





SCIMAP: A Python Toolkit for Integrated Spatial Analysis of Multiplexed Imaging Data

Ajit J. Nirmal ^{1,2,3} and Peter K. Sorger ^{2,3}

1 Department of Dermatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States of America **2** Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA, United States of America **3** Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, United States of America  Corresponding author

DOI: [10.21105/joss.06604](https://doi.org/10.21105/joss.06604)

Software

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

Editor: [Charlotte Soneson](#)  

Reviewers:

- [@jluethi](#)
- [@kevinyamauchi](#)
- [@NMontanaBrown](#)

Submitted: 02 April 2024

Published: 28 May 2024

License

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

Summary

Multiplexed imaging data are revolutionizing our understanding of the composition and organization of tissues and tumors (["Catching up with Multiplexed Tissue Imaging," 2022](#)). A critical aspect of such "tissue profiling" is quantifying the spatial relationships among cells at different scales from the interaction of neighboring cells to recurrent communities of cells of multiple types. This often involves statistical analysis of 10^7 or more cells in which up to 100 biomolecules (commonly proteins) have been measured. While software tools currently cater to the analysis of spatial transcriptomics data ([Liu et al., 2022](#)), there remains a need for toolkits explicitly tailored to the complexities of multiplexed imaging data including the need to seamlessly integrate image visualization with data analysis and exploration. We introduce SCIMAP, a Python package specifically crafted to address these challenges. With SCIMAP, users can efficiently preprocess, analyze, and visualize large datasets, facilitating the exploration of spatial relationships and their statistical significance. SCIMAP's modular design enables the integration of new algorithms, enhancing its capabilities for spatial analysis.

Statement of Need

A variety of methods have been introduced for highly multiplexed imaging of tissues, including MxIF, CyCIF, CODEX, 4i, mIHC, MIBI, IBEX, and IMC ([Angelo et al., 2014](#); [Gerdes et al., 2013](#); [Giesen et al., 2014](#); [Goltsev et al., 2018](#); [Gut et al., 2018](#); [Lin et al., 2018](#); [Tsujioka et al., 2017](#)); although these methods differ in their implementations, all enable the collection of single-cell data on 20-100 proteins within preserved 2D and 3D tissue microenvironments. Analysis of high-plex images typically involves joining adjacent image tiles together and aligning channels across imaging cycles (stitching and registration) to generate a composite high-plex image and then identifying the positions and boundaries of individual cells via segmentation. The intensities of individual protein antigens, stains, and other detectable molecules are then quantified on a per-cell basis. This generates a "spatial feature table" (analogous to a count table in sequencing) that can be used to identify individual cell types and states; tools from spatial statistics are then used to identify how these cells are patterned in space from scales ranging from a few cell diameters ($\sim 10 \mu\text{m}$) to several millimeters.

Spatial feature tables provide the quantitative data for analysis of high-plex data but human inspection of the original image data remains essential. At the current state of the art, many of the critical morphological details in high-resolution images cannot be fully and accurately quantified. Segmentation is also subject to errors identifiable by humans, but not fully resolvable computationally ([Baker et al., 2024](#)). As a consequence, computation of spatial features and relationships must be performed in combination with visualization of the underlying image data. Humans excel at identifying tissue features that correspond to

classical histo-morphologies; they are also effective at discriminating foreground signals from variable background (Nirmal et al., 2023) using a process of “visual gating” (perception of high and low-intensity levels while visualizing an image). More generally, effective integration of visualization and computation enables nuanced interpretation of cellular organization in relation to established tissue architectures.

While packages such as squidpy (Palla et al., 2022), Giotto (Dries et al., 2021), and Seurat (Hao et al., 2024) have the potential to manage multiplexed imaging data, their functionalities are primarily optimized for spatial transcriptomics data. In contrast, SCIMAP is specifically designed to address the unique requirements of multiplexed imaging data analysis, offering features such as image-based visual gating and the integration of prior knowledge for cellular phenotyping, among others. SCIMAP uses the Python-based Napari (Ahlers et al., 2023; Chiu et al., 2022) image viewer to leverage these capabilities by providing a seamless interface to inspect and annotate high-plex imaging data alongside computational analysis. For example, we have implemented an image-based gating approach that allows users to visually determine the threshold that discriminates background from a true signal at both a whole-specimen and single-cell level. Users can also select specific regions of interest (ROIs) for selective or deeper analysis. This involves drawing ROIs over images (freehand or geometric) and then selecting the underlying single cell data for further analysis. This capability is essential for incorporating histopathological information on common tissue structures (e.g., epidermis, dermis, follicles), immune structures (e.g., secondary and tertiary lymphoid structures), tumor domains (e.g., tumor center, boundary, tumor buds), and tumor grade or stage (e.g., early lesions, invasive regions, established nodules). It also allows for excluding regions affected by significant tissue loss, folding, or artifactual staining (Baker et al., 2024). SCIMAP then performs statistical and spatial analyses on individual ROIs or sets of ROIs. Spatial analysis, including the measurement of distances between cells, analysis of interaction patterns, categorization into neighborhoods, and scoring of these patterns, is crucial for elucidating the cellular communications that underpin the functional aspects of the biology being studied. SCIMAP offers various functions to facilitate these analyses.

Lastly, a single high-plex whole slide image can exceed 100GB and 10^6 cells, necessitating optimized functions for handling large matrices and images. SCIMAP employs the well-established AnnData object structure, complemented by Dask and Zarr for efficient image loading in Napari. This approach facilitates seamless viewing of images with overlaid data layers, thus enabling effective analysis of large datasets. To date, SCIMAP has been used in the analysis of over 5 datasets from 8 tissue and cancer types (Gaglia et al., 2023; Maliga et al., 2024; Nirmal et al., 2022; Yapp et al., 2024).

Availability and Features

SCIMAP is available as a standalone Python package for interactive use, in Jupyter Notebooks for example, or can be accessed via a command-line interface (CLI; only a subset of functions that do not require visualization) for cloud-based processing. The package can be accessed at [GitHub](https://github.com/scimaps) and installed locally through pip. Installation, usage instructions, general documentation, and tutorials, are available at <https://scimap.xyz/>. See [Figure 1](#) for a schematic of the workflow and system components.

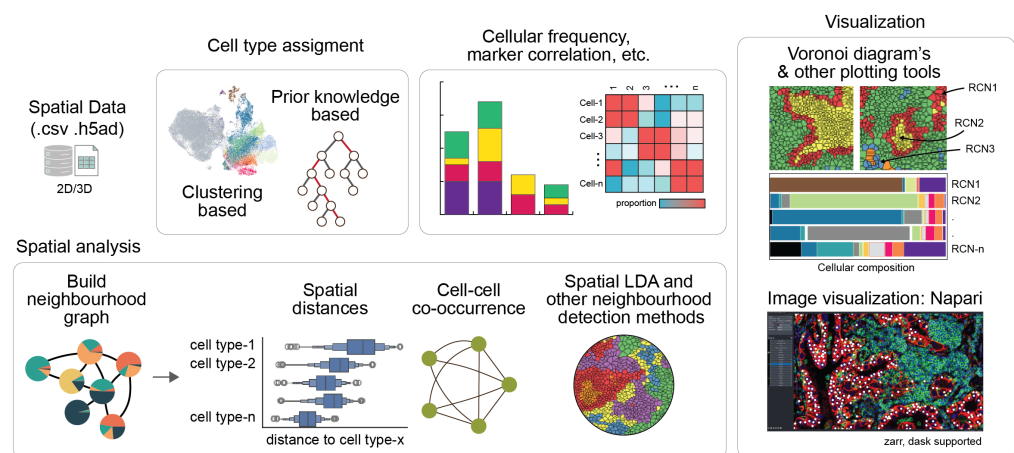


Figure 1: SCIMAP Workflow Overview. The schematic highlights data import, cell classification, spatial analysis, and visualization techniques within the SCIMAP tool box.

SCIMAP operates on segmented single-cell data derived from imaging data using tools such as cellpose (Stringer et al., 2021) or MCMICRO (Schapiro et al., 2022). The essential inputs for SCIMAP are: (a) a single-cell expression matrix and (b) the X and Y coordinates for each cell. Additionally, multi-stack OME-TIFF or TIFF images can be optionally provided to enable visualization of the data analysis on the original raw images.

SCIMAP comprises four main modules: preprocessing, analysis tools, visualization, and external methods. The preprocessing tools include functions for normalization, batch correction, and streamlined import from cloud processing pipelines such as MCMICRO (Schapiro et al., 2022). The analysis tools offer standard single-cell analysis techniques such as dimensionality reduction, clustering, prior knowledge-based cell phenotyping (a method through which cells are classified into specific cell types based on patterns of marker expression defined by the user), and various spatial analysis tools for measuring cellular distances, identifying regions of specific cell type aggregation, and assessing statistical differences in proximity scores or interaction frequencies. SCIMAP also includes neighborhood detection algorithms that utilize spatial-LDA (Wang & Grimson, 2007) for categorical data (cell types or clusters) and spatial lag for continuous data (marker expression values). All tools within the SCIMAP package for spatial analysis are compatible with both 2D and 3D data. Most analysis tools come with corresponding visualization functions to plot the results effectively. Additionally, the external methods module facilitates the integration of new tools developed by the community into SCIMAP, further extending its utility and applicability to both 2D and 3D data.

Acknowledgements

This work was supported by NCI grant R00CA256497 to A.J.N, U01-CA284207 to P.K.S and by Ludwig Cancer Center at Harvard.

We would like to thank our many users and members of the Laboratory for Systems Pharmacology at Harvard Medical School for their invaluable feedback.

References

- Ahlers, J., Althviz Moré, D., Amsalem, O., Anderson, A., Bokota, G., Boone, P., Bragantini, J., Buckley, G., Burt, A., Bussonnier, M., Can Solak, A., Caporal, C., Doncila Pop, D., Evans, K., Freeman, J., Gaifas, L., Gohlke, C., Gunalan, K., Har-Gil, H., ... Yamauchi, K.

- (2023). *Napari: A multi-dimensional image viewer for python* (Version v0.4.18). Zenodo. <https://doi.org/10.5281/zenodo.8115575>
- Angelo, M., Bendall, S. C., Finck, R., Hale, M. B., Hitzman, C., Borowsky, A. D., Levenson, R. M., Lowe, J. B., Liu, S. D., Zhao, S., Natkunam, Y., & Nolan, G. P. (2014). Multiplexed ion beam imaging of human breast tumors. *Nature Medicine*, 20(4), 436–442. <https://doi.org/10.1038/nm.3488>
- Baker, G. J., Novikov, E., Zhao, Z., Vallius, T., Davis, J. A., Lin, J.-R., Muhlich, J. L., Mittendorf, E. A., Santagata, S., Guerriero, J. L., & Sorger, P. K. (2024). *Quality control for single cell analysis of high-plex tissue profiles using CyLinter*. bioRxiv. <https://doi.org/10.1101/2023.11.01.565120>
- Catching up with multiplexed tissue imaging. (2022). *Nature Methods*, 19(3), 259–259. <https://doi.org/10.1038/s41592-022-01428-z>
- Chiu, C.-L., Clack, N., & the napari community. (2022). Napari: A python multi-dimensional image viewer platform for the research community. *Microscopy and Microanalysis*, 28, 1576–1577. <https://doi.org/10.1017/S1431927622006328>
- Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., George, R. E., Pierson, N., Cai, L., & Yuan, G.-C. (2021). Giotto: A toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology*, 22(1), 78. <https://doi.org/10.1186/s13059-021-02286-2>
- Gaglia, G., Burger, M. L., Ritch, C. C., Rammos, D., Dai, Y., Crossland, G. E., Tavana, S. Z., Warchol, S., Jaeger, A. M., Naranjo, S., Coy, S., Nirmal, A. J., Krueger, R., Lin, J.-R., Pfister, H., Sorger, P. K., Jacks, T., & Santagata, S. (2023). Lymphocyte networks are dynamic cellular communities in the immunoregulatory landscape of lung adenocarcinoma. *Cancer Cell*, 41(5), 871–886.e10. <https://doi.org/10.1016/j.ccell.2023.03.015>
- Gerdes, M. J., Sevinsky, C. J., Sood, A., Adak, S., Bello, M. O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R. J., Hollman, D., Kamath, V., Kaanumalle, S., Kenny, K., Larsen, M., Lazare, M., Li, Q., Lowes, C., McCulloch, C. C., ... Ginty, F. (2013). Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences*, 110(29), 11982–11987. <https://doi.org/10.1073/pnas.1300136110>
- Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolmund, D., Buhmann, J. M., Brandt, S., Varga, Z., Wild, P. J., Günther, D., & Bodenmiller, B. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods*, 11(4), 417–422. <https://doi.org/10.1038/nmeth.2869>
- Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., & Nolan, G. P. (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell*, 174(4), 968–981.e15. <https://doi.org/10.1016/j.cell.2018.07.010>
- Gut, G., Herrmann, M. D., & Pelkmans, L. (2018). Multiplexed protein maps link subcellular organization to cellular states. *Science*, 361(6401), eaar7042. <https://doi.org/10.1126/science.aar7042>
- Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, 42(2), 293–304. <https://doi.org/10.1038/s41587-023-01767-y>
- Lin, J.-R., Izar, B., Wang, S., Yapp, C., Mei, S., Shah, P. M., Santagata, S., & Sorger, P. K. (2018). Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife*, 7, e31657. <https://doi.org/10.7554/eLife.31657>

- Liu, B., Li, Y., & Zhang, L. (2022). Analysis and visualization of spatial transcriptomic data. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.785290>
- Maliga, Z., Kim, D. Y., Bui, A.-T. N., Lin, J.-R., Dewan, A. K., Jadeja, S., Murphy, G. F., Nirmal, A. J., Lian, C. G., Sorger, P. K., & LeBoeuf, N. R. (2024). Immune profiling of dermatologic adverse events from checkpoint blockade using tissue cyclic immunofluorescence: A pilot study. *The Journal of Investigative Dermatology*, S0022-202X(24)00107-6. <https://doi.org/10.1016/j.jid.2024.01.024>
- Nirmal, A. J., Maliga, Z., Vallius, T., Quattrochi, B., Chen, A. A., Jacobson, C. A., Pelletier, R. J., Yapp, C., Arias-Camison, R., Chen, Y.-A., Lian, C. G., Murphy, G. F., Santagata, S., & Sorger, P. K. (2022). The spatial landscape of progression and immunoediting in primary melanoma at single-cell resolution. *Cancer Discovery*, 12(6), 1518–1541. <https://doi.org/10.1158/2159-8290.CD-21-1357>
- Nirmal, A. J., Yapp, C., Santagata, S., & Sorger, P. K. (2023). *Cell spotter (CSPOT): A machine-learning approach to automated cell spotting and quantification of highly multiplexed tissue images*. bioRxiv. <https://doi.org/10.1101/2023.11.15.567196>
- Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., Rybakov, S., Ibarra, I. L., Holmberg, O., Virshup, I., Lotfollahi, M., Richter, S., & Theis, F. J. (2022). Squidpy: A scalable framework for spatial omics analysis. *Nature Methods*, 19(2), 171–178. <https://doi.org/10.1038/s41592-021-01358-2>
- Schapiro, D., Sokolov, A., Yapp, C., Chen, Y.-A., Muhlich, J. L., Hess, J., Creason, A. L., Nirmal, A. J., Baker, G. J., Nariya, M. K., Lin, J.-R., Maliga, Z., Jacobson, C. A., Hodgman, M. W., Ruokonen, J., Farhi, S. L., Abbondanza, D., McKinley, E. T., Persson, D., ... Sorger, P. K. (2022). MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nature Methods*, 19(3), 311–315. <https://doi.org/10.1038/s41592-021-01308-y>
- Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, 18(1), 100–106. <https://doi.org/10.1038/s41592-020-01018-x>
- Tsujikawa, T., Kumar, S., Borkar, R. N., Azimi, V., Thibault, G., Chang, Y. H., Balter, A., Kawashima, R., Choe, G., Sauer, D., El Rassi, E., Clayburgh, D. R., Kulesz-Martin, M. F., Lutz, E. R., Zheng, L., Jaffee, E. M., Leyshock, P., Margolin, A. A., Mori, M., ... Coussens, L. M. (2017). Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis. *Cell Reports*, 19(1), 203–217. <https://doi.org/10.1016/j.celrep.2017.03.037>
- Wang, X., & Grimson, E. (2007). Spatial Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, 20. https://papers.nips.cc/paper_files/paper/2007/hash/ec8956637a99787bd197eacd77acce5e-Abstract.html
- Yapp, C., Nirmal, A. J., Zhou, F. Y., Maliga, Z., Tefft, J., Llopis, P. M., Murphy, G. F., Lian, C., Danuser, G., Santagata, S., & Sorger, P. K. (2024). *Multiplexed 3D analysis of immune states and niches in human tissue*. bioRxiv. <https://doi.org/10.1101/2023.11.10.566670>