Quan  ², Ting-Yu Wang  ², and Tsui-Fen Chou  ^{1,2¶}

⁵ 1 Division of Biology and Biological Engineering, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125 2 Proteome Exploration Laboratory, Beckman Institute, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125 ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#))

⁸ Summary

⁹ Proteomics seeks to characterize protein dynamics by measuring both protein abundance and ¹⁰ post-translational modifications (PTMs), such as phosphorylation, acetylation, and ubiquitination, which regulate protein activity, localization, and interactions. In bottom-up proteomics ¹¹ workflows, proteins are enzymatically digested into peptides that are measured as spectra, ¹² from which these peptide-spectrum matches (PSMs) are aggregated to infer protein-level ¹³ identifications and quantitative abundance estimates. Analyzing the two levels of data at ¹⁴ both the peptide level (short fragments observed directly) and the protein level (assembled ¹⁵ from peptide evidence) in tandem is crucial for translating raw measurements into biologically ¹⁶ interpretable results.

¹⁸ Single-cell proteomics extends these approaches to resolve protein expression at the level of individual cells or microdissected tissue regions. Such data are typically sparse, with many ¹⁹ missing values, and are generated within complex experimental designs involving multiple ²⁰ classes of samples (e.g., cell type, treatment, condition). These properties distinguish single-cell ²¹ proteomics from bulk experiments and create unique challenges in data processing, normalization, ²² and interpretation. The single-cell transcriptomics community has established a mature ²³ ecosystem for managing similar challenges, exemplified by the scanpy package ([Wolf et al., 2018](#)) ²⁴ and the broader scverse ecosystem. Building on these foundations, scpviz extends ²⁵ the AnnData data structure to the domain of proteomics. scpviz is a Python package ²⁶ that streamlines single-cell and spatial proteomics workflows, supporting a complete analysis ²⁷ pipeline from raw peptide-level data to protein-level summaries and downstream interpretation ²⁸ through differential expression, enrichment analysis, and network analysis. At its core, scpviz ²⁹ centers the pAnnData class, an AnnData-affiliated data structure specialized for proteomics. ³⁰ Together, these components make scpviz a comprehensive and extensible framework for ³¹ single-cell proteomics. By combining flexible data structures, reproducible workflows, and ³² seamless integration with the AnnData/Scanpy ecosystem, the package enables researchers to ³³ efficiently connect peptide-level evidence to protein-level interpretation, thereby accelerating ³⁴ methodological development and biological discovery in proteomics.

³⁶ Statement of need

³⁷ Although general-purpose data analysis frameworks such as scanpy ([Wolf et al., 2018](#)) and ³⁸ the broader scverse ecosystem have become indispensable for single-cell transcriptomics, ³⁹ comparable tools for proteomics remain limited. Existing proteomics software often focus on ⁴⁰ specialized tasks (e.g., peptide identification or spectrum assignment) and do not provide a ⁴¹ unified framework for downstream analysis of peptide- and protein-level data within single-cell

42 and spatial contexts.

43 scpviz addresses these gaps by offering an integrated system for the complete proteomics work-
44 flow, from raw peptide-level evidence to protein-level summaries and biological interpretation.
45 It is designed for computational biologists and proteomics researchers working with low-input
46 or single-cell datasets from data sources such as Proteome Discoverer or DIA-NN(Demichev et
47 al., 2020).

48 At the core of scpviz is the pAnnData class, an AnnData-affiliated data structure specialized
49 for proteomics. It organizes peptide (.pep) and protein (.prot) AnnData objects alongside
50 supporting attributes such as .summary, .metadata, .rs matrices (protein-peptide relation-
51 ships), and .stats. This design allows users to move flexibly between peptides and
52 proteins while maintaining compatibility with established Python libraries for data science and
53 visualization.

54 Beyond data organization, scpviz implements proteomics-specific operations, including fil-
55 tering (e.g., requiring proteins supported by at least two unique peptides), normalization
56 and imputation methods tailored for sparse datasets, and visualization tools such as PCA,
57 UMAP, clustermaps, and abundance plots. For downstream interpretation, it integrates with
58 UniProt for annotation and STRING for enrichment and network analysis. The framework also
59 incorporates single-cell proteomics-specific normalization strategies such as directlfq (Ammar
60 et al., 2023), ensuring robust quantification across heterogeneous samples. Finally, pAnnData
61 objects interface seamlessly with scanpy (Wolf et al., 2018) and other ecosystem tools such
62 as harmony (Korsunsky et al., 2019), enabling direct incorporation into established single-cell
63 workflows.

64 The design philosophy of scpviz emphasizes both usability and extensibility. General users can
65 rely on its streamlined API to import, process, and visualize single-cell proteomics data without
66 deep programming expertise, while advanced users can extend the framework to accommodate
67 custom analysis pipelines. The package has already been applied in published manuscripts and
68 preprints (Dutta et al., 2025; Pang et al., 2025; Uslan et al., 2025) as well as manuscripts
69 in preparation, and it has been incorporated into graduate-level training to illustrate how
70 proteomics workflows parallel to those in single-cell transcriptomics.

71 The applications of scpviz span diverse areas of life sciences research, from studying protein
72 dynamics and signaling pathways to integrating proteomics with transcriptomics for multi-
73 omics analysis. By bridging the gap between raw mass spectrometry data and systems-level
74 interpretation, scpviz provides a versatile and reproducible platform for advancing single-cell
75 and spatial proteomics.

76 Acknowledgements

77 We acknowledge contributions from Pierre Walker, and support from the A*STAR Scholarship
78 (BS-PhD) during the genesis of this project.

79 References

- 80 Ammar, C., Schessner, J. P., Willems, S., Michaelis, A. C., & Mann, M. (2023). Accurate
81 Label-Free Quantification by directLFQ to Compare Unlimited Numbers of Proteomes.
82 *Molecular & Cellular Proteomics*, 22(7). <https://doi.org/10.1016/j.mcpro.2023.100581>
- 83 Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2020). DIA-
84 NN: Neural networks and interference correction enable deep proteome coverage in high
85 throughput. *Nature Methods*, 17(1), 41–44. <https://doi.org/10.1038/s41592-019-0638-x>
- 86 Dutta, S., Pang, M., Coughlin, G. M., Gudavalli, S., Roukes, M. L., Chou, T.-F., & Grdinaru,
87 V. (2025, February 11). *Molecularly-guided spatial proteomics captures single-cell identity*

- 88 and heterogeneity of the nervous system. <https://doi.org/10.1101/2025.02.10.637505>
- 89 Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner,
90 M., Loh, P., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of
91 single-cell data with Harmony. *Nature Methods*, 16(12), 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>
- 93 Pang, M., Jones, J. J., Wang, T.-Y., Quan, B., Kubat, N. J., Qiu, Y., Roukes, M. L., & Chou,
94 T.-F. (2025). Increasing Proteome Coverage Through a Reduction in Analyte Complexity
95 in Single-Cell Equivalent Samples. *Journal of Proteome Research*, 24(4), 1528–1538.
96 <https://doi.org/10.1021/acs.jproteome.4c00062>
- 97 Uslan, T., Quan, B., Wang, T.-Y., Pang, M., Qiu, Y., & Chou, T.-F. (2025). In-Depth
98 Comparison of Reagent-Based Digestion Methods and Two Commercially Available Kits
99 for Bottom-Up Proteomics. *ACS Omega*, 10(10), 10642–10652. <https://doi.org/10.1021/acsomega.4c11585>
- 101 Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression
102 data analysis. *Genome Biology*, 19(1), 15. <https://doi.org/10.1186/s13059-017-1382-0>