



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

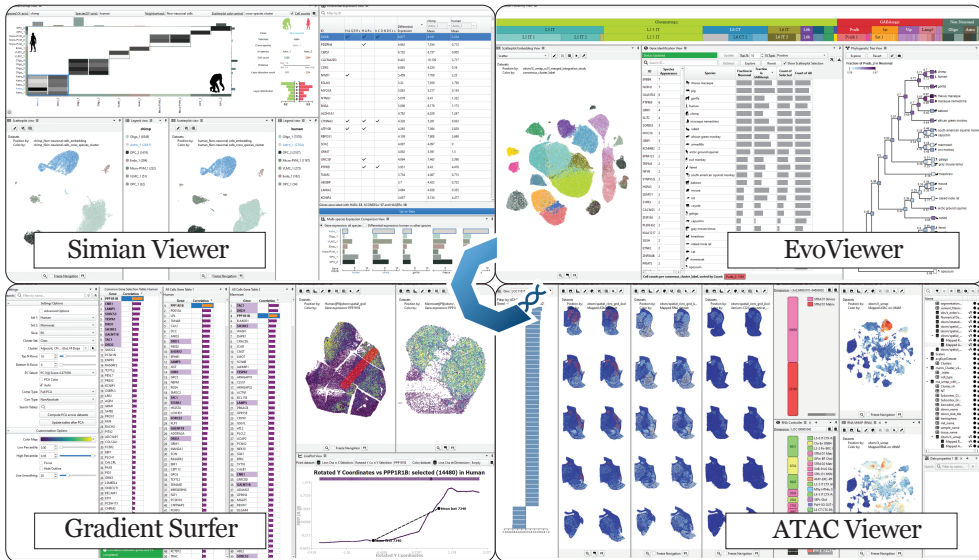
License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4310004>

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

CHAPTER 5

CYTOSPLORE VIEWER



This chapter summarizes the visualization contributions from the published/preprint papers:

N. L. Jorstad, ..., **S. Basu**, ..., and T. Bakken. *Comparative transcriptomics reveals human-specific cortical features*. Science, 382(6667), ade9516, 2023.
doi: [10.1126/science.ade9516](https://doi.org/10.1126/science.ade9516)

M. N. Hewitt, ..., **S. Basu**, ..., and S. C. Seeman. *A cross-species spatial transcriptomic atlas of the human and non-human primate basal ganglia*. bioRxiv, 2025–11, 2025.
doi: [10.1101/2025.11.22.688128](https://doi.org/10.1101/2025.11.22.688128)

S. Dan, ..., **S. Basu**, ..., and F. M. Krienen. *Spatial patterning of transcriptional and regulatory programs in the primate subcortex*. bioRxiv, 2025–11, 2025.
doi: [10.1101/2025.11.22.689869](https://doi.org/10.1101/2025.11.22.689869)



5.1 Introduction

The expansion of single-nucleus and spatial multi-omics technologies has generated extensive brain informatics resources spanning transcriptomic, epigenomic, and spatial modalities across species. Developments in single-nucleus RNA sequencing (snRNA-seq) and spatial transcriptomics have significantly expanded the characterization of brain architecture. Large-scale consortia, including the Brain Initiative Cell Census Network (BICCN⁷) and the BRAIN Initiative Cell Atlas Network (BICAN⁸), have produced comprehensive cellular atlases of human and non-human primate (NHP) brains. While these datasets establish a foundation for comparative mammalian bioinformatics, integrating and systematically comparing such diverse resources poses significant technical and analytical challenges.

As discussed in [Chapter 2](#) and [Chapter 3](#), comparative analysis requires aligning multiple molecular modalities and heterogeneous metadata to support cross-species investigations. Integration is often hindered by inconsistencies in gene coverage, cell type nomenclature, spatial coordinate systems, and experimental technologies. At the data level, mapping cell type correspondences and evolutionary signatures, such as HARs, requires extensive harmonization of these heterogeneous datasets. At the analysis level, current computational frameworks often rely on static, batch-based workflows that present critical limitations. These include the lack of interactive tools for exploring integrated data, challenges in investigating multimodal relationships, and constraints in discovering unexpected biological patterns, such as continuous spatial gradients or rare cell populations.

Interactive visual analytics enable exploratory analysis by allowing researchers to navigate multimodal data, dynamically adjust parameters, and observe the immediate consequences of analytical choices. This exploratory paradigm is consistent with arguments that strictly hypothesis-driven approaches can constrain discovery, as discussed by Yanai and Lercher [9]. This approach is particularly important for comparative brain atlas analysis, where it supports the interactive validation of computational cell type alignments and the exploration of multimodal relationships without requiring commitment to predefined gene sets.

In response to these challenges, we built **Cytosplore Viewer**⁹, an integrated visual analytics system for the comparative, multi-modal exploration of brain atlases. Developed in collaboration with domain experts, this system provides a scalable and extensible framework for linking cell type hierarchies, gene regulatory networks, and spatial distributions.

⁷<https://biccn.org>

⁸<https://bican.org>

⁹<https://viewer.cytosplore.org>



Within the scope of this thesis, Cytosplore Viewer has served as the basis for several specialized modules, which in turn have supported discovery and communication in the corresponding biological research. Simian Viewer and EvoViewer, presented in [Chapter 2](#) and [Chapter 3](#), respectively, are modules for cross-species transcriptomic comparison and the interactive examination of gene expression conservation and divergence. Gradient Surfer, presented in [Chapter 4](#), is a module for spatial gene expression gradient alignment, enabling the comparison of non-congruent tissue samples. Finally, we developed the ATAC Viewer, a module that enables comparison of chromatin accessibility and gene expression by projecting average feature profiles onto spatial data.

The system focuses on methodological utility, visualization capability, and analytical accessibility. By providing a generalizable framework, the system facilitates hypothesis generation and the interactive examination of complex, high-dimensional, and cross-species datasets. In the following, we illustrate the system's utility through key findings from human and mammalian brain studies.

5.2 Comparative brain atlas analysis

Understanding brain organization and evolution requires systematic comparison across multiple axes. Large-scale brain informatics initiatives have generated extensive multi-modal datasets that enable comparative analysis along several dimensions: across species to identify evolutionary conservation and divergence, across anatomical regions to characterize functional specialization, across cell type hierarchies to map cellular correspondences, and across molecular modalities to link regulatory mechanisms with gene expression.

Cross-species comparison forms the foundation of evolutionary neurobiology. By aligning cell types and molecular profiles across primates (from marmosets and macaques to great apes and humans), researchers can identify lineage-specific adaptations and conserved developmental programs. This requires mapping orthologous genes, establishing consensus cell type taxonomies, and accounting for differences in brain morphology and experimental protocols across datasets.

Cross-region comparison addresses the anatomical organization of the brain. Comparing molecular profiles across cortical areas, subcortical structures, and the basal ganglia reveals regional specialization and continuous spatial gradients. However, anatomical boundaries are often gradual rather than discrete, and spatial coordinate systems vary between species, complicating direct alignment.

Cross-modality integration connects different layers of molecular organization. Single-nucleus RNA sequencing (snRNA-seq) defines cellular identity through gene expression, while chromatin accessibility profiling (scATAC-seq) reveals regulatory



elements such as promoters and enhancers. Tissue architecture is preserved in spatial transcriptomics, enabling the interpretation of laminar organization and functional domains. Integrating these modalities is essential for understanding how regulatory mechanisms drive cell type-specific expression patterns.

The comparative analysis of such multi-modal data traditionally involves fragmented computational workflows and static batch scripts. This approach can be laborious and may disconnect the researcher from the biological context, particularly when the underlying structure does not conform to predefined assumptions. Furthermore, manual alignment of spatial gradients across species with distinct brain morphologies requires significant computational effort. Consequently, there is a need for interactive visual analytics systems that allow for unbiased exploration and rapid examination of molecular patterns across these comparative axes without requiring prior knowledge of specific gene targets or cell type correspondences.

5.3 Impact through Cytosplore Viewer

[Chapter 2-Chapter 4](#) describe the design process and technical development of different Cytosplore Viewer modules, supporting the exploration and communication of complex cross-species multi-omics datasets. Beyond developing novel visualization techniques, a central goal of this work was to contribute to neuroscience research through close collaboration with domain experts. Throughout my PhD, we collaborated extensively with neuroscientists at the Allen Institute and within the Human and Mammalian Brain projects.

This chapter reflects on the process of continuous end-user involvement during these collaborations and describes how the developed Cytosplore Viewer modules contributed to several neuroscience studies. Importantly, the impact of these tools was not limited to generating the original findings. In several cases, the visualization modules supported the interactive exploration, interpretation, and communication of results and enabled additional comparative analyses beyond the original computational workflows.

5.3.1 Supporting Comparative Analysis with Visual Analytics

Single-cell transcriptomic studies rely on computational pipelines to cluster cells, identify marker genes, and compute differential expression statistics [109]. While these workflows generate valuable summary tables and static visualizations, interpreting analysis across attributes and datasets requires iterative exploration. Researchers must contextualize these statistical outputs by cross-referencing them with gene expression distributions, cluster annotations, and associated metadata.



During the collaborations described in this thesis, researchers faced the added complexity of analyzing datasets from multiple species simultaneously, such as humans, macaques, marmosets, gorillas, chimpanzees, and others. Existing literature [68, 85] revealed a lack of specialized tools for such cross-species comparisons, as most visualization tools are tailored to high-detail analysis of a single species. By dynamically linking cell embeddings, gene expression plots, cluster summaries, and metadata, these interfaces allow researchers to seamlessly integrate high-level computational results with the foundational single-cell measurements.

Comparative transcriptomic studies require aligning homologous cell populations across species. In these cross-species projects, the researchers established a shared cell-type taxonomy that linked corresponding cell populations between species. This taxonomy provided the structural basis for the comparisons implemented in the different cross-species comparison modules of the Cytosplore Viewer.

We reviewed existing work on systems supporting cross-species comparison and found several tools that help analyze one species' data in detail. However, only a few tools are available for cross-species comparison, which we also describe in the related work of our earlier chapters (Section 2.3, Section 3.3, and Section 4.1). Thus, we engaged in regular interactions with our collaborators to better understand the data, the current analysis workflow, and problem-solving approaches, which helped inform the development of tools for cross-species analysis. This collaboration lasted more than four years throughout my PhD.

5.3.2 Facilitating pairwise cross-species comparison

My initial collaboration began with the exploration of cross-species differences in single-cell datasets. The goal was to compare cell-by-gene matrices, transcriptomic datasets, and other metadata, such as accelerated regions associated with genes and layer distribution across five different species (human, chimpanzee, gorilla, macaque, and marmoset), to understand how the different attributes in the human dataset differ from those in the other species. This work was adopted by our collaborators and resulted in two collaborative papers across the teams [13, 68] and the Simian Viewer (Chapter 2) module in the Cytosplore Viewer system.

By abstracting key domain-related questions, we developed an interface that integrates several visual representations, including a heatmap to display cell-type distances and help explore cell-type attributes, scatter plots to display low-dimensional embeddings of cells, and a differential expression computation table to explore highly differentially expressed genes and gene expression distribution bar charts.

The system implements an interactive exploratory workflow, allowing users to examine gene expression across species and cell types. Users can iteratively explore



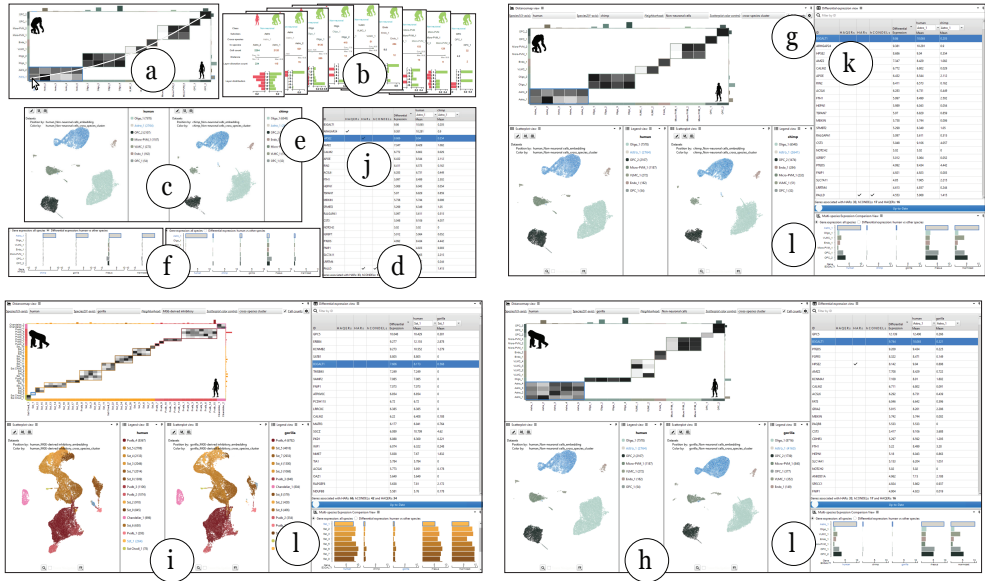


Figure 5.1: Interactive workflow for cross-species transcriptomic analysis in Simian Viewer. The interface enables exploration of cell-by-gene matrices and genomic metadata across primates. It provides an interactive workflow for navigating the data (a). Cell type attributes, including layer distribution and cluster metadata, can be examined (b). Low-dimensional embeddings visualize cellular relationships (c). Differential expression can be computed across species and cell types (d). Selecting *HPSE2* shows its expression in astrocytes across human, chimpanzee, gorilla, macaque, and marmoset (e,f). Multi-gene queries are supported, illustrated with *B3GALT1* (g). Rapid switching between species facilitates comparative analysis (h). Cell type comparisons, such as SST neurons, are also supported (i). Annotations reveal human-accelerated regions for *HPSE2* (j) and their absence for *B3GALT1* (k). The unified interface allows verification of queries and comparison of expression patterns across the dataset (l).

(Figure 5.1a) cell type attributes (Figure 5.1b), embeddings (Figure 5.1c), and compute differential expression across species and cell types (Figure 5.1d). Users can, for example, select a gene such as *HPSE2* (Figure 5.1e) to view its expression profile in astrocytes [110] across multiple species (Figure 5.1f), illustrating the system’s ability to present comparative data. The workflow also supports exploration of multiple genes, such as *B3GALT1* (Figure 5.1g). It also enables users to switch between species (Figure 5.1h) and cell types (Figure 5.1i) to observe similarities or differences and demonstrate how the system facilitates multi-dimensional analysis. Users can also explore associated genomic annotations, such as identifying whether a gene has an accelerated genomic region (e.g., *HPSE2* as shown in Figure 5.1j) or not



(e.g., *B3GALT1* as shown in [Figure 5.1k](#)). By allowing flexible navigation between different queries, comparisons, and annotation layers, the system enables users to verify specific queries, explore alternative paths, and compare expression patterns across species and cell types ([Figure 5.1l](#)), highlighting its capabilities for cross-species analysis.

5.3.3 Scaling cross-species comparison to phylogenies

This collaboration continued when our partners began generating data for twenty-five species, extending from the original five during the previous collaboration ([Section 5.3.2](#)). Through regular meetings, we learned that they were interested in new research questions related to attribute changes in the phylogeny across species. Their existing workflow had been designed for a single species and relied on low-dimensional embeddings to study cellular distribution and the abundance of cell types, requiring inspection of separate scatterplots per species. The biologists began using the resultant tool from this collaboration extensively during their analysis. The collaboration led to a joint paper across the teams [85] and the EvoViewer ([Chapter 3](#)) module in the Cytosplore Viewer system. They are also currently planning to further analyze and release the data and analysis in a future publication.

Through continued discussions, we examined their analysis process in detail. We observed that the biologists had developed a well-defined workflow for one species and were attempting to extend it across all twenty-five species. Our goal was to make this process more effective and to identify additional analysis scenarios that could benefit from improved visualization. At the same time, we aimed to move beyond the pairwise comparison approach used in the Simian Viewer ([Chapter 2](#)) toward enabling comparisons across all twenty-five or more species simultaneously in the EvoViewer ([Chapter 3](#)). During this process, we learned that changes in metadata across the phylogenetic hierarchy of these species were particularly important. For example, the biologists were interested in questions such as how the expression of a gene or the abundance of cells for a particular cell type changes across the phylogeny for the different species. As mentioned earlier ([Section 5.3.1](#)), a common cell type hierarchy generated by our collaborators is always used for comparison. Initially, we attempted to construct hierarchical clusterings of species based on attributes such as cell counts, abundance within a cell type, or mean gene expression. We then compared the resulting tree-like structures with the actual phylogenetic tree to observe how species positions changed depending on the attribute values. For this purpose, we utilized a method proposed by S. Bremm et al. [73]. However, we found that changes in species positions within a hierarchical tree context were not reliable and were confusing for the biologists. As a result, we instead mapped attribute values



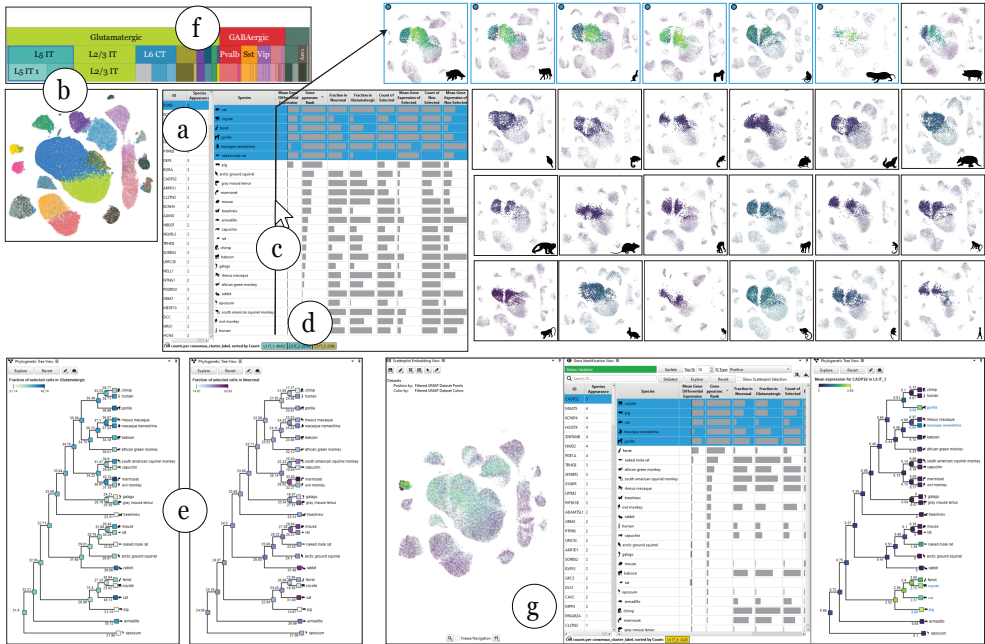


Figure 5.2: Phylogeny-based interactive exploration of cell type attributes across twenty-five mammalian species in EvoViewer. The interface enables the identification of cell-type-specific markers and the evaluation of attribute changes across twenty-five mammalian species. Users can explore a specific marker gene, such as *RORB* (a), within a selected cell-type context, such as *L5 IT* (b), to observe how marker properties are distributed across different species (c). The module facilitates the comparison of cell abundance and other quantitative attributes (d) by mapping them directly onto the branches of a phylogenetic tree (e). Interactive controls support rapid switching between cell types (f) to perform targeted cell-type-specific analysis across the taxonomy (g).

directly onto the phylogenetic tree to better understand differences and similarities across species.

Another goal was to identify marker genes across species. To support the workflow, we developed a module that enables users to zoom into specific cell types and identify cell-type-specific marker genes by comparing expression patterns with other cell types and then extending these comparisons across multiple species. Within this module, attributes can be mapped onto the phylogenetic tree, allowing users to observe trends and potential outliers across species. Users can, for example, explore a gene such as *RORB* (Figure 5.2a) in the *L5IT* cell type (Figure 5.2b) to see how marker gene properties [111] appear across different species (Figure 5.2c), illustrating the system’s ability to highlight patterns and similarities. The module



also allows comparison of abundance and other attributes (Figure 5.2d) across the phylogeny (Figure 5.2e), rapid switching between cell types (Figure 5.2f) to perform cell type-specific analysis (Figure 5.2g), and expansion of analysis to all species in the dataset. Interactive exploration supports inspection of UMAP embeddings, comparison of marker scores, and switching between views and attributes, enabling a more flexible and comprehensive analysis than static charts. Detailed exploration of a gene such as *RORB* in *L5IT* cells across species demonstrates how the system facilitates multi-level inspection, comparison, and cross-species analysis, highlighting its advantages for interactive and flexible data exploration.

5.3.4 Exploring gene expression gradients across species

For single-cell RNA sequencing, we addressed several specific problems through our two modules discussed above (Section 5.3.2 and Section 5.3.3). However, spatial transcriptomics introduces a unique set of analytical challenges. Unlike traditional single-cell RNA sequencing, spatial data inherently relies on preserving the precise anatomical context of cells within the brain. Cross-species integration is further complicated by several compounding factors. First, significant variations in brain size and morphology prevent the direct spatial mapping of cells between species. Second, technical batch effects arising from diverse data generation platforms hinder seamless integration. Finally, due to the limited overlap of homologous marker genes across species, direct genomic comparisons are difficult. Consequently, cross-species alignment must primarily rely on the broader evolutionary conservation of cell types. This is a phenomenon consistent with observations in scRNA-seq datasets.

A review of the current literature indicates that few tools exist to compare spatial transcriptomic datasets across species. While several platforms enable detailed exploration within a single species, cross-species comparison remains a notable gap in the field (Section 4.1). To better understand these technical hurdles, we held regular meetings with collaborators at the Allen Institute. These discussions revealed that the gradient of gene expression across physical space is an essential factor for these datasets. This occurs because cell populations seldom have sharp boundaries and instead exhibit continuous gradients that must be carefully mapped [112, 113, 114].

Typically, when analyzing a single species, researchers identify a specific region, define a trajectory or angle within that area, and isolate genes that exhibit expression gradients along that axis. Although biologists often utilize tailored scripts for individual species, certain technical hurdles persist. For instance, the rapid selection of cells within intricate anatomical structures and the isolation of specific cell populations based on type remain involved processes. Furthermore, extending this workflow across multiple species is often time-intensive and susceptible to inconsistencies,



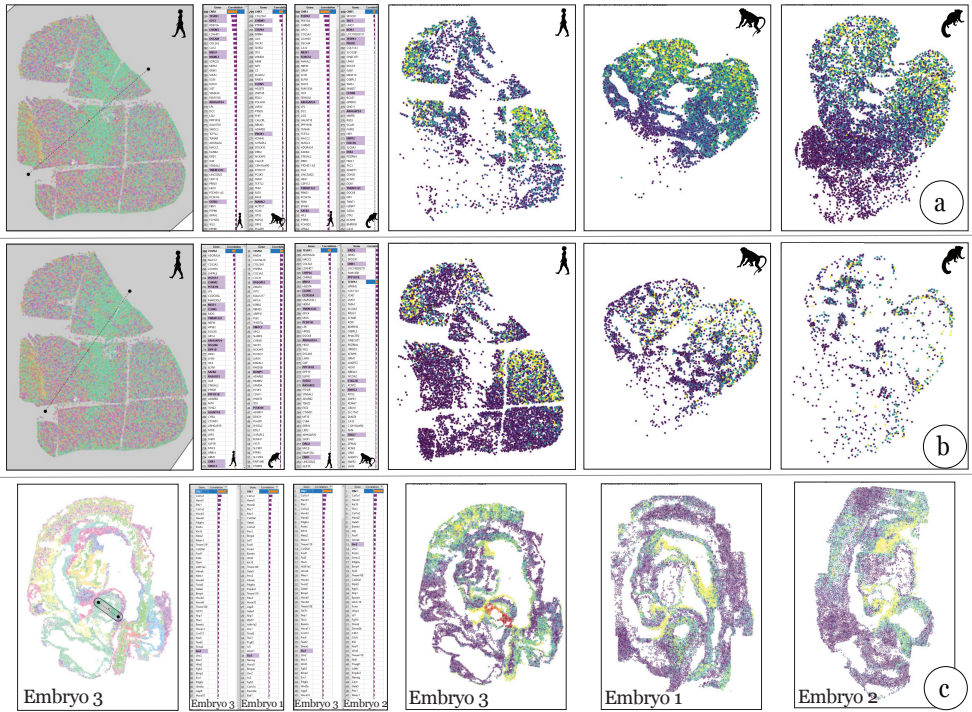


Figure 5.3: Visualizing conserved spatial gene expression gradients with Gradient Surfer. The module provides a unified framework for comparing continuous spatial expression patterns across morphologically distinct species and datasets through an interactive swipe. The interface supports the simultaneous alignment and comparison of expression profiles for genes such as *CNR1* (a) and *TESPA1* (b) across human, macaque, and marmoset brain slices. For developmental studies, the system enables the identification of conserved patterns for genes such as *Dlk1* (c) across multiple embryo datasets. Gradient Surfer facilitates dynamic interaction by allowing users to switch between genes, cell types, and species to view spatial expression gradients within a single interface.

suggesting that a more unified analytical framework would be helpful for these comparisons. This collaborative effort led to a potential high-impact publication [87] and an ongoing project detailed in Chapter 4. Beyond its application in cross-species comparisons, the resulting tool facilitated the analysis of diverse tissue regions in other spatial transcriptomic projects at the Allen Institute. In these instances, analysts integrated their datasets into the system to perform further examination, which demonstrated the broader utility of the module for spatial data exploration.

To address the challenges of cross-species comparison, we developed the Gradient Surfer module (Chapter 4) within the Cytosplore Viewer. This tool allows



researchers to identify conserved spatial gene expression gradients across multiple species through an interactive swipe. For example, in the recently acquired human, macaque, and marmoset spatial datasets from the Human Mammalian Brain Atlas (HMBA¹⁰) consortium, genes such as *CNR1* (Figure 5.3a) and *TESPA1* (Figure 5.3b) show consistent expression gradients across human, macaque, and marmoset brain slices. Identifying these patterns previously required the use of separate scripts for each individual species. Similarly, for the embryo dataset [115], genes such as *Dlk1* (Figure 5.3c) appear conserved across the three embryos. Gradient Surfer supports dynamic interaction by allowing users to switch between genes, cell types, and species to view spatial expression gradients in a single interface.

5.3.5 Investigating the coordination between regulatory landscapes and gene expression

During the development of Gradient Surfer (Chapter 4), we also collaborated with partners to analyze chromatin accessibility datasets generated within the HMBA project. These datasets were produced using single-cell ATAC sequencing (scATAC-seq), which profiles chromatin accessibility to provide insights into the regulatory programs underlying neuronal identity and subcortical patterning. Through regular meetings with biologists, we identified several challenges in their existing workflow, specifically regarding the inherent sparsity of ATAC signals and the difficulty of connecting these regulatory features with gene expression and spatial context. The collaborators' initial approach relied heavily on custom scripts to aggregate sparse peaks and identify correlated elements, which made iterative exploration difficult and required switching between multiple tools. This fragmented process forced researchers to manually export data from command-line environments to separate visualization software, breaking the flow of analysis. Discussions revealed a need for a unified interface to explore the relationships between transcription factors, regulatory elements, and gene expression across cell types and anatomical regions.

To support this workflow, we developed the ATAC Viewer module within the Cytoscore Viewer system. The module integrates chromatin accessibility data with gene expression and spatial information through a unified interface. Because scATAC-seq datasets are highly sparse, we aggregate signals based on established cell type taxonomies and allow users to compute summary representations of feature modules, such as aggregate accessibility signals or the first principal component of peak sets. A bidirectional search engine enables users to identify ATAC peaks associated with transcription factors and explore how these regulatory elements relate to gene expression patterns. This collaboration contributed to a joint publication across the

¹⁰<https://brain-map.org/consortia/hmba>



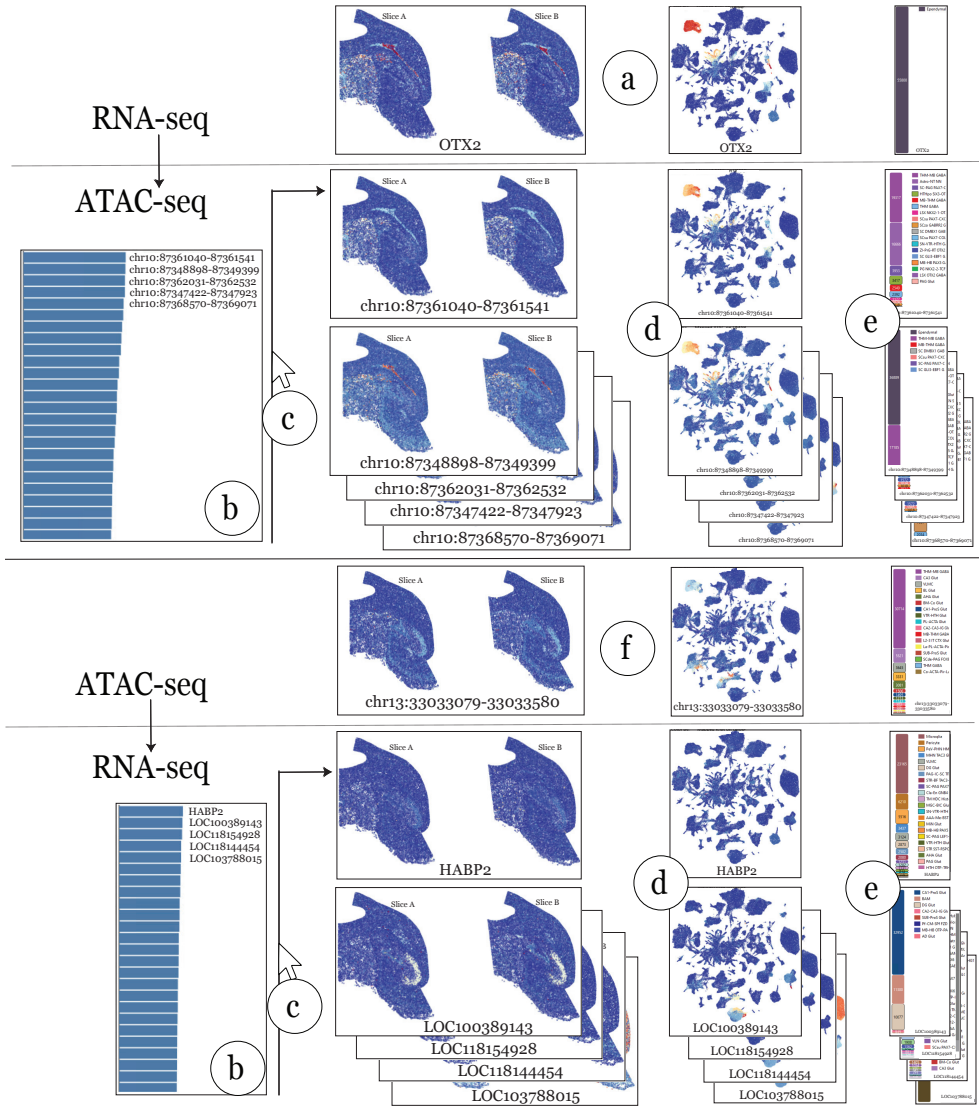


Figure 5.4: Integrative analysis of chromatin accessibility and gene expression in the ATAC Viewer. Selecting a transcription factor, such as OTX2 (a), retrieves associated chromatin accessibility peaks, with top peaks shown in an interactive bar chart (b). Selecting a peak updates spatial maps (c), embedding views (d), and cell type composition charts (e), preserving cellular context. Queries can also start from genomic loci or ATAC peaks, such as *chr13:33033079-33033580* (f), to identify RNA genes with similar spatial distributions across linked views.



teams [116], which investigated the coordination between chromatin accessibility and transcriptional programs in the subcortex. The resulting ATAC Viewer module complements other modules in the Cytosplore Viewer system by extending the analysis workflow from gene expression and spatial gradients to regulatory programs.

Within the system, users can, for example, select a transcription factor such as *OTX2* (Figure 5.4a) and explore associated ATAC peaks across cell types. Through the bidirectional search interface, the system computes correlations between the selected transcription factor and chromatin accessibility peaks across the defined cell type taxonomy. The highest-ranked peaks are presented as an interactive bar chart (Figure 5.4b). Selecting a peak updates all linked visualizations, including spatial maps (Figure 5.4c), embedding views (Figure 5.4d), and cell-type composition charts (Figure 5.4e).

Exploration can also begin from a genomic locus or individual ATAC peak, such as *chr13:33033079-33033580* (Figure 5.4f), allowing users to query RNA genes with similar spatial distributions and visualize them across the coordinated views. Because all visualizations remain linked, researchers can iteratively examine relationships between regulatory elements and transcriptional signals while maintaining their spatial and cellular context. Before the development of this module, such analyses required multiple scripts to compute correlations, aggregate sparse peaks, align spatial coordinates, and generate separate visualizations across different software tools. By integrating these operations into a single environment, the ATAC Viewer enables interactive exploration of regulatory features and gene expression without interrupting the analytical workflow.

5.4 Discussion

Cytosplore Viewer is a modular visual analytics system designed to support comparative analysis of multi-modal brain atlases across species. By integrating coordinated visualizations of embeddings, gene expression, spatial data, and regulatory signals, the system links computational outputs to the underlying single-cell measurements and enables iterative exploration of complex biological datasets. Through close collaboration with neuroscientists, the system was applied in multiple studies investigating evolutionary differences in brain organization and regulatory programs [13, 87, 116]. These collaborations demonstrated the practical value of interactive visual analytics in domains where analysis pipelines traditionally rely on static plots and custom scripts.

Future development will focus on supporting larger atlas-scale datasets and additional omic modalities. As brain informatics resources expand, interactive visual analytics will remain critical for interpreting complex multidimensional data. On-



going refinement will enhance Cytosplore Viewer's ability to support cross-species, multi-modal investigations, contributing to a deeper understanding of molecular and cellular principles underlying mammalian brain organization and evolution.

