Eurographics Workshop on Visual Computing for Biology and Medicine (2025)
T. Gerrits and C. Gillmann (Poster Chairs)

Cytosplore Gradient Surfer: Comparative Visual Analysis of Spatial Gene Expression Gradients Across Datasets

Soumyadeep Basu[1,2†] Chang Li[1†] Thomas Kroes[1] Boudewijn Lelieveldt[1] Brian Long[3] Thomas Höllt[2]

[1]Leiden University Medical Center [2]Delft University of Technology [3]Allen Institute for Brain Science [†]Authors contributed equally

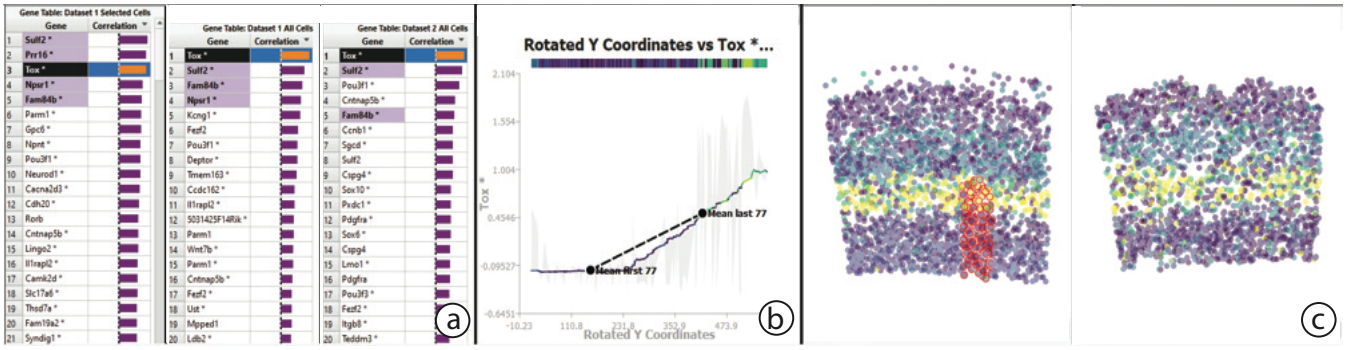

Figure 1: **Cytosplore Gradient Surfer comparing data from two different modalities (MERFISH and smFISH [LMC23])** using linked views: gene correlation tables indicating correlation between a user-defined spatial direction and gene expression (a), line chart showing the gene expression along the selected axis (b), and spatial maps colored by gene expression to visualize spatial gradients across modalities (c).

Abstract

Spatial transcriptomics makes gene expression visible within tissue. This helps reveal tissue structure, cell development, or how genes are regulated. While methods exist to identify spatially patterned genes or align tissue structures, comparative analysis of such patterns across datasets or between different modalities remains a challenging task.

We present Cytosplore Gradient Surfer, an interactive system for identifying and analyzing spatial gene expression gradients across spatial transcriptomics datasets. The system allows users to interactively probe spatial directions of interest and identify genes exhibiting aligned expression patterns across dataset pairs, even when only a subset of genes is shared between them.

CCS Concepts

• ***Human-centered computing*** → *Visualization systems and tools;* • ***Applied computing*** → *Bioinformatics;*

1. Introduction

Spatial transcriptomics technologies measure gene expression variation across tissues while preserving spatial organization, complementing single-cell RNA sequencing. They capture molecular gradients key to understanding tissue architecture, cellular differentiation, and gene regulation patterns. Comparing these gradients across datasets is valuable, e.g., between modalities like MERFISH and smFISH [LMC23] to find consistent or complementary spatial gene expression patterns, or across datasets to study gene expression in similar structures. However, such comparisons are challenging due to variations in tissue morphology, cell arrangements, gene expression levels, and differing gene sets, some unique or varying even among homologous genes, complicating alignment of tissue structures and identification of conserved gradients.

While tools for cross-dataset comparison of single-cell data exist [BWE*25, BEK*23], they lack support for spatial data. On the other hand, a considerable amount of work has been put into tools for spatial data exploration with a focus on single-cell neighborhood analysis [SVUK*21, KGM*21, KBJ*20] and larger scale spatial patterns and gradients [LTK*25, LTK*23]. General single-cell analysis methods address temporal dynamics [MMDP10] or utilize dimensionality reduction techniques [FWS*19, Pat18]. However, none of these tools enable integrated analysis of spatial pattern in cross-dataset comparison.

We present Cytosplore Gradient Surfer, an interactive system for identifying, comparing, and visualizing spatial expression gradients across datasets. While still a work in progress, the system aims to lay the groundwork for more comprehensive comparative spatial transcriptomics analysis across datasets.

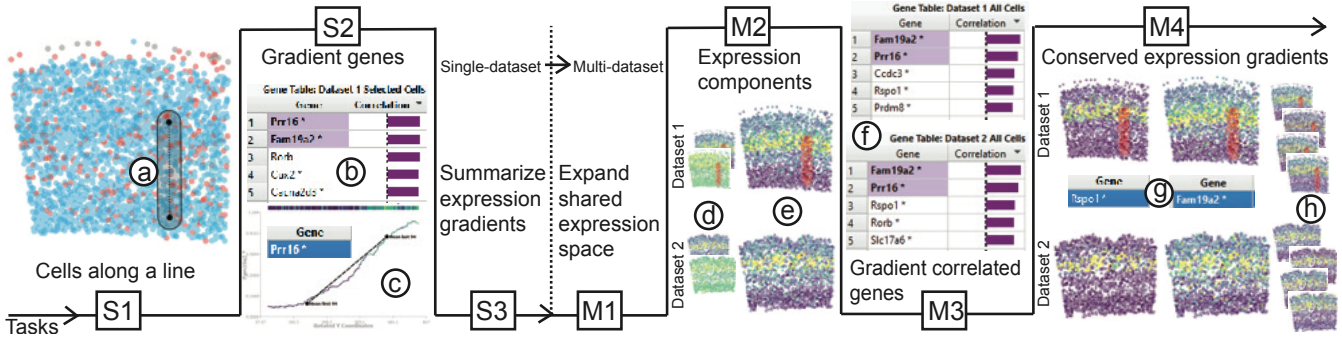

Figure 2: **Cytosplore Gradient Surfer workflow** for exploring spatial gene expression gradients. Line brush to select cells (a). Genes with spatial gradients identified by correlation with the selection (b). A line chart to visualize gene expression along the line (c). PCA loadings reveal dominant spatial patterns (d). Key principal components showing major gradients (e). Ranked gene correlations with PCs in tables and bar charts (f). Cross-dataset comparison of top spatial genes (g). Comparison of conserved or altered spatial patterns across datasets (h).

2. Workflow and Tasks

We developed a workflow (Figure 2) to analyze and compare spatial gene expression gradients across two datasets. Starting with a reference dataset, we select cells along a line and identify genes whose expression correlates with position. The top N genes, shared between datasets, are then used as input for Principal Component Analysis (PCA) space, into which all datasets are projected using the same gene set and loadings. This enables comparison of spatial gradients without requiring identical cell and gene distributions across datasets, assuming a shared underlying signal.

To support this workflow, we defined the following **S**ingle and **M**ulti-dataset tasks:

S1 *Select* cells along a spatial line
S2 *Identify* genes with expression gradients along the line
S3 *Summarize* expression gradient of selected genes and cells

M1 *Derive* data from all datasets into a shared gradient space
M2 *Discover* dataset-specific expression components
M3 *Identify* gradient-correlated genes within each dataset
M4 *Compare* conserved gene expression gradients

3. Cytosplore Gradient Surfer

Cytosplore Gradient Surfer enables users to interactively probe for, extract, and align gradient-based gene expression features to identify spatial patterns in tissues. It identifies conserved or divergent molecular patterns associated with tissue structure or biological function. The system takes as input a cell-by-gene expression matrix and corresponding spatial coordinates to support the analysis.

Biologists often have hypotheses or prior knowledge about spatial gradients in tissues and aim to identify genes that reflect these patterns. To support exploration of spatial gradients, we provide a line brush (Figure 2a) for selecting cells around a user-defined line (**S1**). The cells are projected onto the drawn line and, for each gene, the Pearson correlation is computed between expression levels and the projected cell position. This reveals genes (Figure 2b) whose expression consistently varies with spatial position, reflecting a gradient (**S2**).

To examine these patterns, the system displays the spatial layout of cells in scatterplots, colored by expression level, and a line chart (Figure 2c) summarizing expression along the selected line. Users can identify genes whose expression aligns with the gradient of interest. These genes can then be selected for broader analysis.

To explore broader spatial trends and summarize major patterns of variation between datasets, we apply PCA with the identified spatially varying genes as input. Typically, one of the high-ranked PCs (**S3**) highly correlates with the dominant spatial pattern. PC loadings can be visualized as color overlays on spatial scatterplots similar to the gene expression (Figure 2d) to verify which PC represent the probed gradient best (Figure 2e). Using the results of the PCA, we project the second dataset (**M1**). Correlating the genes of the second dataset with the identified highly-ranked PC now allows to identify genes of interest in the second dataset, too (**M2**).

Next, we compute the correlation of each gene's expression with the selected PC for each dataset (**M3**), linking genes to dominant spatial patterns within and across datasets. The results are presented as tables for each dataset (Figure 2f), with rows representing genes ranked by their alignment with these spatial trends. The user can then browse through the gene lists and select individual genes for comparison. The rank of the genes can then be compared (**M4**) across datasets in the tables(Figure 2g). By overlaying gene expression as color in the side-by-side scatterplots (Figure 2h), the user can evaluate if new genes follow similar patterns and if these patterns are conserved or altered.

4. Conclusion

We defined a work-in-progress workflow to analyze spatial gene expression gradients across datasets and developed Cytosplore Gradient Surfer, built with ManiVault Studio [VKT*23], to support interactive selection, gene correlation, and comparative analysis. By enabling the comparison of spatial gradients across different modalities or datasets, the system supports a diverse range of applications, from cross-modality integration to identifying conserved spatial patterns of gene expression. We are collaborating with experts to incorporate various use cases, including extension to more than two concurrent datasets.

Acknowledgements

This work received financial support from the NWO Gravitation project BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology (NWO: 024.004.012).

References

[BEK*23] BASU S., EGGERMONT J., KROES T., JORSTAD N., BAKKEN T., LEIN E., LELIEVELDT B., HÖLLT T.: Cytosplore Simian Viewer: Visual exploration for multi-species single-cell RNA sequencing data. In *Proc. VCBM* (Eindhoven, NL, 2023), EUROGRAPHICS, pp. 111–120. doi:10.2312/VCBM.20231219. 1

[BWE*25] BASU S., WIRTHLIN M., EGGERMONT J., KROES T., LELIEVELDT B., LEIN E., BAKKEN T., HÖLLT T.: Cytosplore EvoViewer: Visual analytics of conserved evolutionary patterns in multi-species single-cell sequencing data. In *Proc. PacificVis* (Taipei City, Taiwan, 2025), IEEE, pp. 149–159. doi:10.1109/PacificVis64226.2025.00021. 1

[FWS*19] FENG D., WHITEHURST C. E., SHAN D., HILL J. D., YUE Y. G.: Single Cell Explorer, collaboration-driven tools to leverage large-scale single cell RNA-seq data. *BMC Genomics 20*, 1 (2019), article no. 676. doi:10.1186/s12864-019-6053-y. 1

[KBJ*20] KRUEGER R., BEYER J., JANG W.-D., KIM N. W., SOKOLOV A., SORGER P. K., PFISTER H.: Facetto: Combining unsupervised and supervised learning for hierarchical phenotype analysis in multi-channel image data. *IEEE Trans. Visual Comput. Graphics 26*, 1 (2020), 227–237. doi:10.1109/tvcg.2019.2934547. 1

[KGM*21] KELLER M. S., GOLD I., MCCALLUM C., MANZ T., KHARCHENKO P. V., GEHLENBORG N.: Vitessce: Integrative visualization of multimodal and spatially-resolved single-cell data. *OSF Preprints* (2021). doi:10.31219/osf.io/y8thv. 1

[LMC23] LONG B., MILLER J., CONSORTIUM S.: SpaceTx: A roadmap for benchmarking spatial transcriptomics exploration of the brain. *Arxiv* (2023). doi:10.48550/arXiv.2301.08436. 1

[LTK*23] LI C., THIJSSEN J., KROES T., DE BOER M., ABDELAAL T., HÖLLT T., LELIEVELDT B.: SpaceWalker enables interactive gradient exploration for spatial transcriptomics data. *Cell Rep Methods* (2023). doi:10.1016/j.crmeth.2023.100645. 1

[LTK*25] LI C., THIJSSEN J., KROES T., VAN DER BURG X., VAN DER WEERD L., HÖLLT T., LELIEVELDT B.: GeneSurfer enables transcriptome-wide exploration and annotation of gene co-expression modules in 3d spatial transcriptomics data. *iScience* (2025), 112713. doi:10.1016/j.isci.2025.112713. 1

[MMDP10] MEYER M., MUNZNER T., DEPACE A., PFISTER H.: MulteeSum: A tool for comparative spatial and temporal gene expression data. *IEEE Trans. Visual Comput. Graphics 16*, 6 (2010), 908–917. doi:10.1109/tvcg.2010.137. 1

[Pat18] PATEL M. V.: iS-CellR: A user-friendly tool for analyzing and visualizing single-cell RNA sequencing data. *Bioinformatics 34*, 24 (2018), 4305–4306. doi:10.1093/bioinformatics/bty517. 1

[SVUK*21] SOMARAKIS A., VAN UNEN V., KONING F., LELIEVELDT B., HOLLT T.: ImaCytE: Visual exploration of cellular microenvironments for imaging mass cytometry data. *IEEE Trans. Visual Comput. Graphics 27*, 1 (2021), 98–110. doi:10.1109/tvcg.2019.2931299. 1

[VKT*23] VIETH A., KROES T., THIJSSEN J., VAN LEW B., EGGERMONT J., BASU S., EISEMANN E., VILANOVA A., HÖLLT T., LELIEVELDT B.: ManiVault: A flexible and extensible visual analytics framework for high-dimensional data. *IEEE Trans. Visual Comput. Graphics 30*, 1 (2023), 175–185. doi:10.1109/tvcg.2023.3326582. 2