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

Basu, S.

CHAPTER 6

DISCUSSION

This thesis develops an interactive visual analytics system for the comparative analysis of multi-species brain atlases. Advances in single-cell and spatial transcriptomics have enabled the systematic study of gene expression and cell-type organization across species, but the resulting datasets are large, heterogeneous, and difficult to interpret using static, batch-oriented workflows. The central objective of this thesis is to design multiple interactive modalities within a unified analytical system that support exploratory, cross-species analysis of transcriptomic and spatial data.

The thesis is grounded in close collaboration with domain experts and follows a design study methodology. It first characterizes the analytical challenges of comparative brain atlas research, including inconsistent gene coverage, divergent cell-type taxonomies, heterogeneous experimental platforms, and continuous spatial organization that is poorly captured by discrete clustering. These challenges motivate interactive approaches that enable rapid hypothesis generation, iterative refinement, and visual validation of computational results.

Chapter 2 introduces Cytosplore Simian Viewer, a system for pairwise cross-species comparison of single-cell transcriptomic datasets. Through linked views of cell-type hierarchies, gene expression profiles, and species-specific distributions, the system supports interactive identification of conserved and divergent patterns. Validation on primate cortex data demonstrates that interactive analysis can reproduce findings from large comparative studies while enabling additional exploratory insights.

Chapter 3 extends the comparative scope to broader phylogenies with Cytosplore EvoViewer. By integrating evolutionary relationships directly into the visual workflow, the system enables phylogeny-aware exploration of conservation and divergence across multiple species. Interactive branch-based queries support the identification of lineage-specific patterns. Expert evaluation confirms the analytical value of this approach.

Chapter 4 presents Cytosplore Gradient Surfer, an interactive method for comparing spatial transcriptomics datasets. Users can probe spatial directions within a reference dataset, identify gradient-associated genes, and project these gradients onto other datasets through a shared feature space. This approach enables cross-dataset comparison even when tissue geometry and gene panels differ. Validation across multiple spatial modalities demonstrates recovery of known biological gradients and robustness to rotation, subsampling, and partial gene overlap.

Chapter 5 introduces Cytosplore Viewer, a modular system that unifies transcriptomic, evolutionary, spatial, and regulatory components within a single interactive environment. By connecting gene expression, phylogenetic context, spatial organization, and chromatin accessibility through consistent interaction principles, the system enables researchers to iteratively refine interpretations while maintaining biological context.

Together, the contributions of this thesis demonstrate that interactive visual analytics can substantially enhance the interpretability, flexibility, and comparative power of large-scale brain atlases.

A central theme emerging from this thesis is the importance of human-in-the-loop analysis in large-scale biological data exploration. Computational pipelines such as Scanpy [7] and Seurat [2] provide powerful tools for preprocessing, dimensionality reduction, and statistical modeling of single-cell datasets. However, the outputs of these pipelines are often static visualizations and reproducible scripts, which prioritize automation and documentation over interactive exploration.

The systems developed in this thesis introduce an interactive layer that allows researchers to dynamically inspect embeddings, query gene expression patterns, and compare cell populations across species and modalities. This interactive approach supports iterative hypothesis refinement and enables domain experts to directly interrogate computational results. Rather than replacing automated analysis pipelines, this visual analytics system complements them by allowing researchers to explore intermediate results, validate assumptions, and generate new biological hypotheses. In this way, interactive visualization becomes a critical component of the analytical workflow for large-scale atlas datasets.

Another key contribution of this work is the development of task-driven visual analytics systems tailored to comparative neuroscience. The design of Simian Viewer (Chapter 2), EvoViewer (Chapter 3), Gradient Surfer (Chapter 4), and ATAC Viewer (Section 5.3.5) was guided by close collaboration with neuroscientists and by the specific analytical tasks involved in cross-species atlas analysis. These tasks include identifying homologous cell populations, comparing gene expression patterns across species, exploring phylogenetic relationships, detecting spatial gradients of gene expression, and investigating the coordination between chromatin accessibility and gene expression patterns. By designing visualization and interaction techniques specifically for these tasks, the systems provide specialized analytical support tailored to comparative biology workflows.

The modular architecture of Cytosplore Viewer (Chapter 5) demonstrates how specialized analytical components can be integrated within a coherent system. Each module addresses a specific analytical problem while sharing a consistent interaction paradigm, enabling researchers to move between different analysis modes without

disrupting their workflow.

Comparative brain atlas analysis requires reasoning across multiple dimensions of biological organization, including gene expression, spatial organization, cell-type taxonomy, and evolutionary relationships. Visual analytics systems must therefore support coordinated exploration across multiple data modalities. Cytosplore Viewer addresses this challenge through coordinated multiple views linking transcriptomic embeddings, spatial transcriptomics, phylogenetic structures, and regulatory information. Interactions in one view propagate to other views, allowing users to connect patterns observed at different biological scales. This coordinated view approach enables researchers to move between levels of abstraction, from individual genes to cell-type populations to anatomical organization. Such multi-level exploration is essential for understanding complex biological datasets where meaningful patterns often emerge from the interaction between different data modalities.

Many biological processes exhibit continuous rather than discrete variation. Spatial transcriptomics datasets often reveal gradual gradients of gene expression across anatomical regions. Clustering-based approaches, while useful for cell-type identification, may not fully capture these continuous patterns. The Gradient Surfer module demonstrates an alternative approach by enabling interactive exploration of spatial gene expression gradients through line-based probing and principal component projection, complementing discrete clustering with continuous gradient analysis.

Cytosplore Viewer was implemented as a standalone desktop application for Windows and macOS environments. While web-based visualization platforms offer convenient browser access and robust capabilities for large-scale data management, practical considerations led us to adopt a desktop architecture for these projects. The desktop implementation provides low-latency access to local compute and storage, which supports interactive feature computation (E.g., differential expression analysis between manual selections or on-the-fly re-embedding of data subsets. Running the software locally can also simplify the handling of sensitive or embargoed datasets by keeping data within the user's computational environment. Finally, the standalone architecture reduces the immediate need for centralized infrastructure and long-term service maintenance, allowing collaborators to run analyses with minimal external dependencies. These choices reflect pragmatic trade-offs rather than a claim of inherent superiority.

Although Cytosplore Viewer was developed for comparative brain atlas research, many of the challenges addressed in this thesis are shared across single-cell biology. Tissue atlases, developmental atlases, and disease atlases all involve integration of heterogeneous datasets, cross-condition comparison, and exploration of high-dimensional molecular measurements. The visual analytics principles developed in this work, including task-driven design, coordinated views, and scalable interaction

techniques, may therefore generalize well to other domains of biological research.

Several promising directions emerge from this research that could enhance the capabilities and usability of interactive visual analytics systems for comparative biology.

AI-assisted exploration and onboarding As visual analytics systems grow in complexity, user onboarding becomes increasingly challenging. Future work could incorporate AI-assisted interfaces that guide users toward appropriate analytical modules, explain visualization parameters, or suggest analysis steps based on research questions. Large language models could provide natural language query interfaces for navigating atlases and generating exploration strategies. However, such AI assistance must be carefully designed to complement rather than replace expert interpretation, as automated systems can generate plausible but incorrect biological conclusions.

Uncertainty quantification in cross-species alignment Cross-species cell-type correspondences are inherently uncertain due to differences in gene expression patterns, sampling density, and annotation strategies. Future systems could incorporate explicit uncertainty-aware visualizations that communicate confidence levels in alignments or allow researchers to compare alternative mapping strategies. Such capabilities would help researchers identify ambiguous correspondences and understand the limits of cross-species generalization.

Scalability and progressive data loading Brain atlas datasets continue to grow in size and complexity. While the systems presented in this thesis demonstrate interactive performance on current datasets, future expansion to hundreds of species and billions of cells may require additional techniques such as out-of-core data management, progressive loading, hierarchical exploration, and distributed computation. Such innovations would maintain interactive responsiveness while preventing cognitive overload when navigating large comparative datasets.

Explicit representation of continuous biological structure Many biological processes exhibit continuous rather than discrete variation. Future visual analytics methods could explicitly represent continuous structures through trajectory-based visualizations, gradient-aware comparison techniques, and topology-preserving embeddings that capture developmental and evolutionary dynamics more naturally than discrete clustering approaches.

The rapid expansion of multi-species and multi-modal brain atlases has transformed the scale and complexity of neuroscience data. While computational pipelines remain essential for processing these datasets, effective interpretation increasingly

depends on visualization approaches that support interactive and comparative reasoning.

By integrating transcriptomic, spatial, phylogenetic, and regulatory information within a modular visual analytics framework, this thesis demonstrates how interactive exploration can enhance interpretability and facilitate discovery in large-scale comparative neuroscience.