Universiteit Leiden
The Netherlands

Comparative visual analytics for multi-modal single-cell and spatial omics data

Basu, S.

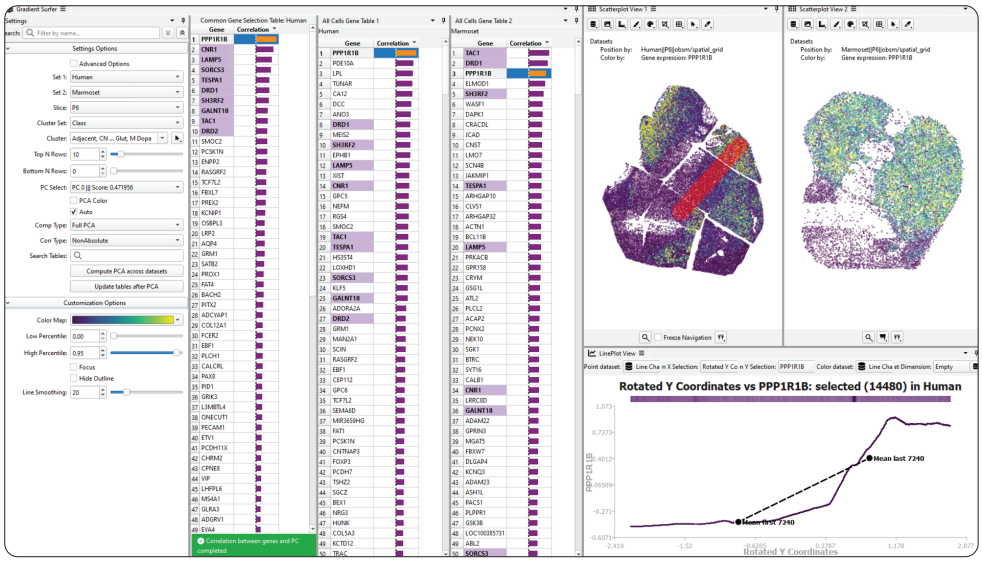
Citation
Basu, S. (2026, September 15). *Comparative visual analytics for multi-modal single-cell and spatial omics data*. Retrieved from https://hdl.handle.net/1887/4310004

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4310004

Note: To cite this publication please use the final published version (if applicable).

Chapter 4

Cytosplore Gradient Surfer



This chapter was adapted from the yet to be published paper:

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

Abstract

Spatial transcriptomics measures gene expression in single cells while preserving tissue structure, making gene expression visible within the tissue. This enables the study of how cells are organized, how they develop, and how genes are regulated within their spatial context. Several experimental platforms are now available, including 10x Genomics Xenium and Visium, Slide-seq, MERFISH, and seqFISH. These differ in resolution and sensitivity, from spot-based capture methods to single-molecule imaging. Computational tools have been developed to detect spatially variable genes (e.g., SpatialDE [79], SPARK [80]), to define spatial domains, and to link spatial data with scRNA-seq to map cell types. However, comparing spatial patterns across samples, conditions, or technologies remains a challenge. Differences in resolution, gene panel, platform-specific biases, and variability in tissue processing all create barriers. We present Cytosplore Gradient Surfer, an interactive system for identifying and analyzing spatial gene expression gradients across spatial transcriptomics datasets with partially overlapping gene panels. The system enables users to interactively explore spatial directions of interest and identify genes exhibiting aligned expression patterns across multiple datasets, even when only a subset of genes is shared between them.

4.1 Introduction

Spatial transcriptomics technologies extend the scope of transcriptomics by measuring gene expression variation within intact tissue contexts. In contrast to single-cell RNA sequencing (scRNA-seq), which dissociates cells and loses spatial information, spatial approaches preserve tissue organization and thus capture molecular gradients that are essential for understanding tissue architecture, developmental processes, and gene regulation. Molecular gradients represent continuous changes in gene expression across tissue regions. By revealing how gene expression varies across space, spatial methods allow researchers to study both fine-grained local neighborhoods of cells and large-scale structural patterns within tissues.

A key next step in this field is the comparative analysis of gradients across datasets [81]. Such comparisons enable researchers to identify conserved or divergent gene expression patterns across biological replicates, experimental conditions, or measurement modalities [82]. For instance, comparing gradients across imaging-based techniques such as MERFISH and smFISH can highlight consistent or complementary spatial organization of gene expression [83]. Similarly, aligning gradients across replicate tissue samples provides opportunities to study recurrent features of tissue structure or biological states [84]. However, these analyses remain difficult due

to variability in tissue morphology, cell arrangements, gene expression levels, and heterogeneity in measured gene sets. Differences in capture resolution and platform-specific biases further complicate alignment and interpretation.

While several computational tools exist for cross-dataset comparison of single-cell transcriptomics data [68, 85], they are not designed to handle comparisons of spatial gene expression gradients. Conversely, many tools have been developed for the exploration of spatial transcriptomics data, including methods for neighborhood-based analysis [46, 48, 47], identification of spatial domains, and detection of macroscopic expression gradients [86, 81]. Yet these approaches typically operate within a single dataset and do not address the problem of comparing gradients across multiple datasets. Existing single-cell methods that study temporal dynamics [40] or rely on dimensionality reduction [42, 43] also fall short when applied to spatial data integration. To address these limitations, we present Cytosplore Gradient Surfer, an interactive system for identifying, analyzing, and comparing spatial gene expression gradients across datasets. The system allows users to explore spatial directions of interest, detect genes exhibiting aligned gene expression patterns across samples, and perform comparative analysis even when only a subset of genes is shared between datasets.

4.2 Results

We developed a workflow to analyze and compare spatial gene expression gradients across two datasets. Here, we define a spatial gene expression gradient as a systematic and monotonic change of gene expression along a given spatial direction in the tissue.

Starting from a reference dataset, we select cells along a line and identify gradient genes whose expression correlates with position along that line. The top N common genes between datasets are then used as input for PCA, into which all datasets are projected using the same gene set. This enables comparison of gradients without requiring identical cell and gene distributions across datasets, assuming a shared underlying gradient signal.

The Cytosplore Gradient Surfer system is built as a plugin for ManiVault Studio [51], an open-source application building platform for visual analytics applications. It is integrated with other ManiVault plugins to extend functionality. The system enables users to interactively probe for, extract, and align gradient-based gene expression features to identify spatial patterns in tissues. From a single-line swipe in one dataset, it identifies conserved or divergent molecular patterns associated with tissue structure or biological function across datasets. The system takes as input a cell-by-gene expression matrix and corresponding spatial coordinates to support the

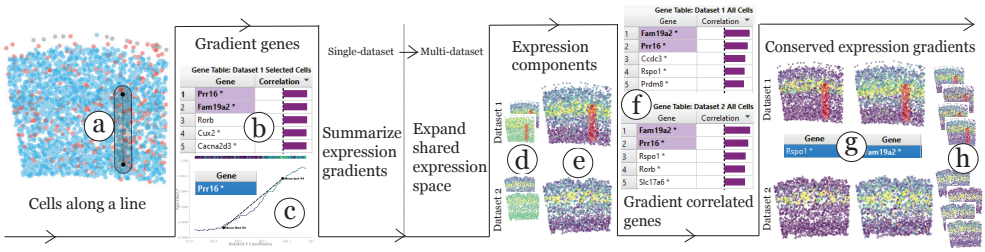


Figure 4.1: **Cytosplore Gradient Surfer workflow** to explore spatial gene expression gradients. Line brush to select cells (a). Genes with spatial gradients identified by correlation with the selected line profile (b). A line chart to visualize gene expression along the line selection (c). PCA loadings 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) .

analysis. Biologists often have hypotheses or prior knowledge about gradients in tissues and aim to identify genes that reflect these patterns.

To support exploration of gradients, we provide a line brush tool. The brush allows interactive selection of cells by quickly dragging a line in the spatial domain (Figure 4.1a). The cells with a user-defined distance from this line are then projected onto the drawn line, and, for each gene, the Pearson correlation is computed between gene expression levels and the projected cell position along the line. This reveals genes whose expression consistently varies with spatial position (Figure 4.1b), reflecting an expression gradient. To visually examine these spatial gradient patterns, the system displays the spatial layout of cells in scatterplots, colored by expression level, and a line chart (Figure 4.1c) 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 further analysis (Figure 4.1b). 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 highly correlates with the dominant spatial pattern. PCs can be visualized as color overlays on spatial scatterplots (Figure 4.1d), similar to gene expression, to verify which PC represents the probed gradient best (Figure 4.1e). Using the results of the PCA, we project the second dataset to identify genes in the target dataset that exhibit similar gradient profiles. Note that we aim to avoid having to draw a cell selection swipe in the target dataset. To achieve this, for each dataset, we compute the Pearson correlation coefficient between each gene's expression and the selected gradient PC, providing a quantitative measure of its alignment with the dominant spatial trend. The results are presented as tables for each dataset (Figure 4.1f), 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 (Figure 4.1g). The rank of the genes can then be compared across datasets in the tables. By overlaying gene expression as color in the side-by-side scatterplots, the user can evaluate if new genes follow similar patterns and if these patterns are conserved or altered (Figure 4.1h).

To evaluate the performance of our method, we analyzed five publicly available spatial transcriptomics datasets: two mouse visual cortex datasets acquired through the SpaceTx consortium [86], human and macaque brain samples [87] from the Vizgen MERSCOPE platform, and marmoset samples [87] from the 10x Genomics Xenium platform.

4.2.1 Interactive probing of spatial directions reveals gene expression gradients

We first assessed whether interactive probing of spatial directions can recover known, biologically meaningful gene expression gradients in spatial transcriptomic datasets. Users define spatial directions corresponding to anatomical axes, such as vertical (cortical depth), horizontal (laminar spread), or oblique orientations aligned with local cytoarchitecture. In particular, we hypothesized that probing directions aligned with known anatomical axes would reveal gene expression gradients corresponding to established tissue organization. We performed two experiments to identify whether we obtain meaningful results across datasets of increasing resolutions.

To test the hypothesis that Gradient Surfer can recover biologically meaningful gene expression gradients from spatial transcriptomic data, we analyzed two datasets from the mouse visual cortex acquired using different experimental modalities. The mouse visual cortex exhibits a well-defined laminar organization along the pial-to-white matter (radial) axis, and the expression of many genes is strongly correlated with cortical depth [88]. This known anatomical and transcriptional structure may provide a suitable reference for evaluating gradient-based gene ranking methods.

We first tested this hypothesis using the smFISH mouse visual cortex dataset [86], consisting of 2360 cells with expression measurements for 22 genes. To explicitly probe the cortical laminar structure, we defined a linear selection oriented perpendicular to the cortical layers, corresponding to the radial axis extending from deep to superficial regions of the cortex. This selection was designed to capture the dominant transcriptional gradient associated with cortical layering.

Applying Gradient Surfer to this selection, we ranked genes by their correlation with position along the radial axis. Several of the top-ranked genes (Figure 4.2a) that have a correlation coefficient greater than 0.35 corresponded to well-established laminar markers reported in the literature [89]. In particular, *Cux2*, which is strongly

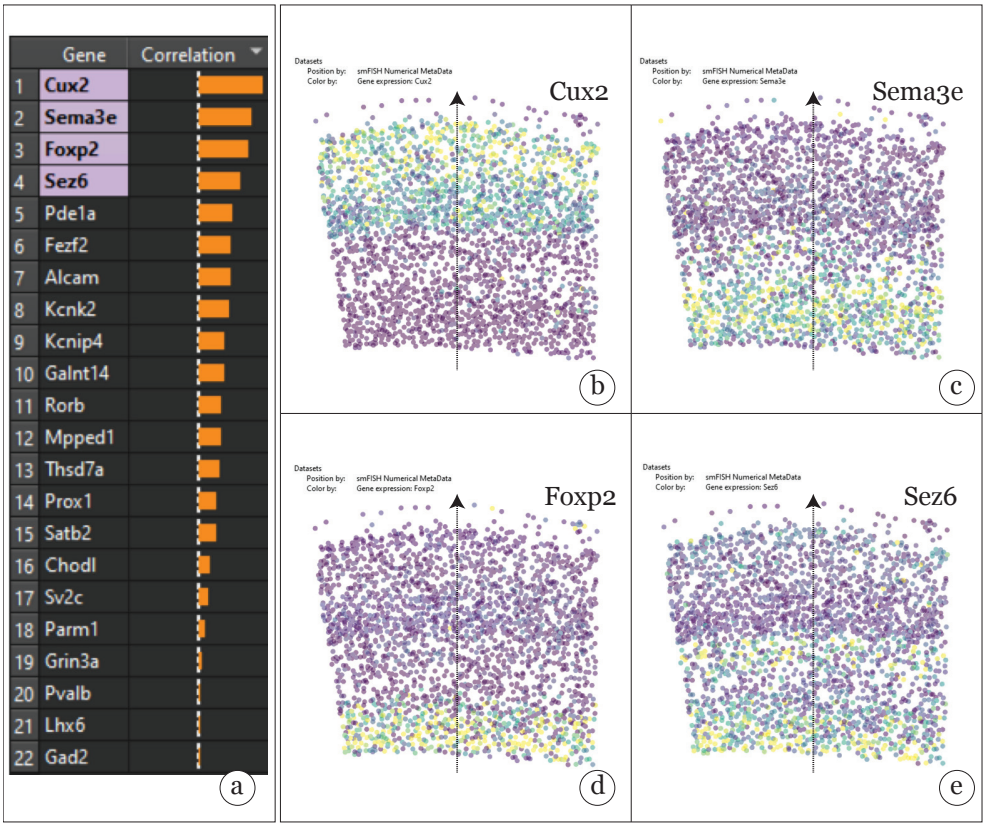


Figure 4.2: **Interactive spatial probing reveals robust laminar gene expression gradients in the SMFISH mouse visual cortex.** Gradient detection across the smFISH mouse visual cortex dataset (22 genes) reveals laminar structure in spatial gene expression. Representative genes (a) with a correlation coefficient greater than 0.35, including *Cux2*, *Sema3e*, *Foxp2*, and *Sez6*, display pronounced layer-specific gradients consistent with previously reported laminar expression profiles in the mouse visual cortex (b-e).

expressed in Layers 2/3 [90] (Figure 4.2b), *Sema3e*, highly specific to Layer 5b/6 [91] (Figure 4.2c), *Foxp2*, marker for Layer 6 [92] (Figure 4.2d), and *Sez6*, found in Layers 5 and 6 [93] (Figure 4.2e) exhibited spatial expression patterns consistent with their known layer-specific localization in the mouse visual cortex, supporting the ability of Gradient Surfer to recover biologically relevant spatial gradients.

We next tested whether the same hypothesis holds across spatial transcriptomic modalities by applying the Gradient Surfer to a MERFISH dataset from the same tissue type [86]. This dataset comprised 2150 cells and 253 genes, providing substantially higher genomic coverage than the previous SMFISH dataset.

Using a comparable linear selection aligned with the radial (white matter-to-pial) axis, Gradient Surfer identified genes whose expression varied systematically with

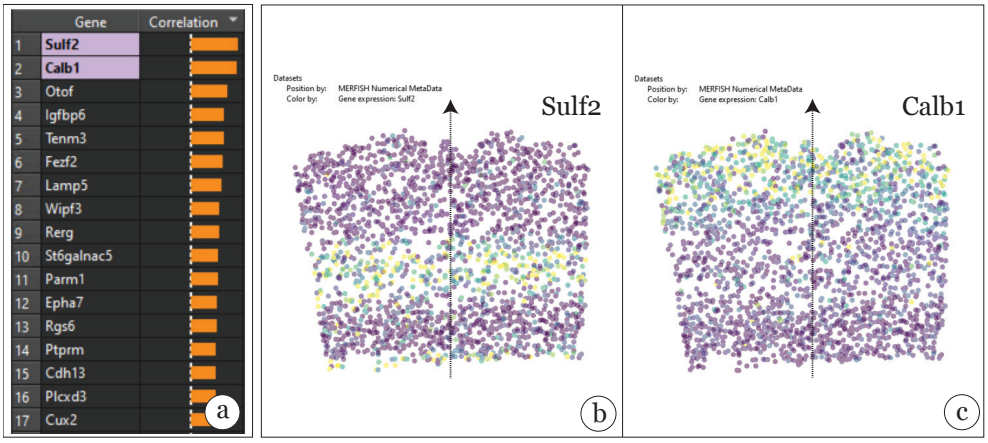


Figure 4.3: **Application of spatial-direction probing to the MERFISH dataset (253 genes) identifies strong layer-specific expression gradients.** Gradient detection highlights pronounced laminar organization across the dataset. Highest-ranked genes (a) with correlation scores greater than 0.35, including *Sulf2* and *Calb1*, exhibit layer-enriched expression patterns consistent with established cortical laminar markers (b-c).

cortical depth. Among the highest-ranked genes with correlation scores greater than 0.35 were well-known laminar markers, such as *Sulf2*, a definitive marker for the deepest cortical layers and subplate [94] (Figure 4.3b), and *Calb1*, which characterizes superficial pyramidal neurons in the mouse visual cortex [95] (Figure 4.3c).

In this first experiment, both datasets contain relatively small gene panels (22 and 253 genes). To test whether our gradient-gene detection method also scales well to transcriptomic datasets with larger gene panels and number of cells, we analyzed an imputed MERFISH mouse visual cortex dataset from the Allen Brain Cell Atlas [96], containing about 8,460 imputed genes and 120,186 cells.

Previous studies [97, 98] have shown that the thalamus is organized by continuous gene expression gradients rather than sharply defined cell boundaries. To test whether Gradient Surfer can recover such gradients in a dataset with substantially higher gene coverage than the cortical smFISH and MERFISH datasets, we examined a linear selection drawn across the thalamus in slice 38 of the dataset.

Applying Gradient Surfer to this selection, we ranked genes by their correlation with position along the selected axis. Among the top-ranked genes were *Maob* [99] (Figure 4.4b) and *Snca* (Figure 4.4c), which are involved in monoaminergic regulation and presynaptic signaling, *Grm2* (Figure 4.4d) and *Stx1a* [100] (Figure 4.4e), which are involved in synaptic transmission, and regionally enriched genes such as *Pced1b* (Figure 4.4f), *Oprm1* [101] (Figure 4.4g), *Sntg2* [102] (Figure 4.4h),

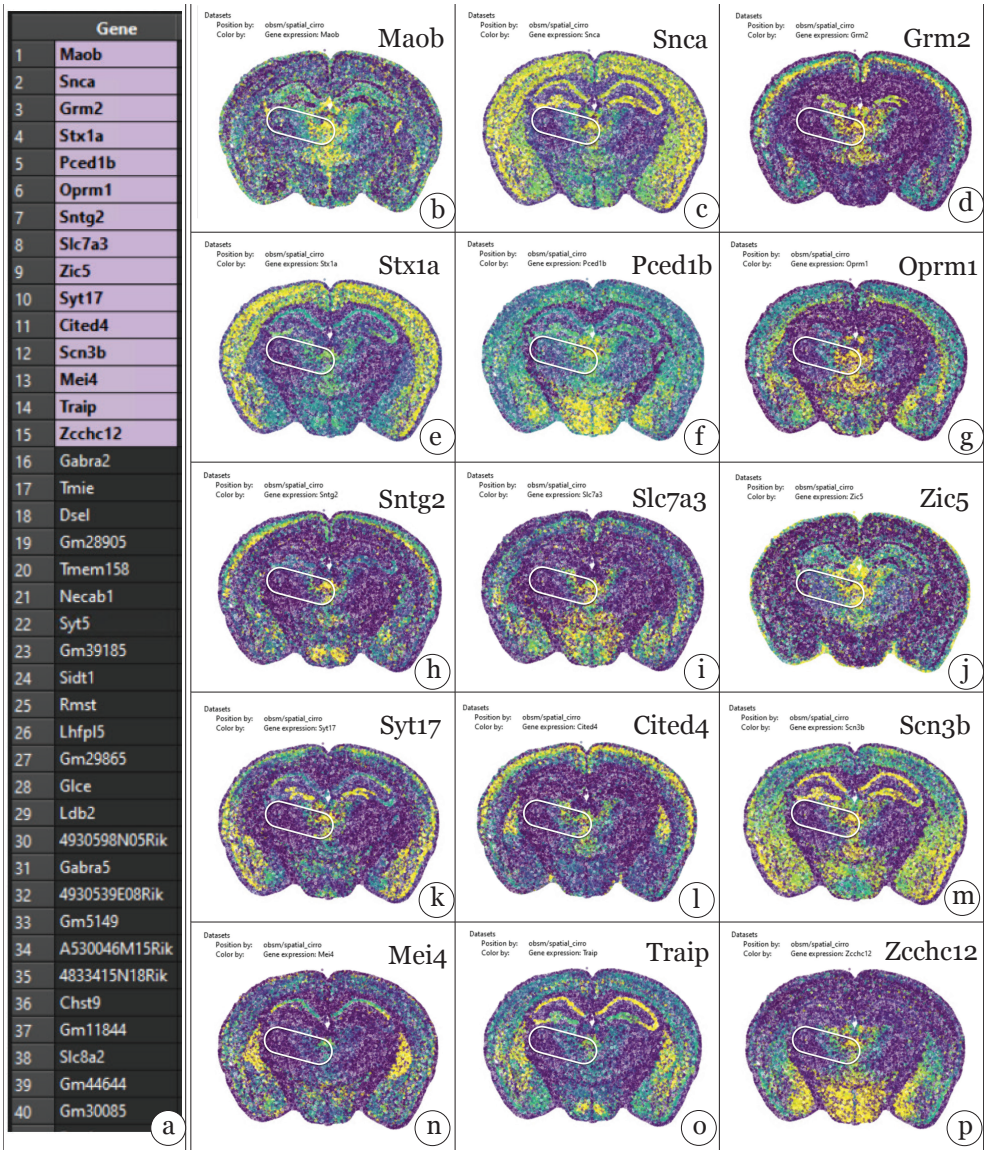


Figure 4.4: **Gradient-based spatial probing applied to an imputed MERFISH dataset (8,460 genes; 120,186 cells) from the Allen Brain Cell Atlas, a line across the thalamus slice 38 identified genes with strong spatial gradients.** Spatial gradient genes across the selection (a). Top gradient genes show pronounced spatial variation consistent with known thalamic organization, including *Maob*, *Snca*, *Grm2*, *Stx1a*, *Pced1b*, *Oprm1*, *Sntg2*, *Slc7a3*, *Zic5*, *Syt17*, and others (b-p).

showed gradient patterns consistent with their reported spatial distributions in the thalamus. Other genes that appeared in the top 15 are *Slc7a3* (Figure 4.4i), *Zic5* (Figure 4.4j), *Syt17* (Figure 4.4k), *Cited4* (Figure 4.4l), *Scn3b* (Figure 4.4m), *Mei4* (Figure 4.4n), *Traip* (Figure 4.4o), and *Zcchc12* (Figure 4.4p). This confirms that the method successfully recovers known, biologically meaningful gene expression gradients even in large-scale datasets.

Together, these results demonstrate that interactive spatial probing in Gradient Surfer can identify biologically meaningful gradients corresponding to known tissue architecture.

4.2.2 Cross-dataset projection is rotationally invariant and highly robust against protocol, cell subsampling, and feature subsampling

Next, we investigated whether the method is robust to technical variations that are expected across diverse experimental settings and spatial transcriptomic (ST) protocols (see Supplemental S1.1). We aimed to assess robustness against ST protocol, sample rotation (see Supplemental S1.1.1), cell sampling density (see Supplemental S1.1.2), and feature set reduction (see Supplemental S1.1.3).

When we rotated one dataset while keeping the other fixed, the highest-ranking gradient genes remained the same, and their spatial patterns simply rotated with the data. This shows that the approach is inherently rotationally invariant and can detect aligned structure regardless of dataset orientation (see Supplemental S1.1.1).

To test robustness to differences in spatial sampling density, we progressively removed fractions of spatial cell locations from one dataset and repeated the analysis. Gene rankings remained highly stable even when 90 percent of cell locations were removed, with only small shifts in rank and strong agreement across all correlation measures. Key gradient genes consistently stayed within the top ranks, indicating that the method is resilient to substantial loss of spatial coverage (see Supplemental S1.1.2).

We also assessed how well the method performs when the two datasets share only a subset of genes. After removing large portions of the feature space using either random or rank-based strategies, all evaluated genes preserved their relative positions within the top set of gradient-associated genes. This demonstrates that the approach remains reproducible even when datasets differ considerably in gene coverage (see Supplemental S1.1.3).

Together, these results show that the method is robust to geometric transformations, variation in spatial cell sampling coverage, and reductions in shared gene space, supporting its use across diverse experimental protocols and dataset configurations.

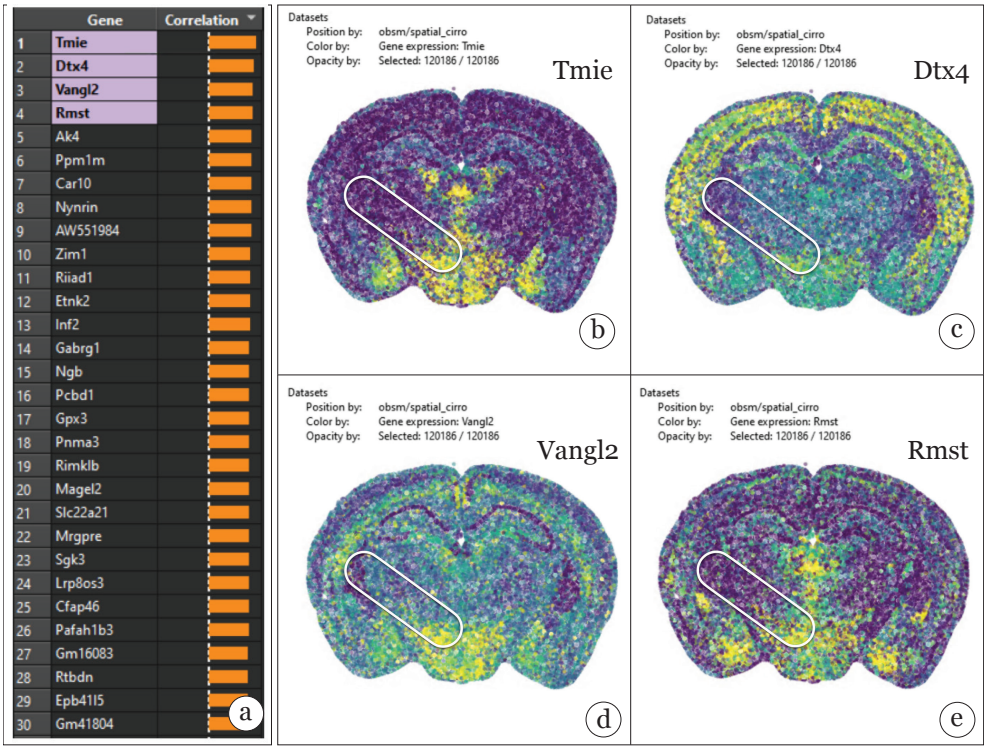


Figure 4.5: **Spatial gradients across all cells along a horizontal line through thalamus, hypothalamus, and striatum (slice 38).** Analysis of 8,460 imputed genes identifies broadly expressed genes dominating the trajectory. The horizontal line captures signals from gradient and non-gradient cells. Top-ranked all-cell genes (a). Examples (*Tmie*, *Dtx4*, *Vangl2*, *Rmst*) show broad gradients across multiple cell types, not specific to excitatory neurons (b-e).

4.2.3 Cell-type-specific gradient decomposition reveals intra-type spatial variation

Spatial gradient analysis of a tissue region using all cells inherently combines expression signals from multiple cell types, which can mask spatial gradients occurring within individual cell populations.

To examine intra-type variation, we focused our analysis by extending our scope of the selection by including the thalamus, hypothalamus, and striatum in slice 38 of the ABC atlas MERFISH dataset, with an imputed panel of 8,460 genes. The striatum and thalamus are ideal for this test because the striatum is primarily composed of resident GABAergic neurons [103, 104], and the thalamus contains a population of glutamatergic cells [105, 106]. In addition, the hypothalamus provides a highly

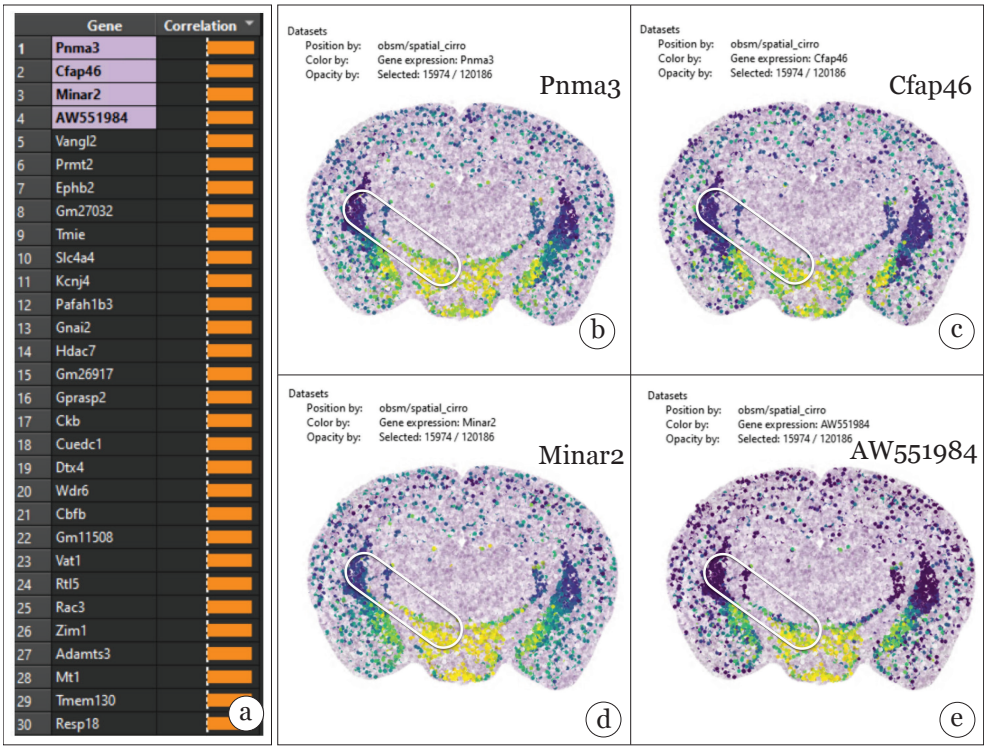


Figure 4.6: **Cell-type-specific gradients in GABAergic neurons.** Analysis of GABAergic neurons along the horizontal trajectory reveals striatum-specific gradients masked in all-cell data. Top-ranked GABA-cell genes (a). Examples (*Pnma3*, *Cfap46*, *Minar2*, *AW551984*) show intra-type spatial patterns (b-e). Genes dominant in bulk lose significance (Table 4.1), showing GABAergic isolation is needed to resolve striatal molecular architecture.

heterogeneous environment containing a mixed population of both types [107, 108].

We defined a linear selection across these structures to maximize the chance of detecting differential gradients between these two populations. Using this selection, we first performed gradient analysis across all cell types (Figure 4.5). The top gradient genes (Figure 4.5a), *Tmie* (Figure 4.5b), *Dtx4* (Figure 4.5c), *Vangl2* (Figure 4.5d), and *Rmst* (Figure 4.5e), exhibited broad spatial gradients across multiple cell types. This confirms that bulk analysis highlights genes that dominate the overall trajectory but fails to resolve cell-type-specific patterns.

Next, we restricted the analysis to GABAergic neurons within the same swipe (Figure 4.6), along the same horizontal spatial trajectory. Because GABAergic neurons are the primary constituents of the striatum in this selection, this restriction effectively isolates striatum-specific gradient genes. This cell-type-specific approach

Gene ID	All-Cell Type Rank	GABA (Striatum) Rank	Glut (Thalamus) Rank
Tmie	1	9	65
Dtx4	2	19	1552
Vangl2	3	5	635
Rmst	4	48	78
Pnma3	18	1	62
Cfap46	25	2	1028
Minar2	82	3	540
AW551984	9	4	1097
Nrtn	5445	6351	1
Kcnab3	7448	7177	2
Pvalb	6705	6927	3
Car2	1014	3404	4

Table 4.1: **Comparative Gene Rank Shifts by Cell Types.** This table tracks how gene rankings differ between all-cell analysis and specific cell type analysis within the same swipe.

revealed a completely different gene ranking: new GABA-specific top gradient genes were identified (Figure 4.6a), including *Pnma3* (Figure 4.6b), *Cfap46* (Figure 4.6c), *Minar2* (Figure 4.6d), and *AW551984* (Figure 4.6e), which displayed coherent spatial expression patterns specific to this population.

Table 4.1 compares gene gradient ranks computed from all cells in the swipe area against gene ranks within specific cell types in the same swipe. For GABA-ergic neurons, the original all-cell top gradient genes ranked substantially lower, whereas low-ranked genes like *Pnma3* rose from rank 18 in the bulk analysis to rank 1 in the GABA (Striatum) specific analysis.

Finally, we restricted the analysis to glutamatergic (Glut) neurons (Figure 4.7) along the same trajectory, which effectively isolates thalamus-specific gradient genes. In this population, we identified Glut-specific top gradient genes (Figure 4.7a) such as *Nrtn* (Figure 4.7b), which became the top-ranked gene, while other Glut-specific genes like *Kcnab3* (Figure 4.7c), *Pvalb* (Figure 4.7d), and *Car2* (Figure 4.7e) emerged.

Table 4.1 highlights the necessity of this decomposition: *Nrtn*, which was buried at rank 5,445 in the all-cell analysis, ascends to Rank 1 in the Glutamatergic population. Markers specific to the opposing GABAergic population (such as *Pnma3*) drop significantly in rank. These results demonstrate that cell-type-specific decomposition is essential to uncover the distinct spatial logic of the striatum and thalamus.

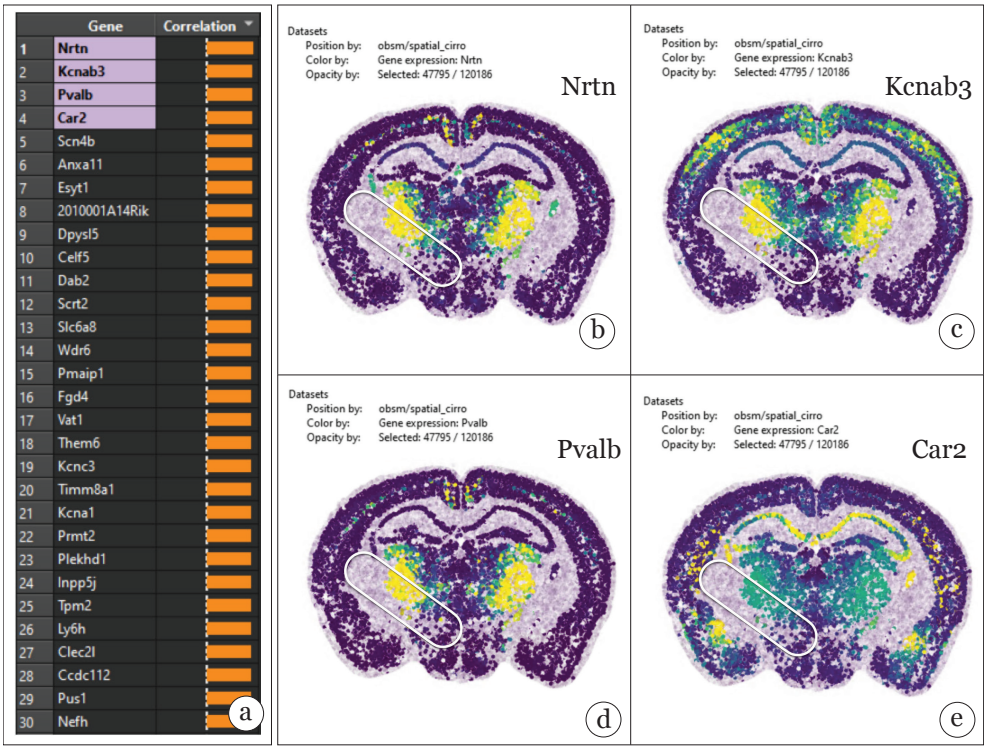


Figure 4.7: **Cell-type-specific gradients in glutamatergic neurons.** Focusing on Glut neurons reveals thalamus-specific spatial gradients. Top-ranked Glut-cell genes (a). Examples (*Nrtn*, *Kcnab3*, *Pvalb*, *Car2*) show spatial patterns defining thalamic nuclei (b-e). *Nrtn* rises to Rank 1 (Table 4.1), showing that cell-type-specific analysis uncovers intra-type gradients masked in multi-cell analysis.

4.2.4 Cross-dataset projection and correlation analysis reveal conserved and divergent spatial gene expression gradients

Next, we investigated whether projecting cells from different species onto a common PCA basis reveals conserved or divergent gene expression gradients. This analysis utilizes data from a recent cross-species spatial atlas of the primate basal ganglia [87], which identified these gradients through quantitative cross-dataset projection and correlation analyses. While the original study relied on script-based methods that required researchers to manually define gradient directions and analyze full datasets for each species individually, we used the Gradient Surfer plugin to interrogate these patterns through interactive exploration.

The Cytosplore Gradient Surfer system provides several methodological advantages for comparative spatial analysis. Instead of probing species individually, our system allows a user to perform a "line-swipe" on a single species to automatically

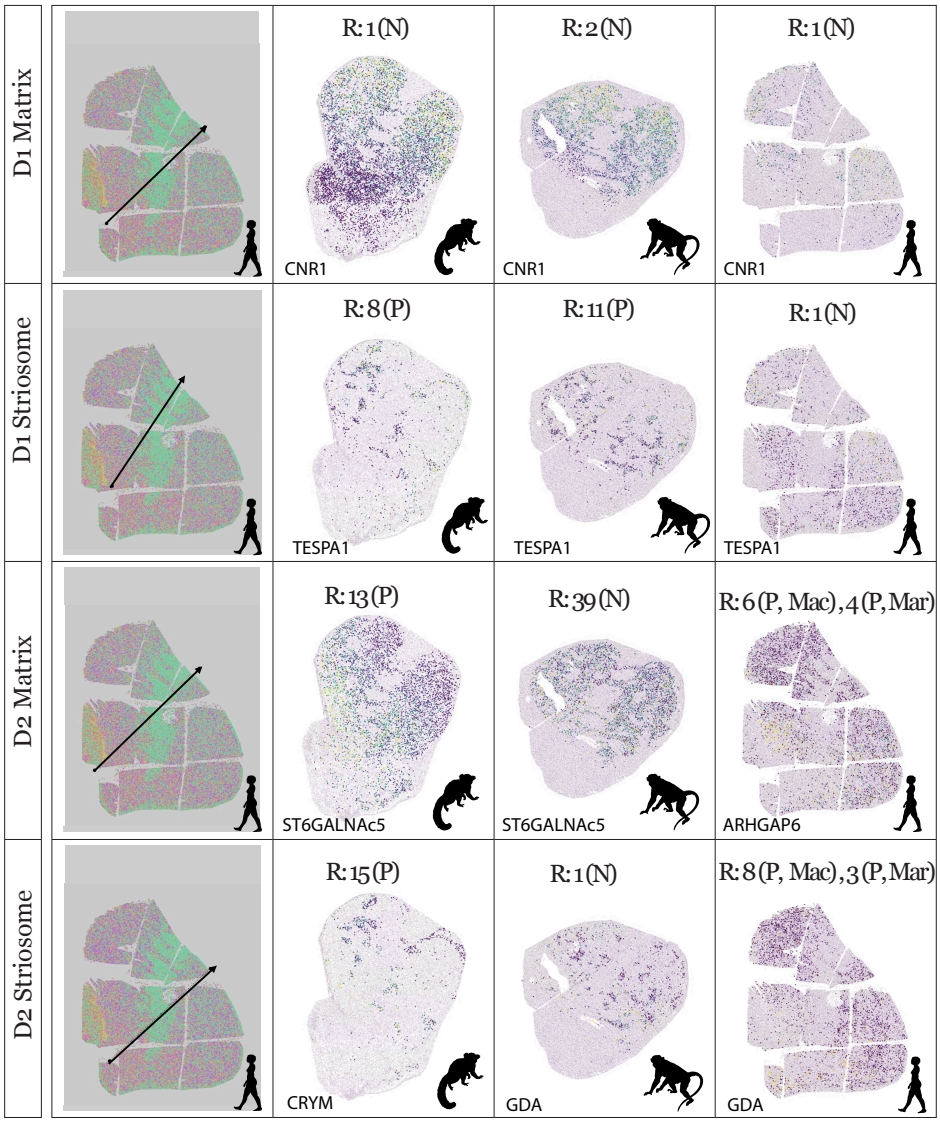


Figure 4.8: **Validation of conserved spatial gene expression gradients across primate species.** Spatial expression maps of representative gradient-associated genes identified by interactive line-swipe analysis across human, macaque, and marmoset brain slices (P7, P9, and P13, respectively), with the line swipe performed on the human slice. In the figure, R denotes rank, P denotes positive expression rank, and N denotes negative expression rank. The figure demonstrates the tool's ability to identify gradient genes in multiple species by swiping on one species. For instance, *CNR1* and *TESPA1* show consistent expression gradients across all three species.

retrieve the top gradient-associated genes across all species through continuous pairwise comparison. Furthermore, the plugin enables the filtering of specific cell types, allowing for focused analysis of populations such as dorsal striatal medium spiny neurons (MSNs). This regional and cell type-specific swiping eliminates the need to process the entire dataset individually for every species.

Despite these methodological differences, our interactive approach successfully recovered the same key gradient-associated genes that are highlighted in the original manuscript. We performed line-swipe analyses across anatomically matched regions in the dorsal striatum (STRd) and ranked genes by the strength of their association with the visualized gradients. These results provide independent support for both the conserved and species-biased spatial patterns identified in the atlas.

For example, *CNR1* is a canonical marker of the dorsolateral to ventromedial striatal gradient. In our analysis, it ranked as the top gradient-associated gene in humans and was consistently prominent across both D1 and D2 MSNs in the human, macaque, and marmoset (Figure 4.8). This is consistent with the manuscript's conclusion that the CNR1 gradient is highly conserved across primates. Similarly, *GDA*, which covaries with the principal dorsal striatal gradient, emerged as one of the most strongly associated genes in our analysis and appeared within the top-ranked genes across all three species.

Our analysis also revealed species-specific expression patterns consistent with the study's main findings. Genes with species-restricted spatial organization showed species-dependent gradient associations, confirming divergent molecular patterns. For instance, *CRYM* was described as a marmoset-specific gradient gene. It was ranked first in our marmoset line-swipe analyses and appeared among the top gradient genes for marmoset MSNs, but it was unavailable in the human and macaque panels used for this comparison.

These results show that Cytosplore Gradient Surfer successfully identifies both conserved and divergent spatial gene expression patterns across species, matching findings from detailed quantitative studies. This confirms the tool's value for rapid visualization, cell-type-specific analysis, and hypothesis-building in spatial transcriptomics.

4.3 Discussion

Spatial gene expression gradients are fundamental to tissue patterning, cell differentiation, and functional specialization. They are critical for understanding both healthy development and disease progression. However, most current spatial transcriptomics pipelines require separate, manual analyses for each species or tissue section. This

often limits the feasibility of large-scale cross-species comparisons and can obstruct the discovery of conserved spatial organization.

We present Cytosplore Gradient Surfer, an interactive system for comparing spatial gene expression gradients across datasets. Rather than relying solely on static, script-based workflows, our approach emphasizes hypothesis-driven analysis through direct interaction and a visual and quantitative feedback loop for interactive analysis. This system was developed in collaboration with domain experts to address the specific needs of researchers working with multiple concurrent datasets.

Gradient Surfer has three novel capabilities compared to existing approaches. First, interactive probing of spatial gradients across tissue architectures enables users to dynamically refine spatial selections, parameters, and gene sets during analysis. This enables on-the-fly comparative investigation. This capability is especially valuable when biological knowledge is uneven across datasets. Gradients identified in a well-characterized reference, such as a well-characterized single-species reference dataset, can guide exploratory analysis in less-annotated datasets from other species. This facilitates the discovery of conserved or divergent spatial patterns missed in static analyses. Second, cell-type-aware gradient analysis by filtering within defined cell populations enables users to examine gradient dynamics at a much finer cellular resolution. This reveals subtle intracluster variations that are often obscured in broader, tissue-level gradient analyses. For instance, in our validation using the primate striatum data from Hewitt et al. [87], we were able to isolate and analyze gradients specifically within Medium Spiny Neuron (MSN) subtypes, such as D1 and D2 Matrix or Striosome cells. Finally, we introduce a swipe-to-identify comparative workflow that advances traditional methods. In previous studies, such as the work by Hewitt et al. [87], researchers had to individually probe each species, manually defining angles to collect cells for analysis across the full dataset. In contrast, Gradient Surfer allows a user to define a spatial axis on a single reference slice and immediately retrieve the corresponding gene expression rankings across all other species through continuous pairwise comparison. This eliminates the need for individual probing and allows established biological knowledge in one organism to inform exploration in another.

We validated these novel functionalities of Gradient Surfer on several publicly available datasets, demonstrating retrieval of biologically meaningful known spatial gradients under varying Spatially Resolved Transcriptomics protocols, cell densities, and feature samplings. We also corroborated several findings from a recent cross-species gradient comparison study on marmoset, macaque, and human spatial transcriptomics of the basal ganglia. Overall, these experiments show that Gradient Surfer offers an interactive visual interface for comparing spatial gradients across species and datasets.

4.4 Limitations and Future Work

While the presented system enables efficient interactive comparison, the current implementation has some limitations. First, gradient comparison is currently optimized for pairwise exploration between two datasets at a time. While this provides a clear and interpretable interface, broader multi-species studies still require an iterative process. Second, the system currently assumes a shared subset of genes across datasets for PC computation. If the comparison involves datasets with significantly different gene sets, noise may be introduced. We plan to resolve this by implementing methods that go beyond pairwise restrictions. This will allow for multi-dataset comparisons without limiting computation to the intersection of genes between just two datasets.

In addition, the current version does not perform formal cross-modal spatial alignment, meaning that tissue shapes are not automatically warped or registered. To further improve analysis, we are exploring automated methods for identifying and aligning corresponding gradient directions in secondary datasets. This would allow the system to suggest candidate spatial axes for comparison automatically.

Analysis is based on a straight-line selection. We plan to extend this to support curved lines to better accommodate complex, folded tissue structures. Furthermore, the system does not yet explicitly distinguish between sharp boundary-driven gradients and smoothly varying intra-domain gradients. Looking ahead, we aim to further differentiate between these types of gradients. Future work will evaluate parameter stability with respect to the number of genes used for PC computation and the width of the selection swipe, ensuring robustness across diverse biological contexts.