Spatial-omics analysis

https://doi.org/10.1038/s41588-023-01646-x

Cellular neighborhood analysis in spatial omics reveals new tissue domains and cell subtypes

Marc J. Ruitenberg & Quan H. Nguyen

Spatial omics enables the molecular profiling of cells with the tissue context preserved. A new analytic approach shows how cellular neighborhood analysis and feature augmentation can spatially connect and cluster millions of cells into higher-order functional units.

Living tissues are made up of diverse communities of cells that coexist in complex spatial arrangements. Groups of tissue-specialized cells work together as higher-order functional units to support organ homeostasis and function. Analyzing spatial context is therefore key to understanding tissue biology. Advances in spatial-omics technologies have given us the tools to profile the transcriptomes (and/or proteomes and/or metabolomes) of thousands to millions of cells in a spatially resolved manner—that is, at their native location within a given tissue or organ[1]. One of the biggest challenges, and opportunities, in the field has been to effectively make use of the spatial information, discovering new biology beyond what may already be discernible from non-spatial methods[2]. The ability to decode tissue domains and/or find regional specializations of cells (that is, areas with distinct cellular compositions or cell states) is an important first step toward resolving this challenge; cell and neighborhood features are captured in spatial-omics data, but current methods to recognize and cluster cells (or cell subsets) accordingly have been lacking in accuracy and sensitivity[3]. In this issue, Singhal et al.[4] present BANKSY, a new spatial clustering algorithm that implements a creative feature-augmentation approach to better map domains by integrating the transcriptional profiles of individual cells with their physical distances and tissue neighborhood context.

The fundamental biological assumption of BANKSY feature augmentation is that the state of a cell is best captured when its own transcriptome is considered in conjunction with that of the surrounding microenvironment. Algorithmically, BANKSY uses two spatial augmentations (kernels) to do so, one of which computes a mean of the gene expression that is weighted by the physical distance between nearest neighboring cells (a Gaussian weighting envelope) and the other using a so-called azimuthal Gabor filter (AGF) (Fig. 1). Gabor filters[5] are widely used for feature detection in image processing, and they are cleverly applied here to estimate the 'spatial texture' (or gradient) of gene expression within the cellular neighborhood (referred to as 'azimuthal', from azimuth—the horizontal angle measured from a given reference direction). By concatenating the raw, non-spatial gene-expression matrix with the average neighborhood expression and AGF matrices, BANKSY constructs an augmented matrix in which the relative contribution of a cell's own transcriptome and that of its neighborhood is tunable to prioritize cell typing or domain mapping. BANKSY achieves either task through controlling the mixing parameter lambda and does not assume (or require) neighboring cells to have similar expression. Cells that are transcriptionally very similar but physically distant will still cluster together if more weighting is placed on the cell's own features (low lambda values). Conversely, when the local microenvironment is emphasized (high lambda values), cells will segregate (or cluster) by tissue domains. The authors demonstrate the potential of both options, comprehensively assessing the performance of BANKSY against existing methods. They demonstrate that BANKSY correctly clusters various neural cell types across the mouse cerebellum in an unsupervised manner and without the need for a reference. BANKSY performed as well as the top deconvolution approach, robust cell-type decomposition or RCTD, which still requires a single-cell RNA-seq reference dataset for annotations. These findings suggest broad applicability for BANKSY in scenarios where no good reference information is available. The authors further show that BANKSY can distinguish oligodendrocytes in gray versus white matter structures of the mouse brain. Functional diversity of cell subsets is a well-recognized concept in immunology, but is less established for glial cells in the central nervous system. BANKSY's ability to detect cellular subtypes unlocks avenues for exploring regional specializations, and for probing whether observed transcriptional differences truly amount to functional states that are biologically significant. The authors went on to show that BANKSY also works well on a diseased tissue, colorectal cancer. Here, it was able to resolve intermixed cell types and identify a population of cycling epithelial cells that no other method could. Overall, BANKSY elegantly utilizes spatial information to address two important analysis questions in one algorithm: cell type identification and domain clustering.

Computationally, BANKSY is fast and easily scales to millions of cells because of the quick computation and concatenation of its neighbor-augmented matrices. This provides a significant advantage over other tools, especially those that require parameter optimization[6–9]. BANKSY's scalability is important as the resolution of spatial-omics platforms is rapidly moving from the multi- to the single-cell level, which exponentially increases data size (for example, from around 5,000 spots with Visium technology to millions of cells and/or bins for other platforms such as STomics, Visium HD and PhusionCycler). The authors show that BANKSY only takes around 1 hour to process a dataset of over 2 million cells, which is 10–1,000 times faster than other methods. They further demonstrate that the software is relatively robust to parameter selection, such as the mixing parameter lambda and the number of nearest neighbors used for augmentation. This should make BANKSY results reproducible, although further work is needed to determine whether these parameters might be more variable for different tissue types and/or cell distribution densities; a parameter sensitivity analysis may be needed in such instances. Another attractive feature of BANKSY is that it is implemented in both R and Python, and is compatible with many popular

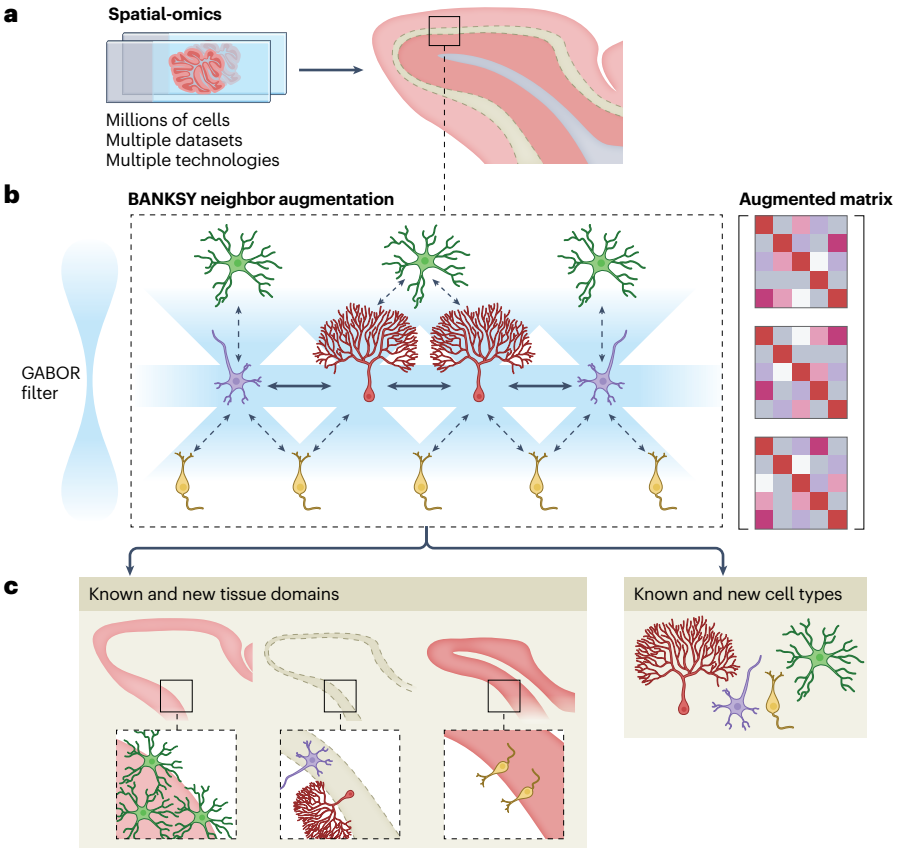

Fig. 1 | BANKSY neighbor augmentation. a, BANKSY is applicable to data from different spatial-omics platform, and is scalable for analyzing millions of cells and for integrating multiple datasets. **b**, Gene feature augmentation is performed for each cell, based on neighbor (spatially nearest) cells, computed by GABOR filter and weighted averaging. **c**, BANKSY's augmentation approach allows for more accurate and sensitive identification of tissue domains and cell types. Figure created with BioRender.com.

software platforms for analyzing single-cell data such as Seurat, Scanpy and SingleCellExperiment.

The overall advances presented by Singhal et al.[4] are significant, but several limitations and/or questions remain. First, the neighbor-augmented matrix in BANKSY improves performance and provides an effective way to integrate different spatial datasets. However, the distance measure that is used for feature augmentation does not yet consider three-dimensional cytoarchitecture and/or cellular relationships within tissues. Better recognition of these is particularly important for organs such as the brain where an interconnectedness between distant structures (or tissue domains) via neural pathways adds a whole other layer of complexity. It would be of interest in that regard to evaluate whether BANKSY can temporally detect (or predict) distal neurodegeneration and/or gliosis following brain injury, clustering regions that are synaptically connected. Other unique tissue features may also require further consideration. For example, the local brain microenvironment will contain RNAs from distant cellular sources, delivered through axons, dendrites and/or synapses. Whether these can analytically and/or biologically influence the assignment of perceived niche-dependent states remains an open question. Furthermore, the BANKSY neighbor-augmented matrix does not make use of tissue morphological images, an important part of spatial data that has been shown to improve neighborhood analysis[10,11]. BANKSY has otherwise mostly been tested on brain tissues so far, with some extension to other data types such as the healthy human colon and intestine, and colorectal cancer; all have rather distinct spatial organizations, and additional testing of BANKSY on more heterogeneous and/or diseased tissues by the user community will be needed to prove its applicability more broadly.

In conclusion, Singhal et al.[4] provide us with a tool to simultaneously and more precisely map out cell types and tissue domains within spatial-omics data. With that, the field can now move on to probing questions regarding what defines a domain in terms of size and cellular composition, and also how spatial domains add biological information to higher-order functional units through interactions between the cells that define them.

Marc J. Ruitenberg [1] **& Quan H. Nguyen** [2,3] ✉
[1]School of Biomedical Science, Faculty of Medicine, The University of Queensland, Brisbane, Australia. [2]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. [3]QIMR Berghofer Medical Research Institute, Brisbane, Australia.
✉e-mail: quan.nguyen@qimrberghofer.edu.au

Published online: 27 February 2024

References

1. Lewis, S. M. et al. *Nat. Methods.* **18**, 997–1012 (2021).
2. Atta, L. & Fan, J. *Nat. Commun.* **12**, 5283 (2021).
3. Charitakis, M. et al. *Genome Biol.* **24**, 209 (2023).
4. Singhal, V. et al. *Nat. Genet.* https://doi.org/10.1038/s41588-024-01664-3 (2024).
5. Gabor, D. *J. Inst. Elec. Eng.* **93**, 429–457 (1946).
6. Long, Y. et al. *Nat. Commun.* **14**, 1155 (2023).
7. Hu, J. et al. *Nat. Methods* **18**, 1342–1351 (2021).
8. Zhao, E. et al. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
9. Chidester, B. et al. *Nat. Genet.* **55**, 78–88 (2023).
10. Zubair, A. et al. *Nucleic Acids Res.* **50**, e80 (2022).
11. Pham, D. et al. *Nat. Commun.* **14**, 7739 (2023).

Acknowledgements

M.J.R. is supported by a Wings for Life Project Grant and a National Health and Medical Research Council (NHMRC) Project Grant (1163835). Q.N. is supported by an Australian Research Council DECRA Award (DE190100116), NHMRC Investigator Grant (GNT2008928) and NHMRC Project Grant (2001514).

Competing interests

The authors declare no competing interests.