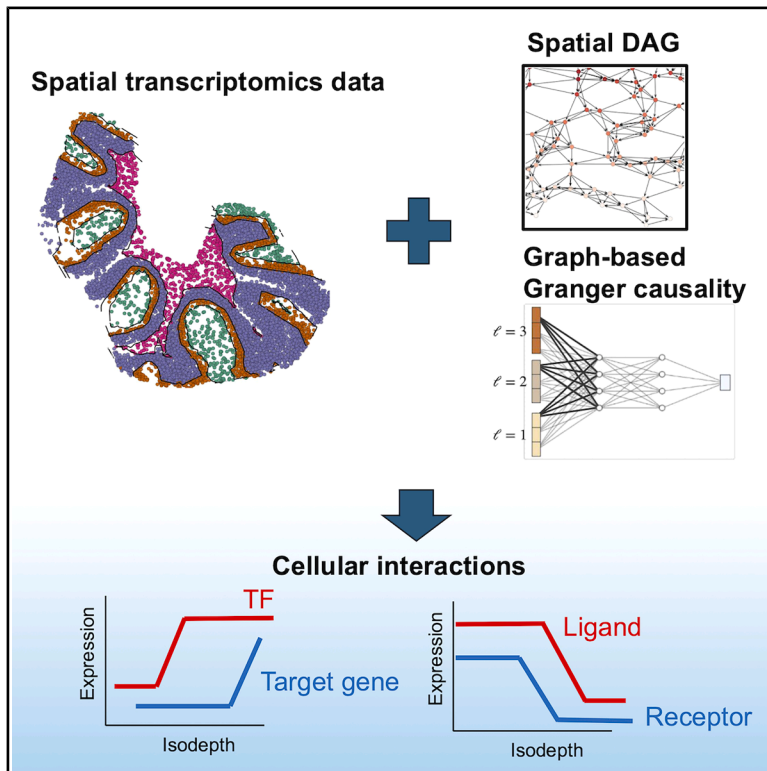


Decoding the causal drivers of spatial cellular topology

Graphical abstract



Authors

Prannav Shankar, Huan Liang,
Uthsav Chitra, Rohit Singh

Correspondence

uthsav@jhu.edu (U.C.),
rohit.singh@duke.edu (R.S.)

In brief

Integrative aspects of cell biology;
Biocomputational method; Systems
biology; Transcriptomics

Highlights

- GLACIER introduces spatial Granger causality to infer spatial regulatory programs
- Combines the GASTON isodepth coordinate with single-cell DAG-based causal inference
- Uncovers within-tissue signaling and distinct regulation at tissue boundaries
- Reveals bi- and uni-directional ligand-receptor and TF-target communication



Article

Decoding the causal drivers of spatial cellular topology

Prannav Shankar,¹ Huan Liang,¹ Uthsav Chitra,^{2,3,5,*} and Rohit Singh^{1,4,5,6,*}¹Department of Bioinformatics and Biostatistics, Duke University, Durham, NC 27705, USA²Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA³Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA⁴Department of Cell Biology, Duke University, Durham, NC 27705, USA⁵Senior author⁶Lead contact*Correspondence: uthsav@jhu.edu (U.C.), rohit.singh@duke.edu (R.S.)<https://doi.org/10.1016/j.isci.2026.115794>

SUMMARY

Decoding how cells influence and communicate with each other in space is fundamental for understanding tissue organization. However, existing approaches either overlook spatial context entirely or rely solely on local cell-cell adjacency, failing to capture how global tissue topology shapes cellular communication. Here, we present GLACIER, which introduces *spatial Granger causality* to infer transcriptional and signaling relationships that emerge from tissue organization. By combining GASTON's global isodepth coordinate with Velorama's graph-based causal inference framework, we enable bidirectional inference of regulatory relationships along spatial axes, identifying transcription factor-target interactions and ligand-receptor pairs that operate across spatial domains. Applying GLACIER to spatial transcriptomics data from the mouse cerebellum, we identify both continuous within-cell-type regulatory gradients and discontinuous drivers at layer interfaces, while distinguishing between forward and backward cellular communication along the isodepth axis. Our approach reveals how tissue architecture directs patterns of cellular communication, providing a framework for understanding spatially encoded regulatory programs.

INTRODUCTION

Multicellular organisms depend on precisely orchestrated cell-cell interactions that unfold across multiple spatial scales, including both direct, short-range interactions and indirect, longer-range communication. For example, cells directly communicate with one another by releasing ligand molecules that bind to receptors on neighboring cells, triggering immediate signaling changes. More broadly, transcription factors (TFs) can initiate regulatory programs that cascade across cellular neighborhoods and lead to coordinated changes in gene expression programs in larger domains of a tissue.¹ Understanding these spatially encoded cellular interactions is crucial for decoding many fundamental biological processes, including tissue development,² immune response,³ wound healing,⁴ and cancer progression.⁵

Recent advances in spatial transcriptomics (ST) offer unprecedented potential for characterizing cellular interactions in tissues. Unlike traditional single-cell RNA sequencing (scRNA-seq), which dissociates cells before measurement and loses spatial information, ST technologies measure both gene expression and the spatial location of individual cells in tissue slices. However, there are two key challenges in the inference of cellular interactions from current ST datasets. First, transcript counts in ST data are highly sparse, making it difficult to reliably estimate

short-range, *local* correlations in gene expression.^{6,7} For example, whole-transcriptome technologies such as Slide-SeqV2⁸ and Stereo-seq⁹ measure a median of less than 500 unique molecular identifiers (UMIs) per cell. Second, and more fundamentally, understanding how cells communicate over long scales requires knowledge of the *global* geometry of the tissue and the spatial axis along which regulatory signals propagate.

Current computational methods for inferring cellular or regulatory interactions from ST data do not leverage global tissue organization. Nearly all methods implicitly assume a uniform interaction likelihood between neighboring cells by representing these local relationships through cell-cell adjacency graphs, and they look for correlated TF-gene or ligand-receptor expression along edges of such an adjacency graph.^{6,7,10–13} However, this oversimplified view fails to capture how intercellular signals propagate in specific directions determined by tissue architecture. For instance, endothelial cells communicate along paths defined by tissue cytoskeleton,¹⁴ while the Notch signaling pathway operates bidirectionally along particular spatial axes during development.¹⁵ Accurate inference of cellular interactions thus requires incorporating global tissue structure to understand the directionality of gene expression variation.

GASTON¹⁶ addresses this need by learning a “topographic map” of tissue from ST data. This map is defined by a spatial



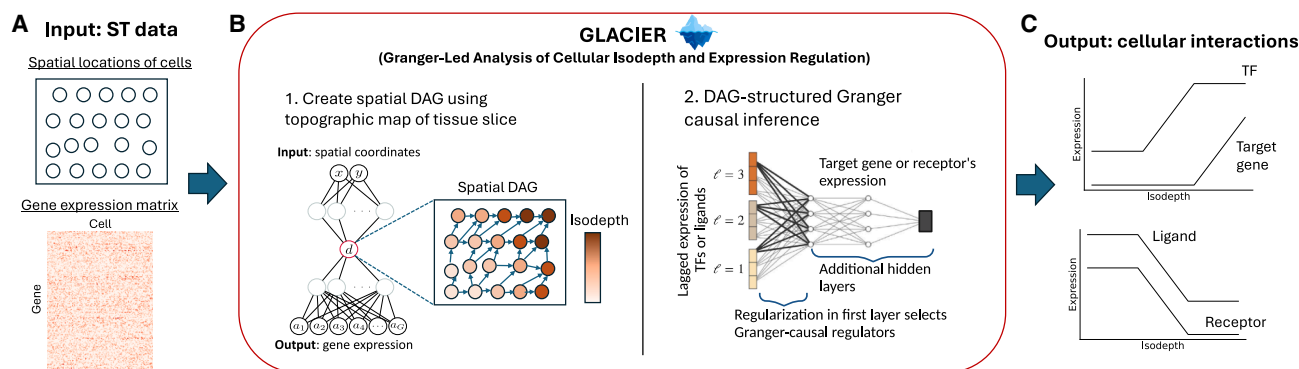


Figure 1. GLACIER uses spatial Granger causality to infer transcription factor-target and ligand-receptor interactions.

(A) The input to GLACIER is spatial transcriptomics (ST) data from a 2-D tissue slice, consisting of the spatial locations of measured cells and the gene expression matrix.

(B) Granger-led analysis of cellular isodepth and expression regulation (GLACIER) first learns a topographic map of the tissue slice defined by an isodepth coordinate, and uses the topographic map to form a spatial DAG. GLACIER then performs directed acyclic graph (DAG)-structured Granger causal inference.

(C) The output of GLACIER is cellular interactions: transcription factors (TFs) that are Granger causal for target genes, and ligand genes that are Granger causal for receptor genes.

coordinate called “isodepth,” which varies globally across the tissue slice, and measures the direction of maximum variation in gene expression. It allows inference of directional cellular interactions by structuring the tissue into a hierarchical framework. This approach helps distinguish between local interactions and broader regulatory signals dictated by global tissue topology. In the mouse cerebellum, GASTON successfully revealed both discrete cellular domains (layers) and continuous gradients of gene activity, providing the first comprehensive map of spatial gene expression in this tissue. However, GASTON models each the variation of gene independently across the isodepth coordinate and does not account for gene-gene interactions, leaving open the challenge of inferring cellular communication and regulatory relationships from ST data.

We introduce *spatial Granger causality*, an approach for learning cellular interactions from ST data. The key insight is that tissue organization, as captured by GASTON’s isodepth, creates natural directions of information flow that can be represented as a directed acyclic graph (DAG). While classical Granger causality required data to follow a linear time series, recent advances^{17–19} have enabled Granger causal inference on partial orderings represented by DAGs. Building on this foundation, we present Granger-led analysis of cellular isodepth and expression regulation (GLACIER), which combines GASTON’s global spatial coordinate with Velorama’s DAG-based nonlinear Granger causality to identify TF-gene and ligand-receptor relationships that propagate along spatial axes. GLACIER systematically captures how regulatory information flows through tissue structure, enabling directional inference of cellular signaling interactions. Applied to Slide-SeqV2 data of the mouse cerebellum,^{8,20} GLACIER identifies directional transcriptional programs within layers, revealing that oligodendrocytes predominantly regulate neuronal connectivity in one direction. At the boundary between the granular and Purkinje-Bergmann (PB) layers, GLACIER reports regulatory activity that is distinct from what is seen in either layer individually, suggesting that inter-layer communication operates through specialized mecha-

nisms rather than a simple gradient. These findings demonstrate how integrating spatial organization with causal inference can reveal fundamental principles of tissue architecture and cellular communication.

RESULTS

We used GLACIER (Figure 1) to identify cellular interactions in the mouse cerebellum using ST data measured using the Slide-SeqV2 technology.^{8,20} The expression of 23, 096 transcripts was measured at 9, 985 spatial locations. The mouse cerebellum has a layered geometry and consists of four distinct layers—the oligodendrocyte, granular, PB, and molecular layers—as identified in previous studies^{16,20} (Figure 2A). To investigate TF-target and ligand-receptor gene interactions, we selected 39 TFs and 773 target genes, 23 ligand genes, and 26 receptor genes, respectively, from the GASTON spatially variable genes.¹⁶

We used GLACIER to focus on two biological questions, which were previously out of reach for existing methods studying cellular interactions.

1. What is the *directionality* of cellular interactions—and more generally, information flow—within a layer?
2. Is there a *discontinuity* in information flow across layers?

To answer the first question, we used GLACIER to identify cellular interactions *within* each cerebellum layer (e.g., granule layer, Figure 2B), which involves forming a spatial DAG (Figure 2C) oriented according to the GASTON isodepth d , which smoothly varies within each layer. Importantly, by *inverting* the isodepth (methods, Figure 2D), GLACIER identifies cellular interactions in the opposite direction and thus determines the *directionality* of cellular interactions. For the second question, we used GLACIER to identify cellular interactions that are unique to cells at the *boundary* of a cerebellum layer (Figure 2E).

To address these questions, we analyzed the data at multiple levels. We identified frequently-regulated target genes and

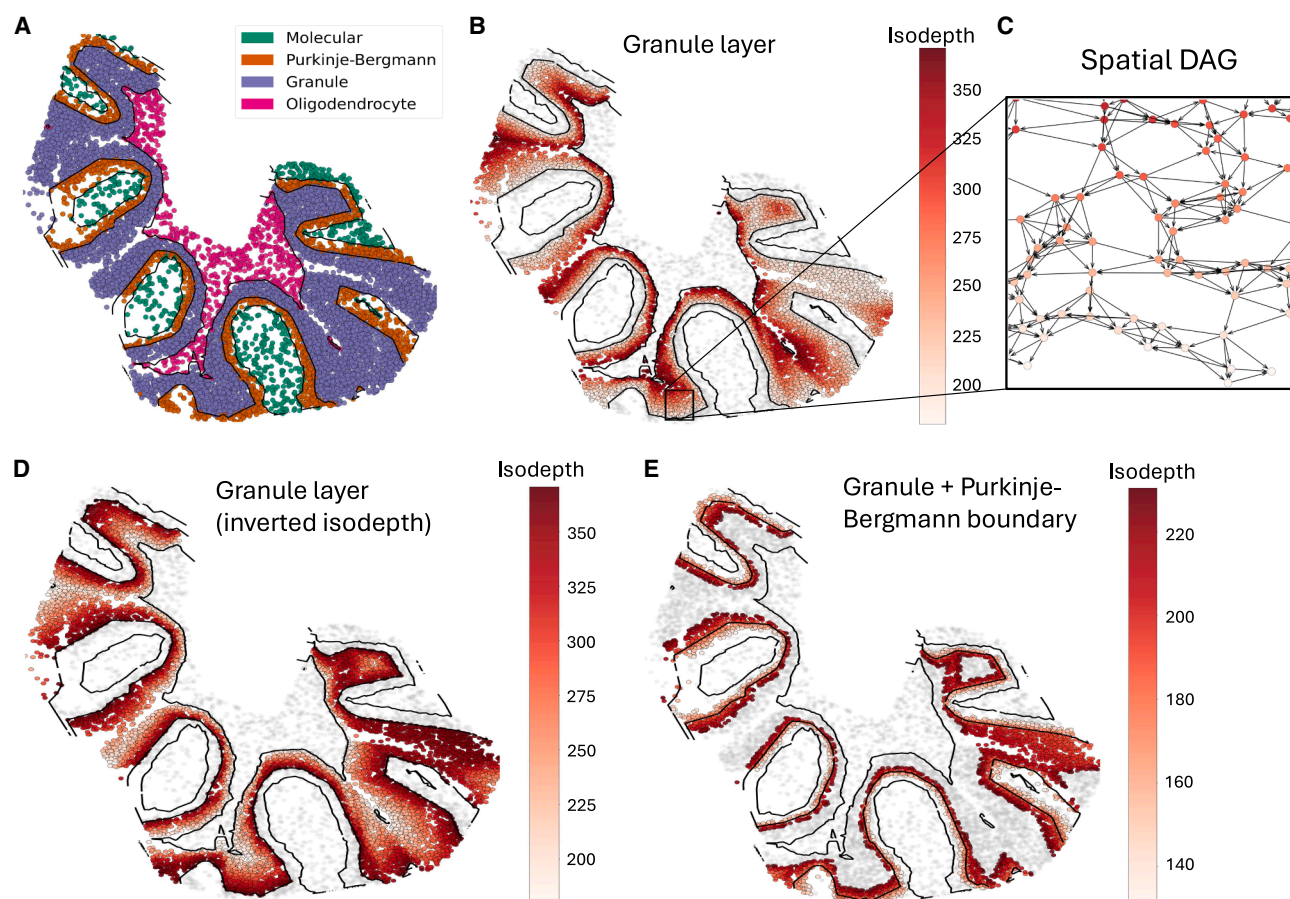


Figure 2. Isodepth and spatial DAG for different layers of the mouse cerebellum

(A) Slide-SeqV2 mouse cerebellum dataset with each cell colored by its layer assigned by GASTON.¹⁶ Each layer is named according to its dominant cell type as in refs.^{20,21}

(B) Cells in the granular layer colored by GASTON isodepth. Black curves denote GASTON-identified layer boundaries.

(C) Spatial DAG $G = (V, E)$ constructed by GLACIER in a granular layer, with directed edges (i, j) oriented from high isodepth to low isodepth, i.e., edge $i \rightarrow j$ if $d_i > d_j$.

(D) Inverted isodepth d_i' (Equation 7) in a granular layer, which allows for the identification of *bidirectional* cellular interactions by GLACIER.

(E) Cells with isodepth d_i within 50 μm of the layer boundary between the granular layer and the Purkinje-Bergmann layer, colored by isodepth d_i , allowing for GLACIER to identify cellular interactions *between* cell types.

performed gene set enrichment analysis (GSEA) to determine which biological processes were enriched among these targets.²² We further examined specific TF-target and ligand-receptor gene pairs to explore key regulatory interactions at fine resolution. Lastly, we also exploited GLACIER-derived lag estimates to identify ligands with short- and long-ranging activity at layer boundaries.

Gene sets

Direction of information flow in oligodendrocytes: We examined gene set enrichments for oligodendrocytes in both orientations of isodepth. We observed that increasing isodepth, i.e., where the DAG leads away from the granular layer, exhibited a much more coherent and interpretable set of enriched biological processes. This direction was strongly associated with axonogenesis, synaptic transmission, and neuronal connectivity, suggesting that oligodendrocytes in this orientation actively shape neuronal function rather than simply responding to

their environment (Figures 3A and 3B). Function enrichments included axon development, regulation of neurotransmitter secretion, and neuron projection morphogenesis, all of which align with known roles of oligodendrocytes in supporting and refining neural circuits.²³

At the gene level, this enrichment was driven by regulators of axon guidance and synaptic adhesion, such as neurexin-1 (*Nrxn1*) and neurofascin (*Nfasc*), which mediate neuron-glia communication, and *Mag*, which plays a key role in stabilizing axonal connections. In contrast, the decreasing isodepth (inverted) direction showed weaker and less functionally coherent enrichments, with a shift toward intracellular and cytoskeletal remodeling rather than neuron-facing processes. Genes such as *Cfil1* and *Phip*, which regulate actin dynamics²⁴ and intracellular signaling, were more prominent in this direction, suggesting a greater emphasis on structural maintenance or adaptation.

These findings raise the possibility that oligodendrocytes' instructive role proceeds in the direction of isodepth, with less

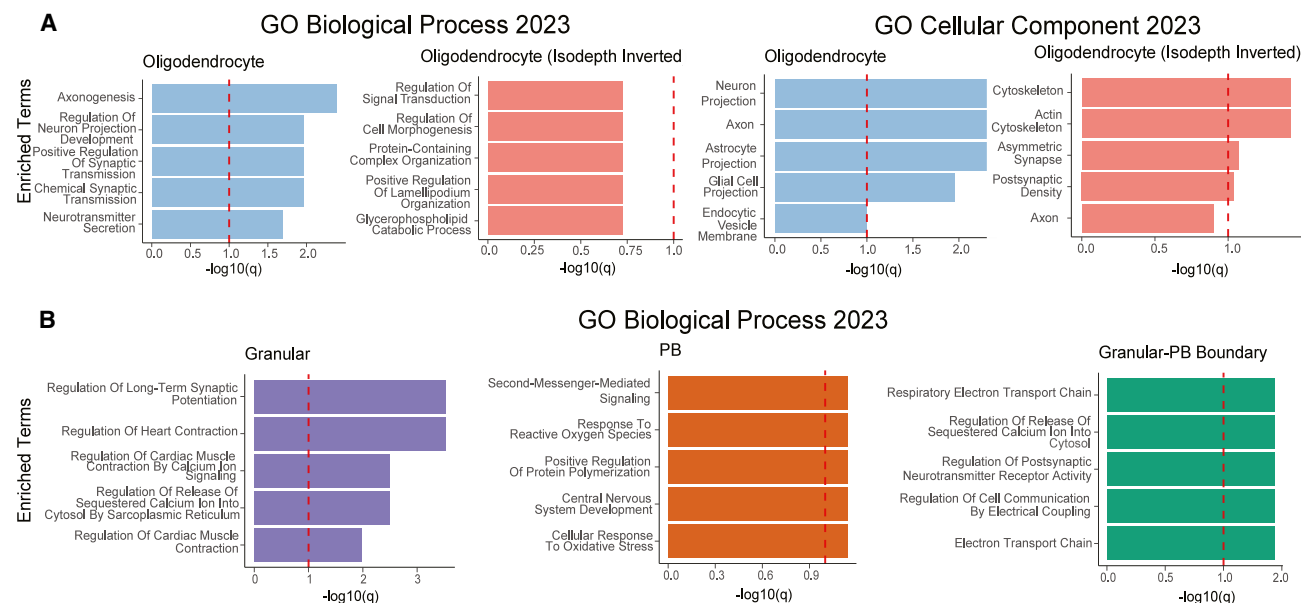


Figure 3. GLACIER reveals biological processes corresponding to global cellular interactions in the different layers of the mouse cerebellum
(A) Gene set enrichment analysis (GSEA) with Enrichr²² for cellular interactions identified by GLACIER in the oligodendrocyte layer with isodepth d and inverted isodepth d' isodepth orientations for biological process (left) and cellular component (right) gene sets. X-axis indicates q-values from GSEA.
(B) GSEA for cellular interactions identified by GLACIER in the granular layer (left), Purkinje-Bergmann (PB) layer (middle), and the boundary between the granular and PB layers (right). More information about Enrichr can be found in [STAR Methods](#).

regulatory control exerted in the other direction. Further experimental work would be needed to investigate this hypothesis. Indeed, the value of GLACIER is precisely in helping generate such hypotheses for further investigation.

Regulatory discontinuity at the boundary of granular and PB layers: Gene set enrichment at this boundary revealed a distinct functional signature that differs from either layer alone. While the granular layer was enriched for processes related to long-term synaptic potentiation and calcium-mediated excitability (Figure 3C, left), and the PB layer showed enrichments for oxidative stress response and second-messenger signaling (Figure 3C, middle), the boundary exhibited a distinct focus on synaptic modulation, electrical coupling, and metabolic activity (Figure 3C, right). This suggests that cells at the interface of the granular and PB layers play a specialized role in mediating communication and bioenergetic demands between the two layers.

One of the strongest enrichments at the boundary included the regulation of postsynaptic neurotransmitter receptor activity and cell communication by electrical coupling, suggesting rapid synaptic transmission. We also observed the enrichment of electron transport chain activity, possibly due to metabolic interactions. The gene set included *Dlgap1* and *Nptx1*, both of which are involved in synaptic scaffolding and activity-dependent synapse formation, reinforcing the idea that the boundary plays a role in synaptic refinement and communication.²⁵ In addition, *Ndufa5* and *Sdha*, key components of the mitochondrial respiratory chain,²⁶ indicate that cells in this region experience high metabolic demand and are likely to support sustained neurotransmission.

These results suggest that the PB and granular cell boundary is not simply a gradient between two transcriptional programs, but rather a specialized interface with its own regulatory demands. Experimental validation could determine whether these transcriptional differences are critical to sustaining high-frequency signaling and synaptic communication in the cerebellum.

Gene pairs

GLACIER reveals several biologically meaningful pairs of genes (TF-target or ligand-receptor) with spatial Granger causal relationships within individual layers. For example, GLACIER identifies that the TF *Hspa5* is spatially Granger causal for the target gene *Pink1* (Figure 4A), with *Pink1* expression closely following *Hspa5* expression. *Hspa5* maintains endoplasmic reticulum (ER) homeostasis²⁷ while *Pink1* performs mitochondrial quality control,²⁸ suggesting that the activation of general ER regulation results in recruitment of proteins that regulate specific cellular functions. GLACIER also identifies that the ligand *S100b* is spatially Granger causal of the receptor *Ptprz1* (Figure 4B). *S100b* has been suggested to support neuron migration²⁹ while *Ptprz1* has been observed to modulate cell migration in some tissues,³⁰ suggesting that neurons in the granular layer may be moving in the isodepth direction toward the adjacent oligodendrocyte layer.

Further, GLACIER identifies spatially Granger causal pairs of genes that are unique to the boundary between two layers. For example, GLACIER identifies an interaction between the TF *Chd2* and target gene *Calb1* at the boundary between the molecular and PB layers (Figure 4C). *Chd2* has been observed to broadly regulate neuronal function³¹ while *Calb1* is a marker gene for Purkinje cells,³² suggesting that *Chd2* may be regulating

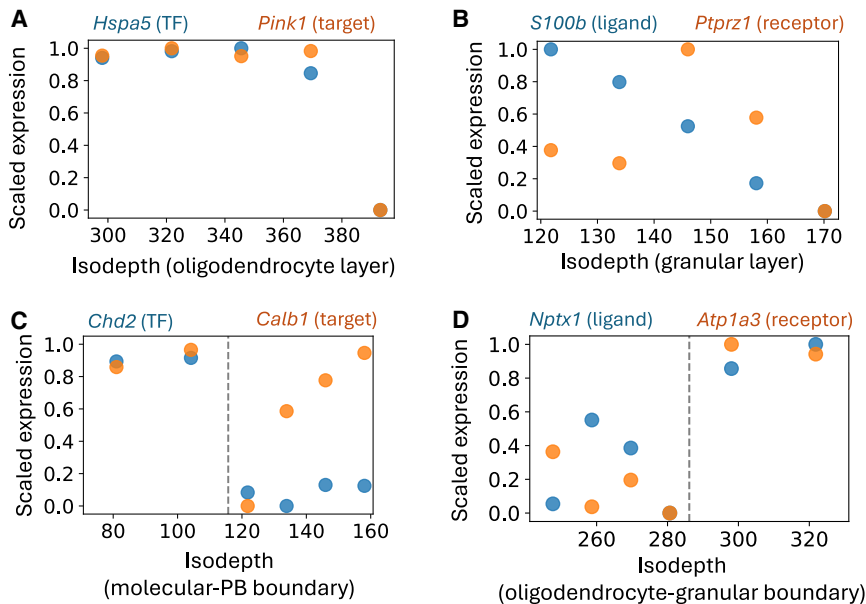


Figure 4. GLACIER learns transcription factor-target and ligand-receptor interactions within tissue layers and across layer boundaries

Isodepth versus scaled expression for (A) transcription factor *Hspa5* and target gene *Pink1* in the oligodendrocyte layer, (B) ligand *S100b* and target gene *Ptptr1* in the granular layer, (C) transcription factor *Chd2* and target gene *Calb1* at the boundary between the molecular and Purkinje-Bergmann (PB) layers, and (D) Ligand *Nptx1* and receptor *Atp1a3* at the boundary between the oligodendrocyte and granular layers.

the function of Purkinje cells specifically at the molecular-PB layer boundary. GLACIER also identifies an interaction between the ligand *Nptx1* and the receptor *Atp1a3* that uniquely occurs at the boundary of the oligodendrocyte and granular layers; both *Nptx1* and *Atp1a3* are known to be involved in neuronal signaling and excitation,^{33,34} suggesting potential communication between oligodendrocytes and granular cells at the boundary.

In this way, GLACIER identifies cellular interactions within and across layer boundaries that cannot be found by existing approaches, which do not incorporate the global tissue topography.

Spatial range estimation via GLACIER lag estimates

GLACIER's causal inference framework reports not only the regulatory interactions but also the lag at which the interaction operates (methods). This allows us to estimate the spatial range of a regulator gene (i.e., a TF or ligand), with bigger lags corresponding to a larger spatial range of a regulator. We investigated some settings where GLACIER reported bimodal lags across TF or ligand regulators (Supplemental Information), denoting substantial variations in the spatial ranges of regulators.

Molecular Layer/PB Boundary: The boundary between the molecular and PB layers displayed particularly pronounced bimodality in the ligand distance distribution. The longest-ranging ligand at this interface was *Apoe*, a gene well-known for its role in lipid transport, synaptic plasticity, and neuronal maintenance³⁵ *Apoe*'s ability to bind lipoproteins and facilitate cholesterol uptake is crucial for neuronal repair and may underlie its long-range effects across these layers. Conversely, the shortest-acting ligand we identified was *Gdf10*, a member of the TGF- β superfamily involved in neuronal regeneration and repair.³⁶ *Gdf10*'s localized expression and role in modulating cell growth suggest its influence is confined to more proximal targets in neighboring cells.

PB/Granular Layer Boundary: We also observed substantial spatial-range variation between ligands at the PBGranular layer

boundary. Here, the furthest-acting ligand was *Gng13*, a G-protein gamma subunit that partners with alpha and beta subunits to regulate a variety of downstream signaling pathways.³⁷ Its presence in long-range signaling contexts suggests that *Gng13* may orchestrate interactions spanning multiple cell types or spatial zones in the cerebellar architecture. While there was no single dominant short-range ligand at this boundary, the *Thy1* receptor was repeatedly seen. *Thy1* (also known as *Cd90*) is a cell-surface glycoprotein involved in cell adhesion, neurite outgrowth, and T cell activation.³⁸ Its role in modulating cell-cell contact may explain its consistent appearance among ligands operating over limited spatial distances, hinting that *Thy1*-dependent interactions mediate localized communication in these cerebellar layers.

Taken together, these findings underscore how spatial organization can drive bimodal patterns of ligand-receptor signaling: Some factors, such as *Apoe* and *Gng13*, can reach distant targets, while others, such as *Gdf10*, *Gnas*, and *Thy1*, function in tight, localized niches. By systematically exploring these "lag distributions," GLACIER reveals how cerebellar architecture shapes the breadth and specificity of cellular communication.

DISCUSSION

GLACIER introduces a framework for inferring causal transcriptional and signaling relationships in ST by leveraging a global tissue topology. Unlike previous approaches that rely on local cell-cell adjacency, GLACIER constructs a DAG based on a global spatial coordinate, allowing us to infer how regulatory programs unfold along natural tissue axes. By orienting regulatory interactions along inferred spatial axes, we move beyond local spatial correlations and instead model how gene regulation and cell signaling propagate through *global* tissue structure. Applied to the cerebellum, GLACIER uncovered directional transcriptional programs within layers and distinct regulatory activity at layer boundaries, demonstrating its ability to identify spatially organized cellular interactions.

GLACIER requires tissues where a global spatial coordinate, such as isodepth, can be inferred. However, this is not a fundamental restriction—user-guided segmentation of complex tissues can enable locally defined spatial coordinates, allowing

GLACIER to be applied more broadly. Moreover, one can also use a supervised approach for determining such spatial coordinates, such as in refs^{39,40}. Another challenge is transcript sparsity, a common issue in ST that may become more pronounced with technologies such as xenium.⁴¹ However, GLACIER naturally mitigates this issue by aggregating over multiple cells (i.e., cells with the same isodepth), increasing robustness to sparsity. Additionally, GLACIER's bootstrap-based hyperparameter selection provides a systematic way to ensure robust inferences.

While we applied GLACIER's spatial Granger causality approach to brain layers, the methodology generalizes to other structured tissues. Future work could leverage this framework to study how signaling cascades change radially from blood vessels or how the interior and periphery of a tumor microenvironment exhibit distinct regulatory states. Moreover, our framework could be used to identify *anomalous* signaling patterns across different spatial datasets.⁴² Looking forward, integrating GLACIER with spatial proteomics⁴³ and multimodal single-cell datasets⁴⁴ could provide a more comprehensive view of spatially structured regulatory programs, further enhancing our ability to decode the organization of complex tissues.

Limitations of the study

There are several key limitations of our work. First, GLACIER cannot be applied to all tissues. This is because GLACIER relies on the isodepth learned by GASTON, which makes certain assumptions on tissue structure in order to learn the isodepth (e.g., the tissue is layered). Second, GLACIER relies on manually curated lists of TFs, ligands, and receptor genes, and thus cannot be used to discover new TFs, ligands, or receptor genes. Third, GLACIER does not incorporate protein, chromatin accessibility, or other modalities, thus limiting the regulatory programs that we are able to discover.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Rohit Singh (rohit.singh@duke.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- **Key resources table:** Table S1.
- **Data:** This study uses publicly available data, which is available at https://singlecell.broadinstitute.org/single_cell/study/SCP1663.
- **Code:** Our code is available at <https://github.com/rohitsinghlab/glacier>. The version of code used in this work is available at Zenodo (<https://doi.org/10.5281/zenodo.19500378>).
- **All other items:** The GLACIER codebase requires GASTON (<https://github.com/raphael-group/GASTON>) and Velorama (<https://github.com/rs239/velorama>) as dependencies.

ACKNOWLEDGMENTS

U.C. was supported in part by funding from the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard. H.L. and R.S. were partially supported by the Chan Zuckerberg Initiative. R.S. also acknowledges support

of the Whitehead Scholarship from the Duke School of Medicine and the Foundation for Prader-Willi Research.

AUTHOR CONTRIBUTIONS

P.S.: data curation, formal analysis, investigation, software, validation, visualization, writing – original draft, and writing – review and editing. H.L.: data curation, formal analysis, software, validation, writing – original draft, and writing – review and editing. U.C.: conceptualization, investigation, methodology, project administration, supervision, writing – original draft, and writing – review and editing. R.S.: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing – original draft, and writing – review and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
 - Granger causality in time-series
 - Spatial Granger causality
- **METHOD DETAILS**
 - Steps of GLACIER
 - Hyperparameter selection
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Robustness testing

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2026.115794>.

Received: June 10, 2025

Revised: September 12, 2025

Accepted: April 15, 2026

REFERENCES

1. Su, J., Song, Y., Zhu, Z., Huang, X., Fan, J., Qiao, J., and Mao, F. (2024). Cell-cell communication: new insights and clinical implications. *Signal Transduct. Targeted Ther.* 9, 196. <https://doi.org/10.1038/s41392-024-01888-z>.
2. Mayr, U., Serra, D., Liberali, P., Klein, A., and Treutlein, B. (2019). Exploring single cells in space and time during tissue development, homeostasis and regeneration. *Development* 146, dev176727. <https://doi.org/10.1242/dev.176727>.
3. Yeh, C.Y., Aguirre, K., Laveroni, O., Kim, S., Wang, A., Liang, B., Zhang, X., Han, L.M., Valbuena, R., Bassik, M.C., et al. (2024). Mapping spatial organization and genetic cell-state regulators to target immune evasion in ovarian cancer. *Nat. Immunol.* 25, 1943–1958. <https://doi.org/10.1038/s41590-024-01943-5>.
4. Liu, Z., Bian, X., Luo, L., Björklund, Å.K., Li, L., Zhang, L., Chen, Y., Guo, L., Gao, J., Cao, C., et al. (2024). Spatiotemporal single-cell roadmap of human skin wound healing. *Cell Stem Cell* 32, 479–498.e8. <https://doi.org/10.1016/j.stem.2024.11.013>.
5. Lewis, S.M., Asselin-Labat, M.L., Nguyen, Q., Berthelet, J., Tan, X., Wimmer, V.C., Merino, D., Rogers, K.L., and Naik, S.H. (2021). Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods* 18, 997–1012. <https://doi.org/10.1038/s41592-021-01203-6>.

6. Sarkar, H., Chitra, U., Gold, J., and Raphael, B.J. (2024). A count-based model for delineating cell–cell interactions in spatial transcriptomics data. *Bioinformatics* 40, i481–i489. <https://doi.org/10.1093/bioinformatics/btae219>.
7. Bafna, M., Li, H., and Zhang, X. (2023). Clarify: cell–cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Bioinformatics* 39, i484–i493. <https://doi.org/10.1093/bioinformatics/btad269>.
8. Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqV2. *Nat. Biotechnol.* 39, 313–319. <https://doi.org/10.1038/s41587-020-0739-1>.
9. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell* 185, 1777–1792. <https://doi.org/10.1016/j.cell.2022.04.003>.
10. Dries, R., Zhu, Q., Dong, R., Eng, C.H.L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., et al. (2021). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology* 22, 1–31. <https://doi.org/10.1186/s13059-021-02286-2>.
11. Miller, B.F., Bambah-Mukku, D., Dulac, C., Zhuang, X., and Fan, J. (2021). Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res.* 31, 1843–1855. <https://doi.org/10.1101/gr.271288.120>.
12. Pham, D., Tan, X., Balderson, B., Xu, J., Grice, L.F., Yoon, S., Willis, E.F., Tran, M., Lam, P.Y., Raghuram, A., et al. (2023). Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat. Commun.* 14, 7739. <https://doi.org/10.1038/s41467-023-43120-6>.
13. Li, Z., Wang, T., Liu, P., and Huang, Y. (2023). Spatialdm for rapid identification of spatially co-expressed ligand–receptor and revealing cell–cell communication patterns. *Nat. Commun.* 14, 3995. <https://doi.org/10.1038/s41467-023-39608-w>.
14. Wang, C., Baker, B.M., Chen, C.S., and Schwartz, M.A. (2013). Endothelial cell sensing of flow direction. *Arterioscler. Thromb. Vasc. Biol.* 33, 2130–2136. <https://doi.org/10.1161/ATVBAHA.113.301826>.
15. Lai, E.C. (2004). Notch signaling: control of cell communication and cell fate. *Development* 131, 965–973. <https://doi.org/10.1242/dev.01074>.
16. Chitra, U., Arnold, B.J., Sarkar, H., Sanno, K., Ma, C., Lopez-Darwin, S., and Raphael, B.J. (2025). Mapping the topography of spatial gene expression with interpretable deep learning. *Nat. Methods* 22, 298–309. <https://doi.org/10.1038/s41592-024-02503-3>.
17. Singh, R., Wu, A.P., Mudide, A., and Berger, B. (2024). Causal gene regulatory analysis with rna velocity reveals an interplay between slow and fast transcription factors. *Cell Syst.* 15, 462–474. <https://doi.org/10.1016/j.cels.2024.04.005>.
18. Wu, A.P., Singh, R., and Berger, B. (2022). Granger causal inference on DAGs identifies genomic loci regulating transcription. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=nZOUYEN6Wvy>.
19. Wu, A.P., Singh, R., Walsh, C.A., and Berger, B. (2025). Unveiling causal regulatory mechanisms through cell-state parallax. *Nat. Commun.* 16, 8096. <https://doi.org/10.1038/s41467-025-61337-5>.
20. Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Irizarry, R.A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* 40, 517–526. <https://doi.org/10.1038/s41587-021-00830-w>.
21. Cable, D.M., Murray, E., Shanmugam, V., Zhang, S., Zou, L.S., Diao, M., Chen, H., Macosko, E.Z., Irizarry, R.A., and Chen, F. (2022). Cell type-specific inference of differential expression in spatial transcriptomics. *Nat. Methods* 19, 1076–1087. <https://doi.org/10.1038/s41592-022-01575-3>.
22. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* 44, W90–W97. <https://doi.org/10.1093/nar/gkw377>.
23. Simons, M., and Nave, K.A. (2015). Oligodendrocytes: Myelination and axonal support. *Cold Spring Harbor Perspect. Biol.* 8, a020479. <https://doi.org/10.1101/cshperspect.a020479>.
24. Muñoz-Lasso, D.C., Romá-Mateo, C., Pallardó, F.V., and Gonzalez-Cabo, P. (2020). Much more than a scaffold: cytoskeletal proteins in neurological disorders. *Cells* 9, 358.
25. Rasmussen, A.H., Rasmussen, H.B., and Silaharoglu, A. (2017). The dlagp family: neuronal expression, function and role in brain disorders. *Mol. Brain* 10, 43. <https://doi.org/10.1186/s13041-017-0324-9>.
26. Torraco, A., Peralta, S., Iommarini, L., and Diaz, F. (2015). Mitochondrial diseases part i: mouse models of oxphos deficiencies caused by defects in respiratory complex subunits or assembly factors. *Mitochondrion* 21, 76–91. <https://doi.org/10.1016/j.mito.2015.01.009>.
27. Wang, J., Lee, J., Liem, D., and Ping, P. (2017). Hspa5 gene encoding hsp70 chaperone bip in the endoplasmic reticulum. *Gene* 618, 14–23. <https://doi.org/10.1016/j.gene.2017.03.005>.
28. Pickrell, A.M., and Youle, R.J. (2015). The roles of pink1, parkin, and mitochondrial fidelity in parkinson's disease. *Neuron* 85, 257–273. <https://doi.org/10.1016/j.neuron.2014.12.007>.
29. Bluhm, B., Laffer, B., Himet, D., Rothermundt, M., Ambree, O., and Lohr, C. (2015). Normal cerebellar development in s100b-deficient mice. *Cerebellum* 14, 119–127. <https://doi.org/10.1007/s12311-014-0606-z>.
30. Lu, K.V., Jong, K.A., Kim, G.Y., Singh, J., Dia, E.Q., Yoshimoto, K., Wang, M.Y., Cloughesy, T.F., Nelson, S.F., and Mischel, P.S. (2005). Differential induction of glioblastoma migration and growth by two forms of pleiotrophin. *J. Biol. Chem.* 280, 26953–26964. <https://doi.org/10.1074/jbc.M502614200>.
31. Lewis, E., Chapman, G., Kaushik, K., Determan, J., Antony, I., Meganaathan, K., Narasimhan, M., Gontarz, P., Zhang, B., and Kroll, K. (2022). Regulation of human cortical interneuron development by the chromatin remodeling protein chd2. *Sci. Rep.* 12, 15636. <https://doi.org/10.1038/s41598-022-19654-y>.
32. Afshar, P., Ashtari, N., Jiao, X., Rahimi-Balaei, M., Zhang, X., Yaganeh, B., Del Bigio, M.R., Kong, J., and Marzban, H. (2017). Overexpression of human sod1 leads to discrete defects in the cerebellar architecture in the mouse. *Front. Neuroanat.* 11, 22. <https://doi.org/10.3389/fnana.2017.00022>.
33. Coutelier, M., Jacoupy, M., Janer, A., Renaud, F., Auger, N., Saripella, G.V., Ancien, F., Pucci, F., Rooman, M., Gilis, D., et al. (2022). Nptx1 mutations trigger endoplasmic reticulum stress and cause autosomal dominant cerebellar ataxia. *Brain* 145, 1519–1534. <https://doi.org/10.1093/brain/awab407>.
34. Heinzen, E.L., Arzimanoglou, A., Brashear, A., Clapcote, S.J., Gurrieri, F., Goldstein, D.B., Johannesson, S.H., Mikati, M.A., Neville, B., Nicole, S., et al. (2014). Distinct neurological disorders with atp1a3 mutations. *Lancet Neurol.* 13, 503–514. [https://doi.org/10.1016/S1474-4422\(14\)70011-0](https://doi.org/10.1016/S1474-4422(14)70011-0).
35. Mahley, R.W. (2016). Apolipoprotein e: from cardiovascular disease to neurodegenerative disorders. *Journal of Molecular Medicine* 94, 739–746. <https://doi.org/10.1007/s00109-016-1427-y>.
36. Li, S., Nie, E.H., Yin, Y., Benowitz, L.I., Tung, S., Vinters, H.V., Bahjat, F.R., Stenzel-Poore, M.P., Kawaguchi, R., Coppola, G., and Carmichael, S.T. (2015). GDF10 is a signal for axonal sprouting and functional recovery after stroke. *Nat. Neurosci.* 18, 1737–1745. <https://doi.org/10.1038/nn.4146>.
37. Clapham, D.E., and Neer, E.J. (1997). G protein beta gamma subunits. *Annu. Rev. Pharmacol. Toxicol.* 37, 167–203. <https://doi.org/10.1146/annurev.pharmtox.37.1.167>.
38. Rege, T.A., and Hagood, J.S. (2006). Thy-1 as a regulator of cell–cell and cell–matrix interactions in axon regeneration, apoptosis, and cancer progression. *FASEB (Fed. Am. Soc. Exp. Biol.) J.* 20, 1045–1054. <https://doi.org/10.1096/fj.05-5460rev>.

39. Ma, C., Chitra, U., Zhang, S., and Raphael, B.J. (2022). Belayr: Modeling discrete and continuous spatial variation in gene expression from spatially resolved transcriptomics. *Cell Syst.* **13**, 786–797. <https://doi.org/10.1016/j.cels.2022.09.002>.
40. Chitra, U., Dan, S., Krienen, F., and Raphael, B.J. (2025). Gaston-mix: a unified model of spatial gradients and domains using spatial mixture-of-experts. *Bioinformatics* **41**, i523–i532. <https://doi.org/10.1093/bioinformatics/btaf254>. https://academic.oup.com/bioinformatics/article-pdf/41/Supplement_1/i523/63745266/btaf254.pdf.
41. Janesick, A., Shelansky, R., Gottscho, A.D., Wagner, F., Williams, S.R., Rouault, M., Beliakoff, G., Morrison, C.A., Oliveira, M.F., Sicherman, J.T., et al. (2023). High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353. <https://doi.org/10.1038/s41467-023-43458-x>.
42. Hu, G., Gold, J., Chitra, U., Joshi, S., and Raphael, B.J. (2025). Anomaly detection in spatial transcriptomics via spatially localized density comparison. *Bioinformatics* **41**, i493–i501. <https://doi.org/10.1093/bioinformatics/btaf242>.
43. Lundberg, E., and Borner, G.H. (2019). Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302. <https://doi.org/10.1038/s41580-018-0094-y>.
44. Vandereyken, K., Sifrim, A., Thienpont, B., and Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515. <https://doi.org/10.1038/s41576-023-00580-2>.
45. Enders, W. (2008). *Applied Econometric Time Series* (John Wiley & Sons).
46. Granger, C.W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: J. Econom. Soc.* **37**, 424–438.
47. McCracken, M.W. (2007). Asymptotics for out of sample tests of granger causality. *J. Econom.* **140**, 719–752.
48. Shojaie, A., and Fox, E.B. (2022). Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application* **9**, 289–319.
49. Wu, Z., Trevino, A.E., Wu, E., Swanson, K., Kim, H.J., D'Angio, H.B., Pre-ska, R., Charville, G.W., Dalerba, P.D., Egloff, A.M., et al. (2022). Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nat. Biomed. Eng.* **6**, 1435–1448. <https://doi.org/10.1038/s41551-022-00951-w>.
50. Townes, F.W., Hicks, S.C., Aryee, M.J., and Irizarry, R.A. (2019). Feature selection and dimension reduction for single-cell ma-seq based on a multinomial model. *Genome biology* **20**, 1–16.
51. Tank, A., Covert, I., Foti, N., Shojaie, A., and Fox, E.B. (2022). Neural granger causality. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 4267–4279. <https://doi.org/10.1109/TPAMI.2021.3065601>.
52. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using cellchat. *Nat. Commun.* **12**, 1088. <https://doi.org/10.1038/s41467-021-21246-9>.
53. Rodrigues, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467. <https://doi.org/10.1126/science.aaw1219>.
54. Russell, A.J., Weir, J.A., Nadaf, N.M., Shabet, M., Kumar, V., Kambhampati, S., Raichur, R., Marrero, G.J., Liu, S., Balderrama, K.S., et al. (2023). Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature* **625**, 101–109. <https://doi.org/10.1038/s41586-023-06837-4>.
55. Moffitt, J.R., Bambach-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324. <https://doi.org/10.1126/science.aau5324>.
56. Tian, L., Chen, F., and Macosko, E.Z. (2023). The expanding vistas of spatial transcriptomics. *Nat. Biotechnol.* **41**, 773–782. <https://doi.org/10.1038/s41587-022-01448-2>.
57. Rao, A., Barkley, D., França, G.S., and Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220. <https://doi.org/10.1038/s41586-021-03634-9>.
58. Palla, G., Fischer, D.S., Regev, A., and Theis, F.J. (2022). Spatial components of molecular tissue biology. *Nat. Biotechnol.* **40**, 308–318. <https://doi.org/10.1038/s41587-021-01182-1>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Slide-SeqV2 of mouse cerebellum available from Cable et al., Nature Methods (2022).	N/A	https://singlecell.broadinstitute.org/single_cell/study/SCP1663 .
Software and algorithms		
Our code is available at https://github.com/rohitsinghlab/glacier	N/A	https://doi.org/10.5281/zenodo.19500378

Granger causality in time-series

Granger causality is a statistical framework for inferring causal relationships in time-series data by assessing whether past values of one variable $\mathbf{x}$ (predictor) improve the prediction of another variable $\mathbf{y}$ (target) beyond what can be explained by the latter's own past values. The key principle is that a cause must precede its effect, and predictive relationships between variables may therefore indicate potential causality. Granger causal interactions are statistical estimates, best interpreted as causal hypotheses to be prioritized for experimental validation; for example, a Granger causal relationship between $\mathbf{x}$ and $\mathbf{y}$ may be indirect, mediated through a latent and potentially unobserved variable.

Linear Granger causality

Given two time series of length N , $\mathbf{x} = (x_t)_{t=1}^N \in \mathbb{R}^N$ and $\mathbf{y} = (y_t)_{t=1}^N \in \mathbb{R}^N$, we say that they follow a *vector autoregressive (VAR)* model with $L > 0$ lags if

$$y_t = \sum_{\ell=1}^L \alpha_{\ell} y_{t-\ell} + \sum_{\ell=1}^L \beta_{\ell} x_{t-\ell} + \epsilon_t, \quad (\text{Equation 1})$$

where ϵ_t is a noise term, and $\alpha_{\ell}, \beta_{\ell} \in \mathbb{R}$ are coefficients that describe the contribution of the observations $y_{t-\ell}, x_{t-\ell}$, respectively, at the ℓ -th lag, i.e., the $(t - \ell)$ -th time. The value $L > 0$ indicates how far back in time past observations influence the observation y_t , i.e., the *maximum lag*. VAR models are widely used for time-series analysis in finance and economics.⁴⁵

We say that $\mathbf{x}$ linearly *Granger causes* $\mathbf{y}$ if $\beta_{\ell} \neq 0$ for some lag $0 < \ell \leq L$.⁴⁶ The standard approach for determining linear Granger causality is through hypothesis testing, where the null hypothesis is $H_0: \beta_{\ell} = 0$ for all lags $\ell = 1, \dots, L$, e.g., using an F-test or likelihood ratio test.^{47,48}

Non-linear Granger causality

In many scenarios, the time series $\mathbf{x}, \mathbf{y}$ are not linearly related. Thus, we generalize (1) and say that $\mathbf{x}$ and $\mathbf{y}$ follow a *non-linear vector autoregressive (NVAR)* model with L lags if

$$y_t = f(y_{t-1}, \dots, y_{t-L}) + g(x_{t-1}, \dots, x_{t-L}) + \epsilon_t, \quad (\text{Equation 2})$$

where $f, g: \mathbb{R}^L \rightarrow \mathbb{R}$ are functions— potentially parameterized by neural networks— that specify how the observation y_t at time t depends on the previous L observations $y_{t-1}, \dots, y_{t-L}$ and $x_{t-1}, \dots, x_{t-L}$, respectively, and ϵ_t is a noise term.

In a non-linear setting, Granger causality can be detected through two approaches: *ablation-* or *invariance-based*.¹⁷ The appropriate choice depends on the problem setting and the kind of interactions modeled between predictor variables.

- **Ablation-based approach.** Causality is assessed by training two models: a full model $\mathcal{M}_{\text{full}}$ that includes past values of both $\mathbf{x}$ and $\mathbf{y}$, and a restricted model $\mathcal{M}_{\text{reduced}}$ that omits $\mathbf{x}$. If removing $\mathbf{x}$ significantly reduces predictive accuracy, $\mathbf{x}$ is deemed Granger-causal of $\mathbf{y}$.
- **Invariance-based approach.** Instead of explicitly comparing models, this approach applies regularization to $\mathbf{x}$ -dependent terms. If removing $\mathbf{x}$ does not alter the model's predictions (e.g., its coefficients shrink to zero under sparsity constraints), then $\mathbf{x}$ does not Granger-cause $\mathbf{y}$. This method is computationally efficient and allows joint learning of multiple interactions. We note that applying it in a neural network setting requires special architecture and training considerations to encode the required sparsity constraints.¹⁷

Spatial Granger causality

Classical Granger causality approaches require a total ordering on the N observations, and are not immediately applicable to *spatially structured* data where there is no obvious total ordering of locations in space.

Here we leverage recent extensions of Granger causality to *partially ordered* data described by a directed acyclic graph (DAG).^{17–19} To extend Granger causality to spatially structured data, we combine the DAG-structured Granger causality framework with the construction of a *spatial DAG*.

We start by describing the DAG-structured Granger causality framework of GrID-Net and Velorama.^{17–19} Given a DAG $G = (V, E)$ and observations $\mathbf{x} = (x_v)_{v \in V} \in \mathbb{R}^{|V|}$, $\mathbf{y} = (y_v)_{v \in V} \in \mathbb{R}^{|V|}$ defined on each vertex $v \in V$, we say that the observations $\mathbf{x}$, $\mathbf{y}$ follow a *DAG-structured vector autoregressive (DVAR)* model with L lags, modifying (2) as follows:

$$y_v = f(\{y_u\}_{u \in \mathcal{P}_L(v)}) + g(\{x_u\}_{u \in \mathcal{P}_L(v)}) + \epsilon_v, \quad (\text{Equation 3})$$

where $\mathcal{P}_L(v)$ is the set of vertices $u \in V$ that are ancestors of v and are at most distance L from v . That is,

$$\mathcal{P}_L(v) = \{u \in V \mid u \text{ is an ancestor of } v \text{ with } d_G(u, v) \leq L\}, \quad (\text{Equation 4})$$

where $d_G(u, v)$ is the length of the shortest directed path from u to v in G ; and $f, g : 2^{\mathbb{R}} \rightarrow \mathbb{R}$ are functions that map each subset of the observations $\mathbf{y}$, $\mathbf{x}$, respectively, to a real number. As mentioned before, one can then infer causal relationships using an ablation or invariance-based approach. To estimate Granger causal relationships between individual genes and ATAC-seq peaks, GrID-Net used an ablation based approach.^{18,49} In contrast, when seeking to assess and select from multiple TFs (i.e., predictors) acting combinatorially, Velorama used an invariance-based approach with per-TF sparsity constraints.¹⁷

DAG construction

We construct a spatial DAG using a *spatial potential* function, similar to how¹⁷ constructs a DAG from a pseudotime function. Briefly, suppose we are given N spatial locations $\mathbf{s}_1 = (x_1, y_1)$, ..., $\mathbf{s}_N = (x_N, y_N)$ and a real number $d_i \in \mathbb{R}^d$ defined on each spatial location $\mathbf{s}_i$, which we call a *spatial potential*. We form a DAG $G = (V, E)$ whose vertices $V = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ are the spatial locations, and whose edges E are those in the k -nearest-neighbors (k -NN) graph. Each edge (i, j) is oriented in the direction of increasing spatial coordinate; that is for any two arbitrary cells i, j , we include a directed edge $i \rightarrow j$ if $d_i > d_j$. Such a construction guarantees that G is a DAG.¹⁷

Isodepth

While our approach is generalizable to any spatial potential d_i , in practice we use the spatial potential d_i learned by GASTON.¹⁶ The GASTON spatial potential d is called the *isodepth* and is a 1-D latent coordinate describing the direction of maximum spatial variation in gene expression. For example, in layered tissues (e.g., the cerebellum tissue slice that we analyze in the Results and Figures 2 and 4), the isodepth d corresponds to the *layer depth*.

We briefly describe GASTON's model of spatial gene expression and its derivation of the isodepth d from spatial transcriptomics data. We aim to learn a gene expression function $\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^G$ where $\mathbf{f}(x, y)$ describes the gene expression vector for all G genes at spatial location (x, y) . GASTON assumes this gene expression function can be factored as $\mathbf{f}(x, y) = \mathbf{h}(d(x, y))$ where $d(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a continuously differentiable function which we call the *isodepth*, and $\mathbf{h} : \mathbb{R} \rightarrow \mathbb{R}^G$ is a 1-D gene expression function. We assume the 1-D gene expression function $\mathbf{h}$ is piecewise linear with L pieces and breakpoints $b_1 < b_2 < \dots < b_{L-1}$. Given spatial coordinates $\mathbf{S} = [\mathbf{s}_i] \in \mathbb{R}^{N \times 2}$ for N spots where $\mathbf{s}_i = (x_i, y_i)$ is the coordinate for spot $i \in \{1, \dots, N\}$, and gene expression counts $\mathbf{C} = [c_{ig}] \in \mathbb{R}^{N \times G}$ where c_{ig} is the observed gene expression for gene g in spot i , we learn the isodepth d and 1-D expression function $\mathbf{h}$ by solving the following optimization problem:

$$\begin{aligned} \arg\max \quad & d \in C^1(\mathbb{R}^2, \mathbb{R}) \\ & b_1 < b_2 < \dots < b_{L-1} \\ \mathbf{h} = (h_1, \dots, h_G) \in & \mathcal{L}(b_1, \dots, b_{L-1}) \end{aligned} \quad \sum_{g=1}^G \left(\sum_{i=1}^N \log \mathbb{P}(c_{ig} | h_g(d(x_i, y_i))) \right), \quad (\text{Equation 5})$$

where $C^1(\mathbb{R}^2, \mathbb{R})$ is the space of continuously differentiable functions from $\mathbb{R}^2$ to $\mathbb{R}$; $\mathcal{L}(b_1, \dots, b_{L-1})$ is the space of piecewise linear functions with breakpoints $b_1, \dots, b_{L-1}$, i.e., a piecewise linear function with L pieces. The problem in 5 can be solved with different probability distributions $\mathbb{P}(a_{ig} | f_g(x_i, y_i)) = \mathbb{P}(a_{ig} | h_g(d(x_i, y_i)))$ for the gene expression values a_{ig} . Following,¹⁶ we model the gene expression c_{ig} with a Poisson distribution of the form $c_{ig} \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(U_i \cdot \exp(f_g(x_i, y_i)))$, where U_i is the total UMI count at spot i . In practice, due to the large number G of genes, we solve 5 using the top GLM-PCs⁵⁰ with a Gaussian error model, which accounts for the Poisson distribution while enabling dimensionality reduction. See the GASTON manuscript¹⁶ for more details.

We also emphasize that the isodepth d is one example of a spatial potential, and our method can be applied directly to perform causal analysis in the setting of other spatial potentials (e.g., radial distance from a blood vessel, germinal center, or tumor mass).

Inference

Since our inference objectives (i.e., identifying combinations of genes as regulators) are similar to Velorama, we follow its invariance-based approach to extend Granger causality to spatial transcriptomics. For each candidate regulator $\mathbf{x} = (x_v)_{v \in V}$ (i.e., gene, out of N_G total candidates) and vertex $v \in V$ (i.e., cell) in G , we accumulate its expression across the $\mathcal{P}_L(v)$ ancestors of v . Together, these form the input to a neural network that predicts the expression of a target gene $\mathbf{y} = (y_v)_{v \in V}$. To enforce sparsity in the selection of regulators, we incorporate a regularization on the first layer of the neural network (Figure 1B, right). This acts as a feature selection mechanism, ensuring that only a subset of candidate regulators contribute to predicting the target gene. Specifically, the first hidden layer takes the form

$$h^{(1)} = \sigma \left(\sum_{\ell=1}^L (A')^{\ell} X W_1^{\ell} + b_1 \right), \quad (\text{Equation 6})$$

where $(A')^{\ell}$ is the ℓ -th power of the adjacency matrix of G ; $X = [\mathbf{x}] \in \mathbb{R}^{|V| \times N_G}$ is the regulator expression matrix, where each column is a candidate regulator $\mathbf{x}$; $W_1^{\ell} \in \mathbb{R}^{N_G \times d}$ are trainable weight matrices with a lasso penalty; d is the dimensionality of the hidden layers; and $\sigma(\cdot)$ is a non-linear activation function.

Following ref¹⁷, we apply a hierarchical regularization scheme that penalizes W_1^{ℓ} . Longer lags are more heavily penalized than shorter ones. This encourages robustness in selecting lagged interactions while reducing sensitivity to the choice of the maximum lag L .⁵¹ To ensure sparsity, the model is trained using proximal gradient descent, which is well-suited for enforcing sparsity constraints. The regularization weight is specified by the hyperparameter λ . To select an appropriate λ , we perform a grid search over a range of values and ensemble the results across valid λ settings where some, but not all, weights remain nonzero. This enables GLACIER to infer spatially structured regulatory relationships that capture both local and long-range dependencies dictated by tissue architecture. We say that the *lag* of an interaction is the maximum number of non-zero weights. See ref¹⁷ for details.

METHOD DETAILS

We present GLACIER, an algorithm for identifying Granger causal interactions from single-cell, spatial transcriptomics (ST) data (Figure 1A).

Input. The input to GASTON is ST data $(\mathbf{A}, \mathbf{S})$ which consists of.

- an $N \times G$ gene expression matrix $\mathbf{A} = [a_{i,g}] \in \mathbb{R}^{N \times G}$, where $a_{i,g}$ is the expression of gene $g \in \{1, \dots, G\}$ in spot/cell $i \in \{1, \dots, N\}$,
- an $N \times 2$ spatial location matrix $\mathbf{S} = [\mathbf{s}_i] \in \mathbb{R}^{N \times 2}$, where $\mathbf{s}_i = (x_i, y_i)$ is the 2-D coordinate of spot/cell i for $i = 1, \dots, N$, and a list $\mathcal{L} \subseteq G \times G$ of putatively interacting genes; e.g., from a database of experimentally observed ligand-receptor interactions such as CellChatDB.⁵²

The ST data $(\mathbf{A}, \mathbf{S})$ is the standard output from applying a spatial sequencing or imaging technology to a 2-D tissue slice. Such spatial technologies include 10x Genomics Visium, Slide-Seq/Slide-SeqV2/Slide-tags,^{8,53,54} and MERFISH.⁵⁵ Different spatial technologies differ in the numbers N, G of spots and genes, respectively, that are measured, as well as their spatial resolution. See^{56–58} for details. We also note that the list $\mathcal{L}$ of putatively interacting genes is incomplete, i.e., not all interacting ligands and receptors are included in $\mathcal{L}$, and may also be biased toward genes that are well-studied (i.e., ascertainment bias).

Steps of GLACIER

GLACIER consists of the following steps. First, GLACIER uses GASTON¹⁶ to compute the isodepth d_i for each cell $i = 1, \dots, N$ and forms a spatial DAG G , as described in the previous section (Figure 1B, left). For each pair $(g, h) \in \mathcal{L}$ of candidate interacting genes, GLACIER uses the spatial DAG to test and report whether the observed gene- g expression values $\mathbf{a}_g = (a_{i,g})_{i=1}^N$ Granger cause the observed gene- h expression values $\mathbf{a}_h = (a_{i,h})_{i=1}^N$ (Figure 1B, right), and reports all pairs (g, h) where this Granger causality relationship holds (Figure 1C).

We test for Granger causality using a version of Velorama¹⁷ adapted to spatial data. We construct the spatial DAG using a k -NN graph (with $k = 7$) on each cell's spatial coordinates, oriented using isodepth. Notably, isodepth has the unique property that it can be *inverted* by redefining each cell's isodepth as

$$d'_i = \max_{1 \leq j \leq N} (d_j) - d_i \quad (\text{Equation 7})$$

which reverses the ordering of isodepth among cells, allowing for analysis of information flow in the opposite direction.

Hyperparameter selection

Due to the sparsity of transcript counts from current ST technologies, several hyperparameter adjustments are necessary to ensure Velorama is able to run smoothly. We adjusted key hyperparameters as follows.

- **Learning Rate:** A higher learning rate than the default used in Velorama was sometimes necessary to achieve stable convergence. To balance this higher learning rate, we also implemented a learning rate scheduler.
- **Number of Epochs:** We also needed to increase the number of training epochs to allow the model sufficient time to reach an optimal solution due to the number of genes and learning rate scheduling.
- **Regularization (λ):** Due to the sparsity of spatial transcriptomics data, we needed to sample from a broader range of λ values to ensure valid settings were obtained. As defined in,¹⁷ valid settings are those where between 5% and 95% of regularization weights remain nonzero.

To validate the robustness of our hyperparameter choices, we conducted a bootstrap-style analysis in which the dataset was randomly partitioned into three subsets. We selected hyperparameters that led to robust inference results across all three runs

(Figure S1). We provide guidance on hyperparameter selection with the source code release. Running GLACIER on the full mouse cerebellum dataset, which consists of 9,985 cells, typically required ≈ 400 min on a single Nvidia A6000 GPU.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene set enrichment analysis was performed with Enrichr,²² using their web server <https://maayanlab.cloud/Enrichr/>. Briefly, Enrichr is a web-based platform for gene set enrichment analysis that evaluates gene lists against diverse functional categories to identify statistically overrepresented pathways and functions. It provides detailed enrichment statistics and visualizations to facilitate biological interpretation.

Robustness testing

Velorama sweeps over a range of regularization hyperparameters (λ), averaging TF-target gene interactions from models where the percentage of non-zero interactions fall within a given threshold (by default, 0.01 to 0.95). To find λ values meeting this threshold for each layer and isodepth orientation, we increased the number of λ s tested from the default of 10–30. We also narrowed the search range from (0.01, 10) to (0.001, 0.1), since we found suitable λ s was always below 0.1 empirically. We used a higher learning rate ($\eta = 0.1$) for TF-target gene analysis and a lower learning rate ($\eta = 0.01$) for ligand-receptor analysis. The changes in learning rate were made across all runs, and the number of iterations per λ during training was significantly increased from the default of 1000 to between 10,000 and 30,000, depending on the run.

We verified other hyperparameters by testing on stratified thirds of the dataset. Each stratified third was constructed by evenly selecting cells from cell type-specific subsets and then combining them, ensuring that the cell type distribution was preserved. We tuned hyperparameters so that results were consistent across stratified thirds and the overall dataset (for both normal and inverted isodepth). Finally, we analyzed the loss for each run to confirm smooth training behavior and convergence at the end of each model training run.